ISSN: 2414-1895 DOI: 10.6919/ICJE.202110_7(10).0029

Screen Image Quality Evaluation based on Text Detection

Yuxiao Yu[1,*], Meng Yuan[1,a]

[1]College of computer science and technology, Qingdao University, Qingdao 266000, China.

*yuxiaoyu_yu@163.com, [a]yuanmeng_926@163.com

Abstract

Unlike the natural images we are familiar with, the screen content image contains two parts: text and pictures, the algorithms for processing natural images cannot handle the screen content images well. This paper proposes a quality evaluation algorithm for non-reference screen content images. Use text detection to cover the text part of an image to generate an image containing only the picture part. For the picture part, use the same processing method. This can increase the number of data, solve the problem of insufficient data, and improve the performance of the algorithm. Two images with different content are put into two networks for training, and the predicted quality scores of the text part and the picture part are weighted and averaged as the quality score. The experimental results show that the algorithm in this paper has excellent performance in subjective quality assessment of screen content images.

Keywords

Image Quality Evaluation; Vgg16; Screen Content Image.

1. Introduction

Image quality evaluation algorithms are generally divided into two categories, one is the quality evaluation algorithm for natural images, and the other is the quality evaluation algorithm for screen content images [1]. In recent years, the quality evaluation algorithm of natural images has been well developed. However, natural image quality evaluation algorithms with better performance are difficult to transfer from natural images to screen content. Because the screen content image is composed of pictures and text, and the natural image is mainly composed of pictures [2]. This leads to a large natural statistical difference between the picture and the text area in the screen content image [24], which cannot be obtained by the natural image evaluation algorithm. However, screen content images are ubiquitous in various multimedia devices, such as smart phones, computers, tablets, etc. [8,9,10,11,12,13]. Therefore, the quality evaluation algorithm of the screen content image has aroused great attention and research in academia.

From the perspective of a multi-client communication system, due to limitations in network bandwidth, images received on multimedia are usually distorted in the process of acquiring, storing, transmitting, and encoding. Therefore, the quality evaluation algorithm of the screen content image has high research value [25]. The early non-reference quality evaluation algorithm was developed based on the classic regression algorithm [3]. Researchers try to design some manual features that can distinguish distorted images, and then predict image quality by training a regression model. With the development of quality evaluation technology, evaluation algorithms for certain distortions [4, 5] have emerged. Such algorithms usually use prior knowledge of one or several distortion types to predict image quality. Recently, in order to evaluate image quality without prior knowledge of distortion, related personnel have proposed a non-characteristic distortion algorithm [26]. BLIINDS-II [6] uses a simple Bayesian model to predict the image quality score of a given image. BRISQUE [7] quantified the loss of "naturalness" of the distorted image by calculating the local normalized

ISSN: 2414-1895 DOI: 10.6919/ICJE.202110_7(10).0029

brightness coefficient of the distorted image based on the algorithm of scene statistics, thereby evaluating the quality of the image. The above algorithm improves the quality evaluation performance of the screen content image, but does not take into account the large differences in the characteristics of different areas of the screen content image. And with the continuous improvement of model performance, the problem of insufficient training data has also been revealed. In order to solve the above problems, the SSEQ [14] algorithm proposes to divide the image, calculate the features of each block separately, and then perform statistical pooling on all the features, and finally obtain the joint features of the image. Although this algorithm solves the above problems, the performance of the algorithm does not improve very well.

Although the method of extracting local features by patch takes into account the differences in different regions, it still does not fully consider the differences between the text content and the picture content in the screen content image. Therefore, in this article, first extract both the text and the picture in the data set image, and then extract the different features of the text part and the picture part respectively. This not only takes into account the differences between the two parts, improves the performance of the model's prediction quality, but also increases the amount of data and solves the problem of insufficient data in training.

2. Related works

2.1 Network structure

The network structure proposed in this article is simple. As shown in Figure 1, it consists of two sub-networks. Each sub-network uses VGG16 [15] as the backbone network to extract regional features, and the fully connected layer behind VGG16 is responsible for score regression. The two sub-networks correspond to different parts of the image, respectively, evaluate the quality of the text and the image, and finally perform a weighted average to obtain the quality score of the entire image.

