Degraded Reference Image Quality Assessment

by

Xinyu Guo

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2021

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Images/videos are playing a more and more important role in the 21st century. The perceived quality of visual content often degrades during the process of acquisition, storage, transmission, display and rendering. Since subjective evaluation of such a large amount of visual content is impossible, the development of objective evaluation methods becomes highly desirable. Traditionally, there are three well established Image Quality Assessment (IQA) paradigms. They are Full Reference (FR) IQA which needs full access to the pristine quality reference, Reduced Reference (RR) IQA which requires partial information from the pristine reference and, No Reference (NR) IQA which does not require any reference information. While the strict requirement prohibits FR IQA from wide usage in many applications, RR and NR IQA methods can not produce comparable performance. In the thesis, we aim to address this problem by exploring the Degraded Reference (DR) paradigm which makes no requirement on pristine reference but on reference of degraded quality, and at the same time, outperforms RR/NR methods.

We address this problem in three steps. Firstly, we develop an FR model built upon a Deep Neural Network (DNN) that can handle multiply distorted images. The model structure of this FR model is then utilized to design DNN-based DR IQA models. We further improve the DR DNN model by adjusting the network structure. Finally, we use a two-step framework, which utilizes an NR model and an FR model as base modules followed by a regressor to create a single DR prediction for a given image.

We test our models on subject-related datasets in IQA field. The testing results show that our FR model has state-of-the-art performance when handling multiply distorted images, and meanwhile produces great performance when handling singly distorted images. Our DR model developed using the two-step framework gives better performance than RR/NR models when the reference is not pristine.

Acknowledgements

I would like to thank all the little people who made this thesis possible.

Words cannot express how grateful I am to my supervisor, Dr. Zhou Wang. Dr. Wang has been very patient, nice and supportive during my MASc studies. Thanks very much Dr. Wang for having faith in me. In this MASc journey, I have experienced many frustrated situations, thanks much for Dr. Wang for supporting me during these situations. Thank you very much Dr. Wang, for giving great suggestions to my research directions, and giving me the flexibility to explore them by myself. Thank you very much Dr. Wang for inviting excellent researchers to share their ideas with me. It has been an honor to work with you Dr. Wang.

I am honored to have Dr. Sherman Shen and Dr. Krzysztof Czarnecki as my thesis committee members, thanks very much for their time they took to review my thesis and for giving valuable suggestions in the seminar.

I would like to thank the students in the the Image and Vision Computing Lab. Thanks very much Wentao Liu, Zhengfang Duanmu, ZhuoranLi, Zhongling Wang and Jinghan Zhou for providing valuable suggestions to my research.

I want to thank my parents for supporting me during my MASc studies, thanks a lot for their help when I was upset.

While I have named just a few people above, I want to thank all the people who made this thesis possible, thank you!

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACO Ant Colony Optimization 16

AGGD Asymmetric Generalized Gaussian Model 17

BIQI Blind Image Quality Index 18

BLIINDS Blind Image Integrity Notator using DCT Statistics 17

BRISQUE Blind/Referenceless Image Spatial Quality Evaluator 17, 30

CIDIQ Colourlab Image Database Image Quality 34, 35

CNN Convolutional Neural Network 25

CORNIA Codebook Representation for No-Reference Image Assessment 18, 19, 30

CRQA Corrupted Reference Quality Assessment 22

CSF Contrast Sensitivity Function 14

CSIQ Categorical Image Quality 34, 35, 37

CV Computer Vision 10

DBCNN Deep Bilinear Convolutional Neural Network 20

DCT Discrete Cosine Transform 17, 18

DIIVINE Distortion Identication-based Image Verity and INtegrity Evaluation 18

DIQaM-NR Deep Image QuAlity Measure-NR 19

DISTS Deep Image Structure and Texture Similarity 11

DL Deep Learning 2, 4, 5, 10, 11, 20, 26, 30, 31, 33, 38

DMOS Difference Mean Opinion Score 35, 36

DNN Deep Neural Network iii, 19, 20

DNT Divisive Normalization Transform 15

DR Degraded Reference iii, 2, 4, 5, 22, 24, 26–33, 38–40, 42, 45, 46

EONSS End-to-end Optimized deep neural Network using Synthetic Scores 24, 25, 31, 39, 40, 42, 45

FC Fully Connected 19, 20, 26, 32

FD Final Distorted 24–27, 29, 32

FFT Fast Fourier Transform 13

FR Full Reference iii, 2–7, 10, 12, 15, 19, 22, 24, 25, 27, 30–34, 37–40, 42, 45

FR-EONSS Full Reference End-to-end Optimized deep neural Network using Synthetic Scores 24–27, 31, 33, 37–40, 42, 45

FSIM Feature Similarity 9

FSIMc Feature Similarity Color 9

GDN General Divisive Normalization 25, 26

GGD Generalized Gaussian Density 14, 18

GM Gradient Magnitude 9

GMSD Gradient Magnitude Similarity Deviation 9

GSIM Gradient Similarity 9

HIQM Hybric Image Quality Metric 13

HSE Hash-based 2D Spatial Embedding 21

HVS Human Visual System 2–4, 7, 13, 15

IQA Image Quality Assessment iii, viii, 1–7, 9, 10, 12, 13, 15, 16, 18–22, 24–26, 30–35, 37, 40, 42, 45, 46

IW-SSIM Information content Weighted Structural Similarity 9, 37

JNB Just Noticeable Blur 16

JP2 Joint Photographic Experts Group 2000 31, 32, 34, 36, 46

JPEG Joint Photographic Experts Group 31, 32, 34, 36, 38, 40, 46

KLD Kullback-Leibler Distance 14, 15

LIVE R2 LIVE Release2 34, 35

LIVEMD LIVE Multiply Distorted 34, 35, 38–40, 42

LQM Local Quality Map 9

MDID Multiply Distorted Image Database 34, 36

MDID2013 Multiply Distorted Image Database2013 34, 36

MDIVL Multiple Distorted IVL 34, 36, 40, 42

MEON Multi-task End-to-end Optimized deep Neural network 20

MOS Mean Opinion Score 35–37, 40, 42

MPEG2 Motion Picture Experts Group 2 13

MS-SSIM Multi-scale Structural Similarity 8, 9, 30

MSCN Mean Subtracted Contrast Normalized 17

MSE Mean Square Error 2, 7, 13, 22

MUSIQ Multi-scale Image Quality Transformer 21

NLP Natural Language Processing 21

NN Neural Network 20

NR No Reference iii, 2–6, 10, 15, 16, 18–22, 24, 25, 27, 28, 30–33, 38–40, 42, 45

NSS Natural Scene Statistics 2, 13–17

PC Phase Congruency 9

PLCC Pearson Linear Correlation Coefficient 34

PR Pristine Reference 5

PSNR Peak Signal to Noise Ratio 2, 7, 13

QA Quality Assessment 12

RBF Radial-Basis Function 18

RR Reduced Reference iii, 2–4, 6, 12–15

SCE SCale Embedding 21

SQB Synthetic Quality Benchmarky 34

SRCC Spearman's Rank Correlation Coefficient 33, 37, 39

SSEQ Spatial-Spectral Entropy-based Quality 16

SSIM Structural Similarity 7, 8, 11, 30

SVM Support Vector Machine 16, 18

SVR Support Vector Regression 17, 18

TDNN Time Delay Neural Network 13

TID2013 Tampere Image Database 2013 34, 35, 37

VCLFER Video Communications Laboratory @ FER 34, 35

ViT Vision Transformer 2, 21

VQA Video Quality Assessment 46

WaDIQaM-FR Weighted Average Deep Image QuAlity Measure-FR 10, 19, 24, 25, 37

WSSI Wavelet Structural Similarity Index 9

Chapter 1

Introduction

An image can be defined as a two-dimensional function $f(x, y)$, where x and y are the spatial coordinates, and $f(x, y)$ is the intensity level at (x, y). Natural images usually have three channels, e.g. RGB, which define the color of a point[2]. In the 21st century, images and videos are playing an ever more important role in our daily life. These digital contents have contributed to the vast majority of the increasing global Internet traffic. For example, there is about one billion hours of videos are watched daily on Youtube[1].

