

PAPER • OPEN ACCESS

A Document Image Dataset for Quality Assessment

To cite this article: Zhenghao Li *et al* 2021 *J. Phys.: Conf. Ser.* **1828** 012033

View the [article online](#) for updates and enhancements.

You may also like

- [A Comprehensive Review on Document Image \(DIBCO\) Database](#)
W A Mustafa, Wan Khairunizam, I Zunaidi et al.
- [Binarization of Document Image Using Optimum Threshold Modification](#)
Wan Azani Mustafa and Mohamed Mydin M. Abdul Kader
- [Document Image Extraction System Design](#)
N I Widiastuti and K E Dewi

PRIME
PACIFIC RIM MEETING
ON ELECTROCHEMICAL
AND SOLID STATE SCIENCE

HONOLULU, HI
Oct 6-11, 2024

Abstract submission deadline:
April 12, 2024

Learn more and submit!

Joint Meeting of
The Electrochemical Society
•
The Electrochemical Society of Japan
•
Korea Electrochemical Society

A Document Image Dataset for Quality Assessment

Zhenghao Li¹, Cihui Yang^{1*}, Qingyun Shen¹ and Shiping Wen²

¹School of Information and Engineering, Nanchang Hangkong University, NanChang, China

²Australian artificial Intelligence institute, University of Technology Sydney, Australia

*Email: 1263212577@qq.com, yangcihui@nchu.edu.cn

Abstract. Mobile device plays an very important role in capturing document image. However, the quality of the captured image is influenced by many factors, such as device quality and shooting conditions. In this context, it is necessary to automatically assess the quality of captured document image. Although there has a lot of work in the filed of image quality assessment (IQA), insufficient attention has been paid to the establishment of document images dataset. Thus, we propose a large dataset of document images containing 19,943 images which are collected by mobile devices. During the process of image acquisition, many factors such as light intensity, distortion type, document material are considered. After capturing images, multiple volunteers participated in the evaluation and collection of Mean Opinion Score (MOS) of the document images. We use two no-reference image quality assessment algorithms to test the proposed dataset. The experimental results show the validity of our dataset and the reliability of MOS. The proposed dataset can be used in the field of image quality assessment and Optical Character Recognition.

1. Introduction

With the proliferation of camera on smartphones, a growing number of people prefer to capture document image by using mobile devices [1, 2]. Although it can bring more convenience, the quality of the captured image is not steady, and it is limited by light conditions, the material of paper (or books), blur due to limited depth of field, out-of-focus blur and so on. Among these factors, the most easily perceived by the human visual perception and the most influential one is the blur of the image, and the main types of blur include the out-of-focus blur, the motion blur and blur due to the depth of field. Due to the influence of above factors, the quality of captured images may not meet the requirements we want. Thus, Image Quality Assessment (IQA) [3] is an indispensable preparatory step in image processing system. Image quality dataset is an important basis for IQA research. Many studies are now using deep learning methods, especially in the research of IQA [4-7], the image quality dataset with a large amount of data is needed as the support.

In recent years, a lot of work has been done with image quality datasets. Some of these are datasets of natural scene images, these datasets include: LIVE dataset [8], TID2008 [9], TID 2013 [10], IVC dataset [11], CID:IQ dataset [12], categorical image quality (CSIQ) dataset [13], Toyoma-MICT dataset [14], and Cornell-A57 dataset [15]. Others are datasets of document images [16, 17]. However, these datasets still cannot meet the research needs of IQA. To the extent of our knowledge, the distortion in each image of these datasets is uniformly distributed. But in practical application, there is a key problem that the image of each area of the distortion degree is not equal. Meanwhile, the sample



size of current IQA datasets is not sufficient for training a deep neural network. For example, The Smart-IQA document dataset which was proposed in 2015 by Nibal Nayef contains 4260 images captured from 30 different paper documents. The number of images of Live database and TID2008/2013 datasets both do not exceed 5000. Due to insufficient sample size of dataset, most of the image quality evaluation methods based on neural network divide the raw image into multiple patches as input rather than using the whole image as input which will cause the absence of local ground truth targets [18]. Taking all above-mentioned into account, it is necessary to propose a dataset that can reflect the real image acquisition and has large sample size. Thus, in this work, we proposed a new document image dataset which has the following prominent features compared to previous datasets:

- (1) Having 19,943 captured images which is much larger than other IQA datasets.
- (2) Considering adding documents with different materials and books as new categories.
- (3) Considering the condition that the texture on paper is inherently blurry.
- (4) Images reflects the influence of various distortion factors during acquisition.
- (5) Provide subjective assessment scores for images.
- (6) Containing English and Chinese texts.