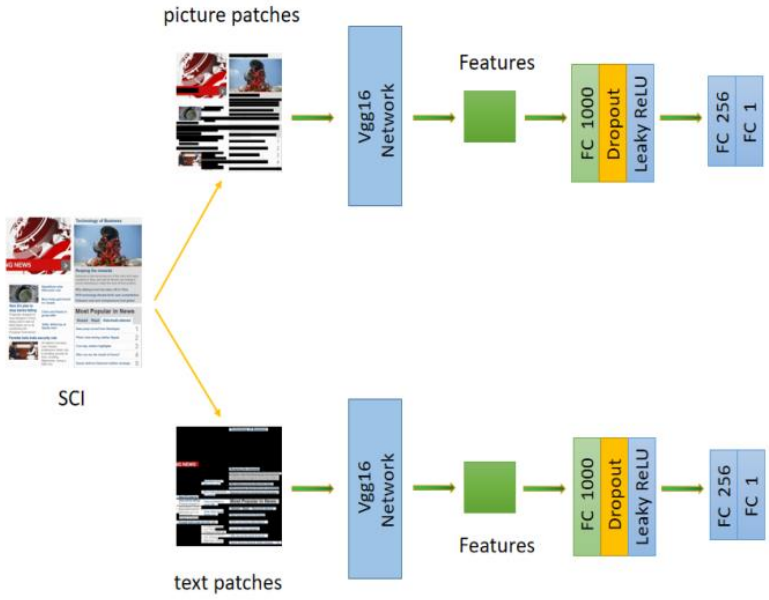

Fig. 1 Network structure

2.2 Data processing

As shown in Figure 2, Figure 3, this is the result of processing the data in this article.

In the text, the text detection algorithm is used to detect the text part and the picture part in the image respectively. First, turn the image part into black (all pixel values are zero), and only keep the text part of the image to get a new image, as shown in Figure 2. Then turn the text part of the image into black (the pixel value is all zero), and only keep the picture part in the image to get another new image, as shown in Figure 3.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202110_7(10).0029

