
A Survey on Image Quality Assessment

Lanjiang Wang

University of Electronic Science and Technology of China

Abstract

Image quality assessment(IQA) is of increasing importance for image-based applications. Its purpose is to establish a model that can replace humans for accurately evaluating image quality. According to whether the reference image is complete and available, image quality evaluation can be divided into three categories: full-reference(FR), reduced-reference(RR), and non-reference(NR) image quality assessment. Due to the vigorous development of deep learning and the widespread attention of researchers, several non-reference image quality assessment methods based on deep learning have been proposed in recent years, and some have exceeded the performance of reduced-reference or even full-reference image quality assessment models. This article will review the concepts and metrics of image quality assessment and also video quality assessment, briefly introduce some methods of full-reference and semi-reference image quality assessment, and focus on the non-reference image quality assessment methods based on deep learning. Then introduce the commonly used synthetic database and real-world database. Finally, summarize and present challenges.

1 introduction

In an era where visual intelligence products are so widespread today, images occupy a dominant position in the traffic carrier. Image based applications can be found everywhere in our daily life. Image dehazing/deraining proposed by [79, 27, 80, 25, 88, 26, 87, 40, 86] can be important in autopilot of smart cars, object detection methods proposed by [30, 6, 7, 53, 10, 31, 32, 54, 55, 33, 74] can be used in monitoring of transportation hubs and image segmentation methods [98, 62, 15, 97, 66, 91, 94, 93, 65, 43, 41] can be applied in medical imaging. However images are transmitted with varying degrees of degradation in the quality of the images eventually received by the observers - mostly humans - at the receiving end due to the hardware and software conditions of the transmission channel or the receiving device, as well as lossy compression, e.g. image JPEG compression can cause blurring and ringing effects. In this case, Image Quality Assessment (IQA) was created in order to maintain and improve the quality of the images at the receiving end. IQA is of great interest for various applications in all stages of computer image processing. For example, it can be used for image acquisition [89], image fusion [49], face recognition [19], and medical images [9]. IQA methods can be generally classified as subjective and objective [73]. Subjective methods are measured by the final recipient of all media - humans - and are the most accurate and reliable methods. However, subjective IQA requires a lot of human and material consumption, does not meet the requirements of real-time IQA systems, and appears impractical in practical applications. Therefore, the researchers propose an objective method that simulates human assessment of picture quality.

Objective image quality assessment is an important area of image processing research, which automatically predicts image quality by means of mathematical models designed to approximate human prediction. Objective image quality assessment can be divided into three categories according to whether the reference image is complete and available: full-reference, reduced-reference and no-reference image quality assessment. Full-reference IQA methods use the entire undistorted reference image to compare with the distorted image and measure the difference between them, while reduced-

reference IQA methods use part of the information in the reference image, such as the extraction of handicraft features. However, in practical applications, the reference image is difficult to obtain, making the above two methods inapplicable. Therefore, it is becoming increasingly important to develop effective NR-IQA methods. This paper will focus on the development of NR-IQA methods.

The rest of this paper is structured as follows: Section 2 provides an introduction to the concept of image and video quality assessment; Section 3 focuses on the NR-IQA method; Section 4 introduces the datasets commonly used in IQA; and finally Section 5 concludes this paper and provides an outlook.

2 IQA

2.1 Definition

Factors affecting image quality come from several sources, such as brightness, contrast, composition, noise and so on. The degree of distortion caused by the above factors varies for different images, so it is difficult to determine which factor plays a major role. According to [17] image quality is defined as consisting of three factors, namely fidelity, perception and aesthetics. Fidelity is the accuracy of the distorted image relative to the original image; perception is derived from the human visual system and such metrics consider visual attention [3], contrast masking [21], etc.; and aesthetics is subjective and may include [17] visual constancy, visual attention and visual fatigue.

2.2 Classification

Image quality assessment can be divided into subjective quality evaluation and objective quality evaluation depending on the subject of the prediction. Subjective quality assessment is the most reliable method for assessing image quality, as human observers are in most cases the ultimate recipients of the image transmission system [46]. Mean opinion score (MOS) is a metric for subjective quality assessment that requires a large number of observers to evaluate. MOS is considered to be the best image quality metrics. Subjective quality assessment is a traditional method of measuring image quality and requires a group of subjects who must rate the quality of the image in a controlled testing environment. Each individual's rating may be different. The results are finalized by processing a weighted average of each individual's results. It can provide accurate results, but is slow in practical applications and expensive to use in practice [75].

Objective image quality assessment can be performed automatically and accurately by means of mathematical models designed to fit the human evaluation of the input image, which can save a lot of human and material resources. Further, objective image quality assessment can be classified into three: full-reference, reduced-reference and no-reference image quality evaluation.

Full-reference image quality assessment (FR-IQA) can use the entire undistorted image and obtain an image quality score for the distorted image based on the difference between the distorted image and the original image. Figure 1 shows the flowchart for full-reference image quality assessment. Efforts to investigate full-reference image quality assessment include [78, 77, 63, 23, 100].