In our dataset, we have taken the following conditions into account:

- (1) The focal length of the camera on the smartphone.
- (2) The light condition.
- (3) The type of paper.
- (4) The type of Smartphone.

In addition, we hope the proposed dataset can be useful in the field of IQA or Optical Character Recognition [19]. In Section 2, we will introduce the details of our dataset, the specific experimental process of image collection and the method of obtaining MOS of images [20, 21]. And the result in Section 3 shows the performance of two no-reference image quality assessment algorithms on our dataset.

2. Description of the Proposed Image Dataset

2.1. Type of Document

Our dataset contains a total of 19,943 images, of which 7,083 are book images and the rest are document paper images. There are 6 types of book images with some differences in the color and texture. Meanwhile book images contain some colour or gray illustrations with varying degrees of distortion. Book images are shown in Figure 1. The 12,860 document samples were collected from 70 raw document images, which were composed of four different textures with background colors of white, red, blue, and yellow. There are also different levels of ink on the paper of the document. At the same time, We consider a special case in which the image itself is blurred rather than caused by the shooting conditions. document images are shown in Figure 2.

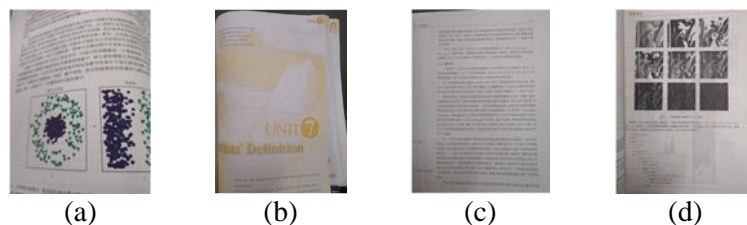


Figure 1. Books with different kinds of images or without images.

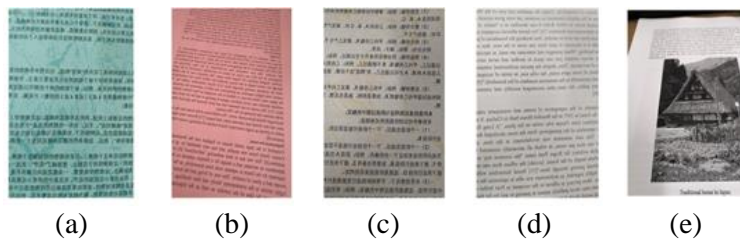


Figure 2. (a)-(d) show documents of different materials, (e) shows document with blurred image.

2.2. Type of Document

All images were taken by the following three smartphones:

(1) Xiaomi 6. It is equipped with dual camera. One camera has 12MP (Million Pixels) wide-angle lens and its aperture size is f/1.8. Its equivalent focal length is 27mm. The other camera has 12MP telephoto lens and its aperture size is f/2.6, and its equivalent focal length is 52mm.

(2) Xiaomi 8. It is equipped with dual camera. One camera has 12MP wide-angle lens and its aperture size is f/1.8. Its equivalent focal length is 27mm. The other camera has 12MP telephoto lens and its aperture size is f/2.4, and its equivalent focal length is 56mm.

(3) Huawei Honor v10. It is equipped with dual camera. The main camera has 16MP telephoto lens and its aperture size is f/1.8. The other has Black and White camera with 20MP.