For people watching images and videos, the quality of the visual content is of great importance. However, visual content usually undergoes lots of distortion processes before reaching its end-users. The acquisition, storage, transmission, rendering and display processes introduce various kinds of distortions to images and videos. In some extreme cases, for example, when the network condition is bad, there might be some serious blocking artifacts inside frames. There is no doubt that improving the quality of the images and videos can greatly improve the experience of end-users. The first step to improve the quality of images and videos is to design objective IQA algorithms that can evaluate the quality of given images or video frames.

There are mainly two broad categories of methods to evaluate the quality of images, subjective IQA and objective IQA. Subjective IQA is expensive in time and cost. Also, subjective IQA can not be done in real-time. Therefore, objective IQA methods are desirable in many real-world application senarios.

Traditionally and intuitively, measuring the quality can be done by directly comparing the pixels at the same spatial locations. Mean Square Error (MSE) and Peak Signal to Noise Ratio (PSNR) are two objective IQA methods that are widely used. However, these methods do not correlate well with Human Visual System (HVS)[46]. As a result, many algorithms are developed, including FR, RR and NR IQA methods. In the FR domain, structural similarity is the dominate paradigm[48]. The structural similarity can be done in spatial domain or frequency domain. Other successful ideas include comparing the similarity of gradients and the phase congruency of the reference and distorted images[59, 25]. Recently, a notable trend is to incorporate Deep Learning (DL) technology into FR IQA domain, producing many high performance models[3, 35, 9]. In the RR domain, researchers mainly use statistical methods to model parameters derived in spatial or frequency domains[23, 50]. In the NR field, early models are usually developed to handle specific distortions[12, 36]. Later, efforts have been made to combine specific distortion models into more general models that can take more distortion types into consideration[62, 26]. Recently, more and more general-purpose NR IQA models appear in which aim to assess the quality of images undergoing general or mutiple distortion steps. Natural Scene Statistics (NSS) based models[29, 38], general DL based models[28, 60] and Vision Transformer (ViT) [10] based models[17] are among the most popular ones.

However, none of the existing objective IQA methods consider a more general situation, in which the pristine reference image is not accessiable but another distorted image of degraded quality is available as a reference. For example, the video feed of a video transcoder often has degraded quality, which is available in the evaluation of the video at the transcoder output. We call such a reference Degraded Reference (DR). DR images are more common reference images in real world due to the fact that complete pristine reference images are not available in most situations. On the other hand, compared with NR IQA, DR IQA can provide better performance since we still have some reference information available. We aim to design a unified IQA framework that bridges FR and NR methods. Also, we aim to design a scalable and flexible model that can be used given any kind of reference images. The developed DR IQA algorithm should have performance between FR and NR IQA when a DR image is given.

1.1 Background and Motivation

1.1.1 Challenges in Image Quality Assessment

Images are two dimensional signals, and can be further categorized into grey scale images and color images. We mainly consider color images as input to IQA algorithms. Given a commonly met 256 by 256 image, it can have 65536 dimensions, in each dimension, we have three color values ranges from 0 to 255. The high dimension of color images make IQA intrinsically a difficult problem.

On the other hand, designing IQA algorithms is quite different from other image related problems due to the fact that humans are the final judges of the quality of images. Algorithms that take HVS into consideration should outperform those algorithms that do not. HVS is an extremely complex systems and is not fully understood. The complex nature of HVS has also made the development of IQA algorithms difficult.

1.1.2 Limitations of Existing Image Quality Assessment Algorithms

The IQA algorithms are traditionally divided into three groups, FR IQA, RR IQA and NR IQA algorithms. FR algorithms generally have the highest performance because more information is used. One assumption of FR methods is that a pristine reference is needed to evaluate its distorted counterpart. In the real world situation, an image is distorted at the time it was born. For example, in images created by photographers using optical cameras, the noise is added at the sensor. Motion blur and other blur might be also added. What's more, consider a situation an image is transmitted to its end-user. To apply FR algorithms in this situation, the bandwidth required to transmit the pristine reference image is much more than that of transmitting the distorted image itself. Because we cannot provide FR IQA algorithms with pristine reference images, the performance of FR algorithms may not be applicable. Moreover, the bandwidth problem makes FR algorithms hard to implement in the real world situation.

In RR and NR IQA, much less information is accessible, and thus the performance drops significantly. Generally speaking, given the same testing condition, RR and NR algorithms' performance is worse than their FR counterparts. Recent NR IQA algorithms use DL technologies which need a large amount of training images that are hard to obtain. NR IQA problem can be solved using supervised learning technology which needs ground truth labels. The most reliable way to label a distorted image is to let human observers give scores in a subjective experiment. Subjective experiments are very time-consuming, especially when the number of images to be labelled is large.

To summarize, current IQA methods have many limitations. Most of them can not be solved in the traditional FR/RR/NR framework

This motivates us to work on a new IQA paradigm named Degraded Reference (DR) IQA. The main difference between DR IQA and FR IQA is that DR IQA algorithms do not make strict assumption on reference images. The reference images can be distorted. To summarize, DR IQA algorithms aim to assess the quality of distorted images when reference images are of degraded quality. .

1.2 Objectives

Our objectives are three folds. Firstly, we aim to develop a DR IQA algorithm that avoids strict requirements on pristine reference images. Secondly, we aim to develop a DR IQA algorithm that outperforms current state-of-the-art NR algorithms. Finally, we aim to develop a general-purpose DR IQA algorithm that is able to handle distorted images with multiple distortion types and distortion levels.

1.3 Thesis Outline

In Chapter 2, an overview of FR, RR, NR and DR algorithms are provided. For FR algorithms, the pixel-wise algorithms, HVS based algorithms, structural similarity base algortihms, DL based algorithms are reviewed. For RR IQA, we mainly focus on statistical modeling based RR algorithms. For NR algorithms, distortion specific modeling algorithms, multiple distortion oriented algorithms and general-purpose NR algorithms are reviewed. Several recently proposed DR algorithms are also discussed.

In Chapter 3, we first develop a FR model which can handle multiply distorted images. We then layout several architectures for deep learning based DR IQA models. Finally, we propose a novel two-step DR IQA algorithm.

In Chapter 4, the performance of the developed models is assessed using carefully selected databases for training and testing. We evaluate the performance of the proposed FR model compared to current state-of-the-art FR models. Next, we compare the performance of DL-based DR architectures when Pristine Reference (PR)/DR images are given. Finally, We compare the performance of DR IQA models with FR and NR models and demonstrate that the proposed DR IQA model, which do not have access to pristine reference but to DR images only, is able to achieve comparable performance against FR models that have full access to pristine reference images.

Chapter 2

Literature Review

This chapter provides a literature review of previous studies that are closely related to our work, including an overview of the current IQA algorithms. In this field, algorithms are traditionally divided into FR IQA, RR IQA and NR IQA algorithms[1]. We illustrate the connections between FR IQA, RR IQA and NR IQA in Figure 2.1.