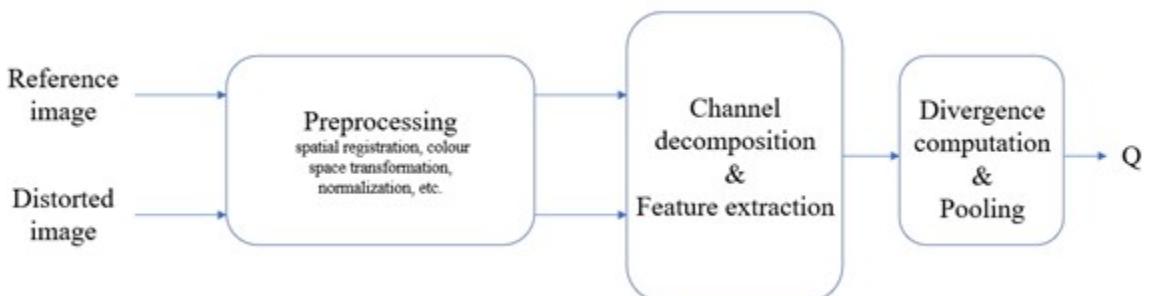


Figure 1: Framework of FR-IQA

Reduced-reference image quality assessment (RR-IQA) is designed to help evaluate the quality of distorted images using only manually extracted features of the reference image when the reference

image is not completely available. The idea of reduced-reference image quality assessment was originally proposed in the context of tracking the degree of visual quality degradation of video transmitted over communication networks [17]. The framework of reduced-referential image quality evaluation is shown in Figure 2. Efforts to study semi-referential image quality evaluation include [56, 76, 29, 14, 81].

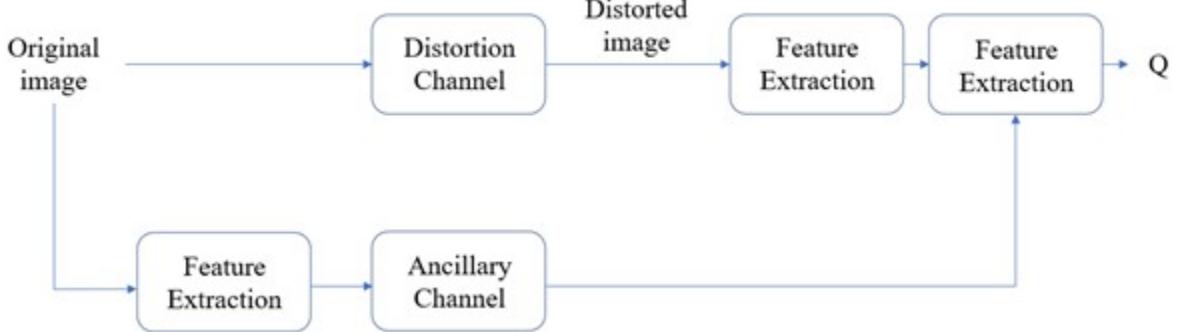


Figure 2: Framework of RR-IQA

In most practical applications it is difficult to obtain information about the reference image and often the receiver only receives a distorted image. For example, when we take a picture with a mobile phone, we can only obtain a picture with various possible truths, but not a reference image. Therefore reference-free image quality assessment (NR-IQA), also known as blind image quality assessment (BIQA), seems so important. In fact, this is a very difficult task, about which we will focus in the next section. Figure 3 shows the basic framework for most of the non-referenced image quality evaluations.

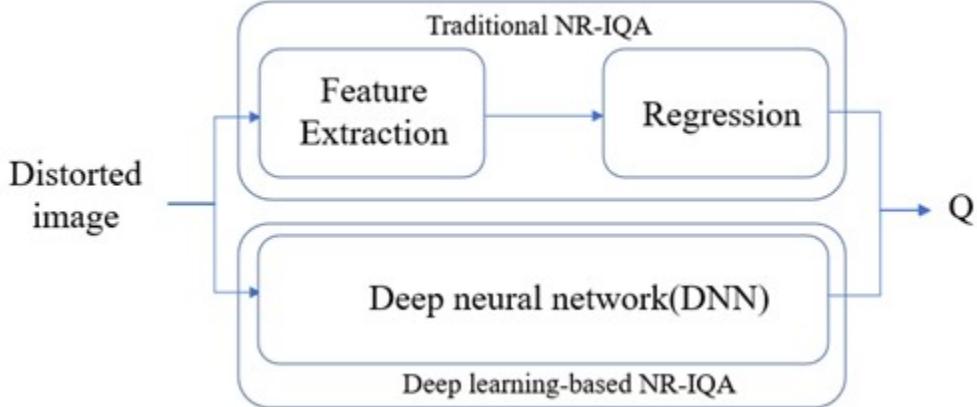


Figure 3: Framework of NR-IQA

2.3 VQA

Video quality assessment (VQA) aims to build models for evaluating video quality in the streaming and social media industries. Same as IQA, VQA can be divided into full-reference (FR), reduced-reference (RR) and no-reference(NR) VQA. While (FR) VQA research is maturing and several algorithms[78, 34] are widely used, more recent attention has turned to creating better no-reference (NR) VQA models that can be used to predict and monitor the quality of synthetically and authentically distorted videos. Many researchers have investigated and proposed possible solutions to the NR VQA problem [44, 58, 90, 45, 13, 70, 24], and one simple but reasonably effective strategy is to compute quality scores for each frame and then represent the evolution or relative importance over time by applying a time pool to the quality scores. Simple time-averaged pooling is widely used for enhancing

FR [60, 1, 72] and NR VQA models[58, 45, 70]. Other types of pooling used include harmonic averaging [35], Minkowski averaging [57, 61], percentile pooling [47, 5] and adaptive weighted sums [48]. More sophisticated pooling strategies consider memory effects such as prioritization, recency [57, 61, 2] and hysteresis[90, 24, 59, 8]. However, to date, the general applicability of these pooling models has not been thoroughly validated in the general context of real-world UGC video NR VQA models, although some more focused research has been conducted [57, 61].

3 No-reference IQA

No-reference image quality assessment (NR-IQA), also known as blind image quality assessment (BIQA), aims to construct a computational model for accurately and automatically predicting the quality of images as perceived by humans without the need for additional information. As can be seen in Figure 3, the first type of NR-IQA methods are based on traditional regression methods, where manually designed features are first extracted from the image and then a trained regression network is used to predict the image quality to obtain the final quality score. Earlier studies used a priori knowledge for predicting specific distortion types [28, 38] by extracting and exploiting specific distortion features. Li et al [38] proposed an NR-IQA method based on fuzzy distortion. They used a gradient image to describe the ambiguity and segmented the gradient image into blocks, extracting the energy features of each block associated with the ambiguity distortion. Finally, the image quality is obtained by normalization. The disadvantage of this type of method is that it is difficult to find efficient features for quality assessment if the image shows distortions that have not been manually designed.

