

# Quality Evaluation of Contrast-Enhanced Images: Central Asian Perspectives

Altynay Kadyrova

Department of Computer Science, KIMEP University, Almaty, Kazakhstan  
E-mail: a.kadyrova@kimep.kz

Marius Pedersen

Department of Computer Science, Norwegian University of Science and Technology, Gjøvik, Norway

---

**Abstract.** Culture can play a significant role in evaluating image quality. Therefore, this work considered one of the least studied cultural regions of observers, examining the impact of Central Asian culture on image quality evaluation. More specifically, it investigated how they evaluate the quality of contrast-enhanced images. It was found that observer evaluations vary and can be divided into groups. These groups may have their individual preferences for the quality of contrast-enhanced images. Therefore, the personalization factor should be incorporated into the quality evaluation of (contrast-) enhanced images. Furthermore, the results were compared with another population and differences were found in the overall outcomes of the two observer groups. The variations observed could be due to cultural differences. In addition, this study introduced the Central Asian Contrast-Enhanced Image Quality Dataset (CACEIQD). A variety of image quality metrics, including deep learning techniques, were tested on the dataset. The results indicate that the dataset is challenging and highlight an area for metric improvement. This dataset can be helpful for future research in the field of enhanced image quality evaluation.

**Keywords:** image quality, image enhancement, contrast, culture, metrics

© 2026 Society for Imaging Science and Technology.

[DOI: 10.2352/J.ImagingSci.Technol.2026.70.2.020403]

---

## 1. INTRODUCTION

Image quality has been widely studied across various domains, including computer graphics, color reproduction, material appearance, and printing. Its evaluation depends on multiple attributes—such as color, gloss, and naturalness—and can be influenced by external factors such as illumination, viewing distance, sample geometry, and cultural background. Among these factors, culture has been shown to play a significant role, as observers from different cultural groups often interpret and judge image quality differently [1].

Despite increasing interest in cross-cultural differences, most existing image quality studies rely on observers from Western countries, East Asia, or South Asia. Some recent online studies do not report demographic information at all. As a result, one major cultural region—Central Asia—remains largely absent from the literature, with only a very small number of observers from this region included in published

image quality evaluations. This lack of representation is particularly evident in studies that focus on enhanced images, where cultural differences may influence how improvements or artifacts are perceived.

The Central Asian region encompasses ethnicities from Kazakhstan, Turkmenistan, Uzbekistan, Tajikistan, and Kyrgyzstan. We hypothesize that these populations have been significantly underrepresented in prior work, and therefore current conclusions about perceived image quality or preferred enhancement levels may not generalize to this cultural group. To address this gap, our goal is to investigate how observers from Central Asia evaluate image quality, with particular emphasis on their perception of enhanced images. We introduce what is, to the best of our knowledge, the first dataset dedicated to enhanced image quality ratings from Central Asian observers. This dataset expands the cultural diversity of existing resources and enables researchers to draw more inclusive and representative conclusions about image enhancement. Furthermore, we examine whether distinct subgroups exist within the Central Asian observer population. We also compare the Central Asian observer results with another population, Norwegians, to find whether there are differences between different populations. Finally, we investigate whether existing image quality metrics are able to predict the judgments of the observers.

The structure of this paper is as follows. We first review related work on image enhancement, overenhancement, quality evaluation of enhanced images, and cultural influences. We then describe our methodology before presenting our results and discussion. Finally, we conclude with a summary and outline future research directions.

## 2. RELATED WORKS

### 2.1 Image Enhancement Techniques

Images can be distorted or enhanced. There are many works dedicated to image quality assessments via image distortions and degradation [2–8] compared to image enhancement. Image enhancement is commonly referred to as a processing step that can improve the quality of an image [9].

There are works that review image enhancement techniques [10–12]. Liu et al. [10] covered previous surveys, existing classification of image enhancement techniques,

---

Received Aug. 1, 2025; accepted for publication Mar. 17, 2026; published online Apr. 24, 2026. Associate Editor: Chunhui Kuo.

1062-3701/2026/70(2)/020403/11/\$25.00

and current enhanced image databases. They suggested some perspectives on the development of future image enhancement techniques. They also discussed the challenges related to image enhancement. Their proposed classification of image enhancement methods is as follows: image contrast enhancement, image sharpness enhancement, image color correction, image de-artifacting, and image enhancement for multiple quality attributes.

They observed that newly created enhancement approaches tend to use machine learning frameworks and consider a human visual system. Their conclusion was that it is desirable to come up with a universal Image Quality Metric (IQM) and databases for performance assessment of image enhancement approaches.

A year earlier, Qi et al. [11] had conducted a survey of image enhancement methods in terms of three aspects: unsupervised methods, supervised methods, and quality evaluation. Similar to Liu et al. [10], they concluded that deep learning based methods are the dominant models.

Low-light image enhancement techniques can be applied to images captured under poor illumination conditions to enhance the visual effect of such images [12]. Wang et al. [12] classified such techniques into seven categories: gray transformation, histogram equalization, Retinex, frequency-domain, image fusion, defogging model, and machine learning methods. They concluded that the selection of the suitable image enhancement algorithm is application dependent.

Image enhancement techniques have a wide range of application areas that signify their demand. For example, image enhancement techniques can be used for underwater [13–16], medical [17–19], satellite [20–22], and natural images [23, 24], among others. Next, we discuss overenhancement of images that could happen in practice.

## 2.2 Image Overenhancement

It is important to highlight the term overenhancement. When using image enhancement methods, overenhancement can occur, and therefore the quality of images can decrease. Nonetheless, a study by Azimian et al. [25] showed that their observers had agreed on being able to detect when overenhancement occurs. Their Subjective Enhanced Image Dataset (SEID) contains 30 reference images, and the contrast stretching technique was applied to produce high- and low-contrast versions of the reference images. Fifteen observers participated in their experiment, and ethnicity information of the observers was not mentioned.

To find whether the images were enhanced well or overenhanced, quality evaluation needs to be performed. Therefore, we next provide the literature on quality evaluation of images.

## 2.3 Quality Evaluation

The quality assessment of enhanced images is considered a challenging task [26, 27]. It can be conducted subjectively via psychophysical experiments and/or objectively via IQMs. In the case of subjective evaluation, images (e.g., enhanced)

are judged by observers in a controlled (e.g., in a lab) or uncontrolled (e.g., on field, online) environment, thereby producing subjective data. In the case of objective evaluation, existing IQMs can also evaluate the same images judged by the observers, thereby producing objective data. Eventually, IQM scores and subjective scores of observers can be checked for correlation.

The IQMs applied to distorted images prevail over those applied to enhanced images [26]. Hence, IQMs are typically designed to assess image distortion, and there are fewer methods to assess image enhancement [28]. Amirshahi et al. [29] evaluated the performance of IQMs on contrast-enhanced images. Twenty-eight IQMs were evaluated to check their suitability to assess contrast-enhanced images. Their dataset contains 26 original images, and four contrast enhancement methods (Retinex, s-shaped contrast correction, Contrast Limited Adaptive Histogram Equalization [CLAHE], and Natural Rendering of Color Image using Retinex) were used. A paired comparison method was used to judge the quality of the images by 15 observers under dark-room conditions. The ethnicity of the observers is not mentioned in this work. Overall, they found that the tested IQMs did not correlate well with the perceived contrast-enhanced image quality. This again proves that current IQMs that are mostly designed for distorted images are incapable of working with enhanced images. Gu et al. [30] highlighted that such IQMs did not yield satisfactory results when applied to enhanced images.

Five attributes (brightness, contrast, saturation, sharpness, and warmth) were used to enhance 16 natural color images in the dataset introduced by Kadyrova et al. [31]. They conducted an online experiment to collect subjective scores on enhanced images using a forced-choice paired comparison method. They had 45 observers; however, their ethnicity is not mentioned. They tested 38 IQMs on their dataset images and concluded that it is a difficult task for IQMs to process enhanced images.