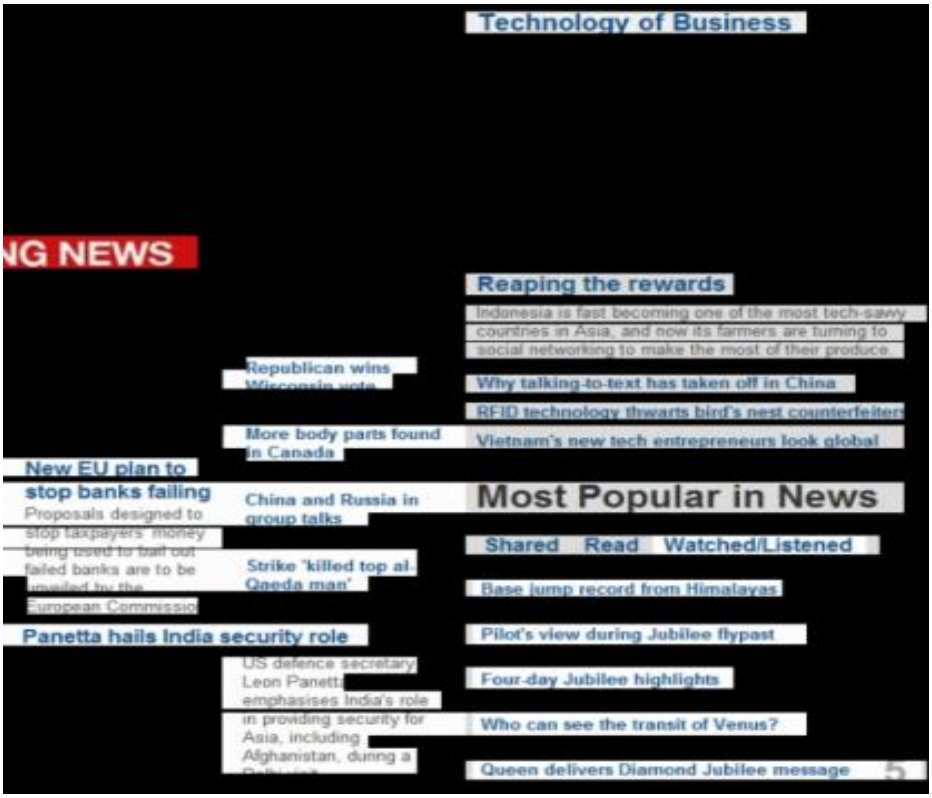

Fig. 2 The part of text

Fig. 3 The part of picture

Using the above methods to process data can not only double the data in the original data set, solve the problem of insufficient training data, but also take into account the problem of different features contained in different areas of the screen content image. The features of the text part and the picture part are extracted separately to characterize the features of the entire image, which can better predict the quality of the entire image and avoid the interference of different features in different regions on the overall image quality prediction.

2.3 Final result

Through the quality evaluation of the two sub-networks, the predicted quality scores of the text part and the picture part can be obtained respectively, and the weighted average of them can be used to

ISSN: 2414-1895 DOI: 10.6919/ICJE.202110_7(10).0029

obtain the predicted quality score of the entire image. The quality score of the final image is calculated as shown in formula 1:

$$\text{Score} = \frac{\text{Score}_{text}+\text{Score}_{pic}}{2} \quad (1)$$

Among them, $Score$ represents the quality score of the entire image, $Score_{text}$ represents the quality score of the text part of the image, $Score_{pic}$ represents the quality score of the picture part of the image.

The MSELoss loss function used in this article can be calculated as:

$$\text{Loss}(x_i, y_i) = (x_i - y_i)^2 \quad (2)$$

Among them, x_i represents the GT value of SCI and y_i represents a predicted score of SCI.

3. Experimental Analysis

3.1 Datasets

This article conducts experiments on two publicly available screen content image dataset, namely SIQAD and SCID.

The SIQAD [16] data set contains 7 common distortion types, and each type is divided into 7 levels according to the actual situation, a total of 980distortion images. Unlike SIQAD, the SCID [17] data set contains 1800 distorted images. There are 9 types of distortion, each of which is divided into 5 levels.

The 6 distortion types (GN, GB, MB, CC, JC, and J2C) at 5 different levels are considered to be the most common in the two datasets. In this paper, 1200 distorted images are randomly selected from the SCID data set as the test set for cross-data set experiments.

3.2 Evaluation

In the image quality evaluation algorithm, there are three generally accepted performance evaluation indicators:

1) Spearman rank correlation coefficient

2) Pearson correlation coefficient

3) Root mean square error

This article also uses these three indicators to evaluate the algorithm

SRCC is defined as follows:

$$SRCC = 1 - \frac{6\sum_{i=1}^{N} e_i^2}{N(N^2-1)} \quad (3)$$

e_i is the difference between the subjective score and the objective score.

PLCC is defined as follows:

$$PLCC = \frac{\sum_{i=1}^{N}(o_i-\bar{o})(s_i-\bar{s})}{\sqrt{\sum_{i=1}^{N}(o_i-\bar{o})\times\sum_{i=1}^{N}(s_i-\bar{s})}} \quad (4)$$

For the i image in the dataset (N images in total), o_i and s_i represent the objective quality score and the predicted subjective quality score, respectively, and $\bar{o}$ and $\bar{s}$ represent the mean value of o_i and s_i.

RMSE is defined as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(o_i-s_i)^2}{N}} \quad (5)$$

When SRCC=1, PLCC=1, and RMSE=0, the objective score is exactly the same as the subjective score, the quality evaluation algorithm is the best. Therefore, the closer the values of PLCC and SRCC are to 1, and the closer the RMSE value is to 0, the better the performance of the quality evaluation algorithm.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202110_7(10).0029

3.3 Cross-dataset experiments

In this section, experiments on SIQAD and SCID datasets are used to prove that the method in this article improves the performance of the model. The experimental results using the original data in the SIQAD and SCID datasets are compared with the experimental results of the method in this paper. The experimental results are shown in Table 1 and Table 2.