To address the problem of assessing image quality without a distortion prior, researchers have used natural statistics (NSS)-based methods to extract reliable features which assume that natural images share certain statistical information and that distortion may alter these statistics [84, 101, 83, 95, 20, 85, 102]. Wu et al. select statistical features extracted from binary patterns of local image structures. Zhang et al. used multivariate Gaussian (MVG) models, Yang et al. used generalized Gaussian distribution (GGD) model to extract image features, and Wu et al. used multi-channel fused image features to simulate the hierarchical and trichromatic properties of human vision. Then, k-nearest neighbor (KNN) based models were used to evaluate image quality. The following table shows the various representative evaluation models. The feature extraction methods, regression methods and datasets used by the three representative IQA models are listed in the table.

Table 1: Comparison of representative models.

Method	Feature Extraction	Regression	Databases
IL-NIQE[101]	MVG	Pooling	TID2013,CSIQ,LIVE,MD1,MD2
BWS[95]	GGD	SVR	LIVE, TID2008, CSIQ
TCLT[83]	Multichannel Feature Fusion	KNN	LIVE II, CSIQ,TID2008

However, these NR-IQA methods are limited to hand-crafted features that may not adequately represent complex image structures and distortions. The rise of deep learning in recent years has piqued the interest of researchers. Deep neural networks (DNNs) can automatically extract deep features relevant to quality assessment and optimize these features by using back-propagation methods to improve predictive performance [96]. As a result, DNNs can be applied to various image quality assessment (IQA) methods such as [82, 43] and becomes a promising option for solving NR-IQA tasks. It is well known that deep learning techniques have been highly successful in solving various image recognition and target detection tasks [22, 68, 16, 67]. The main reason is that it relies heavily on large-scale annotated data, such as the ImageNet dataset, which is oriented towards image recognition. However, for NR-IQA tasks, due to the relatively small size of the dataset, direct training with DNNs leads to overfitting, i.e., the trained model has perfect performance for the training set, but unreliable performance for the tested data. Researchers have therefore focused on exploring effective DNN-based NR-IQA methods to address this problem. Some of the prominent work is presented below.

3.0.1 DeepRN

Varga et al. [71] propose DeepRN, a blind image quality assessment (BIQA) method based on deep learning of convolutional neural networks (CNN). DeepRN uses a residual deep learning network (ResNet-101) as the backbone network to extract features and adds an adaptive spatial pyramidal pooling layer at the end of the network so that the backbone network can output fixed size features, so that images of arbitrary size can be processed. Training was performed on a new, large, labelled dataset of 10073 images (KonIQ-10k) that contained histograms of quality ratings in addition to mean opinion scores (MOS). DeepRN obtained the leading performance at the time.

3.0.2 TS(Two-Stream) model

Yan et al. [92] propose a two-stream convolutional network that takes an image and a gradient map as input respectively thereby acquiring different levels of information from the input and alleviating the difficulty of extracting features from a single input. The distorted local non-uniform distribution in the image is also considered by adding a region-based full convolution layer for using the information around the center of the input image block. The final score for the overall image is calculated by averaging the block scores. Experimental results on a range of benchmark datasets, such as LIVE, CISQ, IVC, TID2013, and Waterloo Exploration datasets, show that the dual-stream network performs reliably.

3.0.3 RankIQA

Liu et al. [39] proposed RankIQA, an innovative application of ranked training to learn image quality assessment criteria. To address the problem of limited dataset size, RankIQA uses a twin network (Siamese Network) trained on a large synthetic dataset (using synthetic distortions of known relative image quality, which can be generated automatically without manual annotation) to rank distortion levels. Once trained, a branch of the network is used to fine-tune it on the IQA dataset, mapping the features to the annotated values. Experiments on the TID2013 benchmark showed that RankIQA improved performance by over 5%. There is also a lot of work on NR-IQA that borrows the idea of Learning-to-Rank, such as DipIQ [42] proposed by Ma et al. and PieAPP [52] proposed by Prashnani et al. Both have achieved good results by learning to rank image pairs to obtain an image quality-aware model.

3.0.4 Hallucinated-IQA

Lin et al [37] applied Generative Adversarial Networks (GAN) to NR-IQA to solve the NR-IQA problem from a novel perspective. They first generated an illusionary reference image from the distorted image to compensate for the missing real reference image, and then fed the difference between the distorted image and the illusionary reference to a regression network to generate quality predictions. The network consists of three components, including a quality-aware generation network G, a specially designed judgement network D and a quality regression network R. The training strategy is to first train the GAN network to generate a large number of illusionary images, which are similar to the reference images in the IQA database. Then, the R network is trained to predict the image quality scores. In the GAN network, the D network is first trained to distinguish between the pseudo-reference images in the IQA database and the reference images. Then, the G network is trained to generate images that are similar to the real reference images in the IQA database. Finally, the image quality score can be predicted by optimizing the loss of the R network. The model gives good results on several artificial, real distortion datasets.

3.0.5 TRIQ

First proposed in 2017 and replacing RNNs in the field of natural language processing (NLP), Transformer has also attracted research interest in the field of computer vision [4, 11]. You et al [99] investigated the application of Transformer in image quality (TRIQ) evaluation in the visual Transformer (ViT) the Transformer encoder uses adaptive positional embedding techniques to process images of arbitrary resolution. The model was trained using the large datasets LIVE-wild [12] and KonIQ-10k [36] Different settings of the Transformer architecture were tried on publicly available image quality databases and the results showed that the TRIQ architecture achieved excellent performance.

