

**VIETNAM NATIONAL UNIVERSITY, HANOI
UNIVERSITY OF ENGINEERING AND TECHNOLOGY**



Nguyen Trung Nghia

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

Major: Computer Science

HA NOI - April 24, 2019

**VIETNAM NATIONAL UNIVERSITY, HANOI
UNIVERSITY OF ENGINEERING AND TECHNOLOGY**

Nguyen Trung Nghia

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

Major: Computer Science

Supervisor: Assoc. Prof., Ph.D. Le Thanh Ha

Co-Supervisor: Assoc. Prof., M.Sc. Pham Thanh Tung

HA NOI - 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 | 6 |
| List of Tables | 7 |
| 1 Introduction | 9 |
| 1.1 Motivation | 10 |
| 1.2 Contributions | 11 |
| 1.3 Thesis Outline | 12 |
| 2 Background | 13 |
| 2.1 Image Quality Assessment | 13 |
| 2.2 Neural Networks | 13 |
| 2.3 Training of Neural Networks | 13 |
| 2.4 Convolutional Neural Networks | 13 |
| 3 Methodology | 14 |
| 3.1 Database Construction | 14 |
| 3.1.1 Testing image database | 14 |
| 3.1.2 Database creation | 15 |
| 3.1.3 Testing methodology | 16 |
| 3.1.4 Benchmark Analyses | 17 |
| 3.2 Deep Image-Patch Quality Assessment | 18 |
| 3.2.1 Architecture | 18 |
| 3.2.2 Feature Fusion | 19 |
| 3.2.3 Training Method | 19 |
| 4 Evaluation | 20 |
| 4.1 Evaluation Method | 20 |
| 4.2 Experiment results | 21 |
| 4.2.1 HMII Benchmark Analysis | 21 |
| 4.2.2 DIPQA Performrance Evaluation | 21 |
| 5 Conclusion | 22 |

List of Figures

| | | |
|-----|---|----|
| 1.1 | Caption of this graph. | 11 |
| 3.1 | Selected image patch. | 15 |
| 3.2 | Testing software screenshot. | 17 |
| 3.3 | Deep Image-Patch Quality Assessment Network Architecture. | 18 |

List of Tables

| | | |
|-----|---|----|
| 3.1 | Example of testing image database. | 16 |
| 4.1 | Comparing different IQA algorithms | 21 |
| 4.2 | Comparing different Full-Reference approaches | 21 |

Abbreviations

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 [12, 18, 19, 32]. 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 [4, 5, 20, 30, 31, 13]. 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 [20], FSIM [28], RFSIM [27] 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, 10], 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* [29] 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 [9], CORNIA [24]. 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 [9]. 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. We 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, we propose an FR Image-Patch model to help estimate the ‘ground truth’ quality of patches based on a state-of-the-art CNN architecture.