Either sharpness, contrast, brightness, and color or their combination was edited to enhance 26 color images that were used in an experiment conducted in a dark room by Vu et al. [32]. In their study, nine observers (ethnicity is not mentioned) judged the quality of enhanced images by employing pairwise comparison and multiple-stimulus continuous quality evaluation paradigms. They used three full-reference IQMs in two modes: first mode—original image was input as the reference; second mode (reverse mode)—enhanced image was input as the reference. The results revealed that applying tested IQMs in reverse mode can improve enhanced image quality evaluation.

Qureshi et al. [33] created the Contrast Enhancement Evaluation Database (CEED2016) consisting of 30 original color images. They used six contrast enhancement methods—Adaptive Edge Based Contrast Enhancement, CLAHE, Discrete Cosine Transform, Global Histogram Equalization, Top Hat Transformation, and Multiscale Retinex—and six contrast metrics. They adapted the pairwise preference based ranking protocol (Condorcet

method), and 15 observers participated in their experiment under laboratory conditions. They mentioned that their observers were of different genders, age groups, and backgrounds. However, it is not clear what exactly they mean by different backgrounds. Hence, ethnicity information appears to be absent in this study. Their results showed that some of the metrics tested are inconsistent with subjective scores of the observers.

The Underwater Image Enhancement Benchmark with 950 real-world underwater images was created by Li et al. [34]. They employed 12 image enhancement methods and conducted a paired comparison experiment with 50 observers (ethnicity is not mentioned). Moreover, they proposed an underwater image enhancement network (called Water-Net). Cherepkova et al. [35] found that there can be individual differences in contrast preferences of natural images between observers. In addition, they mentioned that these individual differences in contrast preferences should be considered in image quality evaluations, image enhancement, and related fields. They had 22 observers (ethnicity is not mentioned) and used a Three-Alternative Forced Choice procedure with a modified adaptive staircase algorithm.

In studies where ethnicity information is not mentioned, we assume that the authors did not collect it because, for example, they did not find it important for their study. Furthermore, we assume that Central Asian observers were not present at all or were present in a very limited number in existing studies based on the location information of the articles.

Therefore, we next provide the literature on culture.

## 2.4 Culture

We used a narrowed definition for culture similar to that by Senthilkumar et al. [36]: culture is determined by geopolitical boundaries (e.g., countries, continents).

In 2006, Aslam [37] stated that most of the works have a Western focus. There have also been several studies showing differences between cultures.

Color is one of the most important attributes in imaging applications. There are works that show that considerable differences exist between cultures in color preferences [38–40]. There are also considerable differences between cultures in terms of color semantics [37, 41, 42].

Color emotion was evaluated in a set of countries (Argentina, Spain, Sweden, France, Germany, and others) via psychophysical experiments using semantic scales: heavy–light, warm–cool, active–passive, and like–dislike [43]. Argentinian observers' responses differed from others in the like–dislike scale. Argentinians preferred passive color pairs more than others based on factor analysis. Ou et al. [43] noted that the effects of gender, age, and professional background are also present along with cultural differences.

There have been works related to image quality and the impact of culture. Lin and Patterson [1] found a difference between Taiwanese and American subjects when assessing the image quality of mobile devices.

In a study that investigated the differences in mobile display color appearance, Europeans preferred a lower color temperature than Asians over the entire range of illuminants that they tested [44]. This study provided a cultural-sensitive approach via their two regression equations (one for Europeans and the other for Asians) to improve the appearance of products. In this way, mobile displays can acquire accurate colorimetric reproduction of images, which in turn can positively impact the image quality process.

Fernandez et al. [45] found that the cultural background of observers causes preference variability, and it was demonstrated to be statistically significant in color preference reproduction. They defined colorimetric adjustment dimensions by combining five (hue naturalness, mid-tone lightness accuracy, mid-tone detail, image naturalness, mid-tone chroma correctness) of the most important image or color quality terms based on the authors' expertise. Their gamma and chroma adjustment dimensions showed the most considerable preference variation between cultures. For instance, lighter image reproduction was preferred by Japanese while Americans preferred slightly less chromatic reproduction than others. They concluded that differences are present between cultures for some color reproduction preferences.

A recent study by Saupe and Pin [46] explored differences at the national level in quality evaluation using crowd-sourced datasets that contain responses from Japan, Serbia, Venezuela, Russia, India, the USA, and Brazil. They found considerable cross-cultural variations in terms of rating behavior.

It is worth noting the extreme response style where a group of observers tend to select the most extreme option on the scale. For example, Americans tend to select extreme options in comparison to those from East Asian countries [47].

In summary, it can be clearly seen that the existing work did not focus on recruiting observers with diverse ethnicities or cultural backgrounds in the quality evaluation of enhanced images. As a result, there is a demand for this study, which focuses on the Central Asian population for enhanced image quality evaluation.

## 2.5 Gap and Motivation

Across enhancement evaluation studies, cultural representation is negligible: participant ethnicity is often unreported, and Central Asian observers appear absent or minimal. At the same time, cross-cultural work strongly suggests that preferences and rating styles vary by culture. Therefore, conclusions about enhanced image quality drawn from Western/East/South Asian samples may not generalize to Central Asia.

The problem we address in this work is “How do observers from Central Asia perceive the quality of enhanced images, and how well do existing IQMs perform in predicting their judgments?”

This motivates our contribution: we introduce (i) a dedicated dataset of enhanced image quality scores from Central Asian observers, (ii) analysis benchmarking IQMs on

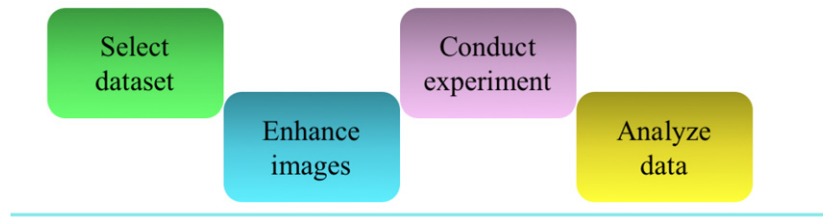


Figure 1. Our methodology workflow. The steps proceed from left to right.

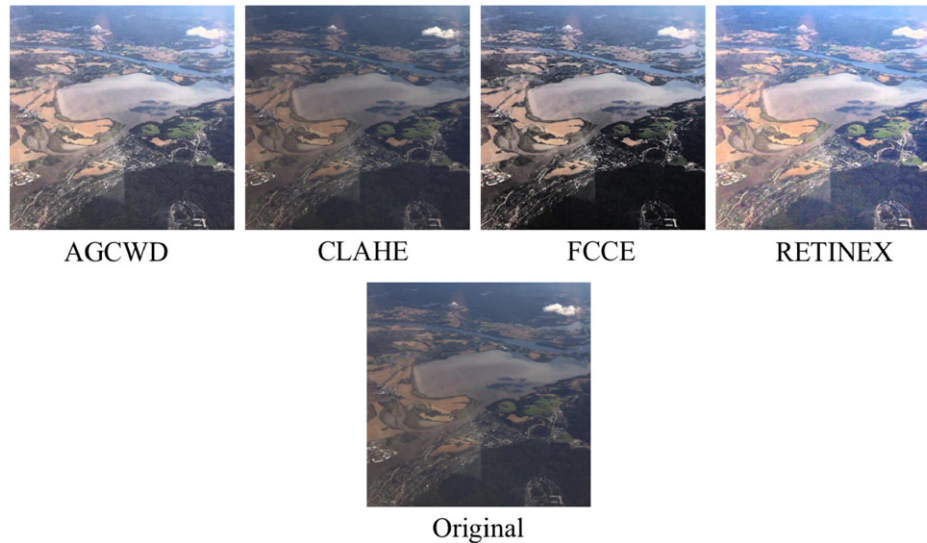


Figure 2. Image enhancement methods: original image (bottom) and four enhanced versions (top).

the collected subjective scores, and (iii) comparative analysis of Central Asian observer data with respect to another population.