4 Datasets and Performance Metrics

Datasets and performance metrics are used to evaluate the performance of various algorithms. The datasets and performance metrics commonly used in IQA are described as follows.

4.1 Datasets

The most used datasets in the IQA are LIVE [64], TID2008 [50], TID2013 [51], CSIQ [23], LIVE MD [18], LIVE In the Wild Image Quality Challenge Database [12], KonIQ-10k [36], etc. Among them, LIVE In the Wild Image Quality Challenge Database and KonIQ-10k are natural distortion datasets, and the rest five are artificially simulated distortion datasets.

4.1.1 LIVE

The LIVE (Laboratory for Image & Video Engineering) dataset [64] was created by the LIVE Laboratory at the University of Texas at Austin and is one of the most used datasets for IQA. The dataset contains 29 reference images with resolutions ranging from 438×634 pixels to 512×768 pixels. These images were manually simulated to produce 5 different types of distorted images, 779 in total. Distortion types include JPEG2000 distortion, JPEG distortion, white noise distortion, Gaussian blur distortion and fast Rayleigh decay distortion. The dataset provides the Differential Mean Opinion Score (DMOS) for all distorted images in the range [0, 100], where 0 means no distortion.

4.1.2 TID2008

TID (Tampere Image Database) 2008 [50] was created by Tampere University of Technology, Finland in 2008 and includes 25 color reference images with a resolution of 384×512 pixels, with 17 distortion types, each containing four different levels of distortion, for a total of 1700 images. Artificial distortions include additive Gaussian noise, spatial correlation noise, masking noise, high frequency noise, impulse noise, quantization noise, Gaussian blur, image denoising, JPEG compression, JPEG2000 compression, JPEG transmission error, JPEG2000 transmission error, non-offset pattern noise, varying intensity of local block-by-block distortion, average offset (intensity shift), and contrast variation. The data set provides MOS values and their standard deviations for all tested images, with MOS values in the range [0, 9], where 9 means that the image is distortion-free.

4.1.3 TID2013

TID2013 [51] was created by Tampere University of Technology, Finland in 2013 and includes 25 reference images, 3000 distorted images (including 25 reference images with 24 distortions simulated manually for each image, each with 5 levels of distortion). There are 24 distortion types, an increase of 8 distortion types compared to the TID2008 image dataset, namely altered color saturation index, lossy compression, multiple Gaussian noise, color image quantization, sparse sampling, chromatic aberration and comfort noise. The annotation type for this dataset is DMOS, which is obtained statistically from 524,340 samples observed by 971 observers, with DMOS values ranging from [0, 9], with larger values indicating poorer image quality.

4.1.4 CSIQ

CSIQ (Categorical Subjective Image Quality) [23] was established by Oklahoma State University, USA, which contains 30 reference images with a resolution of 512×512 pixels and six types of distortion, including overall contrast reduction, JPEG compression, JPEG2000 compression, additive Gaussian pink noise, additive Gaussian white noise and Gaussian blur. Additive Gaussian white noise and Gaussian blur, each containing 4 to 5 distortion levels, for a total of 866 distorted images. The data set also provides the DMOS values for all the images tested, obtained from several subjective ratings by 25 testers, with the DMOS values ranging from [0, 1], with higher values indicating poorer image quality.

4.1.5 LIVE MD

The LIVE MD database [18] is the first database containing multiple distorted images created by the LIVE Lab at the University of Texas at Austin, USA. The images are combined into two types of distortion: JPEG compression and Gaussian blur, then white noise distortion is added. It contains 15 reference and 450 distorted images and provides a DMO for each distorted image, taking values in the range [0, 100].

4.1.6 LIVE In the Wild Image Quality Challenge Database

This is the LIVE dataset [12] of 1162 images from the Field Image Quality Challenge, created by the LIVE Lab at the University of Texas at Austin, USA. The types of distortion in these images are not computer simulations, but real images captured using various mobile cameras such as smartphones and tablets. During the imaging process, these images are subject to a variety of randomly occurring distortions and real photography artefacts. In order to ensure that the subjective quality scores of these images are objective, the researchers designed and implemented a new online crowdsourcing system in which 8100 observers conducted IQA and obtained 350,000 opinion scores, which were combined to produce an evaluation result.

4.1.7 KonIQ-10k

KonIQ-10k [36] was built at the University of Konstanz, Germany, to address the problem of too small the true distortion datasets. KonIQ-10k randomly selected approximately 10 million images from the large public multimedia dataset YFCC 100M [69] and then filtered 10,073 images in stages for use in building the dataset. The types of distortion present in these images included noise, JPEG artefacts, blending, lens motion blur, over-sharpening, etc. Based on the collected dataset, the researchers conducted a large-scale crowdsourcing experiment to obtain 1.2 million evaluations from 1467 workers, using statistical methods such as taking the mean value and removing extreme scores to determine the final MOS value. The image size was 1024 pixels by 768 pixels, and the MOS values ranged from [0, 5], with higher values indicating less distortion.

4.2 Performance Metrics

IQA has a number of performance metrics. PLCC, SROCC, KROCC and RMSE are by far the most used, as defined below.

4.2.1 PLCC(Pearson Linear Correlation Coefficient)

$$PLCC(Q_{est}, Q_{sub}) = \frac{cov(Q_{sub}, Q_{est})}{\sigma(Q_{sub})\sigma(Q_{est})}$$

where Q_{est} and Q_{sub} are the sets of predicted and actual subjective scores, respectively, $cov(\cdot)$ denotes the covariance between Q_{est} and Q_{sub} , and $\sigma(\cdot)$ means standard deviation. PLCC describes the correlation between the result of the algorithm and the subjective scoring by the human eye, and can measure the accuracy of the algorithm.