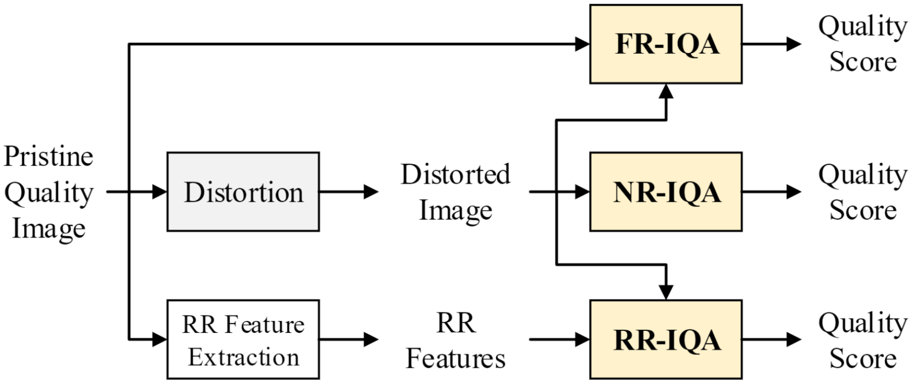


Figure 2.1: Illustration of General IQA Frameworks[1]

As shown in Figure 2.1, FR methods evaluate the quality of a distorted image using its pristine counterpart, RR methods evaluate the quality of distorted image with features extracted from pristine reference image, while NR methods evaluate the quality of distorted image using the distorted image itself without access to the pristine reference images.

2.1 Full Reference Image Quality Assessment

2.1.1 Error based algorithms

To the best of our knowledge, MSE and PSNR are historically the most widely used FR metrics. Denote the distorted image and the pristine reference image by $Y = \{y_i | i = 1, 2, 3, ..., N\}$ and $X = \{x_i | i = 1, 2, 3, ..., N\}$, respectively, where y_i and x_i represent the pixel intensity in the distorted and reference images, respectively. MSE and PSNR can be expressed by:

$$\begin{aligned} \text{MSE} &= \frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)^2 \\ \text{PSNR} &= 10 \log_{10} \frac{L^2}{\text{MSE}} \end{aligned} \tag{2.1}$$

where L is the dynamic range of image pixel intensities. For a typical grey scale image where each pixel is represented by 8 bits, $L = 2^8 - 1$. MSE and PSNR are designed to be a signal fidelity measure, where the quantitative output denotes the level of error. MSE have multiple advantages. It is simple and easy to use. It is directly related to the l_2 norm which holds many convenient mathematical properties, such as non-negativity, identity, symmetry and satisfying triangle inequality. However, MSE does not correlate well with human perception of image quality [46].

Based on the observation that HVS is more sensitive to low frequency distortions than to high frequency ones[8], a PSNR-HVS[11] is proposed which improves the performance of PSNR.

2.1.2 Structural Similarity Based Methods

The Structural Similarity (SSIM) index started a new paradigm of IQA research [48]. Given image X and Y, the luminance comparison is given by:

$$l(\mathbf{X}, \mathbf{Y}) = \frac{2\mu_X \mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1} \tag{2.2}$$

where μ_X and μ_Y are the mean of input image X and image Y. C_1 is a constant to avoid instability. The contrast comparison is given by:

$$c(\mathbf{X}, \mathbf{Y}) = \frac{2\sigma_X \sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2} \tag{2.3}$$

where σ_X and σ_Y are the variance of input image X and image Y. C_2 is a constant to avoid instability. Finally, the structure comparison is given by:

$$s(\mathbf{X}, \mathbf{Y}) = \frac{\sigma_{XY} + C_3}{\sigma_X \sigma_Y + C_3} \tag{2.4}$$

where σ_{XY} is the covariance of input image X and Y. Combining three components above, we have the finial SSIM index:

$$\mathrm{SSIM}(\mathbf{X}, \mathbf{Y}) = \frac{(2\mu_X \mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_X^2 + \sigma_Y^2 + C_2)} \tag{2.5}$$

In order to incorporate image details at different resolutions, Multi-scale Structural Similarity (Multi-scale Structural Similarity (MS-SSIM)) [51] is proposed. In the MS-SSIM framework, an given image is downsampled multiple times, and the SSIM index is computed for images at all scales. The downsampling process is illustrated in Figure 2.2.

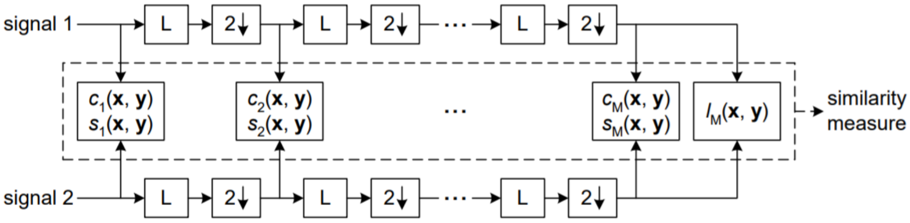


Figure 2.2: Diagram of MS-SSIM [51]

Finally, the MS-SSIM is computed by combining the luminance, contrast and structural comparison in difference scales.

$$\text{MS-SSIM}(\mathbf{X}, \mathbf{Y}) = [l_M(\mathbf{X}, \mathbf{Y})]^{\alpha_M} \cdot \prod_{j=1}^{M} [c_j(\mathbf{X}, \mathbf{Y})]^{\beta_j} [s_j(\mathbf{X}, \mathbf{Y})]^{\gamma_j} \tag{2.6}$$

where $\mathbf{X}$ and $\mathbf{Y}$ are input images, l_M is the luminance comparison at Mth scale, c_j and $s_j(j = 1, ..., M)$ are the contrast comparison and structural comparison at jth scale. $\alpha_M, \beta_j, \gamma_j$ are weight parameters for luminance, contrast and structural comparison respectively.

MS-SSIM takes viewing conditions into account, but weights different parts of the content the same. To compensate this, Information content Weighted Structural Similarity (IW-SSIM) is proposed [49], where different parts in a image are weighted based on the shared information of that part between the reference and distorted images.

Apart from the previous mentioned metrics which measure the similarity between two inputs in spatial domain directly. Some metrics are proposed to use this paradigm in other domains. These metrics include Feature Similarity (FSIM)[59], Wavelet Structural Similarity Index (WSSI)[37], Gradient Similarity (GSIM)[25], and Gradient Magnitude Similarity Deviation (GMSD)[54].

FSIM is designed for grey scale image quality assessment. Firstly, Phase Congruency (PC) has been chosen as the main feature since visually discernable features coincide with the point where different frequency Fourier waves have congruent phases. A complementary feature Gradient Magnitude (GM) has been included considering that local contrast has contribution to visual quality. An Feature Similarity Color (FSIMc) method is also proposed to take color information into consideration. WSSI performs structural similarity in wavelet domain. The images are transformed to wavelet domain before a structural similarity measure is applied. Based on the idea that gradient can convey important information of images which is important for scene understanding, GSIM uses gradient similarity comparison to capture structural and contrast changes. Apart form structural and constrast information, luminance is also a very important component of images, GSIM assesses the quality of distorted image using gradient similarity and luminance similarity.

GMSD models most of the IQA methods as a two-step framework, as illustrated in Figure 2.3, Firstly a Local Quality Map (LQM) is computed using a local quality computation, then a pooling strategy is chose to pool the LQM to a single qulity score. Considering that different local structures in a distorted image suffer from different types and degrees of degradations, GMSD replaces the commonly used average pooling to global variation of LQM.

