Is the development of objective image quality assessment methods keeping pace with technological developments?

Preliminary Communication

Md. Abdur Rahman*
Hongik University,
Department of Electronic & Electrical Engineering
Mapo-gu, Seoul, Republic of Korea
mdabdur.rahman.1995@ieee.org

Hanif Bhuiyan
Monash University,
Monash Data Futures Institute
Clayton, Victoria, Australia
hanif.bhuiyan@monash.edu

*Corresponding author

Abstract – *Visual data proliferation with technological innovations is progressively moving forward, bringing automation, digitization, and smartness to various aspects of daily lives. These visual data-dependent technologies need to be assessed during their development to ensure the quality of services. Consequently, when appropriate quality assessment metrics are unavailable, the advancement of image processing technologies is hindered. This preliminary work addresses the disparity between the rapid advancement of image processing-based algorithms and the lag in developing image quality assessment (IQA) methods while briefly discussing future prospects. The shortcomings in IQA development are discussed by broadly categorizing application areas into human visual system (HVS)-based and computer vision-based, while highlighting the trends and necessity of developing application-specific IQA methods. Despite the existence of several emerging IQAs, this preliminary communication indicates significant opportunities for objective IQA developments in both application areas to support technological advancements. Hence, this work can serve academic and industrial communities by providing guidance for expediting IQA developments.*

Keywords: *image quality assessment (IQA), human vision, computer vision*

Received: January 13, 2024; Received in revised form: April 26, 2024; Accepted: April 26, 2024

1. INTRODUCTION

The widespread adoption of visual media is facilitated by an increasing number of smart devices and Internet of Things (IoT) based systems. The availability of visual media, such as images and videos, has paved the way for new applications and widened the aspects of the existing application areas in various sectors while offering more facilities for consumers. Quality evaluation metrics play a crucial role in ensuring the quality of services for these applications. Since humans are the primary customers or end-users of visual media in most cases, subjective quality evaluation is a well-known approach for evaluating the quality of images and videos during any process or application. Despite receiving ample acceptance, subjective evaluation is less desirable for real-time application due to its expense, time consumption, impracticality for real-time process, and implementational complexity [1, 2]. The difficulties associated with employing subjective metrics for image quality assessment (IQA) have led to the development of objective metrics [2, 3]. Also, even for a specific application, the subjective evaluation process is very complicated, making it challenging to ensure generalized quality assessments. The difficulty of employing subjective IQA arises from the fact that humans' responses to any visual media can vary depending on the environment of observations, the subject's age, display systems, and other factors. Consequently, although subjective metrics are used to validate objective metrics in most cases, objective metrics are usually applied when developing algorithms for optimizing them [3]. Hence, this work focuses on objective metrics to address the gap in IQA development compared to technological advancements.

The objective IQA metrics were initially developed based on mathematical algorithms, such as mean squared error (MSE) and peak signal-to-noise ratio (PSNR), to evaluate the difference between original/source/reference images and processed/modified/transformed images in terms of mathematical errors and signal fidelity, respectively. Though they gained popularity initially, their shortcomings in evaluating image quality became apparent due to their sole focus on the mathematical perspective. As the human visual system (HVS) perceives images in a more complex manner, this motivated the incorporation of HVS-related theories in developing objective IQA metrics. Till now, several objectives have been and are being developed considering spatial and frequency domain behaviors of HVS, which are summarized in existing review studies for a specific area as well as from a broad perspective [3-7].

Despite the availability of several theories to explain the behaviors of HVS, most IQA metrics consider some of those theories to resemble HVS when evaluating image quality. The IQAs are application-specific in most cases [8], which indicates the need for generalized methods to be implementable in any case, irrespective of application area. Additionally, it is essential to note that the source/original image is not always available as a reference for evaluating image quality. Based on the availability of reference images, objective IQAs can be categorized into full-reference (FR) IQA, reduced-reference (RR) IQA, and no-reference (NR) or blind IQA metrics.

As the developments of IQAs are always driven by the necessities or problems faced in different scenarios, they are hardly keeping up with the progressive developments of image processing algorithms. The image processing applications still need sophisticated IQAs, where the application areas of image processing can be broadly categorized into HVS-based and computer vision-based, considering the end-user. Irrespective of the IQA category, objective IQAs for HVS-based applications aim to ensure perceptual quality. In contrast, objective IQAs for computer vision-based applications prioritize meeting the machine-operational requirement to achieve predefined goals. The existing literature reviewing these objective metrics focuses on a single IQA category or multiple IQA categories in most studies. However, no study has adequately addressed the lag in IQA developments relative to technological advancements.