In the process of capturing images, we set the camera mode of the smartphone to manual mode, and all the adjustable parameters except focal length are set to automatic. In order to get different degrees of blurred images, we only manually change the focal length of the camera and fixed the distance between the mobile device and the subject. The capture interfaces of the three mobile phones are shown in Figure. 3, and the specific parameter list of the images taken by the three devices is shown in Table 1.

Table 1. Properties of image taken by three mobile phones.

Parameter	XIAOMI 6	XIAOMI 8	HOSNOR V10
Resolution	4032×3096	3024×4032	3456×4608
Width	4032	3024	3456
Height	3096	4032	4608
Horizontal resolution	72dpi	72dpi	96dpi
Vertical resolution	72dpi	72dpi	96dpi
Bit depth	24	24	24
Color space	sRGB	sRGB	sRGB

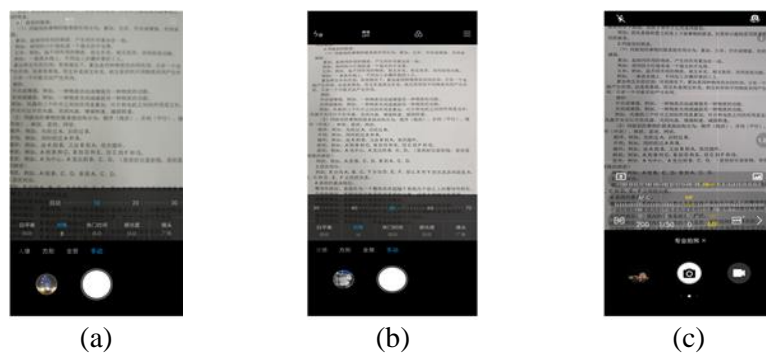


Figure 3. Screenshot of three different mobile phones.

2.3. Lighting Condition

Light sources will have an impact on the quality of the captured image, because different light sources have different color temperature. The color of the document image collected under different color temperature conditions will also have great differences. When the color temperature is too low, the overall image will appear yellow, while when the color temperature is too high, the image itself will appear blue (The color temperature scale table is shown in Figure 4).

In our work, we used the following three different light sources:

- (1) Daylight (about 5900K).
- (2) Tubular fluorescent lamps (about 5000K).
- (3) Soft white incandescent lamps (about 2400K).

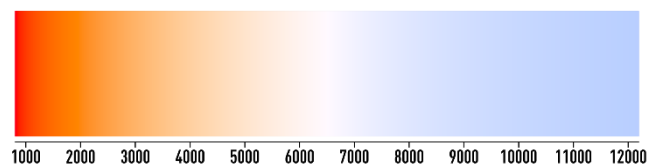


Figure 4. The color temperature on a linear scale.

2.4. MOS Obtaining

A strategy for evaluating image quality is provided in literature [22], which divides images into the following categories according to their quality: “Excellent”, “Good”, “Fair”, “Poor” and “Bad”. In our experiments, we take an approach to assess the quality of image distortion that convert the image quality of five categories (Excellent, Good, Fair, Poor and Bad) into a continuous range of 0 to 10, where 0 is the worst level, 10 is the best level and every 2 points is divided into one interval (See in Figure 5). Such experimental method also has certain defects that the observer ratings of the judgment may be different in different stage. In fact, in many cases, the observer evaluated a distorted image A as “Excellent” at begin phase. But he may evaluate quality of another distorted image B that has same quality as image A as “Fair” after he evaluated different degrees of distorted images. To reduce such deviations and ensure the quality of MOS, multiple observers should participate in the experiment.

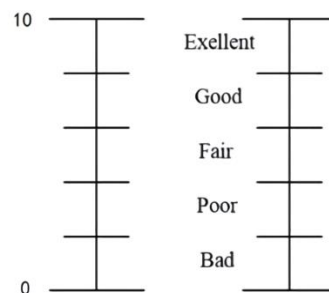


Figure 5. Grading scale.