5 Conclusion

In this review, we paid attention to the work on audio synthesis and audio-visual multimodal processing. The text-to-speech(TTS) and music generation tasks, which plays a crucial role in the maintenance of the audio synthesis field, were comprehensively summarized respectively. In the TTS task, two-stage and end-to-end methods were distinguished and introduced. As for the music generation task, symbolic domain and raw audio domain generative models were presented respectively. In the field of audio-visual multimodal processing, we focused on four typical tasks: lipreading, audio-visual speech separation, talking face generation and sound generation from video. The frameworks related to these tasks were introduced. Finally, several widely adopted datasets were also presented. Overall, this review provides considerable guidance to relevant researchers.

5.0.1 SROCC(Spearman Rank-Ordered Correlation Coefficient)

$$SROCC(Q_{est}, Q_{sub}) = 1 - \frac{6 \sum(d_i)}{m(m^2 - 1)}$$

where d_i is the grade difference between the i_{th} sample of Q_{est} and Q_{sub} , and m is the number of images of the surrogate evaluation database species. The SROCC is primarily used to measure the monotonicity of the algorithm's predictions.

5.0.2 KROCC(Kendall Rank Order Correlation Coefficient)

$$KROCC = \frac{2n_c - n_d}{n(n - 1)}$$

Where n_c the number of consistent element pairs in the dataset; n_d is the number of inconsistent element pairs in the dataset. KROCC is also a good measure of the monotonicity of the algorithm.

5.0.3 RMSE(Root Mean Squared Error)

$$RMSE = [\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2]^{\frac{1}{2}}$$

where x_i is the subjective MOS value and y_i is the quality prediction score. RMSE is a direct measure of the absolute deviation between a person's subjective score and the algorithm's predicted score.

6 Conclusion and Overlook

This paper summarizes and reviews the basic concepts and classifications in the field of IQA, lists representative methods for each type of application, and focuses on the NR-IQA approach. NR-IQA has now become the focus of research. NR-IQA has a wide range of applications and is of significant research value, but it is also more difficult to study and requires more problems to be solved. The existing methods perform well on synthetic distortion datasets, but do not achieve good results on more challenging datasets such as LIVE Wild, KonIQ-10k and other real distortion datasets. As the distortion types are complex in most practical applications, it is difficult to generate corresponding distortion images using computer simulations, so there is more scope for realistic distortion-oriented IQA applications. There is still much scope for research into this type of evaluation task. In future research work, it is possible to expand the data set and design IQA models that are more robust to realistic distortion; to create performance metrics that are more consistent with the visual properties of the human eye, to make the evaluation results consistent with human subjectivity, etc.

References

- [1] Christos G Bampis, Praful Gupta, Rajiv Soundararajan, and Alan C Bovik. Speed-qa: Spatial efficient entropic differencing for image and video quality. *IEEE signal processing letters*, 24(9):1333–1337, 2017.
- [2] Christos George Bampis, Zhi Li, Anush Krishna Moorthy, Ioannis Katsavounidis, Anne Aaron, and Alan Conrad Bovik. Study of temporal effects on subjective video quality of experience. *IEEE Transactions on Image Processing*, 26(11):5217–5231, 2017.
- [3] Ali Borji and Laurent Itti. State-of-the-art in visual attention modeling. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):185–207, 2012.
- [4] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European Conference on Computer Vision*, pages 213–229. Springer, 2020.
- [5] Chao Chen, Mohammad Izadi, and Anil Kokaram. A perceptual quality metric for videos distorted by spatially correlated noise. In *Proceedings of the 24th ACM international conference on Multimedia*, pages 1277–1285, 2016.

- [6] Xiaoyu Chen, Hongliang Li, Qingbo Wu, King Ngi Ngan, and Linfeng Xu. High-quality r-cnn object detection using multi-path detection calibration network. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(2):715–727, 2020.
- [7] Xiaoyu Chen, Hongliang Li, Qingbo Wu, Fanman Meng, and Heqian Qiu. Bal-r2cnn: High quality recurrent object detection with balance optimization. *IEEE Transactions on Multimedia*, 2021.
- [8] Lark Kwon Choi and Alan Conrad Bovik. Video quality assessment accounting for temporal visual masking of local flicker. *Signal Processing: image communication*, 67:182–198, 2018.
- [9] Li Sze Chow and Raveendran Paramesran. Review of medical image quality assessment. *Biomedical signal processing and control*, 27:145–154, 2016.
- [10] Jisheng Ding, Linfeng Xu, Jiangtao Guo, and Shengxuan Dai. Human detection in dense scene of classrooms. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 618–622. IEEE, 2020.
- [11] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *CoRR*, abs/2010.11929, 2020. URL <https://arxiv.org/abs/2010.11929>.
- [12] 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.
- [13] Deepti Ghadiyaram and Alan C Bovik. Perceptual quality prediction on authentically distorted images using a bag of features approach. *Journal of vision*, 17(1):32–32, 2017.
- [14] Ke Gu, Guangtao Zhai, Xiaokang Yang, and Wenjun Zhang. A new reduced-reference image quality assessment using structural degradation model. In *2013 IEEE international symposium on circuits and systems (ISCAS)*, pages 1095–1098. IEEE, 2013.
- [15] Jiangtao Guo, Linfeng Xu, Jisheng Ding, Bin He, Shengxuan Dai, and Fangyu Liu. A deep supervised edge optimization algorithm for salt body segmentation. *IEEE Geoscience and Remote Sensing Letters*, 2020.
- [16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [17] Lihuo He, Fei Gao, Weilong Hou, and Lei Hao. Objective image quality assessment: a survey. *International Journal of Computer Mathematics*, 91(11):2374–2388, 2014.
- [18] Dinesh Jayaraman, Anish Mittal, Anush K Moorthy, and Alan C Bovik. Objective quality assessment of multiply distorted images. In *2012 Conference record of the forty sixth asilomar conference on signals, systems and computers (ASILOMAR)*, pages 1693–1697. IEEE, 2012.
- [19] Pengjun Ji, Yuchun Fang, Zhonghua Zhou, and Jiazen Zhu. Fusion of mssim and svm for reduced-reference facial image quality assessment. In *Chinese Conference on Biometric Recognition*, pages 75–82. Springer, 2012.
- [20] Weiping Ji, Jinjian Wu, Man Zhang, Zuozhi Liu, Guangming Shi, and Xuemei Xie. Blind image quality assessment with joint entropy degradation. *IEEE Access*, 7:30925–30936, 2019.
- [21] Stanley A Klein, Thom Carney, Lauren Barghout-Stein, and Christopher W Tyler. Seven models of masking. In *Human Vision and Electronic Imaging II*, volume 3016, pages 13–24. International Society for Optics and Photonics, 1997.
- [22] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.

