https://doi.org/10.2352/issn.2169-2629.2021.29.83

How Good is Too Good? A Subjective Study on Over Enhancement of Images

***Sahar Azimian*[*]*, Farah Torkamani Azar*[*]*, Seyed Ali Amirshahi*[†]**
[*] ***Faculty of Electrical Engineering, Shahid Beheshti University, Tehran, Iran***
[†] ***The Norwegian Colour and Visual Computing Laboratory, Norwegian University of Science and Technology, Gjøvik, Norway***
sahar.azm8@gmail.com, f-torkamani@sbu.ac.ir, s.ali.amirshahi@ntnu.no

Abstract

For a long time different studies have focused on introducing new image enhancement techniques. While these techniques show a good performance and are able to increase the quality of images, little attention has been paid to how and when over-enhancement occurs in the image. This could possibly be linked to the fact that current image quality metrics are not able to accurately evaluate the quality of enhanced images. In this study we introduce the Subjective Enhanced Image Dataset (SEID) in which 15 observers are asked to enhance the quality of 30 reference images which are shown to them once at a low and another time at a high contrast. Observers were instructed to enhance the quality of the images to the point that any more enhancement will result in a drop in the image quality. Results show that there is an agreement between observers on when over-enhancement occurs and this point is closely similar no matter if the high contrast or the low contrast image is enhanced.

Introduction And Related Works

In today's world, images and their applications play a huge role in our daily life. The rapid spread of digital imaging and communication network technology provides access to millions of digital photos freely accessible and shared on the internet and other platforms. Having access to such a large pool of images from different types and categories has made the role of image quality a critical matter. That is why image creators and vendors would like to provide the best quality image to their audience. As a result, Image Quality Assessment (IQA) and image enhancement have drawn much attention and are now regarded as essential image processing operations. For many practical applications in image processing, such as recognition and object detection, images frequently require adequate visual quality enhancement, visibility, and contrast, among others.

Contrast enhancement methods are divided into direct and indirect methods [1]. By defining a standard definition for contrast measurement in the direct methods, image contrast can be improved further by enhancing the criterion. In the indirect method, the dynamic range of grey levels is extended for image contrast. Indirect methods which have received greater attention in recent years are classified into four categories,

- Decomposition-based methods that decompose the high and low-frequency components of an image [2].
- Transformation-based methods [3].
- Histogram modification methods [4].
- Soft calculations-based methods [5].

An important occurrence to be aware of during image enhancement is the risk of over-enhancing the image. In most cases, over-enhancing images result in unwanted loss of edges and borders, intensity scale, textures, and fine details. Furthermore, over-enhanced images could affect other subsequent processes such as segmentation and classification performed on images. While image contrast enhancement algorithms have received much attention in recent years, the possibility of over-enhancement has often been overlooked.

In general, when it comes to over-enhancing images two issues can occur; structure loss, and new structure creation (artifacts) [6]. The majority of indirect contrast enhancement methods based on histogram modification, particularly global histogram equalization, which enhances the overall contrast of an image by reconstructing intensity values and smoothness of the histogram, are linked to over-enhancement. These image enhancement methods lose or destroy the edges and borders without considering local information (some adjacent areas of similar intensity may merge). This type of over-enhancement is classified as structure loss. Other algorithms which introduce artifacts can result in improper texture or edges in the image.

To the best of our knowledge, there exists no quantitatively reliable, standard method to detect over-enhancement of images. Since over-enhancement of images result in a drop in the quality of an image, in an ideal case one approach to detect over-enhancement in images is the use of objective Image Quality Metrics (IQMs). Depending on whether an IQM is in need of the reference (original) image in its calculation three types of IQMs exist, Full Reference (FR), Reduced Reference (RR), and No Reference (NR). This reliance on the reference image in IQMs is based on the assumption that in all cases the reference image has the best quality (distortion-free and/or the ideal image) and so any reproduction of the image (what is referred to as the test image) would have a lower quality and should be evaluated in comparison to the reference image [7, 8]. Such an approach is clearly in contrast with the goal of image enhancement where our aim is that our test image has a higher quality compared to the reference image. This issue can link to the subpar performance of most image quality metrics when used on enhanced images [9, 10]. To avoid over-enhancement, several image enhancement algorithms limit the range of intensity. However, a vital downside of these methods is under-enhancement, resulting in under-enhancement in some areas of the image or even over-enhancement in others.