In our experiments, 5 volunteers were invited as observers to participate in the image quality assessment experiment independently. The experiment is divided into several rounds, and the observer will rest for ten minutes before proceeding to the next round of experiments. Each experiment will take no more than 30 minutes, and the first 5 minutes were remained for the observer to train. In the training procedure, multiple images with different distortion degrees are given, which allows the observer to evaluate these distorted images first. At the formal experimental stage, the observer fully evaluates the image quality using the software, and gives a score between 0 and 10. This result is not recorded immediately, but is filtered after all the scores have been recorded. If there is an abnormal score which is more than 3 points behind the mean value of others, the outlier will be removed.

In the resulting MOS, the number of images with each score is concentrated between 1500 and 2500 which varies little. This shows that adjusting the focal length of the camera proportionally during image acquisition can effectively affect the quality of the image, which makes the image quality distribution relatively uniform. At the same time, the value of the image quality is concentrated in the two values of 3 and 7, which form two peaks, it indicating that the observer is more willing to give 3 instead of 0 for images with bad quality, and 7 instead of 9 or 10 for images with good quality.

Due to the resolution of the image captured by the smartphone reached 3456×4608 , and the resolution of the display device used by each of our observers was 1600×900 , the lack of resolution leads to the loss of part of the image information, which has a great impact on the subjective judgment of the observer. Therefore, we designed an auxiliary evaluation software, and the screenshot of this software is shown in Figure 6. The main display area of this software is divided into two parts. The left side is the overall display area of the image, and the right side can be used to zoom in, zoom out and pan the image, so that the observer can observe the details of different areas of the image. The final MOS is shown in Figure 7. The complete dataset can be downloaded from <http://cvbrain.cn>.

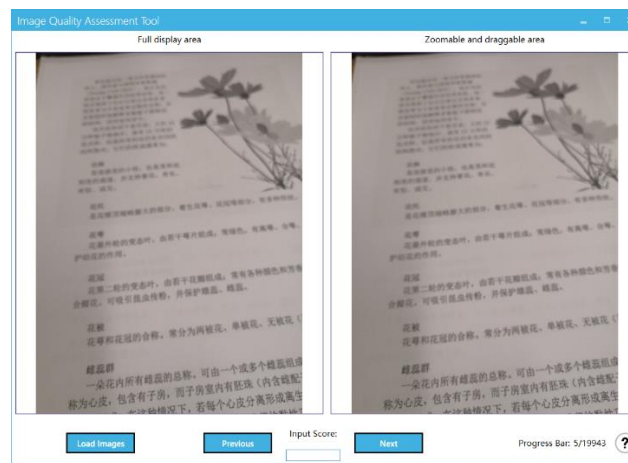


Figure 6. The screenshot of evaluation software.

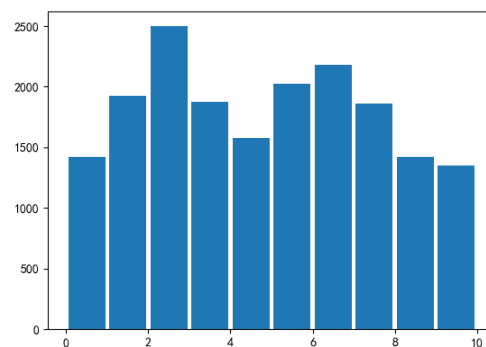


Figure 7. MOS histogram of our dataset

3. Image Quality Assessment Using the Dataset

In this section, we will select two no-reference quality assessment methods proposed recently to test our dataset. The following is a brief introduction of these two methods and experimental process and results.

3.1. Evaluation Metrics

We use the Spearman Rank Order Correlation Coefficient (SROCC) and Linear Correlation Coefficient (LCC) to evaluate the performance of this method on this dataset [23, 24]. SROCC and LCC are two metrics that are often used to evaluate the performance of methods in the field of image quality assessment. SROCC assesses how well the relationship between two variables can be described using a monotonic function, which is the monotonicity between the predicted value and the ground-truth. LCC is the covariance of the two variables divided by the product of their standard deviations. Therefore, it can well represent the accuracy of predicted value.

3.2. Experimental Methods

RankIQA. Liu et al [25] proposed a no-reference image quality assessment approach that learns from rankings. In the case of insufficient training data, the performance of this method is better than that of other methods without reference image quality evaluation, and even exceeds that of some methods with full reference image. We used the python version of the implementation code provided by the author, 80% of our dataset are used for training, and 20% are used for test.