### 3. METHODOLOGY

Our workflow is illustrated in Figure 1. We first start with the relevant dataset selection. Afterwards, image enhancement methods are applied to enhance the images. Next, a psychophysical experiment is conducted, and we further analyze the data.

As there are already existing datasets focused on image enhancement, we chose to work with the original images from the SEID [25] dataset. This dataset, ‘Central Asian Contrast-Enhanced Image Quality Dataset (CACEIQD)’, is a mixture of images from the following two datasets: CEED [48] and the Colourlab Contrast Enhanced Image Dataset [29]. In this way, the authors of the SEID dataset aimed to have diversity in the dataset in terms of image colorfulness and visual contents.

We applied several image enhancement methods on the original images (the original images had a resolution of  $512 \times 512$  pixels). Adaptive Gamma Correction with Weighting Distribution (AGCWD) [49] was selected because it produces enhanced images of higher quality as demonstrated by experimental results. The AGCWD method enhances the contrast of images and improves the brightness through

gamma correction and probability distribution of luminance pixels. The CLAHE [50] and Retinex [51] methods were selected because they are among the most common methods for image enhancement. Fuzzy-Contextual Contrast Enhancement (FCCE) [52] was chosen as it preserves the natural characteristics of the image and enhances contrast. Figure 2 demonstrates the image enhancement methods used.

Compared to existing datasets focusing on contrast enhancement, namely [29, 48, 51], the inclusion of FCCE and AGCWD is new, as they overlap with CLAHE and Retinex.

In Figure 3, we show contrast levels in the images using the RAMMG metric [53]. The contrast is enhanced in all the 30 scenes compared to the original image.

After the images were enhanced, we conducted a psychophysical experiment with 30 observers (12 males, 18 females) with an average age of 24.5 years. The observers had normal color vision. A Snellen chart and an Ishihara test were used to check their visual acuity and color vision, respectively. The recruited Central Asian observers were Kazakhs except for four observers (three from Tajikistan, one from Kyrgyzstan). All observers can be considered non-experts (i.e., without previous experience in image quality). Compared to existing datasets, which have observers of mixed background or have not stated the background of the observers, our dataset, to the best of our knowledge, is the only one with a majority of Central Asian observers.

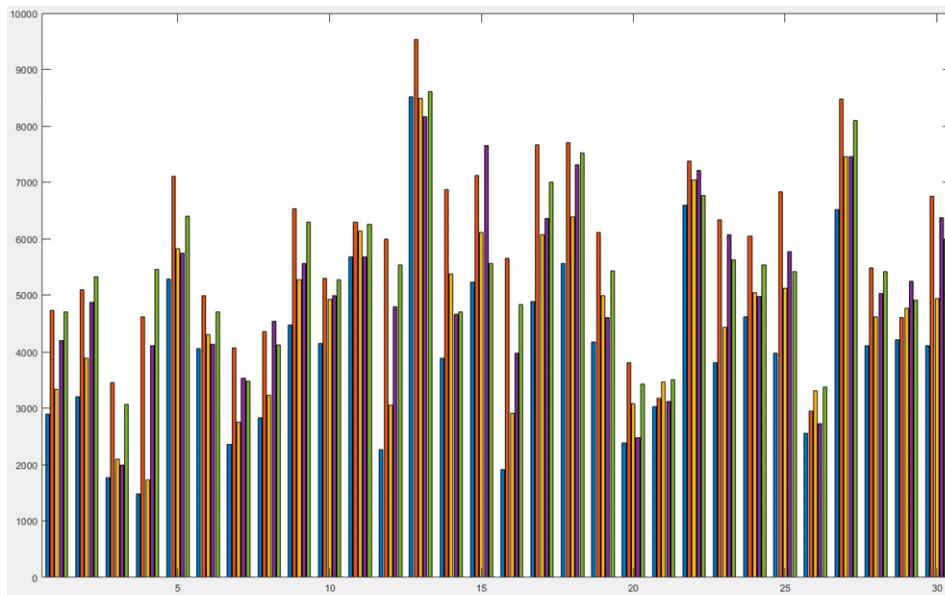


Figure 3. The contrast levels in the images via RAMMG metric. Blue = original, orange = AGCWD, yellow = CLAHE, purple = FCCE, and green = Retinex. The Y-axis shows RAMMG metric values (higher values indicate greater contrast).

Before starting the experiment, consent was obtained from the observers. The experiment was conducted in a dark room (AOC 24 LCD Monitor) with the following instruction: “Please choose the image with the highest quality.” Quick-Eval [54], a web platform, was used for the experiment. We chose to conduct the experiment with the paired comparison method as it was considered to be the easiest for the observers. The original images were included in the dataset. The observers were not informed that the original images were included and that contrast was varied to prevent potential bias.

The distance between the observer’s eyes and the monitor was around 50 cm. There was no time restriction, and the average duration was approximately 17 minutes per observer.

Furthermore, to compare the results of the Central Asian observers, we have conducted an additional experiment with eight Norwegian observers. The experiment was carried out under similar conditions to that with Central Asian observers but on a Dell U2419HC Monitor. Results from the two experiments are compared.

We process the data from the experiment to  $z$ -scores [55]. Relevant IQMs are tested for correlation between observer data and IQM data.

## 4. RESULTS AND DISCUSSION

### 4.1 Subjective Scores of Observers

Figure 4 shows the  $z$ -scores [55] of all images for four enhancement methods and the original, plotted with a 95% confidence interval according to Montag [56]. The higher the  $z$ -score, the higher the quality. When considering individual pictures, images enhanced by different methods including the original were ranked as having higher and lower quality in different scenes. However, when considering all the image scenes together, the images enhanced by CLAHE were rated

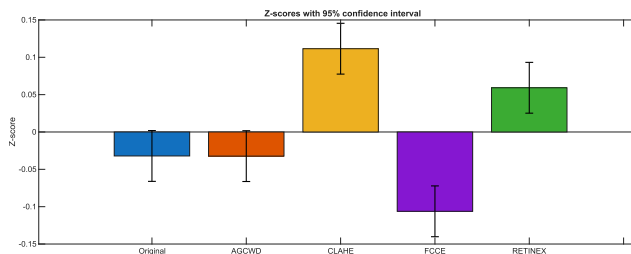


Figure 4. The  $z$ -scores for all image scenes together for Central Asian observers. The scores are plotted with a 95% confidence interval.

the highest quality followed by Retinex. The original images were not rated the highest quality while FCCE-processed images were ranked as having the lowest quality (Fig. 4). The FCCE-enhanced images were rated the lowest quality probably due to overenhancement and related artifacts. During contrast enhancement, artifacts such as halo effects, ringing, blocking, and color shift might appear.

We have also used the Bradley–Terry model in the overall frequency matrix, which is a probability model for the outcome of pairwise comparisons between items. We report the ability ( $\beta$ ) of an item to win in a paired comparison. Table I indicates that CLAHE has the highest value, consistent with the  $z$ -score plot in Fig. 4. Furthermore, we calculate the pairwise significance matrix ( $p$ -values, Table II), which shows that there is a statistically significant difference between CLAHE and the other methods.