Over the years different enhanced image datasets have been proposed. As an example the Digitally Retouched Image Quality (DRIQ) contrast changed dataset [13] consists of 26 reference images, 78 enhanced images, and also provides the subjective ratings recorded from 9 participants. In the dataset, for each reference image three enhanced images were created. Using Adobe Photoshop, various combinations of color, saturation, brightness, and sharpness are enhanced in each image. The Contrast Changed Image Dataset (CCID) includes 15 references and 655 enhanced images [14]. Another dataset for subjective evaluation of sharpened images was proposed in [15]. In this study the performance of seven no-reference S3 metrics was adapted and

Figure 1. Reference images in the SEID dataset. Numbers below each image correspond to the name of the image in our dataset. Images 1-18 are from the CEED dataset [11] while images 19-30 are from the CCEID dataset [12].

compared to the subjective scores collected. In all three works mentioned the enhanced image is considered as the reference image and the original image as the distorted (test) image.

The Contrast Enhancement Evaluation Dataset (CEED) [11] includes 30 references and 180 enhanced images (six different enhanced image per reference image). For image enhancement, six different contrast enhancement techniques were applied. Finally, the Colourlab Contrast Enhanced Image Dataset (CCEID) [12] consists of 26 reference and 104 enhanced images. Four different contrast enhancement techniques were used in this study. In this work different NR color IQMs, FR color IQMs, NR greyscale IQMs, and FR greyscale IQMs were calculated for each image and their results were compared to the subjective scores collected. As expected results show that current IQMs are not able to evaluate the quality of enhanced images with a high accuracy.

While the mentioned datasets could be used in studies focused on the quality evaluation of enhanced images, to the best of our knowledge no dataset is focused on over-enhancement of images. In this work we introduce the Subjective Enhanced Image Dataset (SEID) which is specifically focused on image over-enhancement and is available to download at `www.colourlab.no/cid`. SEID was developed to address the flaws in image quality enhancing algorithms and to provide a quantitative evaluation criteria for enhanced image quality. This dataset could not only be used in studies focused on detecting over-enhancement in images but it could also be used for proposing new IQMs for enhanced images. 15 observers have participated in our subjective experiments which was performed on 30 reference images.

In the rest of the paper we first introduce the SEID dataset. The subjective experiment performed in this study is then introduced followed by analysing the subjective scores given to different images in the dataset. Finally, a conclusion and the future direction of the work are presented.

SEID Dataset

The SEID dataset consists of 30 reference images (Figure 1). To select the reference images in our dataset we focused on complementing the already available datasets in the field of image contrast enhancement. Keeping this goal in mind, the reference images in the SEID dataset were selected from already available images in the CEED [11] (18 images) and CCEID [12] (12 images) datasets. This image selection allowed us to guarantee the diversity in the dataset with regards to visual contents and colorfulness in the images. As mentioned in the previous section the primary goal of our study is to investigate over-enhancement in images. That is, to find the point in which

- an increase in contrast in the case of low contrast images and
- a decrease in contrast in the case of high contrast images

will result in the degradation of image quality. To this end, a subjective experiment was designed.

Subjective Experiment

For each reference image in the SEID dataset a high contrast (Figures 2(a)-(e)) and a low contrast (Figures 2(f)-(j)) test image was produced resulting in a total number of 60 test images. To produce the images with the low and high contrast we used a contrast stretching technique. In the case of a color image $\mathscr{I}$ in the RGB color space where each channel has a size of $U \times V$ pixels,

$$\mathscr{R}(u,v)_{new} = \begin{cases} \mathscr{R}(u,v) + \mathscr{R}(u,v) \times P_i, & \text{if } 0 \leq \mathscr{R}(u,v) \leq 220 \\ \mathscr{R}(u,v), & \text{if } 220 < \mathscr{R}(u,v) \leq 255 \end{cases} \tag{1}$$

will result in an increase in the contrast of the red color channel ($\mathscr{R}$), while

$$\mathscr{R}(u,v)_{new} = \begin{cases} \mathscr{R}(u,v), & \text{if } \mathscr{R}(u,v) \leq 10 \\ \mathscr{R}(u,v) - \mathscr{R}(u,v) \times P_d, & \text{if } 10 \leq \mathscr{R}(u,v) \leq 255 \end{cases} \tag{2}$$

will result in a decrease in the contrast in the red channel. To calculate the contrast enhanced image Eq. (1) and Eq. (2) are calculated for the green ($\mathscr{G}$) and blue ($\mathscr{B}$) channel as well. P_i and P_d in Eqs. (1) and (2) respectively correspond to the percentage of contrast changes made when increasing or decreasing the image contrast.