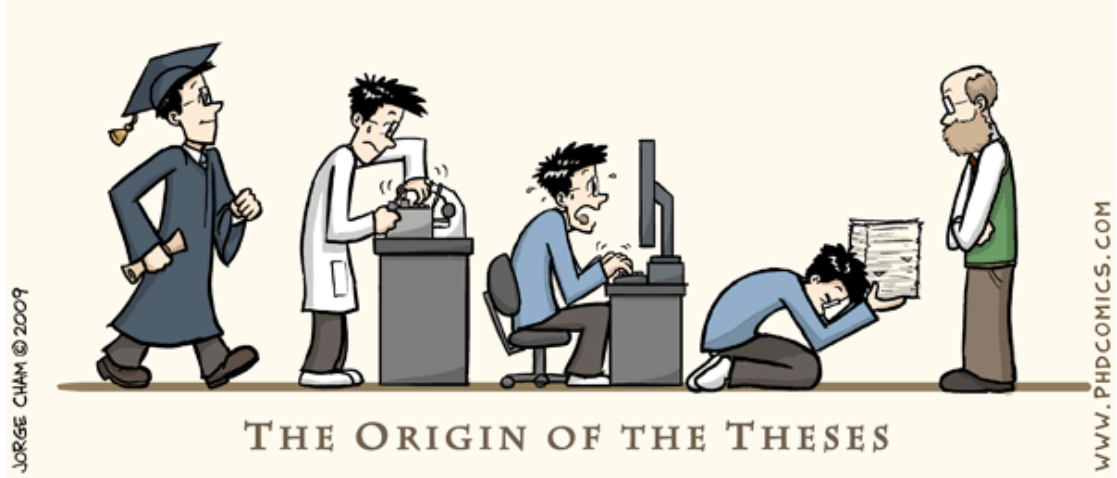


FIGURE 1.1: Caption of this graph.

citing to the graph above [1.1](#).

1.2 Contributions

This thesis provides the following contributions:

1. Image-Patch Quality Assessment dataset

To our knowledge, this dataset is the first one constructing to provide benchmark for compressed image patch quality assessment, and also benefit for perceptual-based image compression. The existing databases with coarse-grained quality are inefficient to evaluate IQA algorithms especially patch-based methods on images with fine-grained quality differences. In perceptual-based image compression problem, for each coding block there are many coding modes to select according their rate-distortion costs. Therefore, the proposed dataset can help researchers in image compression community to select the best IQA method to do the perceptual based image optimization. 7 well-know IQA algorithms are evaluated and analyzed on the proposed database to reveal some limitations of the existing algorithms.

2. Deep Image-Patch Neural Network Design

We also investigate different FR methods to model the relationship between the image patch and patch quality score. After multiple of experiments, Deep Image-Patch Quality Assessment

(DIPQA) is proposed to address the problem in and end-to-end optimization. We adapt the concept of Siamese networks known from classification task [2, 6] that allow for a joint regression of the features extracted from the reference and distorted patch using a deep convolution neural network.

1.3 Thesis Outline

The rest of this thesis is organized as follows. After this introduction, we present the literature review in Chapter 2 in which we introduce the fields of image quality assessment and deep learning.

Chapter 2

Background

2.1 Image Quality Assessment

2.2 Neural Networks

2.3 Training of Neural Networks

2.4 Convolutional Neural Networks

Chapter 3

Methodology

3.1 Database Construction

All available image quality benchmark databases are only suitable for evaluating the quality of images as a whole and not able to investigate which parts of the testing image contribute to the testing results or the score for a particular patch of image. In this project, we set up an experimental database to evaluate the quality that human perceive for each image patch.

3.1.1 Testing image database

A good database for testing is critical to be the success of the research. Due to the research orientation for video encoding, testing images are cropped from extracted frames in the video test sequence and noise types are added to the original video by H265/HEVC compression before extracting. In this database, we randomly select several patches from each image so that the database includes at least three attributes: smooth texture, complex texture and edge texture.