This preliminary study aims to highlight this issue by discussing the lag in IQA developments for different applications. As the selection of the category of objective metrics for an application typically depends on the application requirements and corresponding available resources, this study addresses the gaps in IQA developments without emphasizing any specific category. Although a detailed categorization of IQAs based on their application areas is feasible, this work discusses the gaps in IQA development by broadly categorizing the areas into two groups based on end-users, including HVS-based and computer vision-based areas. The contribution of this work is outlined as follows.

- A preliminary study is presented to investigate the trend of objective IQA development in response to technological advancements.
- Limitations in IQA developments in HVS-based and computer vision-based application areas are identified.
- Guidelines for future research are provided to harmonize IQA developments with technological advancements.

The rest of the manuscript is organized as follows. Sections 2 and 3 address the trends of the existing IQA development to keep pace with the technological innovations in each application area while pointing out some notable limitations. This work is concluded in Section 4 with a brief discussion of future prospects.

2. HVS-BASED APPLICATIONS

HVS-based applications are one of the primary motivators for developing various IQAs to ensure the quality of services for HVS. Considering the requirements of existing and emerging applications, several factors need to be accounted for when developing IQAs, as presented below.

2.1. DISTORTIONS OR ARTIFACTS

In HVS-based applications, IQAs are frequently employed to assess image quality across multiple stages. These stages encompass various processes, including image capture, compression, transmission, decompression, and enhancement, among others. Fig. 1 illustrates the image processing pipeline, from capture to display for end-users.

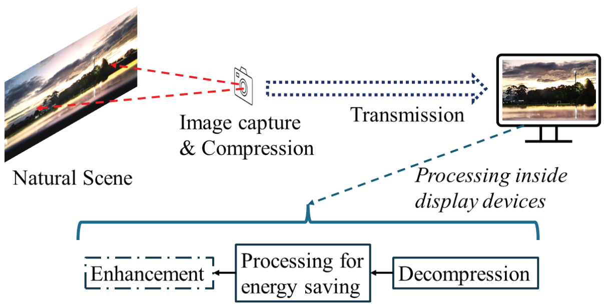

Fig. 1. A block diagram of image processing from source to end users.

Due to the possibility of different distortions during these steps, several conventional IQA algorithms have been developed, focusing on them. However, conventional IQAs may fall short in accurately assessing quality when multiple distortions are present simultaneously in an image. As a solution, multi-distortion IQA is developed [9]. Additionally, distortions or artifacts can be caused by various energy-saving technologies employed within the display devices or systems, presenting a challenge for existing IQAs. Although distortions stemming from external processes at various stages differ from internal distortions, IQAs capable of accounting for all these cumulative distortions are yet to be developed. Hence, the development of new IQAs is essential to effectively assess image quality while considering internal and external distortions.

On the other hand, the requirements of new applications and the problems in assessing the existing applications have created the urge to develop new methodologies. Several emerging IQAs, alongside conventional ones, have been discussed in [3]. While these applications-driven emerging IQAs are promising, they highlight the ongoing race in IQA developments to keep pace with new technologies. Moreover, there is a pressing need for new IQA methodologies, especially in areas involving distorted or degraded image processing. For example, dehazing IQA [10] is developed to evaluate image quality during the dehazing process, as a hazy source image has less information as a reference. A new NR IQA is discussed in [10], as conventional IQAs perform inadequately, and synthetic images are insufficient reference for this image reconstruction process. Similar to dehazing IQAs, proper IQAs are required to assess images enhanced from various sources, currently evaluated by NR IQAs or cumbersome subjective scores. For instance, images captured in night vision mode by the closed-circuit television (CCTV) often require enhancement for better visualization and monitoring. Despite its critical role in security and monitoring purposes, optimal IQAs to ensure service quality post-enhancement remain elusive.