BIQI. A. K. Moorthy and A. C. Bovik [26] proposed a new two-step framework for no-reference image quality assessment based on natural scene statistics (NSS) [27, 28]. Once the training is complete, the framework does not require any knowledge of the image distortion. Moreover, the framework is highly extensible and can be applied to any number of distortions. We used this method for cross-dataset [29] testing that LIVE dataset was used for training and our dataset was used for testing.

3.3. Results and Discussion

Table 2 summarizes the results of RankIQA method on our dataset, LIVE dataset and TID2013 dataset. We computed the SROCC and LCC scores on our testing set after training to convergence. This process is repeated ten times and the results are averaged, and the SROCC and LCC scores on Live and TID2013 dataset are provided from the author's paper. In this method, RankIQA performed well on LIVE and our dataset, the closer the value of LCC and SROCC is to 1, the closer the subjective score and the predicted score are similar to previous result, Table 3 shows the SROCC and LCC scores of BIQI method obtained by training on LIVE and testing on LIVE and our dataset. From the table we can see that the SROCC and LCC in LIVE and our dataset are very similar, which indicates that BIQI method has a good adaptability that can learn from arbitrary and non-IQA data.

Table 2 SROCC and PLCC of RankIQA Method on our dataset, LIVE and TID2013 dataset.

Evaluation Metrics	Our Dataset	LIVE	TID2013
SROCC	0.971	0.981	0.780
LCC	0.942	0.982	0.799

Table 3 SROCC and PLCC of BIQI Method obtained by training on LIVE and testing on LIVE and our dataset.

Evaluation Metrics	Our Dataset	LIVE
SROCC	0.824	0.819
LCC	0.825	0.820

4. Conclusions

To address the lack of data sets in mobile phone captured documents, we create a new document image data set, which can be used in the field of no-reference image quality evaluation and OCR. Compared to some previous image datasets, our dataset has the following innovations:

- (1) With more than 19,943 captured images, the size of dataset is far larger than previous datasets.
- (2) Documents with different materials and books are added as new categories.

(3) The condition that the texture on paper is inherently blurred.

In the field of image quality evaluation, many methods encounter the problem of insufficient training data before experiments [25, 30]. Our dataset can mitigate this problem to a large extent. We hope that this dataset can be helpful for future work in the field of image quality assessment. In addition, the ground-truth of images can provide researchers a reference to the relative quality of the images.

In the future, we're planning to add more document types, using more smartphones, and inviting more volunteers to participate in the experiment.

Acknowledgments

The authors acknowledge the financial supports of Science and Technology Department of Jiangxi Province under Grant No. 20202BABL202028, the Educational Bureau of Jiangxi Province under Grant NO. GJJ180516, and Innovative Fund Project for Postgraduates of Nanchang Hangkong University under Grant NO. YC2019029.