The subjective experiment interface was created using MATLAB. Participants were provided with an instruction on how the test is performed. At the start of the experiment the

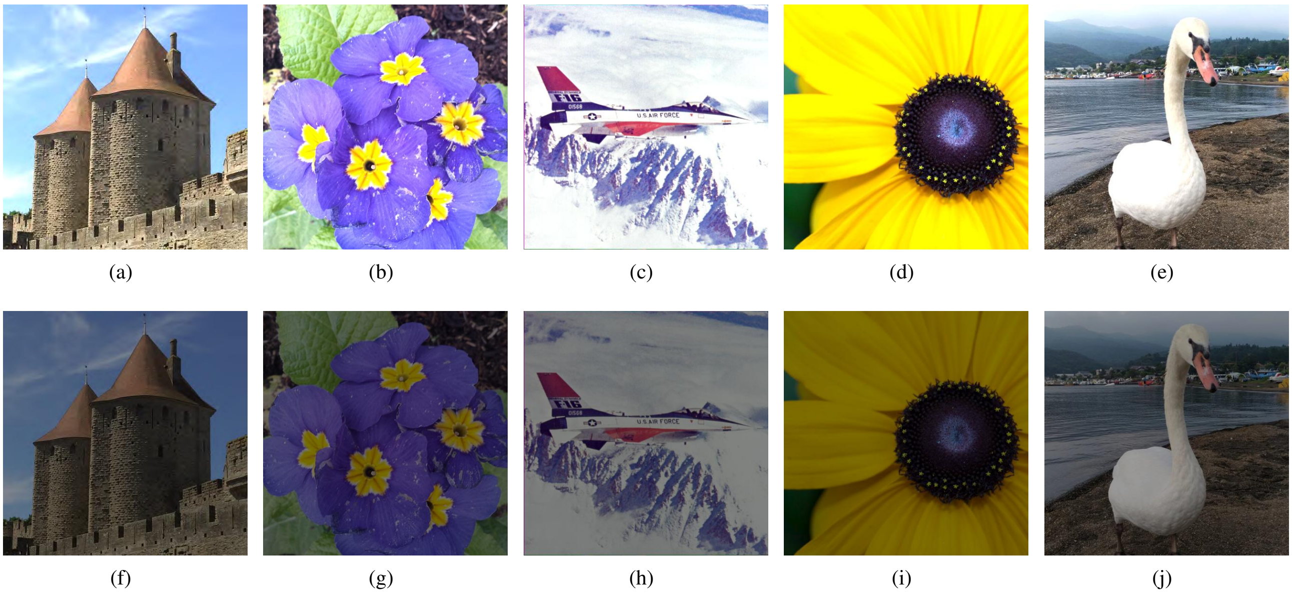
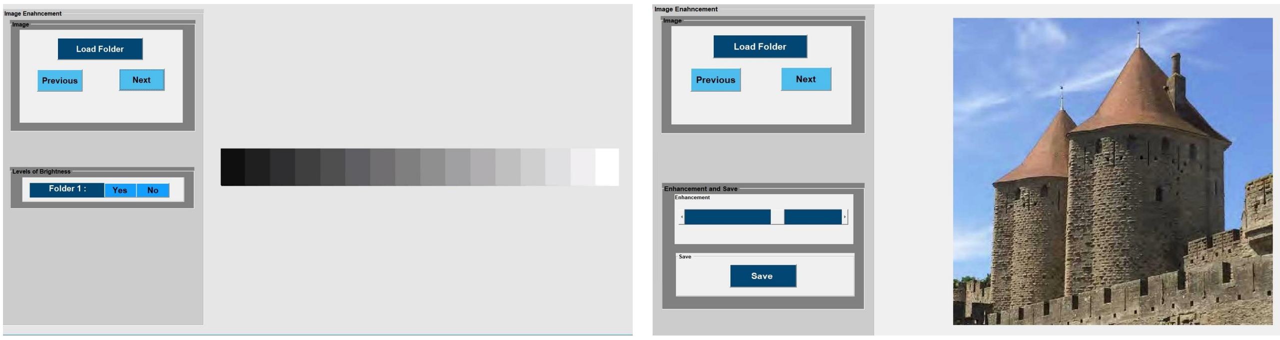