- [23] Eric Cooper Larson and Damon Michael Chandler. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of electronic imaging*, 19(1):011006, 2010.
- [24] Dingquan Li, Tingting Jiang, and Ming Jiang. Quality assessment of in-the-wild videos. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 2351–2359, 2019.
- [25] Hui Li, Qingbo Wu, King Ngi Ngan, Hongliang Li, and Fanman Meng. Region adaptive two-shot network for single image dehazing. In *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2020.
- [26] Hui Li, Qingbo Wu, Haoran Wei, King Ngi Ngan, Hongliang Li, Fanman Meng, and Linfeng Xu. Haze-robust image understanding via context-aware deep feature refinement. In *2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6. IEEE, 2020.
- [27] Hui Li, Qingbo Wu, King Ngi Ngan, Hongliang Li, and Fanman Meng. Single image dehazing via region adaptive two-shot network. *IEEE MultiMedia*, 2021.
- [28] Leida Li, Weisi Lin, Xuesong Wang, Gaobo Yang, Khosro Bahrami, and Alex C Kot. No-reference image blur assessment based on discrete orthogonal moments. *IEEE transactions on cybernetics*, 46(1):39–50, 2015.
- [29] Qiang Li and Zhou Wang. Reduced-reference image quality assessment using divisive normalization-based image representation. *IEEE journal of selected topics in signal processing*, 3(2):202–211, 2009.
- [30] Wei Li, Qingbo Wu, Linfeng Xu, and Chao Shang. Incremental learning of single-stage detectors with mining memory neurons. In *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, pages 1981–1985. IEEE, 2018.
- [31] Wei Li, Hongliang Li, Qingbo Wu, Xiaoyu Chen, and King Ngi Ngan. Simultaneously detecting and counting dense vehicles from drone images. *IEEE Transactions on Industrial Electronics*, 66(12):9651–9662, 2019.
- [32] Wei Li, Hongliang Li, Qingbo Wu, Fanman Meng, Linfeng Xu, and King Ngi Ngan. Headnet: An end-to-end adaptive relational network for head detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(2):482–494, 2019.
- [33] Wei Li, Zhenting Wang, Xiao Wu, Ji Zhang, Qiang Peng, and Hongliang Li. Codan: Counting-driven attention network for vehicle detection in congested scenes. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 73–82, 2020.
- [34] Zhi Li, Anne Aaron, Ioannis Katsavounidis, Anush Moorthy, and Megha Manohara. Toward a practical perceptual video quality metric. *The Netflix Tech Blog*, 6(2), 2016.
- [35] Zhi Li, Christos Bampis, Julie Novak, Anne Aaron, Kyle Swanson, Anush Moorthy, and JD Cock. Vmaf: The journey continues. *Netflix Technology Blog*, 25, 2018.
- [36] Hanhe Lin, Vlad Hosu, and Dietmar Saupe. Koniq-10k: Towards an ecologically valid and large-scale iqa database. *arXiv preprint arXiv:1803.08489*, 2018.
- [37] Kwan-Yee Lin and Guanxiang Wang. Hallucinated-iqa: No-reference image quality assessment via adversarial learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 732–741, 2018.
- [38] Hantao Liu, Nick Klomp, and Ingrid Heynderickx. A no-reference metric for perceived ringing artifacts in images. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(4):529–539, 2009.
- [39] Xialei Liu, Joost Van De Weijer, and Andrew D Bagdanov. Rankiqa: Learning from rankings for no-reference image quality assessment. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1040–1049, 2017.