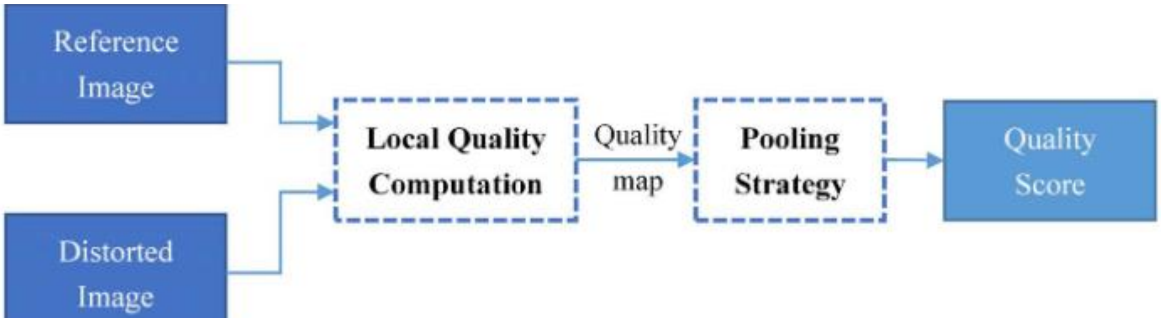


Figure 2.3: Illustration of two-step FR-IQA Framework [54]

2.1.3 Deep Learning Based Methods

Recently, many researchers try to improve FR IQA performance using DL technology. The ability of DL models has been proved in many Computer Vision (CV) tasks [19, 39, 15]. The good performance of these data driven models inspired researchers in IQA fields to use DL models.

A representative method is Weighted Average Deep Image QuAlity Measure-FR (WaDIQaM-FR) model [3]. The authors have developed both a FR verion and a NR version WaDIQaM. Here we mainly introduce the idea behind the WaDIQaM-FR model.

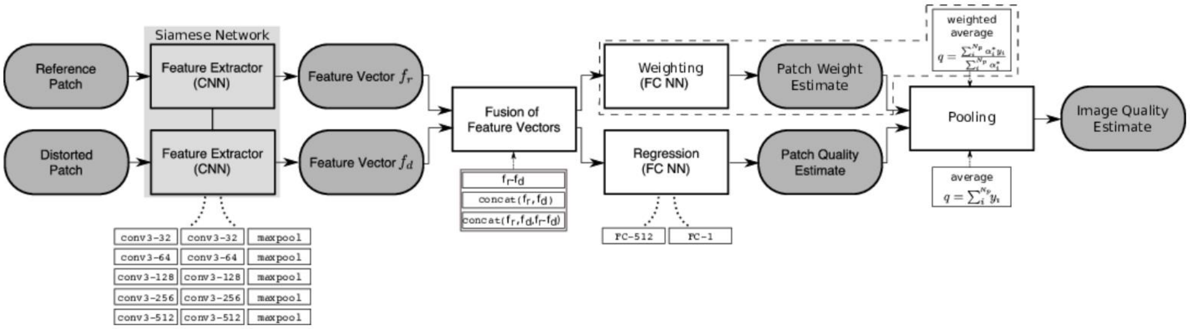


Figure 2.4: WaDIQaM-FR Structure [3]

Figure2.4 shows the network structure of the WaDIQaM-FR model. The authors adopted VGGnet as the backbone for feature extraction. VGGnet is one of the top performance networks in image classification and localization tasks[42]. The WaDIQaM-FR model mainly uses the Siamese structure, where the weights of the feature extractor is shared between the reference images and the distorted images. The quality aware features are then concatenated. Two sub-tasks are designed, the first one is the patch weigh estimation sub-task and the other is the patch score regression sub-task. Finally, patch quality and patch weight are pooled into a single image quality score. The simple idea results in superior performance.

A different DL learning framework is to learn useful image features using a pairwise preference framework. As shown in Figure 2.5. The model tries to learn which pair of images is better than the other.

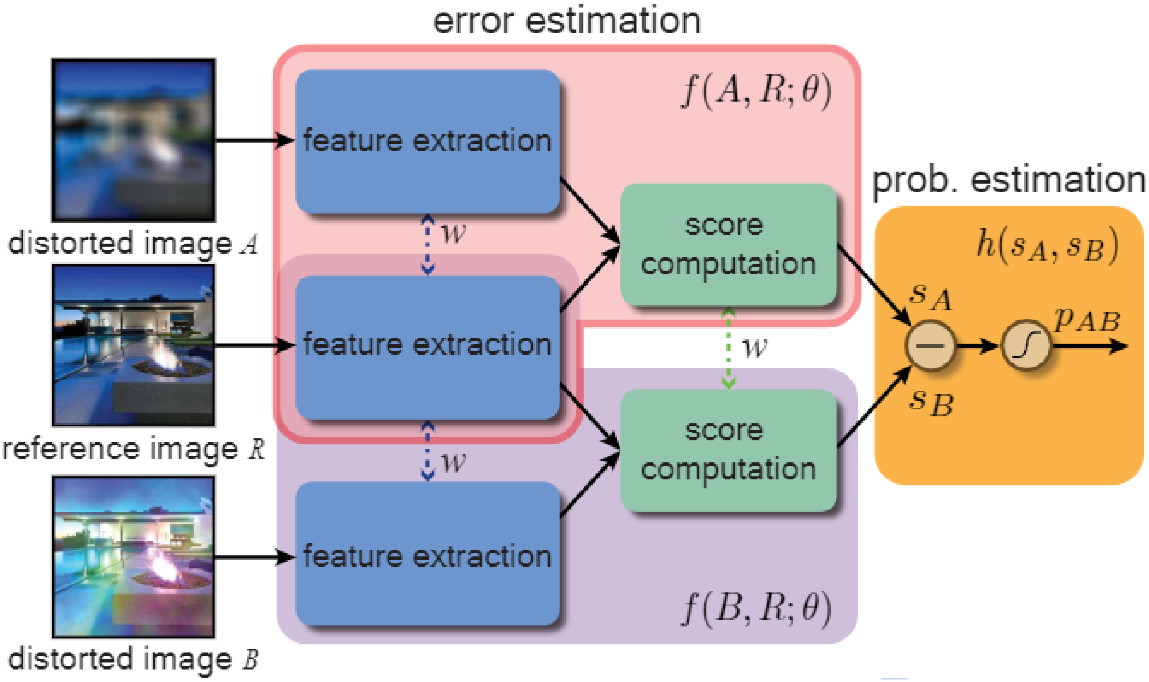


Figure 2.5: Pairwise Preference Framework [35]

Based on the fact that image is formed by both structural regions and texture regions, a model called Deep Image Structure and Texture Similarity (DISTS) is proposed. The authors combine the DL model with structural similariy paradigm, at the same time, texture invariance is also considered [9]. A combined loss is computed based on the features extracted by VGG backbone. For comparing structural similarity, SSIM is adopted. For measuring texture invariance, the Gram matrix[13] is adopted. The Gram matrix is given by:

$$G_{ij}^{l} = \sum_{k} F_{ik}^{l} F_{jk}^{l} \tag{2.7}$$

where F denotes the feature map extracted by VGG backbone. The model gives state-of-the-art performance against FR IQA models.

2.2 Reduced Reference Image Quality Assessment

The idea of RR Quality Assessment (QA) was first proposed in the 1990s [52]. The authors try to use the concept of RR in real-time video quality monitoring. The entire system is divided into sender side and receiver side. The original image/video is sent to the receiver side through the visual communication channel. The RR features extracted from the original image/video are sent to the receiver side through an ancillary channel. Then, at the receiver side, RR quality assessment module takes the RR features and distorted image/video as input and produces the final quality score [47]. The whole process is depicted in the Figure 2.6.