Figure 2. Sample test images from the SEID dataset. (a)-(e) Represent the high contrast images while (f)-(j) correspond to their low contrast images.



(a) (b)

Figure 3. The test was conducted by running the GUI file in MATLAB. (a) Checking that the observer is able to clearly distinguish between the 16 brightness levels in the image. (b) Observer enhances the image using the sliding bar.

program adjusted its size to the full screen and the resolution of the screen was recorded. Next, the observers were shown two different images one with eight and another with 16 different brightness levels and were asked if they can clearly see the different brightness levels in the image (Figure 3(a)). This was done mainly to check the visual acquity of the observers. Next, the 60 test images which all had a size of 512×512 pixels were presented to the observers one by one in a random order and the observers were asked to improve the quality of the image (Figure 3(b)). In this step participants were given the task to enhance the image using the sliding bar provided to them on the side of the image. During the experiment the observers were able to instantly see the changes they made in the image using the sliding bar. The changes (enhancements) made in the image were performed using Eqs. (1) and (2) with P_i and P_d values ranging from $[-100, 100]$. We should point out that to avoid any prior judgment, the observers were given the task of enhancing the quality of the image and at no point the fact that enhancement is purely done based on contrast was communicated to the observers. The use of high and low contrast images in our experiments allowed us to provide the observers with an over-enhanced image and then ask the observers to enhance the image to create the best quality image.

Participants

To perform the subjective experiment, 15 observers with an average age of 30.4, including seven females and eight males were recruited. Twelve participants had previously studied or worked in the field of image processing, while three others were non-experts.

Data Analysis And Image Quality Assessment

To analyse the enhanced images created by the observers the final images by each observer was reconstructed (Figure 4). This resulted in 30 images (15 of which were reconstructed from low-contrast and 15 from high-contrast images) for each of the reference image.

Analysis of the Subjective Assessment

It is no surprise that in the case of low contrast images observers decided to increase the contrast (Figure 5(a)) while in the case of high contrast images the contrast was reduced (Figure 5(b)). While subjectively the output image created by different observers were mostly similar (Figure 4) the similarity was also confirmed using objective IQMs. Finally we investigated the correlation between the 15 observers with regards to percentage of contrast changes made by each observer for each test image (Fig-