Table 1. SIAQD dataset

Type	SRCC↑	PLCC↑	RMSE↓
unprocessed data	0.784	0.791	9.207
processed data	0.838	0.854	7.572

Table 2. SCID dataset

Type	SRCC↑	PLCC↑	RMSE↓
unprocessed data	0.802	0.810	8.824
processed data	0.858	0.872	7.147

From the experimental results, it can be seen that on the two datasets of SIQAD and SCID, there is no data processed by this algorithm, and the network cannot distinguish different features of different regions well, resulting in inaccurate extracted features and poor algorithm performance. This result shows that the data processed by this algorithm is effective in image quality evaluation.

3.4 Performance comparison

The algorithm proposed in the text is compared with nine image quality evaluation algorithms on the public SIQAD and SCID databases. The 9 evaluation algorithms include the non-referenced NI evaluation algorithm and the non-referenced SCI evaluation algorithm. The compared models include BLINDS-II [6], BRISQUE [7], NRLT [18], HRFF [19], CNN-Kang [20], RankIQA+FT [21], WaDIQaM-NR [22] and BQACNN- Yue [23].

The experimental results are shown in Table 3 and Table 4. The quantitative comparison results are shown in detail in Table 3 and Table 4. We mark the indicators with the best performance in the table in bold. As shown in Table 3 and Table 4, the algorithm proposed in this paper has always been in a leading position in performance compared with the other 9 algorithms on the two public datasets.

Table 3. Comparison results of all models in the SIQAD dataset

	BLINDS-II	BRISQUE	NRLT	HRFF	CNN-Kang	RankIQA+FT	WaDIQaM-NR	BQACNN-Yue	Ours
SRCC↑	0.681	0.723	0.820	0.832	0.809	0.851	0.852	0.863	**0.872**
PLCC↑	0.725	0.771	0.844	0.852	0.849	0.878	0.859	0.883	**0.887**
RMSE↓	9.499	8.134	7.595	7.415	7.447	7.005	7.057	—	6.984

Table 4. Comparison results of all models in the SCID dataset

	BLINDS-II	BRISQUE	NRLT	HRFF	CNN-Kang	RankIQA+FT	WaDIQaM-NR	BQACNN-Yue	Ours
SRCC↑	0.708	0.737	—	—	0.813	0.838	0.860	—	0.882
PLCC↑	0.715	0.798	—	—	0.827	0.866	0.872	—	0.890
RMSE↓	9.324	8.016	—	—	7.196	6.914	7.013	—	6.871

4. Conclusion

The VGG16-based screen content image quality evaluation algorithm proposed in this paper processes a screen content image to obtain two new screen content images containing only text and pictures, which not only extracts the features of different areas of the screen content image, but also It also solves the problem of insufficient training data. Put two different images into different

ISSN: 2414-1895 DOI: 10.6919/ICJE.202110_7(10).0029

networks for training, so as to better predict the quality score of the screen content image. The experimental results show that the method in this article is superior in multiple indicators. However, it is not well applicable to individual distortion types. In future work, the model will need to be improved to achieve better performance of the model.

References

[1] Zhu YY, Cao L, Wang X. No Reference Screen Content Image Quality Assessment[J]. Journal of Software, 2018, 29(4).

[2] Zheng L, Shen L, Chen J, et al. No-reference quality assessment for screen content images based on hybrid region features fusion[J]. IEEE Transactions on Multimedia, 2019, 21(8): 2057-2070.

[3] Zeng H, Zhang L, Bovik A C. A probabilistic quality representation approach to deep blind image quality prediction[J]. arXiv e-prints, 2017: arXiv: 1708.08190.

[4] Li L, Lin W, Wang X, et al. No-reference image blur assessment based on discrete orthogonal moments[J]. IEEE transactions on cybernetics, 2015, 46(1): 39-50.