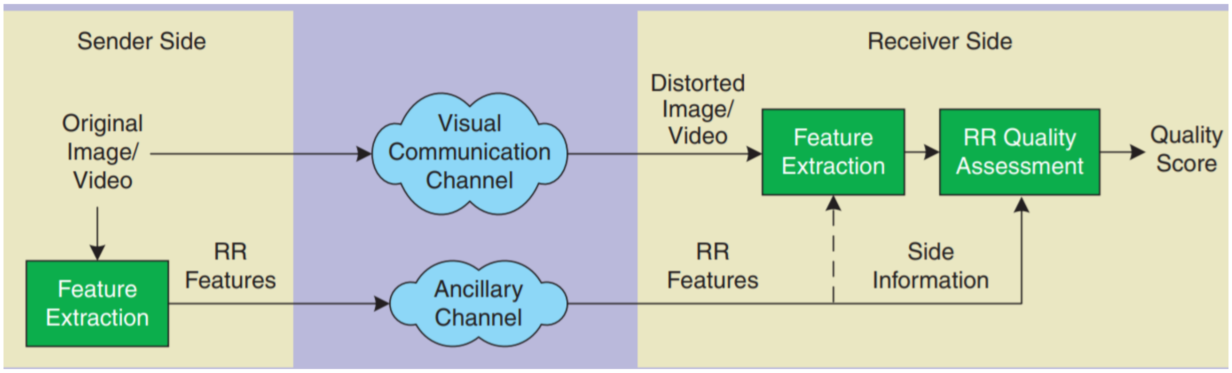


Figure 2.6: RR QA Framework [47]

As shown in the Figure 2.6, the RR QA methods mainly rely on the features extracted from the original image and the received distorted image for quality assessment. The main challenge is to develop a reliable set of features that can effectively represent and differentiate the information of the original and distorted images.

RR IQA algorithms are generally classified into 3 types[23]. The first type of algorithms are developed based on the idea to model image distortions. Most of the RR algorithms in this category are developed for specific distortion types. The second type of RR algorithms are developed based on modeling HVS. The third type of RR algorithms mainly try to model NSS.

A naive idea is to randomly sample pixels from the reference image at the sender side, and compare with the corresponding pixels using MSE or PSNR at the receiver side. This naive approach has multiple drawbacks. For example, the selected pixels might not contain the distorted part of image. The random selection may not provide accurate summarization of the images.

Several RR IQA algorithms are based on modeling image distortions. An RR metric for compressed digital images and videos is proposed by Irwan et al. The local harmonic amplitude information is extracted from edge-detected picture as reduced-reference[14]. Specifically, the authors applied Sobel operator to the input images, and the gradient images are divided into blocks. In each block Fast Fourier Transform (FFT) is applied, and local harmonics strengths are extracted from these FFT blocks. A similar algorithm was proposed for Motion Picture Experts Group 2 (MPEG2) video[22]. Specifically, the reference image or the distorted image is transformed to the YUV color space. For each color channel, 4 features are considered, two for spectrum content, one for spatial content, and one for blocking effect. At the receiver side, 12 features from the reference image and the distorted image are compared using a Time Delay Neural Network (TDNN).The two algorithms above mainly consider compression artifact. To take multiple image distortions into account, a Hybric Image Quality Metric (HIQM) is proposed to combine blocking measurement, blur measurement, and intensity masking detection, etc.[20]. The final quality metric is the weighted linear combination of these measurements.

A few perceptual representation based RR IQA algorithms are proposed. An early work uses a two-stage method to represent the reference or distorted image[5]. In the first stage, a set of low level processes are applied, including color normalization, gamma function, and projection onto some perceptual color spaces. In the second stage, structural features around fixation points are compared for image representations. In [6], the reduced description is built according to the corresponding stages in the HVS. This process contains gamma function adjustment, projection to perceptual color space, luminance normalization, Contrast Sensitivity Function (CSF) weighting, subband decomposition, component masking and feature extraction. The process is shown in Figure 2.7.

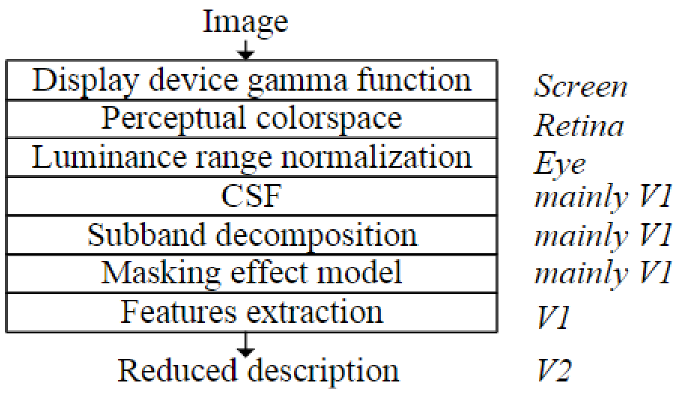


Figure 2.7: Building Reduced Descriptor [6]

NSS based RR methods usually model the images using low level statistic models. An early work considers using wavelet coefficients[50]. The idea is to transform the image from spatial domain to wavelet domain, where the subbands' coefficients are modeled using Generalized Gaussian Density (GGD) model. The Figure 2.8 shows the feature extraction process. The original image is represented by multiple subbands, each subband is modeled using GGD parameters.

The GGD model is given by:

$$p_m(x) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x|/\alpha)^{\beta}} \tag{2.8}$$

where, $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt($ for $a > 0)$ is the Gamma function, and α and β are the parameters of the GGD model. After modeling the subbands of the distorted image and the reference image, the Kullback-Leibler Distance (KLD) is employed to determine the quality of distorted image:

$$d\left(p_m \| p\right) = \int p_m(x) \log \frac{p_m(x)}{p(x)} dx \tag{2.9}$$

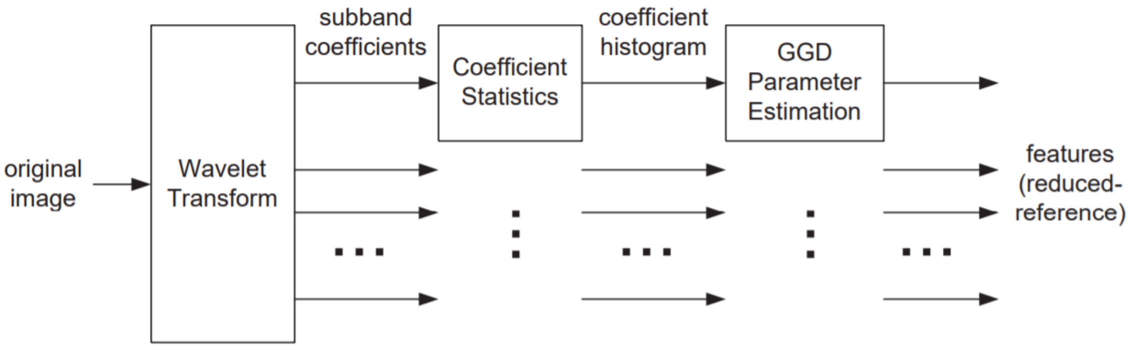


Figure 2.8: Wavelet Feature Extraction [50]

Another NSS based RR IQA method uses Divisive Normalization Transform (DNT) based image representation[23]. For better representation, a DNT is applied to wavelet coefficients to obtain DNT coefficients. The quality is represented by the variations of the marginal probability distribution of the DNT coefficients. The distance between reference marginal distribution and distorted marginal distribution is measured using KLD.

2.3 No Reference Image Quality Assessment

The NR IQA problem is fundamentally different from the FR and RR paradigms, because no information about the pristine reference is available. Early NR IQA methods aim to predict quality for distortion specific images. Recently, more and more general-purpose NR IQA algorithms have appeared.

We first review some distortion specific NR IQA methods that assess the quality of given distorted images based on specific distortion models. These models are able to measure one or multiple distortions such as blur and noise.

Blur artifact can be detected by analyzing the strength and the spread of edges and other fine details. The measurement can be done in either spatial domain or frequency domain. Simple algorithms detect blur by measuring edge width or computing dominant eigen values of covariance matrix[53]. More complex methods take HVS characteristics into consideration. A Just Noticeable Blur (JNB) concept is introduced[12], which accounts for the minimum amount by which a stimulus intensity must be changed relative to background intensity in order to produce a noticeable variation in sensory experience. Based on the concept of JNB, a perceptual blur image quality metric is proposed[18], where a perceptual edge map is constructed from every edge map at different resolutions. Based on the perceptual edge map, the blur artifact is used to determine the perceptual quality of a given image based on its bluriness.