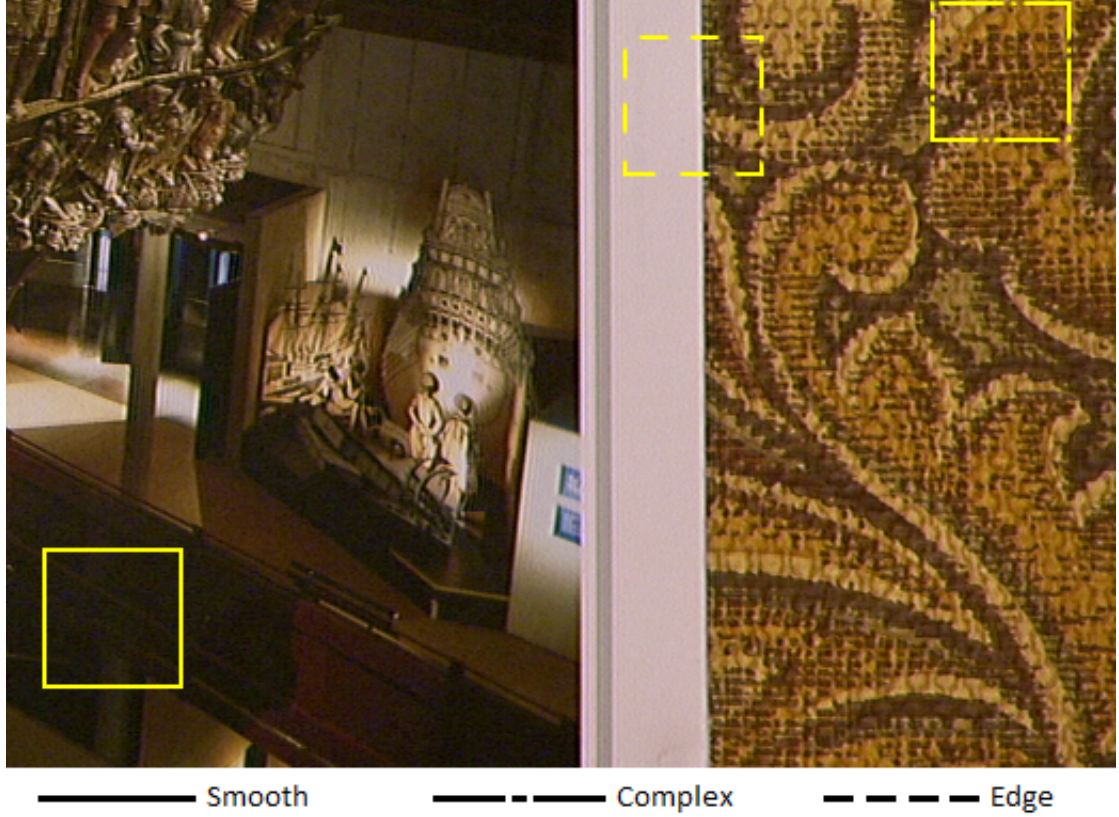


FIGURE 3.1: Selected image patch.

3.1.2 Database creation

There are 40 original source videos of high-definition (1280x720) and full high-definition (1920x1080) being compressed by H.265/HEVC with different Quantization parameters (QPs) with the range from 2 to 50. Testing images are extracted from those testing video sequences. For each video sequence, we select a different number of frames depend on the original video, this number drops in the values from 5 to 15. After that, we select random positions of the image to crop different 128x128 patches of each pair of image. We also crop the center 64x64 patches from the original pair 128x128 to evaluate in the experiments. Finally, we obtain 161,144 images: 40286 pairs of 64x64 patches and 40286 pairs of 128x128 patches.



















| Original | QP | | | | | | |
|---|---|---|---|---|--|---|---|
| | 25 | 30 | 35 | 40 | 45 | 50 | 55 |
|  | | |  |  |  |  |  |
|  | | |  |  |  |  |  |
|  |  |  |  |  |  | | |

TABLE 3.1: Example of testing image database.

3.1.3 Testing methodology

Depending on the nature of the test, observers may be expert or non-expert. Studies have found that systematic differences can occur between different laboratories conducting similar tests [3]. One of the reasons for this is that expert observers have different view in compare with no-experts. Other explanations may include gender, age, and occupation. However, in reality, the majority of consumers should be non-expert observers are chosen for this experiment. Before final selection, all candidates have been checked to ensure that they possess normal visual acuity.

For the purpose of this experiment, 1200 subjects who are undergraduates, graduates, researchers, and lecturers of University of Fire Fighting and Prevention are employed. These subjects have been trained and practiced quality assessment of several sample images.

For the purpose of subject testing methodology, the International Telecommunication Union set the ITU-R BT.500-11 [3] standard. In such standard, there are several popular subjective methodologies for testing such as Single stimulus categorical rating, Double stimulus categorical rating, Ordering by force-choice pairwise comparison and Pairwise similarity judgments. Double stimulus categorical rating is chosen in this practical. In this method, both the test and reference images are displayed for a fixed amount of time. After that, the images disappear from the screen and observers are asked to rate the quality of the test image according to the abstract scale containing one of the five categories: excellent, good, fair, poor or bad. All those images are displayed randomly. At the beginning of each session, an explanation is given to the observers about the type of assessment, the grading scale, the sequence and timing (reference image, grey, test image, voting period).