2.2. CONTENT SPECIFICATION

Differences in image content compared to natural images necessitate the development of new IQAs for their evaluation. For instance, images with textual and pictorial information differ from conventional images, leading to the development of screen content IQA [11]. Another instance involves cartoon images, which prioritize structural and color features. Since conventional IQAs designed to evaluate the perceptual quality of natural images are insufficient for cartoon images, a new IQA tailored specifically for cartoon images is discussed in [12]. Additionally, remote sensing, an important area for image processing applications, requires good IQA to ensure the quality of service while monitoring remote regions. As the existing IQAs are found to be insufficient, a low-level and deep-level feature combination-based IQA has been discussed recently in [13] as a solution. Furthermore, retargeted image quality assessment (RIQA), an important branch of IQA for assessing the quality of content-aware image retargeting in multimedia applications for HVS, is still in the early stage, as addressed in [14]. As technological advancement is still in progress for this application, developing an appropriate IQA is essential. These content-oriented IQA developments address the limitations of existing IQAs and highlight the need for new IQAs tailored to application-specific content requirements for existing applications.

The trend of content-oriented new IQA developments is particularly pronounced in emerging technologies, notably in the realm of 3D visual content. One such example is the emergence of 360° images and videos. The progress in virtual reality (VR) has spurred the creation of 360° images and videos, necessitating new technologies to handle the spherical nature of these visual contents. As addressed in [15], new perceptual and deep learning-based methods (i.e., subjective assessment of multimedia panoramic video quality (SAMPVIQ), Craster parabolic projection PSNR (CPP-PSNR), viewport-based CNN (V-CNN) and others) are developed to ensure the quality of services for this emerging technology. Similarly, the rise of 3D visual media has spurred the development of stereoscopic/3D IQA [16]. Although 3D/stereoscopic IQAs are notable, they may not be suitable enough with new technologies used for 3D view generation. As the concept of stereoscopy has limitations in providing depth information and has vergence-accommodation conflict, multifocal displays [17] are emerging as a promising solution in generating 3D views for which new IQAs will be required.

2.3. SURROUNDING ENVIRONMENTS' EFFECTS

In addition to distortions and visual contents, surroundings that affect the sensitivity of HVS need to be considered for perceptual quality evaluations in different environmental situations. Despite being a critical factor affecting technological advancements, no IQA has incorporated this consideration to support existing and emerging applications. Backlight dimming is an example of an established application area where challenges arise from the absence of appropriate IQAs while developing new image processing algorithms. In [18], an energy-efficient dimming technology is discussed, which takes advantage of changes in HVS sensitivity with variations in viewing distance and ambient light. However, due to the lack of IQAs with surroundings consideration, image quality was assessed by capturing reference and processed images in the same environment for that research work. This approach proves to be inefficient and impractical, especially for large datasets.

Similar problems can be evident in emerging technologies, illustrated by the case of augmented reality (AR) or mixed reality (MR). While VR IQA [19] exists to evaluate the quality of VR content, it can be ineffective for AR or MR. As AR or MR technology merges the real world with the virtual world, virtual visual content needs to be clearly visible amidst real-world objects. An example of such a scenario is an automotive head-up display (HUD), where vital information is provided using transparent displays. Hence, new IQAs are required to evaluate the quality of such virtual content with consideration of the real world in the background. With the emergence of the metaverse concept, the importance of such IQAs can not be ignored.

Nonetheless, beyond the examples outlined in this section, there are several promising fields where the lack of proper IQAs hinders the growth of technologies to ensure the quality of their services while offering various advantages to human users.

3. COMPUTER VISION-BASED APPLICATIONS

Due to computer vision applications prioritizing accuracy over visual quality, various accuracy-oriented metrics are used due to the absence of appropriate IQAs. As the perception of visual media by HVS differs from that of computer vision, the performance of computer vision applications can not be ensured by perceptual quality resembling the perception of HVS. At the same time, accuracy metrics or mathematical algorithms are more suitable for such situations, as computer vision deals with factors like signal quality, data, and features in visual media. Pixel-wise accuracy, mean absolute error (MAE), MSE, and PSNR are examples of popular mathematical algorithms used in computer vision. However, these metrics may prove inadequate to evaluate application-wise requirements. In [20], an objective/goal-oriented IQA (GO-IQA) to support computer vision applications such as image segmentation has been discussed, pointing out the necessity of new IQAs. The research work on GO-IQA has notably addressed the ineffectiveness of focusing on perceptual quality for computer vision applications when the objectives significantly differ from those of HVS-based applications.