[5] Liu H, Klomp N, Heynderickx I. A no-reference metric for perceived ringing artifacts in images[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2009, 20(4): 529-539.

[6] Moorthy A K, Bovik A C. Blind image quality assessment: From natural scene statistics to perceptual quality [J]. IEEE transactions on Image Processing, 2011, 20(12): 3350-3364.

[7] Mittal A, Moorthy A K, Bovik A C. No-reference image quality assessment in the spatial domain[J]. IEEE Transactions on image processing, 2012, 21(12): 4695-4708.

[8] Chen C, Li S, Qin H, et al. Real-time and robust object tracking in video via low-rank coherency analysis in feature space[J]. Pattern Recognition, 2015, 48(9): 2885-2905.

[9] Chen C, Li S, Qin H, et al. Robust salient motion detection in non-stationary videos via novel integrated strategies of spatio-temporal coherency clues and low-rank analysis[J]. Pattern recognition, 2016, 52: 410-432.

[10] Chen C, Li Y, Li S, et al. A novel bottom-up saliency detection method for video with dynamic background[J]. IEEE Signal Processing Letters, 2017, 25(2): 154-158.

[11] Fang Y, Yan J, Li L, et al. No reference quality assessment for screen content images with both local and global feature representation[J]. IEEE Transactions on Image Processing, 2017, 27(4): 1600-1610.

[12] Hu H, Wen Y, Luan H, et al. Toward multiscreen social TV with geolocation-aware social sense[J]. IEEE MultiMedia, 2014, 21(3): 10-19.

[13] Miao D, Fu J, Lu Y, et al. A high-fidelity and low-interaction-delay screen sharing system[J]. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2016, 12(3): 1-23.

[14] Liu L, Liu B, Huang H, et al. No-reference image quality assessment based on spatial and spectral entropies [J]. Signal Processing: Image Communication, 2014, 29(8): 856-863.

[15] Jiang X, Shen L, Yu L, et al. No-reference screen content image quality assessment based on multi-region features [J]. Neurocomputing, 2020, 386: 30-41.

[16] Yang H, Fang Y, Lin W. Perceptual quality assessment of screen content images[J]. IEEE Transactions on Image Processing, 2015, 24(11): 4408-4421.

[17] Ni Z, Ma L, Zeng H, et al. ESIM: Edge similarity for screen content image quality assessment[J]. IEEE Transactions on Image Processing, 2017, 26(10): 4818-4831.

[18] Fang Y, Yan J, Li L, et al. No reference quality assessment for screen content images with both local and global feature representation[J]. IEEE Transactions on Image Processing, 2017, 27(4): 1600-1610.

[19] Zheng L, Shen L, Chen J, et al. No-reference quality assessment for screen content images based on hybrid region features fusion [J]. IEEE Transactions on Multimedia, 2019, 21(8): 2057-2070.

[20] Kang L, Ye P, Li Y, et al. Convolutional neural networks for no-reference image quality assessment [C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 1733-1740.

[21] Liu X, Van De Weijer J, Bagdanov A D. Rankiqa: Learning from rankings for no-reference image quality assessment[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 1040-1049.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202110_7(10).0029

[22] Bosse S, Maniry D, Müller K R, et al. Deep neural networks for no-reference and full-reference image quality assessment [J]. IEEE Transactions on image processing, 2017, 27(1): 206-219.

[23] Yue G, Hou C, Yan W, et al. Blind quality assessment for screen content images via convolutional neural network[J]. Digital Signal Processing, 2019, 91: 21-30.

[24] Chen C, Zhao H, Yang H, et al. Full-reference Screen Content Image Quality Assessment by Fusing Multilevel Structure Similarity[J]. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2021, 17(3): 1-21.

[25] Hou W, Gao X, Tao D, et al. Blind image quality assessment via deep learning[J]. IEEE transactions on neural networks and learning systems, 2014, 26(6): 1275-1286.

[26] Gu K, Zhai G, Lin W, et al. Learning a blind quality evaluation engine of screen content images[J]. Neurocomputing, 2016, 196: 140-149.