The previous image quality assessment methods are only suitable for assess quality of image as a whole. It cannot be directly applied for our testing experiments. Therefore, we modify this image selection method in the standard so that the users can only concentrate and assess the local image patch instead of the whole image. Each pair quality is assessed with the following procedure: The subjects observe the original image within the time T1 at minimum 5s then

click on the observing image patch to observe the compressed image within the time T2. After watching at least twice per image, the observers would score on scale of 5 as in Fig.3.2.

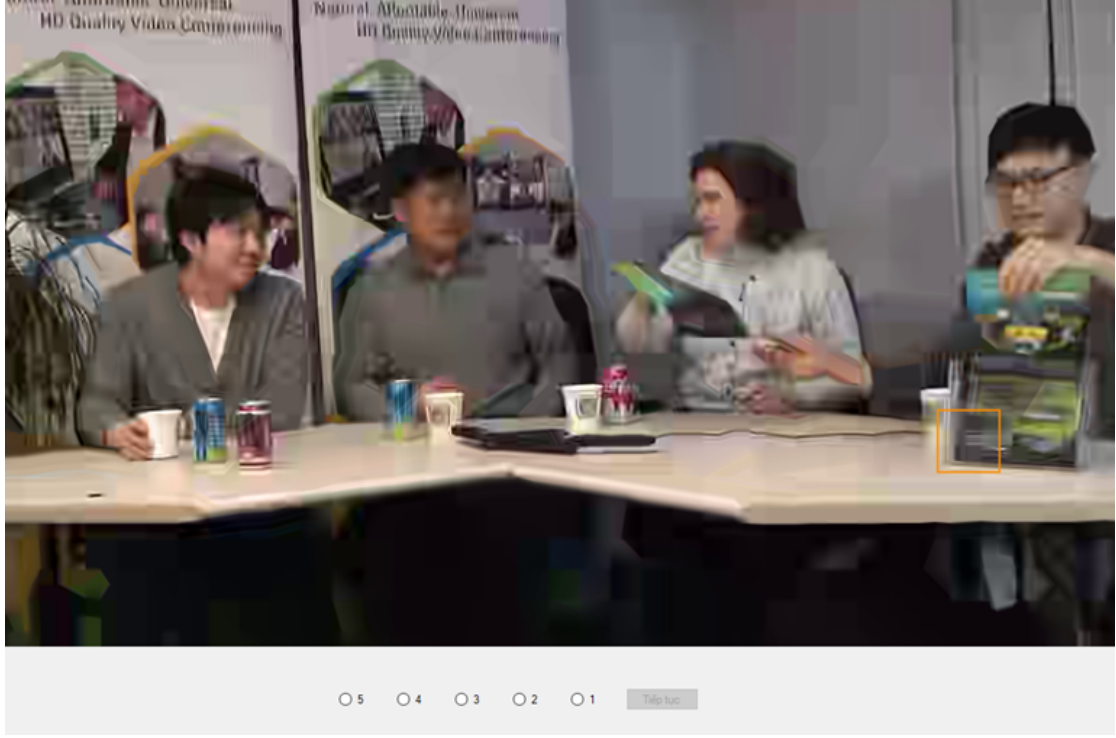


FIGURE 3.2: Testing software screenshot.

At the end of experiment, each pair is scored by the mean of the DMOS that up to 20 subjects give during the experiment.

We carefully select 1511 pairs which are scored by at least 10 people and name this sub-database HMII (Human Machine Interaction Image). HMII is used to evaluate well-known IQA algorithms and our methods in the second contribution.

3.1.4 Benchmark Analyses

To analyze the efficiency of IQA algorithms, we apply 7 state-of-the-art full reference image quality assessment methods on the proposed HMII database, to investigate their performance and demonstrate the new challenges in fine-grained image-patch quality assessment problem. The FR-IQA algorithms include PSNR, SSIM [22], RFSIM [27], FSIM [28], SRSIM [25], UQI [21], VSI [26]. The implementations of all algorithms are obtained from the public websites.