The necessity of task-oriented IQA development becomes evident when considering the broad application areas of computer vision. As illustrated in Fig. 2, computer vision tasks include image classification, image segmentation, object detection and recognition, as well as action and activity recognition, all of which find applications across various domains. The application areas of computer vision include education, healthcare, construction projects, commercial/industrial productions, agriculture, livestock operations, unmanned aerial vehicles, intelligent transportation systems, underwater activities, and so on. These tasks can suffer due to having images with noise, bad quality, inadequate contrast, or degraded structural/feature information. For these applications, whether the collected images are suitable for the desired task can be easily determined if appropriate IQAs are available. Furthermore, leveraging the scores of the appropriate IQAs, the pre-processing step can be invoked to process images according to application requirements. However, till now, most cases focus on distortion-free image collections with sufficient resolution, which indicates focusing on visual quality but may not always be useful, as addressed earlier.

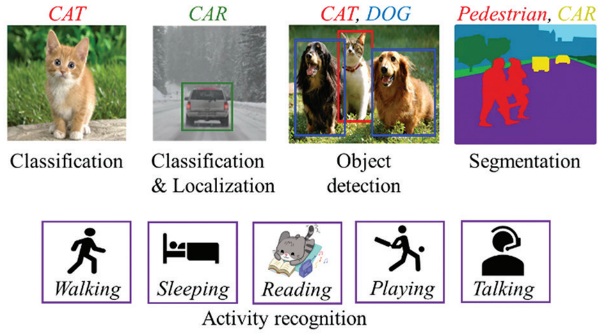

Fig. 2. Examples of different computer vision tasks

Another example highlighting the limitations of relying solely on perceptual IQAs is found in adversarial examples. As adversarial attacks introduce barely visible perturbations to images, HVS-based IQAs usually suggest good quality. However, these perturbations can significantly impact machine learning or deep learning models. If application-oriented IQAs are available for assessing quality from a computer vision perspective instead of HVS, they can identify the poor quality of corrupted images for that specific task. Similarly, steganography, which conceals important information within visual data while maintaining its invisibility, requires IQA to assess its imperceptibility. However, as addressed in [21], existing IQA methods appear inefficient for such an objective, indicating the necessity of appropriate IQA development. Furthermore, similar to HVS-based applications, computer vision-based applications require surrounding considerations. For example, unique metrics such as underwater image contrast measure and underwater image quality measure are suitable for evaluating images collected in a blueish aquatic environment [22].

New technological advancements in computer vision address the necessity for new IQAs. For instance, text-to-image generation, initiated from the urge to improve user experience and facilitate different sectors while presenting an economical solution for the content creators, requires appropriate IQAs. While objective NR IQAs may be used for evaluating artificial intelligence (AI) generated visual content, they often fall short as they primarily focus on real or natural scenarios during their development. Consequently, new IQAs tailored to artificial scenarios have emerged, including Fréchet inception distance (FID) [23] as an algorithmic approach and learned perceptual image patch similarity (LPIPS) [24] as a deep learning-based approach. Synthetic images are another example of AI-generated visual content widely used in computer vision, especially for training AI-based models. As synthetic images differ from natural images, conventional IQAs concerned with real-world images are inadequate. Moreover, as synthetic images are generated for computer vision-oriented tasks, modifications are required even for utilizing perceptual metrics [25].

All these discussions mentioned above highlight the challenges associated with using perceptual metrics for computer vision applications and signify the necessity of developing IQAs considering the specific requirements of computer vision. The discussions also indicate the attempts of IQA development to align with technological developments, even for computer vision-based applications.

4. DISCUSSIONS

Regardless of the application area, image processing applications often outpace the developments of IQAs. This issue can be a significant obstacle to technological innovations, leading to the use of subjective methods in most cases at the initial stage of any technological advancements. However, relying on subjective assessments is a time-consuming and expensive process, and it introduces the possibility of bias due to the personal preferences of the observers. The situation worsens when dealing with a large volume of images. In recent years, deep learning-based approaches have started to reduce the gap between IQA development and technological advancements. As deciding whether to opt for FR or NR IQA can be a concern for any application, a flexible deep learning-based IQA method is discussed in [26], offering the ability to switch between FR and NR IQA as needed. Deep learning-based approaches have also received notable attention for their success as NR IQAs [27], offering an alternative to the conventional complex modeling of IQAs required by theoretical perception-based schemes.