Noise is another artifact that commonly appears in images. To determine the quality of a noisy image, measuring the level of noise is important. There are several useful noise measurement approaches. For example, [36] uses histogram of local image variance to determine the level of noise. While [43] use graph and Ant Colony Optimization (ACO) to determine the noise level. An image is represented using a graph $G(V, E)$, $v \in V$ are vertices representing patches inside an image, and $e \in E$ are edges representing connections between edges. Patches are selected to determine the noise level using the ACO algorithm.

In order to apply NR IQA algorithms in a more complex situation, multiple distortions oriented algorithms are proposed. A basic idea is to model the quality by a weighted combination of noise, blur and other artifacts[62]. In this method, noise level, blur level and blocking level are firstly determined using some commonly used methods. Then the logistics regression is used to find the optimal weight for each distortion measure. A Spatial-Spectral Entropy-based Quality (SSEQ) index [26] use a 2-stage framework to handle multiple distortions. In the first stage, a Support Vector Machine (SVM) is used to determine the distortion type. Then, local spatial and spectral entropy features are used to quantify the quality of a given image. [24] uses three objective measures to characterize three important aspects of visual quality, which jointly determine the quality of a given image.

All the above mentioned NR IQA algorithms measure the visual quality by modeling specific distortions. Some work with one specific type of distortions only. Others combine multiple distortions by regression. Next, an overview of general-purpose NR IQA algorithms is given.

One way to develop general-purpose NR IQA algorithms is to use NSS. NSS based methods usually assume that natural images occupy a small cluster in the entire space of all possible images, and the distortion is quantified as the distance to the natural image cluster. Many efforts are made to utilize spatial or transform domain statistical distribution for determining image quality. An algorithm called Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) extracts local normalized luminance and measures image naturalness by measuring the deviation from a natural image model[29]. In the first stage of the proposed algorithm, the Mean Subtracted Contrast Normalized (MSCN) coefficients are computed using:

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + C} \tag{2.10}$$

where $\mu(i,j)$ and $\sigma(i,j)$ are the mean and standard deviation of the local image. After extracting features from images, the Asymmetric Generalized Gaussian Model (AGGD) is used to fit the features:

$$f_X\left(x;\alpha,\sigma_l^2,\sigma_r^2\right) = \begin{cases} \frac{\alpha}{(\beta_l+\beta_r)\gamma\left(\frac{1}{\alpha}\right)} \exp\left(-\left(\frac{-x}{\beta_l}\right)^{\alpha}\right) & x < 0 \\ \frac{\alpha}{(\beta_l+\beta_r)\gamma\left(\frac{1}{\alpha}\right)} \exp\left(-\left(\frac{x}{\beta_r}\right)^{\alpha}\right) & x \geq 0 \end{cases} \tag{2.11}$$

where α is a shape parameter, σ_l and σ_r are scale parameters, and β_l and β_r are given by:

$$\begin{gathered} \beta_l = \sigma_l \sqrt{\frac{\gamma\left(\frac{1}{\alpha}\right)}{\gamma\left(\frac{3}{\alpha}\right)}} \\ \beta_r = \sigma_r \sqrt{\frac{\gamma\left(\frac{1}{\alpha}\right)}{\gamma\left(\frac{3}{\alpha}\right)}} \end{gathered} \tag{2.12}$$

where γ is the gamma function:

$$\gamma(a) = \int_0^{\infty} t^{a-1} e^{-t} dt \quad a > 0 \tag{2.13}$$

Fitting the extracted features using AGGD model gives a set of final features which are used to train the Support Vector Regression (SVR) model.

Blind Image Integrity Notator using DCT Statistics (BLIINDS) is another NSS based NR IQA method[38]. Unlike BRISQUE, BLIINDS use Discrete Cosine Transform (DCT) Statistics. In order to quantify the sharpness or blur in an image, the statistics of frequency domain parameters are needed, thus DCT is used to transform from original image to frequency domain. To quantify the structural difference between the distorted and natural images, the kurtosis of DCT histograms is used:

$$\kappa(x) = \frac{E(x-\mu)^4}{\sigma^4} \tag{2.14}$$

where x denotes the DCT histograms' value, and μ and σ are the mean and standard deviation of the DCT histograms respectively. To quantify the directional information, for each orientation $(0°, 45°, 90°$ and $135°)$, the anisotropy value is computed based on the Renyi entropy:

$$R_\theta[n] = -\frac{1}{2}\log\left(\sum_k \tilde{P}_\theta[n,k]^3\right) \tag{2.15}$$

where $\tilde{P}_\theta[n,k]$ denotes the normalized DCT coefficients, n is the spatial index, k is the frequency index.

Let $X_i = [x_1, x_2, \ldots x_n]$ denote the ith image' n features, and let $DMOS_i$ be the subjective score given to this image. The probability of $P(X, DMOS)$ is modeled by a multivariate Gaussian distribution and a multivariate Laplacian distribution on a training database. The prediction given by maximizing the quantity $P\left(DMOS_i/X_i\right)$ is equivalent to maximizing $P(X, DMOS)$ because $P(X, DMOS) = P(DMOS/X)p(X)$.

In [32], a Blind Image Quality Index (BIQI) is proposed, which uses a two-step framework to assess the quality of a distorted image. The distortion type is determined by a classification model before quality assessment is applied. The final quality score is given by the weighted summation $BIQI = \sum_{i=1}^{5} p_i \cdot q_i$. Firstly, wavelet transform is used to compare subband coefficients which are modeled using GGD. The variance and shape parameters are used as feature representations of the distorted images. The multiclass SVM with a Radial-Basis Function (RBF) kernel is served as the classifier to predict the distortion types based on image features. After classifying the given distorted images into corresponding distortion categories, SVR method is used to map image features into quality scores. In [33], Distortion Identication-based Image Verity and INtegrity Evaluation (DIIVINE) index is proposed. DIIVINE improves BIQI by replacing the wavelet transform with the steerable pyramid transform.

In [55], unsupervised learning strategy is used to build a general-purpose NR IQA algorithm called Codebook Representation for No-Reference Image Assessment (CORNIA).

The overall learning framework for CORNIA includes local feature extraction, codebook construction, local feature encoding and feature pooling. In the first step, image patches are extracted as feature representations for an image, where the patches are randomly selected. To build a codebook, the unsupervised learning methods such as K-means clustering is used. A matrix $D_{d\times K} = [D_1, D_2, \ldots, D_K]$ is a visual codebook, where $D_{i(i=1,\ldots K)}$ are the centers of clusters, and d is the feature dimension of a image patch. The codebook is then normalized to obtain unit length base vector. The normalized codebook is denoted by $\tilde{D}_{d\times K} = \left[\tilde{D}_1, \tilde{D}_2, \ldots, \tilde{D}_K\right]$. In the local feature encoding process, soft-assignment coding is used. Distance between the local descriptors (x_i) and the visual codewords $(\tilde{D})$ is computed using dot product. The code for a local descriptor x_i is given by the following equation:

$$\begin{aligned} c_i = [&\max\left(s_{i1}, 0\right), \ldots, \max\left(s_{id}, 0\right) \\ &\max\left(-s_{i1}, 0\right), \ldots, \max\left(-s_{id}, 0\right)]^T \end{aligned} \tag{2.16}$$

A coefficient matrix is given by the local feature coding step: $C_{K\times N} = [c_1, c_2, \ldots, c_N]$ which is used to pool to a quality score.