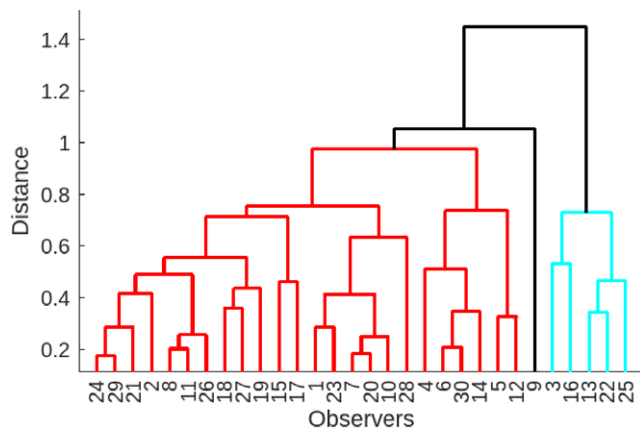
Hierarchical clustering (unweighted average distance was used to calculate distances between the clusters with the Euclidean distance metric) revealed three clusters of observers (Figure 5). Observer 9 evaluated the quality of the images considerably differently from the others.

**Table I.** Bradley–Terry  $\beta$  values. Higher value indicates better performance.

Method	$\beta$
CLAHE	0.17709
Retinex	0.094622
Original	-0.049994
AGCWD	-0.051779
FCCE	-0.16994

**Table II.** Pairwise significance matrix ( $p$ -values).

	Original	AGCWD	CLAHE	FCCE	Retinex
Original	—	0.960	1.265e-10	0.001	4.096e-05
AGCWD	0.960	—	9.065e-11	0.001	3.287e-05
CLAHE	1.265e-10	9.06e-11	—	0	0.020
FCCE	0.001	0.001	0	—	6.861e-14
Retinex	4.095e-05	3.287e-05	0.020	6.861e-14	—

**Figure 5.** The hierarchical clustering that shows how similarly observers evaluated quality of images. The higher the distance, the more dissimilar the evaluation.

Additionally, observers 24 and 29, 8 and 11, 7 and 20, and 6 and 30 evaluated images more or less in the same way, as the heights of distances are smaller in the cluster marked in red. Observers 13 and 22 also evaluated images in a similar way in the cluster marked in blue.

Although images enhanced by CLAHE followed by Retinex were ranked the highest quality when considering all image scenes together, this was not the case when the results of individual observers were analyzed.

From the dendrogram (Fig. 5) and the results of evaluations of each observer for all images, we can divide observers into groups. One group (blue cluster) perceived images enhanced with AGCWD (followed by Retinex) as better quality whereas those with CLAHE and FCCE as lower quality. We can assume that observers in the blue cluster

rated brighter (due to AGCWD) images as better quality and darker (due to FCCE) images as lower quality. Senior (13 and 22) and younger (3 and 16) observers also provided similar evaluations. The observers in the blue cluster are all females except one.

In contrast to the blue cluster, another group (subset 1 of the red cluster—observers 4, 6, 30, 14, 5, and 12) perceived images enhanced with AGCWD and Retinex as lower quality. Observers in subset 2 of the red cluster (observers 1, 23, 7, 20, 10, and 28) perceived images enhanced with FCCE as better quality and the original as lower quality. Subset 1 has one male observer, and subset 2 of the red cluster has two male observers.

The remaining observers in the red cluster (subset 3) did not show a clear pattern or find significant differences. It is worth mentioning that subset 3 of the red cluster contains eight males and four females. The only observer (9, female) in the black cluster found the original images as better quality unlike subset 2 of the red cluster.

In this light, it seems that individual preference for the perceived quality of contrast-enhanced images can be an unavoidable factor. This emphasizes that such individual preferences should be considered in image quality evaluations. This is in line with the findings of Cherepkova et al. [35] that there may be individual differences in contrast preferences among observers.

Moreover, we can propose that there may be differences between genders in contrast-enhanced image quality evaluations among Central Asian observers based on the results. To test this assumption, more work is needed that focuses particularly on gender in the evaluation.

Analyzing each image in the experiment, we note that in only three scenes (numbers 10, 13, and 30) the original has the highest  $z$ -score (Figure 6, plotted with 95% confidence intervals calculated according to Montag [56]). For example, Figure 7 shows the original and enhanced versions of image 30. Based on the RAMMG metric values in Fig. 3, the original version of image 30 has a lower level of contrast compared to the enhanced versions, which can also be perceived from Fig. 7. Image 10 follows a similar pattern while for image 13, the version enhanced via FCCE has a slightly lower contrast compared to the original version.

It is also worth noting here that the difference between the original and one or more of the enhanced images is small. In a few other enhanced images, the differences are clearly greater from the original. We can also observe that for some images, enhancement can significantly decrease quality, such as in images 9 and 12. It is also apparent that none of the enhancement techniques provide the best results for all images.

More in-depth analysis of the enhanced images, such as image 9 (Figure 8), reveals that AGCWD is capable of enhancing details in the shadow region while FCCE exhibits the opposite behavior and does not enhance shadow details. Observers have been highly consistent in their ratings for images with regard to FCCE and AGCWD. A similar observation is made for images 3, 12, and 14, where

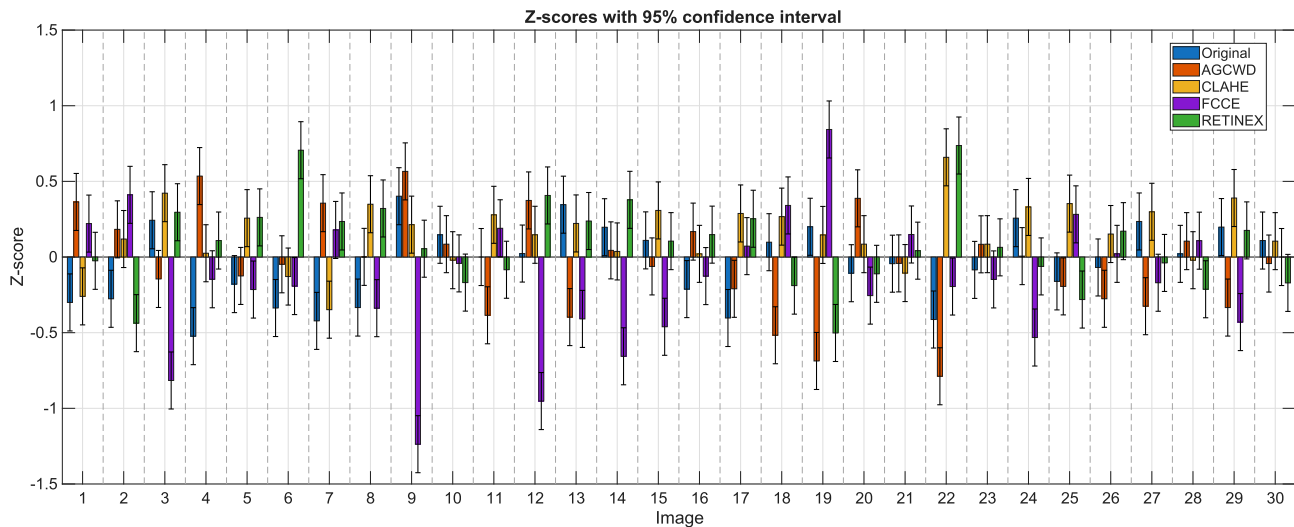


Figure 6. The z-scores for 30 images with 95% confidence intervals for Central Asian observers.