- [40] Hao Luo, Hanxiao Luo, Qingbo Wu, King Ngi Ngan, Hongliang Li, Fanman Meng, and Linfeng Xu. Single image deraining via multi-scale gated feature enhancement network. In *International Forum on Digital TV and Wireless Multimedia Communications*, pages 73–84. Springer, 2020.
- [41] Kunming Luo, Fanman Meng, Qingbo Wu, and Hongliang Li. Weakly supervised semantic segmentation by multiple group cosegmentation. In *2018 IEEE Visual Communications and Image Processing (VCIP)*, pages 1–4. IEEE, 2018.
- [42] Kede Ma, Wentao Liu, Tongliang Liu, Zhou Wang, and Dacheng Tao. dipiq: Blind image quality assessment by learning-to-rank discriminable image pairs. *IEEE Transactions on Image Processing*, 26(8):3951–3964, 2017.
- [43] Fanman Meng, Lili Guo, Qingbo Wu, and Hongliang Li. A new deep segmentation quality assessment network for refining bounding box based segmentation. *IEEE Access*, 7:59514–59523, 2019.
- [44] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing*, 21(12):4695–4708, 2012.
- [45] Anish Mittal, Michele A Saad, and Alan C Bovik. A completely blind video integrity oracle. *IEEE Transactions on Image Processing*, 25(1):289–300, 2015.
- [46] Pedram Mohammadi, Abbas Ebrahimi-Moghadam, and Shahram Shirani. Subjective and objective quality assessment of image: A survey. *arXiv preprint arXiv:1406.7799*, 2014.
- [47] Anush Krishna Moorthy and Alan Conrad Bovik. Visual importance pooling for image quality assessment. *IEEE journal of selected topics in signal processing*, 3(2):193–201, 2009.
- [48] Jincheol Park, Kalpana Seshadrinathan, Sanghoon Lee, and Alan Conrad Bovik. Video quality pooling adaptive to perceptual distortion severity. *IEEE Transactions on Image Processing*, 22(2):610–620, 2012.
- [49] G. Piella and H. Heijmans. A new quality metric for image fusion. In *Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429)*, volume 3, pages III–173, 2003. doi: 10.1109/ICIP.2003.1247209.
- [50] 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*, 10(4):30–45, 2009.
- [51] Nikolay Ponomarenko, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Lina Jin, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al. Color image database tid2013: Peculiarities and preliminary results. In *european workshop on visual information processing (EUVIP)*, pages 106–111. IEEE, 2013.
- [52] Ekta Prashnani, Hong Cai, Yasamin Mostofi, and Pradeep Sen. Pieapp: Perceptual image-error assessment through pairwise preference. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1808–1817, 2018.
- [53] Heqian Qiu, Hongliang Li, Qingbo Wu, Fanman Meng, King Ngi Ngan, and Hengcan Shi. A2rmnet: Adaptively aspect ratio multi-scale network for object detection in remote sensing images. *Remote Sensing*, 11(13):1594, 2019.
- [54] Heqian Qiu, Hongliang Li, Qingbo Wu, Fanman Meng, Linfeng Xu, King Ngi Ngan, and Hengcan Shi. Hierarchical context features embedding for object detection. *IEEE Transactions on Multimedia*, 22(12):3039–3050, 2020.
- [55] Heqian Qiu, Hongliang Li, Qingbo Wu, and Hengcan Shi. Offset bin classification network for accurate object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13188–13197, 2020.

- [56] Abdul Rehman and Zhou Wang. Reduced-reference image quality assessment by structural similarity estimation. *IEEE transactions on image processing*, 21(8):3378–3389, 2012.
- [57] Snjezana Rimac-Drlje, Mario Vranjes, and Drago Zagar. Influence of temporal pooling method on the objective video quality evaluation. In *2009 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, pages 1–5. IEEE, 2009.
- [58] Michele A Saad, Alan C Bovik, and Christophe Charrier. Blind prediction of natural video quality. *IEEE Transactions on Image Processing*, 23(3):1352–1365, 2014.
- [59] Kalpana Seshadrinathan and Alan C Bovik. Temporal hysteresis model of time varying subjective video quality. In *2011 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 1153–1156. IEEE, 2011.
- [60] Kalpana Seshadrinathan and Alan Conrad Bovik. Motion tuned spatio-temporal quality assessment of natural videos. *IEEE transactions on image processing*, 19(2):335–350, 2009.
- [61] Michael Seufert, Martin Slanina, Sebastian Egger, and Meik Kottkamp. “to pool or not to pool”: A comparison of temporal pooling methods for http adaptive video streaming. In *2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 52–57. IEEE, 2013.
- [62] Chao Shang, Qingbo Wu, Fanman Meng, and Linfeng Xu. Instance segmentation by learning deep feature in embedding space. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 2444–2448. IEEE, 2019.
- [63] Hamid R Sheikh and Alan C Bovik. Image information and visual quality. *IEEE Transactions on image processing*, 15(2):430–444, 2006.
- [64] Hamid R Sheikh, Muhammad F Sabir, and Alan C Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing*, 15(11):3440–3451, 2006.
- [65] Hengcan Shi, Hongliang Li, Fanman Meng, and Qingbo Wu. Key-word-aware network for referring expression image segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 38–54, 2018.
- [66] Hengcan Shi, Hongliang Li, Qingbo Wu, and King Ngi Ngan. Query reconstruction network for referring expression image segmentation. *IEEE Transactions on Multimedia*, 23:995–1007, 2020.
- [67] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [68] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [69] Bart Thomee, David A Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li. Yfcc100m: The new data in multimedia research. *Communications of the ACM*, 59(2):64–73, 2016.
- [70] Domonkos Varga. No-reference video quality assessment based on the temporal pooling of deep features. *Neural Processing Letters*, 50(3):2595–2608, 2019.
- [71] Domonkos Varga, Dietmar Saupe, and Tamás Szirányi. Deeprn: A content preserving deep architecture for blind image quality assessment. In *2018 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2018.
- [72] Phong V Vu and Damon M Chandler. Vis3: an algorithm for video quality assessment via analysis of spatial and spatiotemporal slices. *Journal of Electronic Imaging*, 23(1):013016, 2014.

