Modeling Human Visual System in Patch-base Image Quality Assessment using Deep Learning

Nguyen Trung Nghia

April 24, 2019

Authorship

"I hereby declare that the work contained in this thesis is of my own and has not been previously submitted for a degree or diploma at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no materials previously published or written by another person except where due reference or acknowledgement is made."

Signature: ....

Supervisor's approval

"I hereby approve that the thesis in its current form is ready for committee examination as a requirement for the Bachelor of Computer Science degree at the University of Engineering and Technology."
Signature: ....

Acknowledgement

I am grateful to thank the Department of Computational Science and Engineering, and HMI Lab both at the VNU UET for the support.

I would like to express our sincere thanks to our advisor Ph.D. Le Thanh Ha and M.Sc. Pham Thanh Tung for their support and guidance throughout this research work.

Abstract

As humans are the ultimate receivers of the majority of visual signals being processed, the most accurate way of assessing image quality is to ask humans for their opinions of an image's quality, known as the subjective visual quality assessment (VQA). The subjective image quality scores gathered from all subjects are processed to be the mean opinion score (MOS), which is regarded as the ground truth of image quality. Conventionally, a number of full-reference image quality assessment (FR-IQA) methods adopted various computational models of the human visual system (HVS) from psychological vision science research.

Due to the fact that the human visual system (HVS) is differently sensitive to features of image patch, we propose Deep Image Patch Quality Assessment (DIPQA), a novel image patch quality assessment that used deep neural network-based approach. An experimental quality assessment to approach database for image patch has been developed. The network is train end-to-end and comprises 8 convolutional layers and 4 pooling layers for feature extraction, and 2 fully connected layers for regression, which make it significantly deep enough to learn the mean opinion score (MOS) of the developed dataset.

We promise that this project was contributed by all members in our group, which are supervised by Ph.D. Le Thanh Ha and M.Sc. Pham Thanh Tung from HMI Lab. The report contains no materials previously published or written by another person except where due reference or acknowledgement is made.

Contents

List of Figures

List of Tables

Abbrevations

Chapter 1

Introduction

With an increase in the popularity of smartphones, compact cameras, and Internet services like Facebook and Instagram, past few years have seen tremendous growth in the production and sharing of digital images. The journey of a picture begins with it being obtained by a camera, which changes over it into a digital format and compresses it utilizing lossy compression algorithms to meet the onboard storage accessibility. This image is then transmitted over wired or wireless transmission channels and is altered in its resolution to meet the available bandwidth. Finally, the end user receives this image and watches it over devices ranging from smartphones to 4K displays, which require further alterations to its resolution. The end users tend to become more inclined towards the selection of a content provider, a service provider, and a display device that could better satisfy their expectations of image quality at delivery. Thus it becomes crucial for all content providers, service providers, and display providers to optimize these respective technologies towards the provision of perceptually good results, and to do so, perceptual image quality needs to be estimated. Furthermore, this estimation process should be automated, as much as possible, to make it independent from the availability of human observers in order to determine the perceptual quality.

Image quality assessment (IQA) aims to measure the perceived visual signal quality according to its statistical characteristics and human perceptual mechanism, which is widely required in numerous image processing applications. IQA plays a vital role in guiding many visual processing algorithms and systems, as well as their implementation, optimization and verification [7, 9, 10, 20]. In particular, image compression is one of the most representative applications of IQA, which can be utilized in the rate-distortion optimization process to obtain compressed images with better visual quality at the same bit-rate level [2, 3, 11, 18, 19, 8]. The traditional image compression methods mainly utilize the signal-fidelity based quality metrics, which are less correlated with human perceptual quality, e.g., MAE (mean absolute error), MSE (mean square error), SNR (signal-to-noise ratio), PSNR (peak SNR) and their relatives. Although these metrics possess many favorable properties, e.g., clear physical meaning and high efficiency for calculation, they severely hinder the compression performance improvement in further reducing the visual redundancies in images due to their poor consistency with human visual perception.

To obtain more consistent measures with human visual perception, many perceptual quality metrics have been proposed during the recent years. According to the availability of a reference image, these methods can be divided into three categories, i.e., full reference (FR) ones where the pristine reference image is available, reduced reference (RR) ones where partial information of the reference image is available and no reference (NR) ones where the reference image is unavailable. For image compression problem, the reference images are available at the encoder side such that the FR-IQA algorithms are applicable.