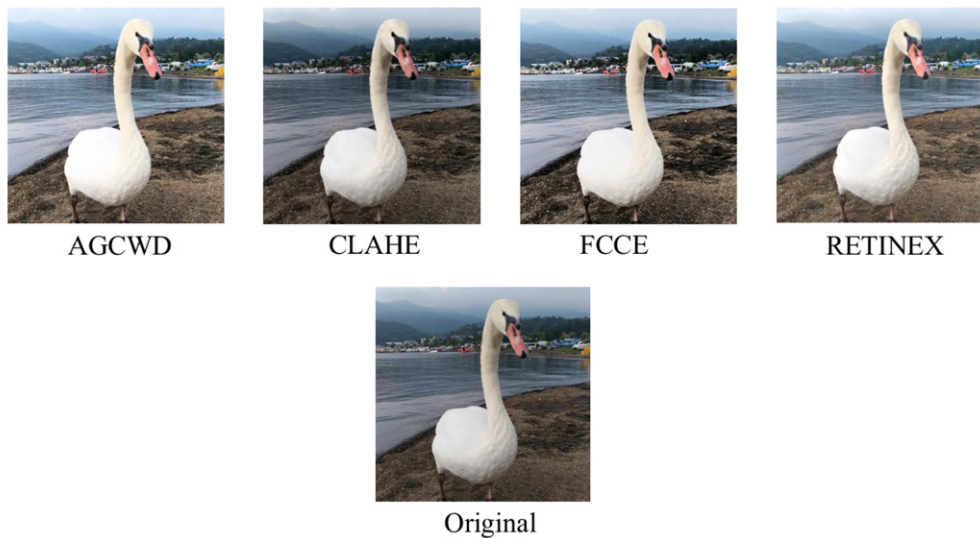


Figure 7. Image 30: original and enhanced versions.

FCCE is unable to enhance shadow details. In image 19, where FCCE has the highest z-score, it produces a natural image while Retinex and AGCWD produce images that are brighter with more pixels being clipped in the highlights compared to FCCE. For image 22, Retinex and CLAHE provide the highest z-scores with images that have acceptable contrast and a natural appearance. Regarding image 22, FCCE produces a darker image while AGCWD renders it excessively bright with both versions being less natural than those generated by Retinex and CLAHE.

Analysis of features of the original image, such as mean lightness, mean saturation, and detail level, and checking their correlation with resulting z-scores of each enhancement method did not reveal a relationship. These simple features do not seem to contain direct information on predicting the enhancement quality.

#### 4.2 Comparison between Central Asian and Norwegian Observers

Figure 9 shows the z-score plots for the eight Norwegian observers for all image scenes together. We can see that the highest z-score is for Retinex, followed by AGCWD, CLAHE, the original, and finally FCCE. Compared to the Central Asian observers who scored CLAHE the highest, Norwegian observers have scored CLAHE lower. Both groups do not prefer FCCE; Norwegian observers have rated this lower than Central Asian observers. Norwegians also rated AGCWD to be slightly above average, whereas Central Asians rated it to be slightly below average. It is worth noting that there are more observers in the Central Asian experiment than in the Norwegian experiment.

Figure 10 shows the z-scores for each image in the dataset. We can see that the Norwegian observers consistently rated FCCE low. This includes image 19—Central

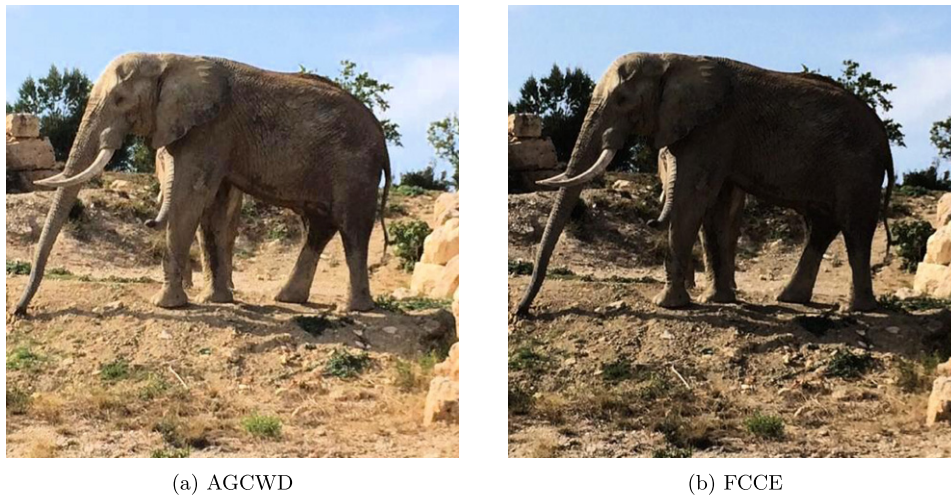


Figure 8. Comparison of AGCWD and FCCE for image 9.

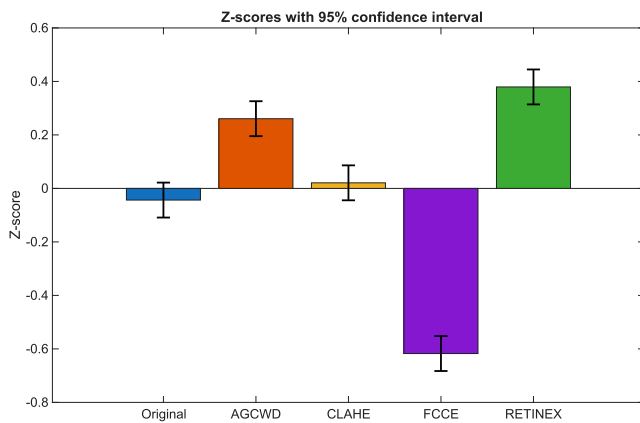


Figure 9. The z-scores for all image scenes together for the Norwegian observers. The scores are plotted with a 95% confidence interval.

Asian observers strongly preferred this image. This is a scene for which Central Asian observers seem to prefer a darker image compared to Norwegians. Moreover, regarding image 28 where Norwegians prefer AGCWD, which is an image with increased lightness, there are smaller differences between the enhancement algorithms for the Central Asian observers. We also see similarities between the observer groups for certain images, which indicates that image content could play a role.

#### 4.3 Objective Scores of IQMs

Given that the original image is part of the evaluated images, we have focused on no-reference IQMs. We have calculated the following IQMs: BRISQUE [57], Language-Image Quality Evaluator (LIQE) [58], CLIPIQA [59], CNNIQA [60], Neural Image Assessment (NIMA) [61], NRQM [62], PIQE [63], PAQ2PIQ [64], ARNIQA [65], ENTROPY [66], Multi-dimension Attention Network for No-Reference Image Quality Assessment (MANIQA) [67], TOPIQ [68], UNIQUE [69], Weighted Average Deep Image Quality

Measure (WADIQAM) [70], LAION-Aesthetics predictor (LAIONAES) [71], Perceptual Image (PI) [72], Fog Aware Density Evaluator (FADE) [73], High Order Statistics Aggregation (HOSA) [74], Natural Image Quality Evaluator (NIQE) [75], Perceptual Sharpness Index (PSI) [76], and Just Noticeable Blur Metric (JNBM) [77]. These metrics span a wide range, including blur metrics, natural image predictors, aesthetics, and general quality.

Investigation of the overall Pearson and Spearman correlation coefficients between the IQMs and subjective scores (z-scores) reveals that none of the IQMs perform well. These results align with previous research showing that it is challenging to assess the quality of enhanced images [29–31]. This might indicate that current IQMs should be customized for a specific application.

We also analyzed the correlation per image for each of the tested IQMs. A boxplot of the Spearman correlation per image is shown in Figure 11. We found that some IQMs at the image level correlate well, but when correlation is calculated overall the performance drops. This implies that, similar to reports in the literature [78–80], scale problems may persist in enhanced images [31]. The highest performing IQM is LIQE. The LIQE is based on vision–language correspondence. The IQM has been trained to simultaneously conduct blind image quality assessment, scene classification, and distortion identification. Despite being trained on datasets with distortions, it performs the best on our dataset. This could be due to the combination of scene classification with identification of enhancement-related distortion.