- [73] Bo Wang, Zhibing Wang, Yupeng Liao, and Xinggang Lin. Hvs-based structural similarity for image quality assessment. In *2008 9th International Conference on Signal Processing*, pages 1194–1197. IEEE, 2008.
- [74] Lanxiao Wang, Chao Shang, Heqian Qiu, Taijin Zhao, Benliu Qiu, and Hongliang Li. Multi-stage tag guidance network in video caption. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 4610–4614, 2020.
- [75] Zhou Wang. Applications of objective image quality assessment methods [applications corner]. *IEEE signal processing magazine*, 28(6):137–142, 2011.
- [76] Zhou Wang and Eero P Simoncelli. Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In *Human vision and electronic imaging X*, volume 5666, pages 149–159. International Society for Optics and Photonics, 2005.
- [77] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pages 1398–1402. Ieee, 2003.
- [78] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.
- [79] Haoran Wei, Qingbo Wu, Hui Li, King Ngi Ngan, Hongliang Li, and Fanman Meng. Single image dehazing via artificial multiple shots and multidimensional context. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 1023–1027. IEEE, 2020.
- [80] Haoran Wei, Qingbo Wu, Hui Li, King Ngi Ngan, Hongliang Li, Fanman Meng, and Linfeng Xu. Non-homogeneous haze removal via artificial scene prior and bidimensional graph reasoning. *arXiv preprint arXiv:2104.01888*, 2021.
- [81] Jinjian Wu, Weisi Lin, Guangming Shi, Leida Li, and Yuming Fang. Orientation selectivity based visual pattern for reduced-reference image quality assessment. *Information Sciences*, 351:18–29, 2016.
- [82] Jinjian Wu, Jichen Zeng, Weisheng Dong, Guangming Shi, and Weisi Lin. Blind image quality assessment with hierarchy: Degradation from local structure to deep semantics. *Journal of Visual Communication and Image Representation*, 58:353–362, 2019.
- [83] Qingbo Wu, Hongliang Li, Fanman Meng, King N Ngan, Bing Luo, Chao Huang, and Bing Zeng. Blind image quality assessment based on multichannel feature fusion and label transfer. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(3):425–440, 2015.
- [84] Qingbo Wu, Zhou Wang, and Hongliang Li. A highly efficient method for blind image quality assessment. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 339–343. IEEE, 2015.
- [85] Qingbo Wu, Hongliang Li, King N Ngan, and Kede Ma. Blind image quality assessment using local consistency aware retriever and uncertainty aware evaluator. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(9):2078–2089, 2017.
- [86] Qingbo Wu, Lei Wang, King N Ngan, Hongliang Li, and Fanman Meng. Beyond synthetic data: A blind deraining quality assessment metric towards authentic rain image. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 2364–2368. IEEE, 2019.
- [87] Qingbo Wu, Li Chen, King Ngi Ngan, Hongliang Li, Fanman Meng, and Linfeng Xu. A unified single image de-raining model via region adaptive coupled network. In *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, pages 1–4. IEEE, 2020.
- [88] Qingbo Wu, Lei Wang, King Ngi Ngan, Hongliang Li, Fanman Meng, and Linfeng Xu. Subjective and objective de-raining quality assessment towards authentic rain image. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(11):3883–3897, 2020.

- [89] Feng Xiao, Joyce E. Farrell, and Brian A. Wandell. Psychophysical thresholds and digital camera sensitivity: the thousand-photon limit. In Nitin Sampat, Jeffrey M. DiCarlo, and Ricardo J. Motta, editors, *Digital Photography*, volume 5678, pages 75 – 84. International Society for Optics and Photonics, SPIE, 2005. URL <https://doi.org/10.1117/12.587468>.
- [90] Jingtao Xu, Peng Ye, Yong Liu, and David Doermann. No-reference video quality assessment via feature learning. In *2014 IEEE international conference on image processing (ICIP)*, pages 491–495. IEEE, 2014.
- [91] Xiaolong Xu, Fanman Meng, Hongliang Li, Qingbo Wu, Yuwei Yang, and Shuai Chen. Bounding box based annotation generation for semantic segmentation by boundary detection. In *2019 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pages 1–2. IEEE, 2019.
- [92] Qingsen Yan, Dong Gong, and Yanning Zhang. Two-stream convolutional networks for blind image quality assessment. *IEEE Transactions on Image Processing*, 28(5):2200–2211, 2018.
- [93] Longrong Yang, Hongliang Li, Qingbo Wu, Fanman Meng, and King Ngi Ngan. Mono is enough: Instance segmentation from single annotated sample. In *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, pages 120–123. IEEE, 2020.
- [94] Longrong Yang, Fanman Meng, Hongliang Li, Qingbo Wu, and Qishang Cheng. Learning with noisy class labels for instance segmentation. In *European Conference on Computer Vision*, pages 38–53. Springer, 2020.
- [95] Xiaohan Yang, Fan Li, Wei Zhang, and Lijun He. Blind image quality assessment of natural scenes based on entropy differences in the dct domain. *Entropy*, 20(11):885, 2018.
- [96] Xiaohan Yang, Fan Li, and Hantao Liu. A survey of dnn methods for blind image quality assessment. *IEEE Access*, 7:123788–123806, 2019.
- [97] Yuwei Yang, Fanman Meng, Hongliang Li, King N Ngan, and Qingbo Wu. A new few-shot segmentation network based on class representation. In *2019 IEEE Visual Communications and Image Processing (VCIP)*, pages 1–4. IEEE, 2019.
- [98] Yuwei Yang, Fanman Meng, Hongliang Li, Qingbo Wu, Xiaolong Xu, and Shuai Chen. A new local transformation module for few-shot segmentation. In *International Conference on Multimedia Modeling*, pages 76–87. Springer, 2020.
- [99] Junyong You and Jari Korhonen. Transformer for image quality assessment. *CoRR*, abs/2101.01097, 2021. URL <https://arxiv.org/abs/2101.01097>.
- [100] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. Fsim: A feature similarity index for image quality assessment. *IEEE transactions on Image Processing*, 20(8):2378–2386, 2011.
- [101] Lin Zhang, Lei Zhang, and Alan C Bovik. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*, 24(8):2579–2591, 2015.
- [102] Yu Zhou, Leida Li, Shiqi Wang, Jinjian Wu, Yuming Fang, and Xinbo Gao. No-reference quality assessment for view synthesis using dog-based edge statistics and texture naturalness. *IEEE Transactions on Image Processing*, 28(9):4566–4579, 2019.