However, it is premature to envision predicting technological innovations and preemptively developing IQAs. Considering the requirement for sufficient datasets for both technological advancements and IQA developments, it is prudent to explore them concurrently whenever feasible. The simultaneous development of new technologies and corresponding IQAs can help to avoid the problems associated with the lack of proper IQAs and to ensure the effective and reliable performance of new technologies. Moreover, during the development of IQAs, it is essential to consider the characteristics of end-users, such as HVS or computer vision, to enhance their effectiveness. Developing a generalized IQA for both HVS and computer vision can be quite challenging. Therefore, one feasible solution is to focus on developing or modifying application-specific IQA development or modification can be one feasible solution. In addition to adhering to existing norms, the following issues can be considered in IQA development efforts.

- Is it tailored to a specific application?
- Does it account for all relevant distortions/artifacts?
- Are the distortions considered to be confined to their mathematical representations?
- Does it consider the sensitivity/responses of end-users (e.g., computer vision, HVS) from a broad perspective?
- Does it take relevant environmental factors into account?
- Is the available dataset comprehensive enough to cover all cases for verification?

As the aforementioned challenges are minimum requirements for IQA development, the research community always needs to find new factors affecting IQAs' effectiveness to keep up with technological advancements. AI-based approaches are promising in this regard due to their ability to learn. In [28], an interesting approach is presented for IQA development, where textual information is considered part of the quality assessment reference. This concept can be a potential pathway for future IQA development. In the future, a generalized AI-based IQA can be developed as a prospective solution, capable of taking text-based guidance or some critical factors into account to find appropriate principles and tune the model according to application requirements. This type of model can be considered as an NR IQA, where textual information is the reference for evaluating an image. The adaptivity of such IQA can be increased by incorporating the ability to extract requirements or guidance from pictorial references in the pre-processing stage to make the model applicable as an FR IQA. Nonetheless, considering the technological innovations, several sectors lack appropriate IQA methods. However, there are different potential ways for developing IQAs to accommodate these sectors, which are broader than the examples addressed in this work.

5. REFERENCES:

[1] H. Sheikh, A. Bovik, "Image Information and Visual Quality", IEEE Transactions on Image Processing, Vol. 15, No. 2, 2006, pp. 430-444.

[2] Z. Wang, "Applications of Objective Image Quality Assessment Methods [applications corner]", IEEE Signal Processing Magazine, Vol. 28, No. 6, 2011, pp. 137-142.

[3] G. Zhai, X. Min, "Perceptual Image Quality Assessment: A Survey", Science China Information Sciences, Vol. 63, 2020, pp. 1-52.

[4] W. Lin, M. Narwaria, "Perceptual image quality assessment: recent progress and trends", Visual Communications and Image Processing 2010, Vol. 7744, 2010, pp. 33-41.

[5] G. Zhai, W. Sun, X. Min, J. Zhou, "Perceptual Quality Assessment of Low-light Image Enhancement", ACM Transactions on Multimedia Computing, Communications, and Applications, Vol. 17, No. 4, 2021, pp. 1-24.

[6] W. Lin, W. Yuxuan, X. Lishi, C. Weiling, Z. Tiesong, W. Hongan, "No-reference quality assessment for

low-light image enhancement: Subjective and objective methods", Displays, Vol. 78, 2023, p. 102432.

[7] S. Cheng, H. Zeng, J. Chen, J. Hou, J. Zhu, K.-K. Ma, "Screen Content Video Quality Assessment: Subjective and Objective Study", IEEE Transactions on Image Processing, Vol. 29, 2020, pp. 8636-8651.

[8] A. George, S. J. Livingston, "A Survey on Full Reference Image Quality Assessment Algorithms", International Journal of Research in Engineering and Technology, Vol. 2, No. 12, 2013, pp. 303-307.

[9] K. Okarma, P. Lech, V. V. Lukin, "Combined Full-Reference Image Quality Metrics for Objective Assessment of Multiply Distorted Images", Electronics, Vol. 10, No. 18, 2021, p. 2256.

[10] X. Lv, T. Xiang, Y. Yang, H. Liu, "Blind Dehazed Image Quality Assessment: A Deep CNN-Based Approach", IEEE Transactions on Multimedia, Vol. 25, 2023, pp. 9410-9424.