## 5. CONCLUSIONS AND FUTURE PERSPECTIVES

We found that Central Asian observers are in agreement with observations from other studies ([26, 27]) that the evaluation of enhanced image quality is a challenging task. Moreover, original images tend to be perceived as lower quality compared to the enhanced versions when considering them overall. This also aligns with the results of existing

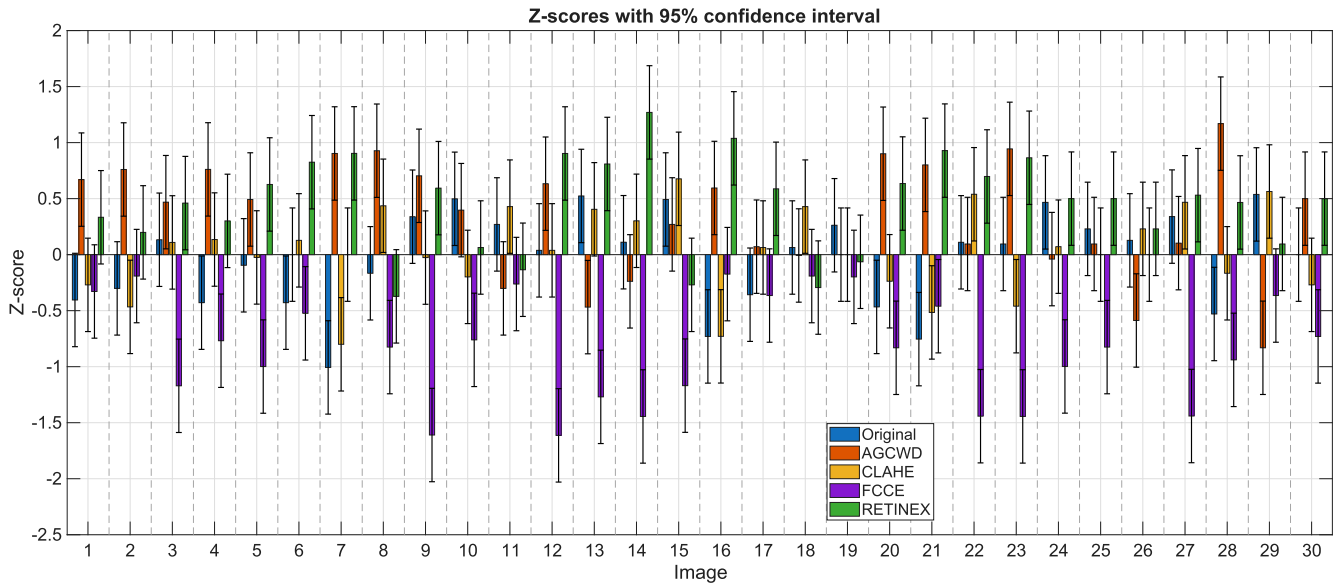


Figure 10. The z-scores for single-image scenes together for the Norwegian observers. The scores are plotted with a 95% confidence interval.

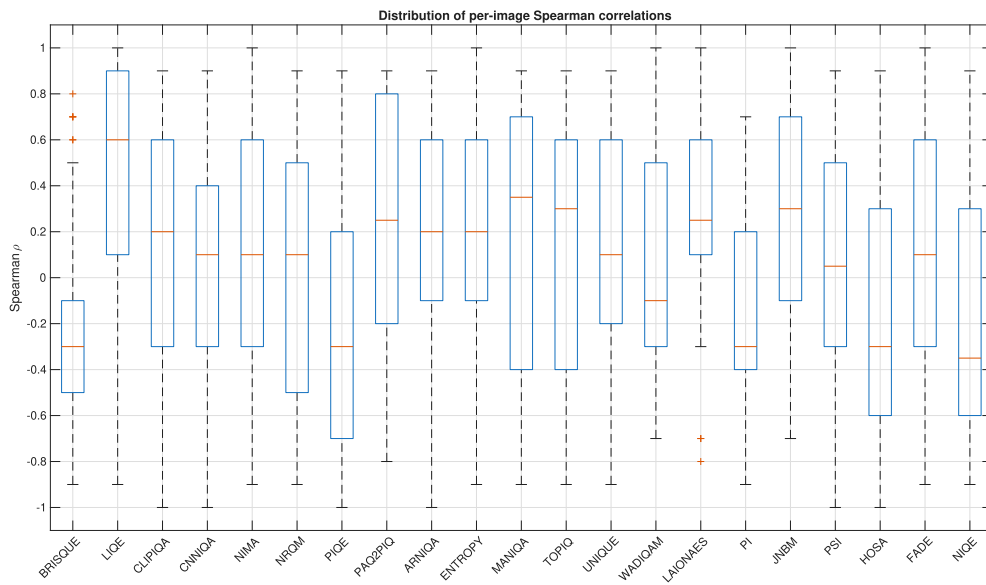


Figure 11. Boxplot of Spearman correlation per image.

studies. We determine that individual preferences might exist in the evaluations of contrast-enhanced images.

The IQMs tested did not correlate with the observers’ perception on our introduced dataset, which indicates that current IQMs find it difficult to measure the quality of enhanced images, consistent with the results of existing work [31]. As a result, we can assume that current IQMs should be customized for a specific application.

In addition, we have conducted another experiment with Norwegian observers. The comparative analysis showed that differences are present between two observer groups. These variations could be due to cultural differences. For

some images, observer groups showed similarities in quality perception. This might illustrate that image content could play a role.

One of the shortcomings of this study is that Kazakh observers were the majority who represented the Central Asian group (due to practical limitations). Therefore, expanding this research with more representatives from Central Asian countries would be the aim of future work.

In conclusion, whether designing a universal IQM or creating customized IQMs for image enhancement evaluation, the dataset developed in this work would be highly beneficial.

## ACKNOWLEDGMENT

The authors would like to thank the observers for their participation in the experiment. Marius Pedersen is supported by the Research Council of Norway through the “Quality and Content” project (Grant Number 324663).