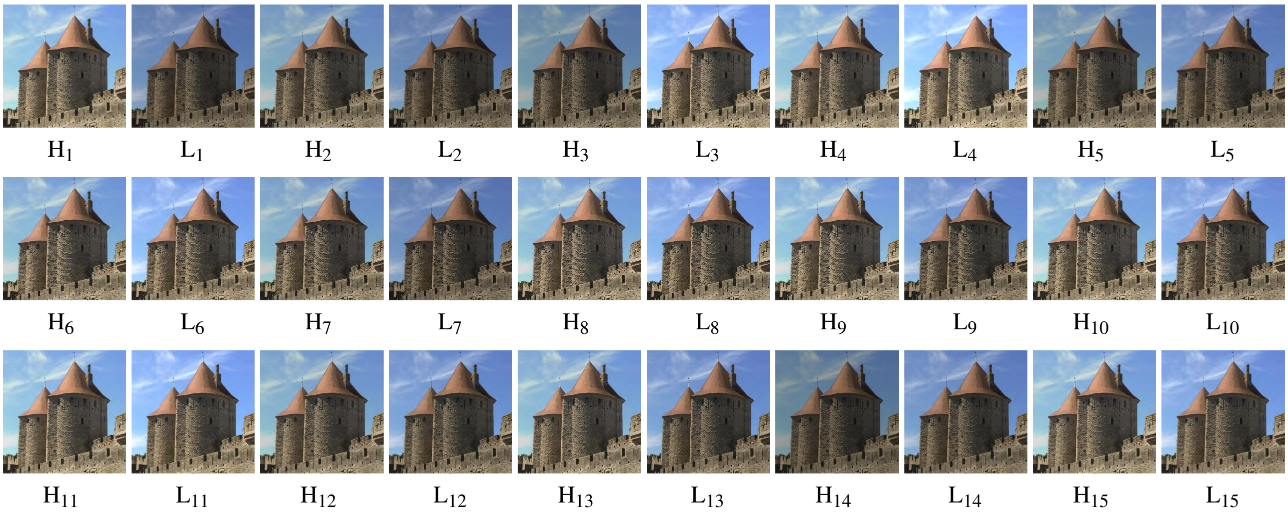
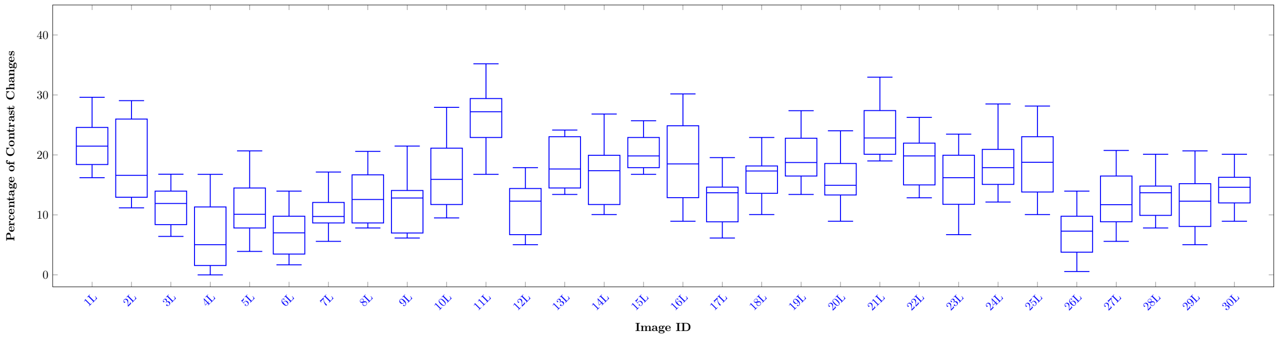
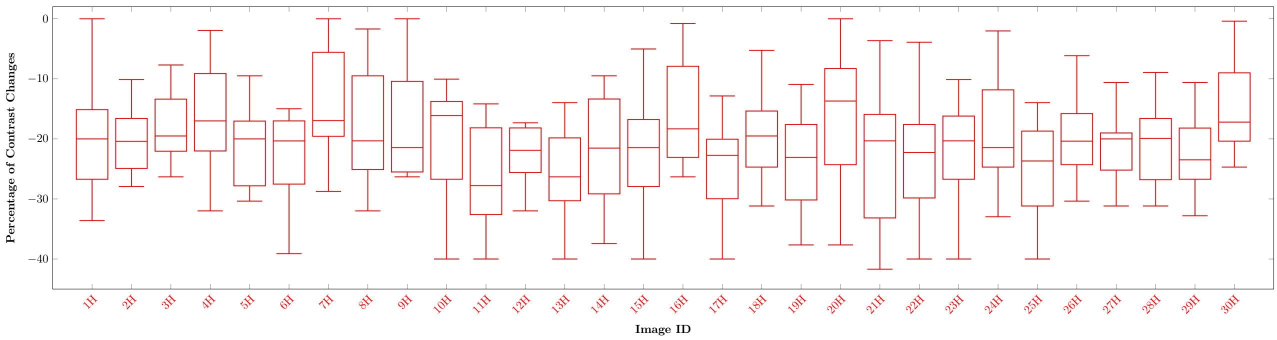


Figure 4. *Final enhanced images created by the 15 participants in our experiments. In the figure H_i corresponds to the enhanced image created by observer i from the high contrast image depicted in Figure 2(a), and L_i corresponds to the enhanced image created by the same observer from the low contrast image depicted in Figure 2(f).*



(a) Low contrast images



(b) High contrast images

Figure 5. *Box-plot illustrating the change in the contrast made by the 15 observers in each of the low (a) and high (b) contrast images.*

ure 6). It is interesting to observe that percentage of change in the contrast values show a high correlation in the case of low contrast images while this is not the case for high contrast images.