3.2 Deep Image-Patch Quality Assessment

3.2.1 Architecture

Being known as a designed architecture to learn the similarity relations between two given inputs, Siamese network has been applied for face verification [6] and signature [2] tasks. The main concept is processing two networks that share the same architecture and weights parallel. In this work, we employ Siamese network for feature extraction. Before feeding the extracted features as input to the regression layers, feature extraction is followed by a feature fusion step. The proposed architecture of DIPQA is sketched in Fig.3.3

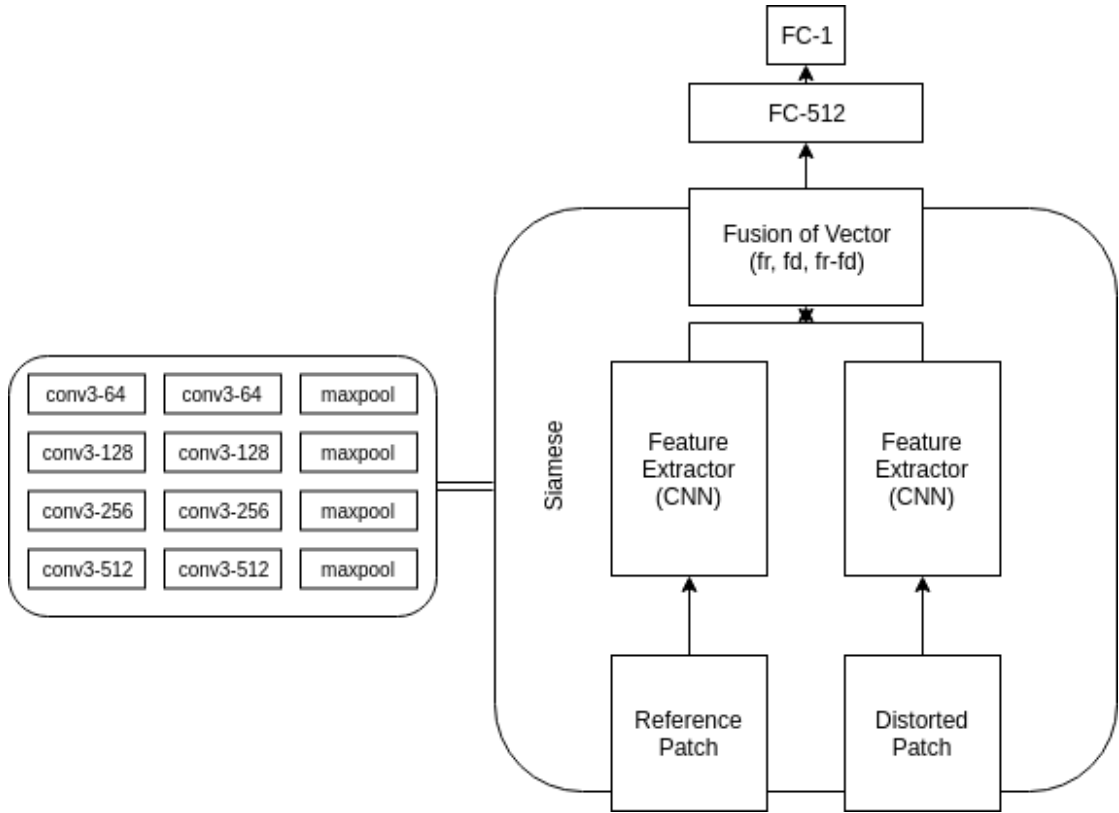


FIGURE 3.3: Deep Image-Patch Quality Assessment Network Architecture.