## REFERENCES

- 1 P.-H. Lin and P. Patterson, “Investigation of perceived image quality and colourfulness in mobile displays for different cultures, ambient illumination, and resolution,” *Ergonomics* **55**, 1502–1512 (2012).
- 2 E. C. Larson and D. M. Chandler, “Most apparent distortion: full-reference image quality assessment and the role of strategy,” *J. Electron. Imaging* **19**, 011006 (2010).
- 3 N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, “Image quality assessment based on a degradation model,” *IEEE Trans. Image Process.* **9**, 636–650 (2000).
- 4 S. Ahn, Y. Choi, and K. Yoon, “Deep learning-based distortion sensitivity prediction for full-reference image quality assessment,” *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition* (IEEE, Piscataway, NJ, 2021), pp. 344–353.
- 5 L. Agnolucci, L. Galteri, M. Bertini, and A. Del Bimbo, “ARNIQA: learning distortion manifold for image quality assessment,” *Proc. IEEE/CVF Winter Conf. on Applications of Computer Vision* (IEEE, Piscataway, NJ, 2024), pp. 189–198.
- 6 X. Min, G. Zhai, K. Gu, Y. Liu, and X. Yang, “Blind image quality estimation via distortion aggravation,” *IEEE Trans. Broadcast.* **64**, 508–517 (2018).
- 7 L. Liu, B. Liu, H. Huang, and A. C. Bovik, “No-reference image quality assessment based on spatial and spectral entropies,” *Signal Process. Image Commun.* **29**, 856–863 (2014).
- 8 H. R. Sheikh, M. F. Sabir, and A. C. Bovik, “A statistical evaluation of recent full reference image quality assessment algorithms,” *IEEE Trans. Image Process.* **15**, 3440–3451 (2006).
- 9 D. M. Chandler, M. M. Alam, and T. D. Phan, “Seven challenges for image quality research,” *Proc. SPIE* **9014**, 901402 (2014).
- 10 X. Liu, M. Pedersen, and R. Wang, “Survey of natural image enhancement techniques: Classification, evaluation, challenges, and perspectives,” *Digit. Signal Process.* **127**, 103547 (2022).
- 11 Y. Qi, Z. Yang, W. Sun, M. Lou, J. Lian, W. Zhao, X. Deng, and Y. Ma, “A comprehensive overview of image enhancement techniques,” *Arch. Comput. Meth. Eng.* **29**, 1–25 (2021).
- 12 W. Wang, X. Wu, X. Yuan, and Z. Gao, “An experiment-based review of low-light image enhancement methods,” *IEEE Access* **8**, 87884–87917 (2020).
- 13 S. Anwar and C. Li, “Diving deeper into underwater image enhancement: a survey,” *Signal Process. Image Commun.* **89**, 115978 (2020).
- 14 M. J. Islam, Y. Xia, and J. Sattar, “Fast underwater image enhancement for improved visual perception,” *IEEE Robot. Autom. Lett.* **5**, 3227–3234 (2020).
- 15 W. Zhang, P. Zhuang, H.-H. Sun, G. Li, S. Kwong, and C. Li, “Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement,” *IEEE Trans. Image Process.* **31**, 3997–4010 (2022).
- 16 C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, and W. Ren, “Underwater image enhancement via medium transmission-guided multi-color space embedding,” *IEEE Trans. Image Process.* **30**, 4985–5000 (2021).
- 17 Z. Ullah, M. U. Farooq, S.-H. Lee, and D. An, “A hybrid image enhancement based brain MRI images classification technique,” *Med. Hypotheses* **143**, 109922 (2020).
- 18 J. Lu, D. M. Healy Jr, and J. B. Weaver, “Contrast enhancement of medical images using multiscale edge representation,” *Opt. Eng.* **33**, 2151–2161 (1994).
- 19 Z. Huang, S. Wang, H. Hu, and Y. Xu, “RetiGAN: a hybrid image enhancement method for medical images,” *2024 5th Int'l. Conf. on Computer Vision, Image and Deep Learning (CVIDL)* (IEEE, Piscataway, NJ, 2024), pp. 25–29.
- 20 H. Demirel, C. Ozcinar, and G. Anbarjafari, “Satellite image contrast enhancement using discrete wavelet transform and singular value decomposition,” *IEEE Geosci. Remote Sensing Lett.* **7**, 333–337 (2009).
- 21 J.-L. Lisani, J. Michel, J.-M. Morel, A. B. Petro, and C. Sbert, “An inquiry on contrast enhancement methods for satellite images,” *IEEE Trans. Geosci. Remote Sens.* **54**, 7044–7054 (2016).
- 22 H. Demirel and G. Anbarjafari, “Discrete wavelet transform-based satellite image resolution enhancement,” *IEEE Trans. Geosci. Remote Sens.* **49**, 1997–2004 (2011).
- 23 S. Lal, M. Chandra, Z.-ur Rahman, D. J. Jobson, and G. A. Woodell, “Efficient algorithm for contrast enhancement of natural images,” *Int. Arab J. Inf. Technol.* **11**, 95–102 (2014).
- 24 Z.-ur Rahman, D. J. Jobson, and G. A. Woodell, “Retinex processing for automatic image enhancement,” *J. Electron. Imaging* **13**, 100–110 (2004).
- 25 S. Azimian, F. Torkamani-Azar, and S. A. Amirshahi, “How good is too good? A subjective study on over enhancement of images,” *29th Color and Imaging Conf.* (IS&T, Springfield, VA, 2021).
- 26 D. M. Chandler, “Seven challenges in image quality assessment: past, present, and future research,” *Int. Scholarly Res. Not.* **2013**, 905685 (2013).
- 27 Y. Cheng, M. Pedersen, and G. Chen, “Evaluation of image quality metrics for sharpness enhancement,” *Proc. 10th Int'l. Symp. on Image and Signal Processing and Analysis* (IEEE, Piscataway, NJ, 2017), pp. 115–120.
- 28 W. Lin, L. Dong, and P. Xue, “Visual distortion gauge based on discrimination of noticeable contrast changes,” *IEEE Trans. Circuits Syst. Video Technol.* **15**, 900–909 (2005).
- 29 S. A. Amirshahi, A. Kadyrova, and M. Pedersen, “How do image quality metrics perform on contrast enhanced images?,” *2019 8th European Workshop on Visual Information Processing (EUVIP)* (IEEE, Piscataway, NJ, 2019), pp. 232–237.
- 30 K. Gu, G. Zhai, W. Lin, and M. Liu, “The analysis of image contrast: from quality assessment to automatic enhancement,” *IEEE Trans. Cybern.* **46**, 284–297 (2015).
- 31 A. Kadyrova, M. Pedersen, B. Ahmad, D. J. Mandal, M. Nguyen, and P. Zimmermann, “Image enhancement dataset for evaluation of image quality metrics,” *IST Int'l. Symp. on Electronic Imaging 2022, Image Quality and System Performance XIX* (IS&T, Springfield, VA, 2022).
- 32 C. T. Vu, T. D. Phan, P. S. Banga, and D. M. Chandler, “On the quality assessment of enhanced images: a database, analysis, and strategies for augmenting existing methods,” *2012 IEEE Southwest Symp. on Image Analysis and Interpretation* (IEEE, Piscataway, NJ, 2012), pp. 181–184.
- 33 M. A. Qureshi, A. Beghdadi, B. Sdiri, M. Deriche, and F. Alaya-Cheikh, “A comprehensive performance evaluation of objective quality metrics for contrast enhancement techniques,” *2016 6th European Workshop on Visual Information Processing (EUVIP)* (IEEE, Piscataway, NJ, 2016), pp. 1–5.
- 34 C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, “An underwater image enhancement benchmark dataset and beyond,” *IEEE Trans. Image Process.* **29**, 4376–4389 (2019).
- 35 O. Cherepkova, S. A. Amirshahi, and M. Pedersen, “Individual contrast preferences in natural images,” *J. Imaging* **10**, 25 (2024).
- 36 N. K. Senthilkumar, A. Ahmad, M. Andretto, V. Prabhakaran, U. Prabhu, A. B. Dieng, P. Bhattacharyya, and S. Dave, “Beyond aesthetics: cultural competence in text-to-image models,” *Adv. Neural Inf. Process. Syst.* **37**, 13716–13747 (2024).
- 37 M. M. Aslam, “Are you selling the right colour? A cross-cultural review of colour as a marketing cue,” *J. Mark. Commun.* **12**, 15–30 (2006).
- 38 T. R. Garth, “The color preferences of five hundred and fifty-nine full-blood Indians,” *J. Exp. Psychol.* **5**, 392 (1922).
- 39 A. Choungourian, “Color preferences and cultural variation,” *Perceptual Motor Skills* **26**, 1203–1206 (1968).
- 40 S. Shoyama, Y. Tochiara, and J. Kim, “Japanese and Korean ideas about clothing colors for elderly people: intercountry and intergenerational differences,” *Color Res. Appl.* **28**, 139–150 (2003).
- 41 T. Oyama, Y. Tanaka, and Y. Chiba, “Affective dimensions of colors a cross-cultural study,” *Japan. Psychological Res.* **4**, 78–91 (1962).
- 42 T. J. Madden, K. Hewett, and M. S. Roth, “Managing images in different cultures: a cross-national study of color meanings and preferences,” *J. Int. Mark.* **8**, 90–107 (2000).
- 43 L. C. Ou, M. Ronnier Luo, P. L. Sun, N. C. Hu, H. S. Chen, S. S. Guan, A. Woodcock, J. L. Caivano, R. Huertas, A. Treméau, and M. Billger, “A cross-cultural comparison of colour emotion for two-colour combinations,” *Color Res. Appl.* **37**, 23–43 (2012).