Analysis of the Objective Assessment

In our experiments we used different IQMs to evaluate the final enhanced images created by different observers. For this goal the SSIM color [16], Feature SIMilarity for color images (FSIMc) [17], Visual Saliency-Induced Index (VSI) [18], SR-SIM [19], and QPSV [20] were calculated.

1. As a first step the quality value between the corresponding enhanced images for the low and high contrast image is calculated for each observer (H_i and L_i for observer i). Results for each of the 30 images (Figure 7) show a high similarity between the final enhanced image for each observer showing that in fact there is an image that each observer sees as the best quality image no matter if this is created by increasing the contrast in a low contrast image or decreasing the contrast in a high contrast image.
2. The final enhanced image for each observer in each of the 60 cases (high and low contrast images for the 30 reference image) were compared to the reference (original) images. This would allow us to have a numeric estimate on how

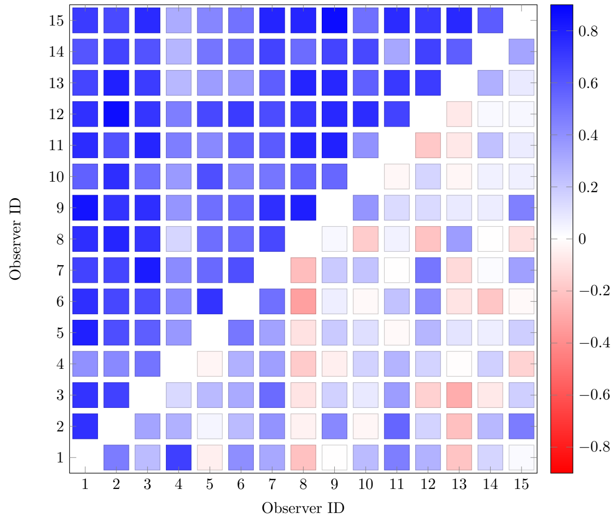


Figure 6. *Correlation between the 15 observers with regards to percentage of contrast changes for each test image. The top left of the map correspond to the low contrast images while the bottom right to the high contrast images.*

different the final image is compared to the original image among different observers. Results show that the difference in the quality value between the enhanced and reference image is closely similar among different observers (Figure 8).

3. As noted in the subjective IQA, the percentage of contrast changes (P_i and P_d values) made by observers was close to each other, and enhanced images created by the 15 observers were similar. This similarity was examined using different IQMs. This is done by comparing the quality value between the enhanced image of an observer and each of the 14 images created by the other observers (Figure 9). From the results it is evident that the final enhanced images created by different observers are closely similar with regards to their quality.

Conclusion and Future works

In this work we introduced the Subjective Enhanced Image Dataset (SEID) which is focused on finding the threshold in which an over-enhancement occurs when enhancing the images. In the study first the quality of 30 different images were degraded by drastically increasing or decreasing the contrast in the image. 15 different observers are then given the task of enhancing the quality of the 60 resulted images. Results show a high inter and intra observer similarity between the images created by each observer showing that indeed there is a common agreement between observers on when an enhancement of an image reaches the point of over-enhancement.

References

[1] Azeddine Beghdadi, Muhammad Ali Qureshi, Seyed Ali Amirshahi, Aladine Chetouani, and Marius Pedersen. A critical analysis on perceptual contrast and its use in visual information analysis and processing. *IEEE Access*, 8:156929–156953, 2020.

[2] Allen Oppenheim, Ronald Schafer, and Thomas Stockham. Nonlinear filtering of multiplied and convolved signals. *IEEE Transactions on Audio and Electroacoustics*, 16(3):437–466, 1968.

[3] Sos Agaian, Blair Silver, and Karen Panetta. Transform coefficient histogram based image enhancement algorithms using contrast entropy. *IEEE Transactions on Image Processing*, 16(3):741–758, 2007.

[4] Qing Wang and Rabab Ward. Fast image/video contrast enhancement based on weighted thresholded histogram equalization. *IEEE Transactions on Consumer Electronics*, 53(2):757–764, 2007.