References

- [1] J.a.B. Kumar R and Ding H and Emmett P, 2013 Mobile Video Capture of Multi-page Documents *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*
- [2] Kumar J, Chen F and Doermann D 2012 Sharpness estimation for document and scene images in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*
- [3] He L, Gao F, Hou W, and Hao L 2014 Objective image quality assessment: a survey *International Journal of Computer Mathematics* **91**(11-12) 2374-88
- [4] Zhang Y, Wang H, Tan F, Chen W, and Wu Z 2019 No-Reference Image Sharpness Assessment Based on Rank Learning *2019 IEEE International Conference on Image Processing (ICIP)* pp 2359-2363
- [5] Lin K Y and Wang G 2018 Hallucinated-IQA: No-Reference Image Quality Assessment via Adversarial Learning
- [6] D. Yang D, Peltoketo V -T, and Kämäräinen J -K 2019 CNN-based cross-dataset no-reference image quality assessment *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)* pp 3913-3921
- [7] Liu X, Weijer J v D, and Bagdanov A D 2019 Exploiting unlabeled data in CNNs by self-supervised learning to rank *2019 IEEE Transactions on Pattern Analysis and Machine Intelligence* **41** 1862-78
- [8] Sheikh H R, 2003 LIVE Image Quality Assessment Database [Online]. Available: <http://live.ece.utexas.edu/research/quality>
- [9] N. a. L. Ponomarenko V and Zelensky A and Egiazarian K and Carli M and Battisti F, 2009 TID2008 - A database for evaluation of full-reference visual quality assessment metrics *Advances of Modern Radioelectronics* **10** 30-45
- [10] Ponomarenko N *et al.* 2013 Color image database TID2013: Peculiarities and preliminary results *European Workshop on Visual Information Processing*
- [11] P. a. A. Le Callet, Florent, Subjective quality assessment IRCCyN/IVC database 2005.
- [12] Liu X, Pedersen M, and Hardeberg J Y 2014 *CID:IQ – A New Image Quality Database* (Springer International Publishing)
- [13] Chandler E. C. L. a. D. M. 2009 Categorical image quality (CSIQ) database [Online]. Available: <http://vision.okstate.edu/csiq>.
- [14] Horita K S Y, and Kawayoke Y, "MICT image quality evaluation database." [Online]. Available: <http://mict.eng.u-toyama.ac.jp/mictdb.html>.
- [15] D. M. C. a. S. S. Hemami, A57 Dataset [Online]. Available: <http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html>.

- [16] Kumar J, Ye P, and Doermann D, 2013 *A Dataset for Quality Assessment of Camera Captured Document Images*. (Springer International Publishing)
- [17] Nayef N, Luqman M M, Prum S, Eskenazi S, and Ogier J M 2015 SmartDoc-QA: A Dataset for Quality Assessment of Smartphone Captured Document Images - Single and Multiple Distortions *The 6th International Workshop on Camera Based Document Analysis and Recognition (CBDAR 2015)*
- [18] Wang Z, Bovik A C, and Lu L, 2011 Why is image quality assessment so difficult? *IEEE International Conference on Acoustics*
- [19] Impedovo S, Ottaviano L, and Occhinegro S, 1991 Optical Character Recognition - A Survey," *International Journal of Pattern Recognition & Artificial Intelligence*
- [20] Huynh-Thu Q, Garcia M-N, Speranza F, Corriveau P, Raake A 2011 Study of rating scales for subjective quality assessment of high-definition video *IEEE Transactions on Broadcasting* **57**(1) 1-14
- [21] S Katsigiannis *et al* 2018 Interpreting MOS scores, when can users see a difference? Understanding user experience differences for photo quality *Quality & User Experience* **3**(1) 6
- [22] Union I T 2002 Methodology for the subjective assessment of the quality of television pictures *ITU-R Recommendation BT 500-11*
- [23] Fieller E C and Pearson E S 1957 Tests for Rank Correlation Coefficients I *Biometrika* **1-2** 1-2
- [24] Myers J L and Well A D 2013 Research Design and Statistical Analysis *Japanese Journal of National Medical Services* **55**
- [25] Liu X, Joost V D W, and Bagdanov A D 2017 RankIQA: Learning from Rankings for No-reference Image Quality Assessment *2017 IEEE International Conference on Computer Vision (ICCV)*
- [26] Moorthy A K and Bovik A C 2010 A two-step framework for constructing blind image quality indices *IEEE Signal Processing Letters* **17**(5) 513-6
- [27] Simoncelli E P and Olshausen B A 2001 Natural image statistics and neural representation," *Annual Review of Neuroence* **24**(1) 1193-1216
- [28] Mittal A, Moorthy A K, and Bovik A C 2012 No-reference image quality assessment in the spatial domain *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society* **21**(12) 4695
- [29] Kang L, Ye P, Li Y, and Doermann D 2014 Convolutional neural networks for no-reference image quality assessment *2014 IEEE Conference on Computer Vision and Pattern Recognition*
- [30] Lee J K a S 2017 Fully deep blind image quality predictor *IEEE Journal of Selected Topics in Signal Processing* **11**(1) 206-20