- 44 K. Choi and H.-J. Suk, "A comparative study of psychophysical judgment of color reproductions on mobile displays between Europeans and Asians," *Proc. SPIE* **9395**, 212–220 (2015).
- 45 S. R. Fernandez, M. D. Fairchild, and K. Braun, "Analysis of observer and cultural variability while generating "preferred" color reproductions of pictorial images," *J. Imaging Sci. Technol.* **49**, 96 (2005).
- 46 D. Saupé and S. H. Del Pin, "Uncovering cultural influences on perceptual image and video quality assessment through adaptive quantized metric models," *J. Perceptual Imaging* **7** (2025).
- 47 C. Chen, S.-ying Lee, and H. W. Stevenson, "Response style and cross-cultural comparisons of rating scales among East Asian and North American students," *Psychological Sci.* **6**, 170–175 (1995).
- 48 A. Beghdadi, M. A. Qureshi, B. Sdiri, M. Deriche, and F. Alaya-Cheikh, "CEED - a database for image contrast enhancement evaluation," *2018 Colour and Visual Computing Symposium (CVCS)* (IEEE, Piscataway, NJ, 2018), pp. 1–6.
- 49 S.-C. Huang, F.-C. Cheng, and Y.-S. Chiu, "Efficient contrast enhancement using adaptive gamma correction with weighting distribution," *IEEE Trans. Image Process.* **22**, 1032–1041 (2012).
- 50 K. Zuiderveld, "Contrast limited adaptive histogram equalization," *Graph. Gems* **4**, 474–485 (1994).
- 51 M. A. Qureshi, A. Beghdadi, and M. Deriche, "Towards the design of a consistent image contrast enhancement evaluation measure," *Signal Process. Image Commun.* **58**, 212–227 (2017).
- 52 A. S. Parihar, O. P. Verma, and C. Khanna, "Fuzzy-contextual contrast enhancement," *IEEE Trans. Image Process.* **26**, 1810–1819 (2017).
- 53 A. Rizzi, T. Algeri, G. Medeghini, and D. Marini, "A proposal for contrast measure in digital images," *Conf. on Colour in Graphics, Imaging, and Vision* (IS&T, Springfield, VA, 2004), pp. 187–192.
- 54 K. Van Ngo, J. Storvik Jr., C. A. Dokkeberg, I. Farup, and M. Pedersen, "QuickEval: a web application for psychometric scaling experiments," *Proc. SPIE* **9396**, 212–224 (2015).
- 55 P. G. Engeldrum, *Psychometric Scaling: a Toolkit for Imaging Systems Development* (Imcotek Press, Winchester, MA, 2000).
- 56 E. D. Montag, "Empirical formula for creating error bars for the method of paired comparison," *J. Electron. Imaging* **15**, 010502–010502 (2006).
- 57 A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.* **21**, 4695–4708 (2012).
- 58 W. Zhang, G. Zhai, Y. Wei, X. Yang, and K. Ma, "Blind image quality assessment via vision-language correspondence: a multitask learning perspective," *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition* (IEEE, Piscataway, NJ, 2023), pp. 14071–14081.
- 59 J. Wang, K. C. Chan, and C. C. Loy, "Exploring clip for assessing the look and feel of images," *Proc. of the AAAI Conf. on Artificial Intelligence* (AAAI Press, Washington, DC, USA, 2023), Vol. 37, pp. 2555–2563.
- 60 L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* (IEEE, Piscataway, NJ, 2014), pp. 1733–1740.
- 61 H. Talebi and P. Milanfar, "NIMA: neural image assessment," *IEEE Trans. Image Process.* **27**, 3998–4011 (2018).
- 62 C. Ma, C.-Y. Yang, X. Yang, and M.-H. Yang, "Learning a no-reference quality metric for single-image super-resolution," *Comput. Vis. Image Underst.* **158**, 1–16 (2017).
- 63 N. Venkatanath, D. Praneeth, S. C. Sumohana, and S. M. Swarup, "Blind image quality evaluation using perception based features," *2015 21st National Conf. on Communications (NCC)* (IEEE, Piscataway, NJ, 2015), pp. 1–6.
- 64 Z. Ying, H. Niu, P. Gupta, D. Mahajan, D. Ghadiyaram, and A. Bovik, "From patches to pictures (PaQ-2-PiQ): Mapping the perceptual space of picture quality," *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition* (IEEE, Piscataway, NJ, 2020), pp. 3575–3585.
- 65 L. Agnolucci, L. Galteri, M. Bertini, and A. Del Bimbo, "Arniqa: learning distortion manifold for image quality assessment," *Proc. of the IEEE/CVF Winter Conf. on Applications of Computer Vision* (IEEE, Piscataway, NJ, 2024), pp. 189–198.
- 66 R. C. Gonzalez, R. E. Woods, and S. L. Eddins, "Digital image processing using MATLAB," *Digital Image Processing Using MATLAB, Chapter 11* (Prentice Hall, New Jersey, 2003).
- 67 S. Yang, T. Wu, S. Shi, S. Lao, Y. Gong, M. Cao, J. Wang, and Y. Yang, "Maniqa: multi-dimension attention network for no-reference image quality assessment," *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition* (IEEE, Piscataway, NJ, 2022), pp. 1191–1200.
- 68 C. Chen, J. Mo, J. Hou, H. Wu, L. Liao, W. Sun, Q. Yan, and W. Lin, "Topiq: a top-down approach from semantics to distortions for image quality assessment," *IEEE Trans. Image Process.* **33**, 2404–2418 (2024).
- 69 W. Zhang, K. Ma, G. Zhai, and X. Yang, "Uncertainty-aware blind image quality assessment in the laboratory and wild," *IEEE Trans. Image Process.* **30**, 3474–3486 (2021).
- 70 S. Bosse, D. Maniry, K. R. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Trans. Image Process.* **27**, 206–219 (2017).
- 71 C. Schuhmann, *LAION Aesthetics Predictor* (2022), online <https://laion.ai/blog/laion-aesthetics/>.
- 72 Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The 2018 PIRM challenge on perceptual image super-resolution," *European Conf. on Computer Vision* (Springer International Publishing, Cham, 2018), pp. 334–355.
- 73 L. K. Choi, J. You, and A. C. Bovik, "Referenceless prediction of perceptual fog density and perceptual image defogging," *IEEE Trans. Image Process.* **24**, 3888–3901 (2015).
- 74 J. Xu, P. Ye, Q. Li, H. Du, Y. Liu, and D. Doermann, "Blind image quality assessment based on high order statistics aggregation," *IEEE Trans. Image Process.* **25**, 4444–4457 (2016).
- 75 A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Process. Lett.* **20**, 209–212 (2012).
- 76 C. Feichtenhofer, H. Fassold, and P. Schallauer, "A perceptual image sharpness metric based on local edge gradient analysis," *IEEE Signal Process. Lett.* **20**, 379–382 (2013).
- 77 R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *IEEE Trans. Image Process.* **18**, 717–728 (2009).
- 78 M. Pedersen and I. Farup, "Improving the robustness to image scale of the total variation of difference metric," *2016 3rd Int'l. Conf. on Signal Processing and Integrated Networks (SPIN)* (IEEE, Piscataway, NJ, 2016), pp. 116–121.
- 79 R. Hlayhel, M. Mobini, B. E. Agossou, M. Pedersen, and S. A. Amirshahi, "Colourlab image database: optical aberrations," *London Imaging Meeting* **5**, 22 (2024).
- 80 T. U. Ahmed, S. A. Amirshahi, and M. Pedersen, "Image demosaicing: subjective analysis and evaluation of image quality metrics," *Electron. Imaging* **35**, 1–6 (2023).