As introduced in Section 2.1, WaDIQaM-FR is a model proposed for FR IQA. There is also a NR version called Deep Image QuAlity Measure-NR (DIQaM-NR) which is designed for NR IQA. DIQaM-NR is a patch based DNN model that contains two concatenated structures. The NR model is divided into a backbone section and a score prediction section. In the backbone section, VGGNet is used because of its good performance in image classification and localization tasks. The perceptual relevance features are the output from the backbone. To compute the quality score based on the perceptual relevance features, a weight estimation network and a quality prediction network are designed. A Fully Connected (FC) network is used to learn the weight for input patch, another FC netowrk is used to estimate the quality score for the corresponding patch. A weighted average pooling is used to obtain the final quality score for input image.

$$\hat{q} = \sum_{i}^{N_p} p_i y_i = \frac{\sum_i^{N_p} \alpha_i^* y_i}{\sum_i^{N_p} \alpha_i^*} \tag{2.17}$$

where N_p is the number of patches in the image, α_i^* is the weight for a patch i, y_i is the quality score for patch i, and p_i is the normalized weight given by:

$$p_i = \frac{\alpha_i^*}{\sum_j^{N_p} \alpha_j^*} \tag{2.18}$$

Multi-task End-to-end Optimized deep Neural network (MEON) is an NR IQA model built by DL technology. In this method, the Neural Network (NN) is trained by two tasks. The first task is distortion type classification which gives a probability vector as the output. The second task combines the distortion task classification result with quality score prediction using dot product. The overall network structure can be divided into two parts. The first part is the backbone structure, which consists of convolutional layers, normalization layers and max pooling layers. The second part of the network takes feature output from the first part as input, and the two subtasks are performed jointly in this second part. Some FC layers are used to map features to probability vectors and quality score vectors. The dot product between the probability vectors and the quality score vectors creates the final quality score for a distorted image.

Deep Bilinear Convolutional Neural Network (DBCNN) is another DL based methods for NR IQA[60]. DBCNN works with both synthetically and authentically distorted images using two streams of DNN. A pre-train method is adopted. For synthetic distortions. S-CNN is designed by modifying VGG-16 network, S-CNN is pre-trained to classify both distortion types and distortion levels for given images. For authentic distortions, the VGG-16 network is pre-trained on an image classification task. Two pre-trained networks are then combined using bilinear pooling:

$$\mathbf{B} = \mathbf{Y}_1^T \mathbf{Y}_2 \tag{2.19}$$

where $\mathbf{Y}_1$ is the output of S-CNN, $\mathbf{Y}_2$ is the output of VGG-16. Finally, a FC layer map the bilinear output to a final quality score.

In [63], Meta-learning is used to develop an NR IQA model. The authors use Meta-learning to learn prior information which is used in the general-purpose NR IQA task. For learning prior information, a meta-learning set is constructed: $\mathcal{D}_{\text{meta}}^{p(\tau)} = \left\{\mathcal{D}_s^{\tau_n}, \mathcal{D}_q^{\tau_n}\right\}_{n=1}^{N}$, where $\mathcal{D}_q^{\tau_n}$ is the query set, $\mathcal{D}_s^{\tau_n}$ is the support set. N is the total number of tasks for learning different distortion specific information, and k tasks are randomly selected form a mini-batch. The model parameters are adjusted based on the task specific gradients. After prior information is learned using meta-learning, the NR model is fine-tuned for general-purpose NR IQA.

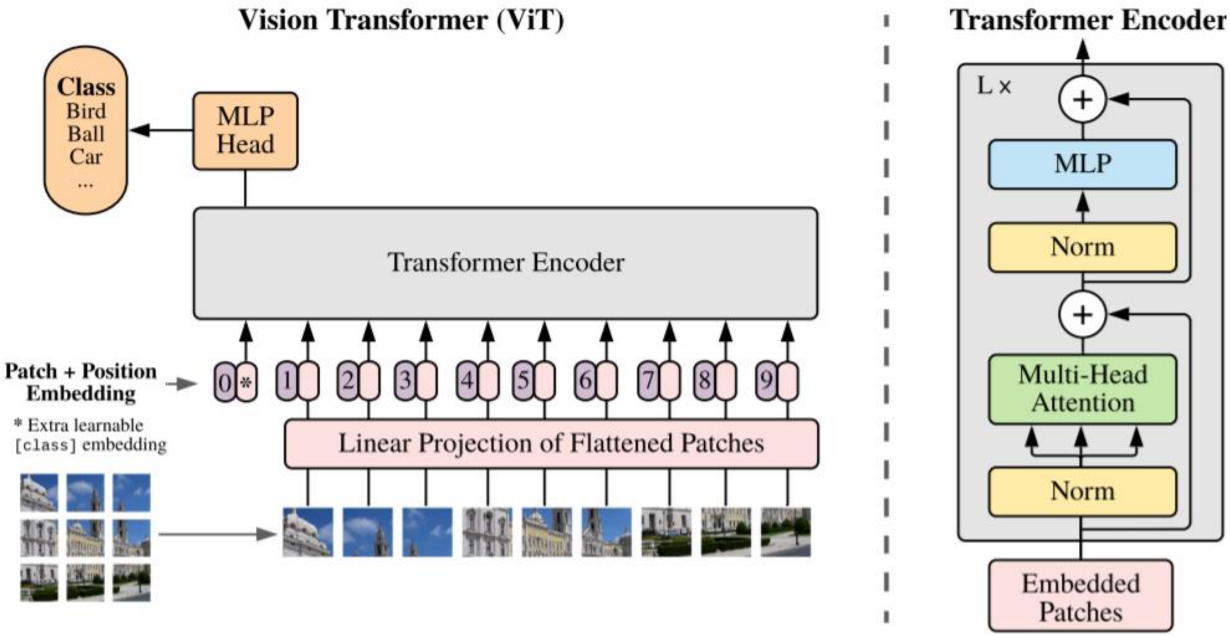


Figure 2.9: Architecture of Vision Transformer [10]

Recently, ViT is proposed [10] that produces good performance of self-attention-based architectures in Natural Language Processing (NLP). The basic idea behind the ViT is to split images into patches and feed linear embeddings of patches into a Transformer. The overall architecture of a ViT is shown in Figure 2.9, where the input patches are mapped to linear embeddings using linear projection. The Transform Encoder is used to project linear embeddings to class labels.

Based on the idea of ViT, Multi-scale Image Quality Transformer (MUSIQ) is proposed for general-purpose NR IQA task. To support various aspect ratios and resolution inputs, MUSIQ adopts a multi-scale representation. Hash-based 2D Spatial Embedding (HSE) and SCale Embedding (SCE) are used in the architecuter. The original classification head is replaced by regression head to produce a quality score.

2.4 Degraded Reference Image Quality Assessment

To the best of our knowledge, the first work relavant to DR-IQA is the Corrupted Reference Quality Assessment (CRQA) method.[7, 58]. In the work, the authors try to assess the quality of a restored image without the help of the pristine reference. This process is illustrated in Figure 2.10, where in the context of image restoration the ”process” represnts an image restoration operator. The CRQA method assesses the quality of the restored image by using the pair of a corrupted image and a restored image.

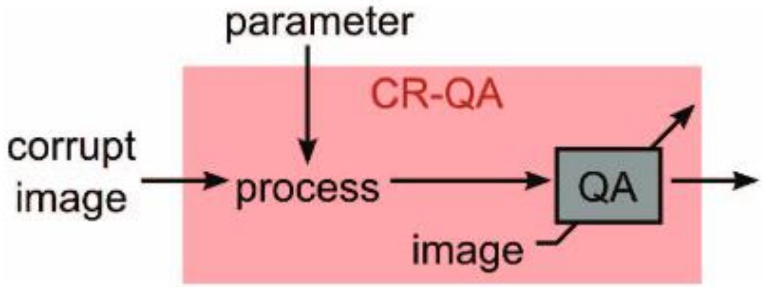


Figure 2.10: CRQA Paradigm [7]

Another highly relavant work is the 2stepQA algorithm[56], which attempts to predict the quality of a compressed-after-distorted image using an NR algorithm and an FR algorithm. The overall structure of the 2stepQA algorithm is shown in Figure 2.11, where the NR IQA algorithm is directly applied on the distorted image, and the FR IQA algorithm is responsible for assessing the quality of compressed-after-distorted image. Then, the two quality score is combined into a final 2stepQA quality score.