Many FR-IQA based algorithms have been proposed over time. One class of these algorithms including SSIM [11], FSIM [16], RFSIM [15] use handcrafted features (attributes (edge, color, etc.) in data (images) that are relevant to the modeling problem) that supposedly captures relevant factors affecting image quality. Although their performance is acceptable, there is still large room for improvement regarding the accuracy with which they reproduce human judgment of quality. Another set of algorithms, including convolutional neural network (CNN) based approaches [1, 6], employ automatic learning of features from the raw image pixels, which are superior and more efficient as they make feature selection automatic and embedded within the system itself.

1.1 Motivation

Most of the existing IQA databases usually contain limited distortion levels (5-6 levels) covering the whole quality range from “Bad” to “Excellent”, which make the images in adjacent distortion levels obviously different and easy to rank. To describe the obvious and subtle quality differences between two images, Zhang *et al* [17] use the terms "coarse-grained" and "fine-grained". More specifically, the images with "coarse-grained" quality differences correspond to the compressed ones generated using the same codec at obvious different bitrates, while the images with "fine-grained" quality difference correspond to the compressed ones generated using different optimization methods at the same bitrate. Therefore, these databases with coarse-grained distortion variations for the same image may not be able to provide sufficient information to further improve the performance of IQA algorithms in evaluating fine-grained quality differences.

Another weakness for the existing IQA databases is that they only contain a few reference images with limited visual content. To solve this problem patch-based methods are gradually used in IQA, e.g. CNN-IQA [5], CORNIA [14] The patch-based learning methods requires the 'ground truth' of patch quality for training but there are only the ground truth quality of images instead of patches in IQA datasets. To deal with this problem, existing works usually assign the image quality score to all patches in this image as their 'ground truth', e.g. CNN-IQA [5]. This approach might introduce much noise in patches labels because in some distortion types the quality of patches in one image varies much and the patches quality score can't be simply assigned as the image quality core.

Based on all these observations, this project promotes IQA in the new challenges of fine-grained quality assessment task by constructing a large-scale Image-Patch Quality Assessment database with fine-grained distortion differences. I also analyze 7 state-of-the-art IQA algorithms on the proposed database and show that there is still a large room to improve the IQA in the prediction of the fine-grained quality preference. Finally, I propose an Image-Patch model to help estimate the 'ground truth' quality of patches based on a state-of-the-art CNN architecture.

1.2 Contributions

1.3 Thesis Outline

Citing to the article [?] and citing to the graph above 1.1.

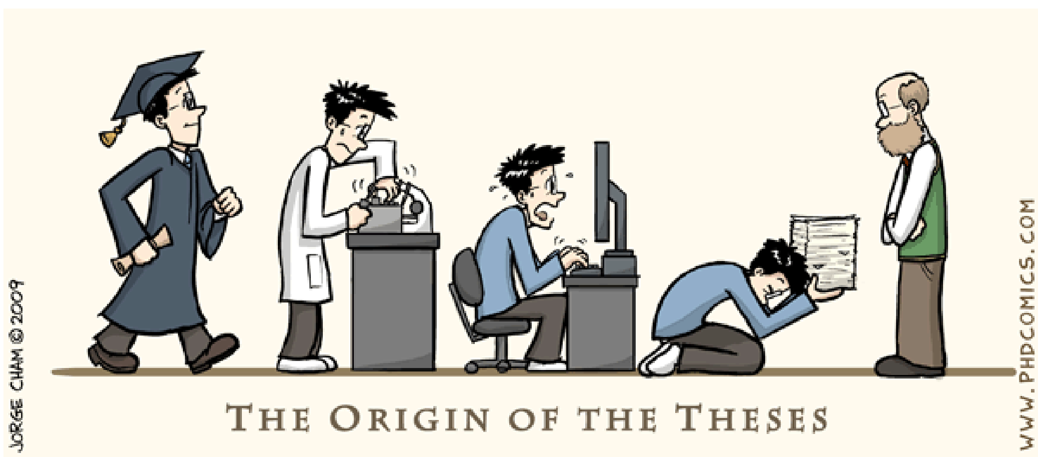

Figure 1.1: Caption of this graph.

Chapter 2

Background

In this chapter, ...

Column 1	*Column 2*	*Column 3*	*Column 4*
Row 1.1	Row 1.2	Row 1.3	Row 1.4
Row 2.1	Row 2.2	Row 2.3	Row 2.4

Table 2.1: Caption of this table.