With the successful adaptation for various computer vision tasks [7, 15], especially in image quality assessment [1], VGGnet [16] was chosen as a base network for the feature extraction. The input of the VGG network is the size of 224 x 224 pixels. For the purpose of adjusting the network for 64 x 64 and 128 x 128 pixels, we have tried to change the architecture of VGG network such as: extend the network by 3 layers [1], cut last 3 layers, last 6 layers or even replace VGG with Resnet. Finally, we choose to cut the last 3 layers of VGGnet and achieve the best performance comparing to the other approaches. Our VGGnet-inspired DCNN comprised 12 weight layers as a feature extraction module and a regression module. The features are extracted in a series of conv3-64, conv3-64, maxpool, conv3-128, conv3-128, maxpool,

conv3-256, conv3-256, maxpool, conv3-512, conv3-512, maxpool, layers. The fused features are regressed by a sequence of two fully connected layers (FC-512, FC-1). This results in about 17.3 million trainable network parameters. All convolutional layers apply 3x3 pixel-size convolution kernels and are activated through a rectified linear unit (ReLU) [14] activation function after being normalized with batch normalization. All max-pool layers have 2 x 2 pixel-sized kernels. In order to prevent overfitting, dropout regularization [17] is applied to the fully connected layers with a ratio of 0.5.

3.2.2 Feature Fusion

The feature extraction layers extract f_r and f_d which are the feature vectors of reference and distorted patch respectively. The regression layers require the network to combine these two vectors in a feature fusion step. A simplest strategy is concatenating f_r and f_d to a unique vector (f_r, f_d) . Besides, $f_r - f_d$ is known as a meaningful representation for distance in feature space. Therefore, concatenating $f_r - f_d$ is expected to contribute to learning to relations between reference and distorted patch. The final output of this state is $(f_r, f_d, f_r - f_d)$.

3.2.3 Training Method

For better convergence of the optimization, the feature extraction parameters are initialized with VGG13-batchnorm weights which is trained on ImageNet dataset. Our network is trained end-to-end by backpropagation, over a number of epochs. The adaptive moment estimation optimizer (ADAM) [11] is employed to alter the regular stochastic gradient descent method. Parameters of ADAM are chosen as recommended in [11] $\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8}$ and the learning rate α is initially set to 5×10^{-4} . The mean loss, PCC, SRCC over images during validation is computed in evaluation mode after each epoch.

Chapter 4

Evaluation

4.1 Evaluation Method

Dataset: This database comprises 1511 quality annotated images based on 1511 source reference image patches that are subject to different distortion levels of compression. Differential mean opinion score (DMOS) for this dataset were computed for each pair, which is in the range 1 to 5.

Evaluation Metrics: To evaluate the performances of the IQA algorithms, we used two standard measures, i.e., Spearman’s rank order correlation coefficient (SRCC) and Pearson’s linear correlation coefficient (PLCC).

Experiment Setup: Both the experiments in this thesis are performed on HMII database.

For the first experiment, the purpose is to evaluate how well an objective metric agrees with subjective preferences of subjects. We carefully select the Matlab implementations of 7 algorithms to predict object scores for the entire database.

For the second one, different models are competed to find the best ‘ground truth’ predictor for patch quality. In this experiment, we use the following models to evaluate with our proposed DIPQA:

- *Image-Patch model:* Zhang *et al.*[29] assume the the curve model to predict image-patch quality is a cubic polynomial function:

$$f(\Phi(\mathbf{d}); \theta) = a\Phi(\mathbf{d})^3 + b\Phi(\mathbf{d})^2 + c\Phi(\mathbf{d}) + d$$

where $\theta = a, b, c, d$ are the parameters for the non-linear function and $\Phi(\mathbf{d})$ represents the feature of patch \mathbf{d} . MSE and SSIM are chosen for the design of features. In our work, we tried top 3 FR-IQA methods from the first experiment: SSIM, FSIM and SRSIM.

- *DIQaM*: Bosse *et al.*[1] present a Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment which obtains superior performance on different IQA benchmarks. We utilize the extractor architecture from this paper to train a Deep Neural Network on our database.

Results reported are based on the average performance of 10 folds cross-validation. Deep learning models converge after 50 epochs.