[5] Sara Hashemi, Soheila Kiani, Navid Noroozi, and Mohsen Ebrahimi Moghadam. An image contrast enhancement method based on genetic algorithm. *Pattern Recognition Letters*, 31(2):1816–1824, 2010.

[6] Cheng and Yingtao Zhang. Detecting of contrast over-enhancement. In *2012 19th IEEE International Conference on Image Processing*, pages 961–964. IEEE, 2012.

[7] Seyed Ali Amirshahi, Marius Pedersen, and Azeddine Beghdadi. Reviving traditional image quality metrics using cnns. In *Color and Imaging Conference*, volume 2018, pages 241–246. Society for Imaging Science and Technology, 2018.

[8] Seyed Ali Amirshahi, Marius Pedersen, and Stella X Yu. Image quality assessment by comparing cnn features between images. *Journal of Imaging Science and Technology*, 60(6):60410–1, 2016.

[9] Seyed Ali Amirshahi, Altynay Kadyrova, and Marius Pedersen. Do contrast measures correlate? a pilot investigation. In *2019 8th European Workshop on Visual Information Processing (EUVIP)*, pages 40–45. IEEE, 2019.

[10] Seyed Ali Amirshahi and Marius Pedersen. Future directions in image quality. In *Color and Imaging Conference*, volume 2019, pages 399–403. Society for Imaging Science and Technology, 2019.

[11] Azeddine Beghdadi, Muhammad Ali Qureshi, Bilel Sdiri, Mohamed Deriche, and Faouzi Alaya-Cheikh. Ceed-a database for image contrast enhancement evaluation. In *2018 Colour and Visual Computing Symposium (CVCS)*, pages 1–6. IEEE, 2018.

[12] Seyed Ali Amirshahi, Altynay Kadyrova, and Marius Pedersen. How do image quality metrics perform on contrast enhanced images? In *2019 8th European Workshop on Visual Information Processing (EUVIP)*, pages 232–237. IEEE, 2019.

[13] Cuong Vu, Thien Phan, Punit Banga, and Chandler Vu. On the quality assessment of enhanced images: A database, analysis, and strategies for augmenting existing methods. In *2012 IEEE Southwest Symposium on Image Analysis and Interpretation*, pages 181–184. IEEE, 2012.

[14] Ke Gu, Guangtao Zhai, and Min Liu. The analysis of image contrast: From quality assessment to automatic enhancement. *Transactions On Cybernetics*, 46(1):284–297, 2016.

[15] Lukáš Krasula, Patrick Le Callet, Karel Fliegel, and Miloš Klíma. Quality assessment of sharpened images: Challenges, methodology, and objective metrics. *IEEE Transactions on Image Processing*, 26(3):1496–1508, 2017.

[16] Nicolas Bonnier, Francis Schmitt, Hans Brettel, and Stephane Berche. Evaluation of spatial gamut mapping algorithms. In *Color and imaging conference*, volume 2006, pages 56–61. Society for Imaging Science and Technology, 2006.

[17] 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.

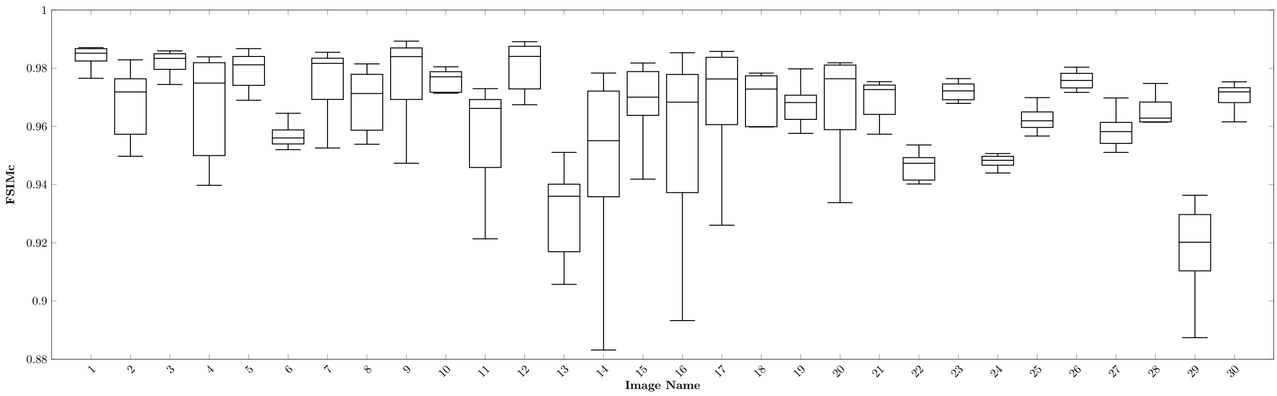


Figure 7. *Box-plot illustrating FSIMc values between enhanced images created from low and high contrast image for different observers per image.*