Citing to the table above 2.1

Chapter 3

Proposed method

Some mathematics formula

$$F(x) = \arg\max_{y \in GEN(x)} w \cdot f(x, y)$$

Chapter 4

Evaluation

In this chapter, ...

Chapter 5

Conclusion

In this chapter, ...

Bibliography

[1] Sebastian Bosse, Dominique Maniry, Klaus Robert Müller, Thomas Wiegand, and Wojciech Samek. Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment. *IEEE Transactions on Image Processing*, 27(1):206–219, 2018.

[2] Sumohana S. Channappayya, Alan Conrad Bovik, and Robert W. Heath. Rate bounds on SSIM index of quantized images. *IEEE Transactions on Image Processing*, 2008.

[3] Zhenzhong Chen, Weisi Lin, and King Ngi Ngan. Perceptual video coding: Challenges and approaches. In *2010 IEEE International Conference on Multimedia and Expo, ICME 2010*, 2010.

[4] Q. Huynh-Thu and M. Ghanbari. Scope of validity of PSNR in image/video quality assessment. *Electronics Letters*, 2008.

[5] Le Kang, Peng Ye, Yi Li, and David Doermann. Convolutional neural networks for no-reference image quality assessment. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014.

[6] Le Kang, Peng Ye, Yi Li, and David Doermann. Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks. In *Proceedings - International Conference on Image Processing, ICIP*, 2015.

[7] Weisi Lin and C. C. Jay Kuo. Perceptual visual quality metrics: A survey. *Journal of Visual Communication and Image Representation*, 2011.

[8] Siwei Ma, Xinfeng Zhang, Jian Zhang, Chuanmin Jia, Shiqi Wang, and Wen Gao. Nonlocal in-loop filter: The way toward next-generation video coding? In *IEEE Multimedia*, 2016.

[9] Haiqiang Wang, Xinfeng Zhang, Chao Yang, and C. C.Jay Kuo. Analysis and Prediction of JND-Based Video Quality Model. In *2018 Picture Coding Symposium, PCS 2018 - Proceedings*, 2018.

[10] Shiqi Wang, Ke Gu, Xinfeng Zhang, Weisi Lin, Siwei Ma, and Wen Gao. Reduced-Reference Quality Assessment of Screen Content Images. *IEEE Transactions on Circuits and Systems for Video Technology*, 2018.

[11] Shiqi Wang, Abdul Rehman, Zhou Wang, Siwei Ma, and Wen Gao. SSIM-motivated rate-distortion optimization for video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 2012.

[12] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 2004.

[13] Wen Heng; Tingting Jiang. FROM IMAGE QUALITY TO PATCH QUALITY : AN IMAGE-PATCH MODEL FOR NO-REFERENCE IMAGE QUALITY ASSESSMENT Wen Heng and Tingting Jiang National Engineering Laboratory for Video Technology , Cooperative Medianet Innovation Center , School of Electronics Engineer. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2017*, pages 1238–1242, 2017.

[14] Peng Ye, Jayant Kumar, Le Kang, and David Doermann. Unsupervised feature learning framework for no-reference image quality assessment. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012.

[15] Lin Zhang, Lei Zhang, and Xuanqin Mou. RFSIM: A feature based image quality assessment metric using Riesz transforms. In *Proceedings - International Conference on Image Processing, ICIP*, 2010.

[16] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. FSIM: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 2011.

[17] Xinfeng Zhang, Weisi Lin, Shiqi Wang, Jiaying Liu, Siwei Ma, and Wen Gao. Fine-Grained Quality Assessment for Compressed Images. *IEEE Transactions on Image Processing*, 28(3):1163–1175, 2019.

[18] Xinfeng Zhang, Shiqi Wang, Ke Gu, Weisi Lin, Siwei Ma, and Wen Gao. Just-Noticeable Difference-Based Perceptual Optimization for JPEG Compression. *IEEE Signal Processing Letters*, 2017.

[19] Xinfeng Zhang, Ruiqin Xiong, Weisi Lin, Jian Zhang, Shiqi Wang, Siwei Ma, and Wen Gao. Low-Rank-Based Nonlocal Adaptive Loop Filter for High-Efficiency Video Compression. *IEEE Transactions on Circuits and Systems for Video Technology*, 2017.

[20] Yabin Zhang, Weisi Lin, Xinfeng Zhang, Yuming Fang, and Leida Li. Aspect Ratio Similarity (ARS) for image retargeting quality assessment. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2016.