4.2 Experiment results

4.2.1 HMII Benchmark Analysis

| | HMII (64x64) | | HMII (128x128) | |
|-----------|--------------|--------------|----------------|--------------|
| | PLCC | SRCC | PLCC | SRCC |
| SSIM[22] | 0.785 | 0.787 | 0.795 | 0.797 |
| RFSIM[27] | 0.774 | 0.757 | 0.789 | 0.759 |
| FSIM[28] | 0.794 | 0.799 | 0.824 | 0.815 |
| PSNR | 0.200 | 0.737 | 0.194 | 0.752 |
| UQI[21] | 0.023 | 0.621 | 0.012 | 0.589 |
| VSI[26] | 0.765 | 0.765 | 0.768 | 0.786 |
| SRSIM[25] | 0.777 | 0.803 | 0.718 | 0.803 |

TABLE 4.1: Comparing different IQA algorithms

4.2.2 DIPQA Performance Evaluation

| | HMII (64x64) | | HMII (128x128) | |
|------------------------|--------------|--------------|----------------|--------------|
| | PLCC | SRCC | PLCC | SRCC |
| IPM (SSIM) | 0.836 | 0.784 | 0.843 | 0.794 |
| IPM (FSIM) | 0.848 | 0.795 | 0.871 | 0.810 |
| IPM (SRSIM) | 0.854 | 0.802 | 0.857 | 0.798 |
| DIQaM | 0.916 | 0.824 | 0.905 | 0.819 |
| DIPQA (VGG extractor) | 0.802 | 0.754 | 0.830 | 0.760 |
| DIPQA (VGG finetuning) | 0.921 | 0.848 | 0.955 | 0.871 |

TABLE 4.2: Comparing different Full-Reference approaches

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] JANE BROMLEY, JAMES W. BENTZ, LÉON BOTTOU, ISABELLE GUYON, YANN LECUN, CLIFF MOORE, EDUARD SÄCKINGER, and ROOPAK SHAH. SIGNATURE VERIFICATION USING A “SIAMESE” TIME DELAY NEURAL NETWORK. *International Journal of Pattern Recognition and Artificial Intelligence*, 2004.
- [3] Itu-r Bt, Question Itu-r, The Itu, and Radiocommunication Assembly. RECOMMENDATION ITU-R BT . 500-11 Methodology for the subjective assessment of the quality of television pictures ANNEX 1 Description of assessment methods Common features. *Methodology*, 2002.
- [4] Sumohana S. Channappayya, Alan Conrad Bovik, and Robert W. Heath. Rate bounds on SSIM index of quantized images. *IEEE Transactions on Image Processing*, 2008.
- [5] 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.
- [6] Sumit Chopra, Raia Hadsell, and Yann LeCun. Learning a similarity metric discriminatively, with application to face verification. In *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, 2005.
- [7] Ross Girshick. Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*, 2015.
- [8] Q. Huynh-Thu and M. Ghanbari. Scope of validity of PSNR in image/video quality assessment. *Electronics Letters*, 2008.
- [9] 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.

- [10] 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.
- [11] Diederik P Kingma and Jimmy Lei Ba. Adam: A method for stochastic gradient descent. *ICLR: International Conference on Learning Representations*, 2015.
- [12] Weisi Lin and C. C. Jay Kuo. Perceptual visual quality metrics: A survey. *Journal of Visual Communication and Image Representation*, 2011.
- [13] 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.
- [14] Vinod Nair and Geoffrey Hinton. Rectified Linear Units Improve Restricted Boltzmann Machines. In *Proceedings of the 27th International Conference on Machine Learning*, 2010.
- [15] Evan Shelhamer, Jonathan Long, and Trevor Darrell. Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.
- [16] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. pages 1–14, 2014.
- [17] Ilya Sutskever, Geoffrey Hinton, Alex Krizhevsky, and Ruslan R Salakhutdinov. Dropout : A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 2014.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] Zhou Wang and Alan C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 2002.
- [22] 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.

- [23] 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.
- [24] 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.
- [25] Lin Zhang and Hongyu Li. SR-SIM: A fast and high performance IQA index based on spectral residual. In *Proceedings - International Conference on Image Processing, ICIP*, 2012.
- [26] Lin Zhang, Ying Shen, and Hongyu Li. VSI: A visual saliency-induced index for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 2014.
- [27] 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.
- [28] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. FSIM: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 2011.
- [29] 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.
- [30] 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.
- [31] 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.
- [32] 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.