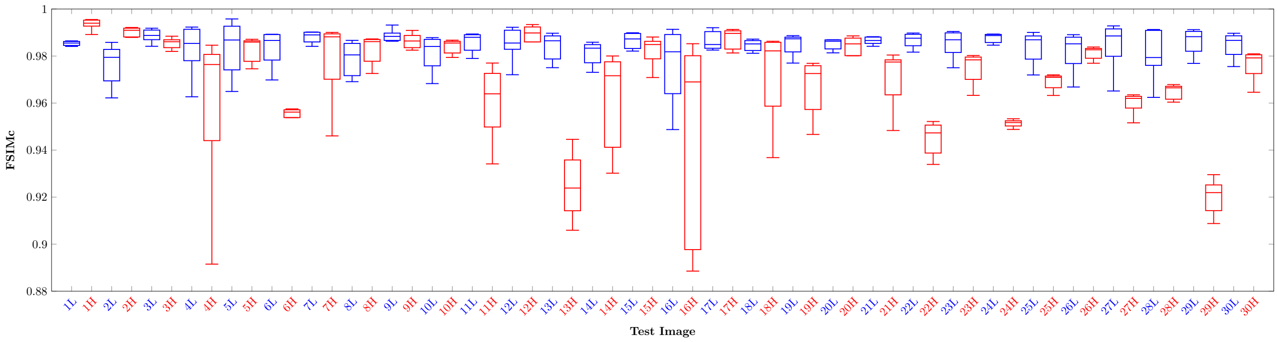
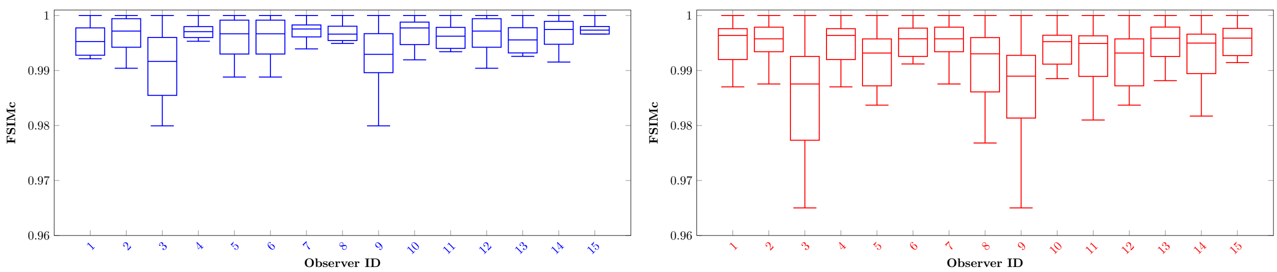


Figure 8. *Box-plot illustrating the FSIMc quality values calculated between each enhanced image by different observer and the reference image.*



(a) 1L (b) 1H

Figure 9. *Box-plot illustrating the results of subjective experiment for 2 test images (1L,1H). Each of the output images were compared with one output image out of 15 output images by FSIMc.*

[18] Lin Zhang, Ying Shen, and Hongyu Li. Vsi: A visual saliency-induced index for perceptual image quality assessment. *IEEE Transactions on Image processing*, 23(10):4270–4281, 2014.

[19] Lin Zhang and Hongyu Li. Sr-sim: A fast and high performance iqa index based on spectral residual. In *2012 19th IEEE international conference on image processing*, pages 1473–1476. IEEE, 2012.

[20] Azadeh Mansouri, Ahmad Mahmoudi Aznaveh, Farah Torkamani-Azar, and J Afshar Jahanshahi. Image quality assessment using the singular value decomposition theorem. *Optical Review*, 16(2):49–53, 2009.