[11] H. Yang, Y. Fang, W. Lin, "Perceptual Quality Assessment of Screen Content Images", IEEE Transactions on Image Processing, Vol. 24, No. 11, 2015, pp. 4408-4421.

[12] H. Chen, X. Chai, F. Shao, X. Wang, Q. Jiang, X. Meng, Y.-S. Ho, "Perceptual Quality Assessment of Cartoon Images", IEEE Transactions on Multimedia, Vol. 25, 2023, pp. 140-153.

[13] Y. Wang, G. Liu, L. Wei, L. Yang, L. Xu, "A Method to Improve Full-resolution Remote Sensing Pansharpening Image Quality Assessment via Feature Combination", Signal Processing, Vol. 208, 2023, p. 108975.

[14] B. Asheghi, P. Salehpour, A. M. Khiavi, M. Hashemzadeh, "A Comprehensive Review on Content-aware Image Retargeting: From Classical to State-of-the-art Methods", Signal Processing, Vol. 195, 2022, p. 108496.

[15] M. Xu, C. Li, S. Zhang, P. Le Callet, "State-of-the-Art in 360° Video/Image Processing: Perception, Assessment and Compression", IEEE Journal of Selected Topics in Signal Processing, Vol. 14, No. 1, 2020, pp. 5-26.

[16] K. Sim, J. Yang, W. Lu, X. Gao, "Blind Stereoscopic Image Quality Evaluator Based on Binocular Semantic and Quality Channels", IEEE Transactions on Multimedia, Vol. 24, 2022, pp. 1389-1398.

[17] T. Zhan, J. Xiong, J. Zou, S.-T. Wu, "Multifocal Displays: Review and Prospect", PhotoniX, Vol. 1, 2020, pp. 1-31.

[18] M. A. Rahman, J. You, "Human visual sensitivity based optimal local backlight dimming methodologies under different viewing conditions", Displays, Vol. 76, 2023, p. 102338.

[19] A. K. R. Poreddy, R. B. C. Ganeswaram, B. Appina, P. Kokil, R. B. Pachori, "No-Reference Virtual Reality Image Quality Evaluator Using Global and Local Natural Scene Statistics", IEEE Transactions on Instrumentation and Measurement, Vol. 72, 2023, pp. 1-16.

[20] S. Kiruthika, V. Masilamani, "Goal Oriented Image Quality Assessment", IET Image Processing, Vol. 16, 2022, pp. 1054-1066.

[21] De R. I. M. Setiadi, S. Rustad, P. N. Andono, G. F. Shidik, "Digital Image Steganography Survey and Investigation (Goal, Assessment, Method, Development, and Dataset)", Signal Processing, Vol. 206, 2023, p. 108908.

[22] N. V. Dharwadkar, A. M. Yadav, M. A. Kadampur, "Improving the Quality of Underwater Imaging using Deep Convolution Neural Networks", Iran Journal of Computer Science, Vol. 5, No. 2, 2022, pp. 127-141.

[23] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter. "GANs trained by a two time-scale update rule converge to a local nash equilibrium", Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4-9 December 2017, pp. 6629-6640.

[24] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18-23 June 2018, pp. 586-595.

[25] J. Chaudhary, D. R. Pant, S. Pokharel, J.-P. Skön, J. Heikkonen, R. Kanth, "Image Quality Assessment by Integration of Low-level & High-level Features:

Threshold Similarity Index", Proceedings of the IEEE 31st International Symposium on Industrial Electronics, Anchorage, AL, USA, 1-3 June 2022, pp. 135-141.

[26] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, W. Samek, “Deep Neural Networks for No-reference and Full-reference Image Quality Assessment”, IEEE Transactions on Image Processing, Vol. 27, No. 1, 2017, pp. 206-219.

[27] X. Yang, F. Li, H. Liu, "A Survey of DNN Methods for Blind Image Quality Assessment", IEEE Access, Vol. 7, 2019, pp. 123788-123806.

[28] Y. Watanabe, R. Togo, K. Maeda, T. Ogawa, M. Haseyama, "Assessment of Image Manipulation using Natural Language Description: Quantification of Manipulation Direction", Proceedings of the IEEE International Conference on Image Processing, Bordeaux, France, 16-19 October 2022, pp. 1046-1050.