Conditional knowledge distillation has also been used to handle the DR IQA problem[61], where, the network is trained to learn from pristine reference features. The MSE loss is used to map the features from the DR images to those from the pristine reference images. The knowledge learned from the training phase is employed in the test phase to assess the quality with degraded reference.

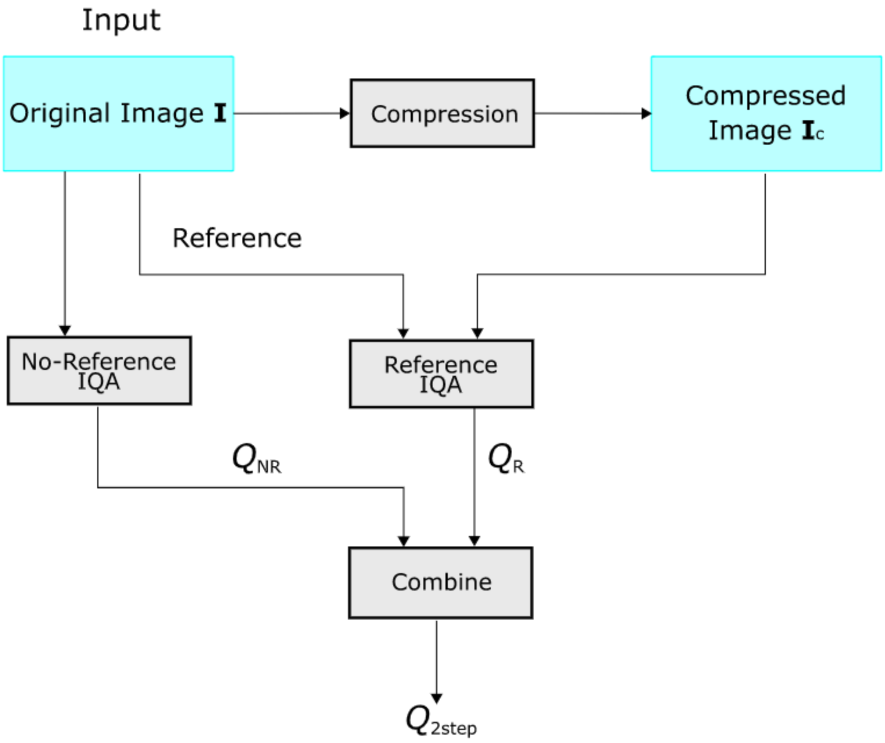


Figure 2.11: 2stepQA Framework [56]

Chapter 3

Degraded Reference Image Quality Assessment

In this chapter, we aim to exploit the architecture of our DR-IQA model. Our exploration starts from End-to-end Optimized deep neural Network using Synthetic Scores (EONSS) [45] which is an NR IQA model designed for multiply distorted images. We combine the backbone of EONSS and the framework of WaDIQaM-FR to create a model called Full Reference End-to-end Optimized deep neural Network using Synthetic Scores (FR-EONSS), which is an FR model developed for multiply distorted image. We then exploit different variations of deep learning architectures for DR IQA. Finally, we explore the two-step framework for DR IQA.

3.1 Full Reference Model for Multiply Distorted Images

As a first step in the development of our DR IQA method, we develop an FR model for multiply distorted images. We call multiply distorted images the Final Distorted (FD) images. We believe developing an FR model that can handle FD images with pristine reference images can help build the fundation of the DR framework.

The design of the FR model starts from EONSS which is an NR model developed for multiply distorted images. The backbone of EONSS is used to develop the FR-EONSS model. We believe this backbone is able to extract perceptual relevance features. In addition, since EONSS is developed for multiply distorted images, it is a good candidate to handle information from FD images.

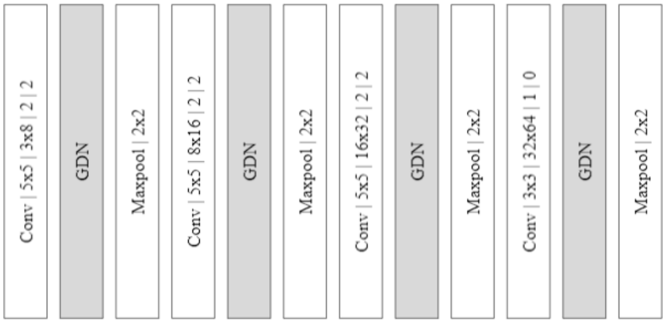


Figure 3.1: FR-EONSS Backbone

The backbone structure is shown in Figure 3.1, where the backbone of the FR-EONSS model consists of convolutional layers, General Divisive Normalization (GDN) layers, and max pooling layers. The convolutional layers and max pooling layers used are similar to the ones used in [19, 39, 15]. The GDN is first introduced to the IQA field in [28]. Given an S dimensional feature map $\mathbf{x}(m,n) = [x_1(m,n), \cdots, x_S(m,n)]^T$, at a spatial location (m,n), the GDN transform is given by:

$$y_i(m,n) = \frac{x_i(m,n)}{\left(\beta_i + \sum_{j=1}^{S} \gamma_{ij} x_j(m,n)^2\right)^{\frac{1}{2}}} \tag{3.1}$$

where $\mathbf{y}(m,n) = [y_1(m,n), \cdots, y_S(m,n)]^T$ is the normalized feature map. γ is the weight matrix, β is the bias vector. These parameters need to be optimized during the training phase.

To use this backbone in FR task, we utilize a siamese network structure demonstrated in WaDIQaM-FR. The whole structure of FR-EONSS is shown in Figure 3.2, where we use a share-weight Convolutional Neural Network (CNN) backbone to extract features from the reference patches and the distorted patches. We then fused the reference features and

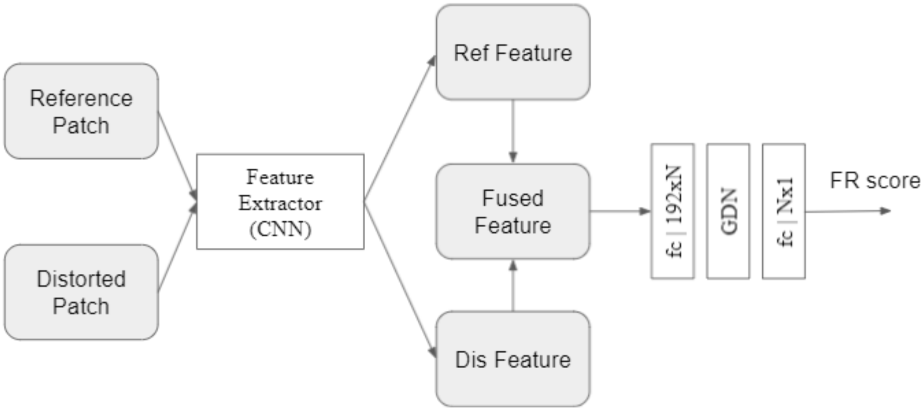


Figure 3.2: FR-EONSS Structure

the distorted features by concatenate these features and their differences. We use a stack of FC layers and GDN layers for the final regression task.

We utilize Waterloo-Exploration2 database[1] to help with the development. Waterloo-Exploration2 database is the largest IQA database so far. There are 3570 pristine images, from which over 3 million distorted images are created with multiple distortion types and distortion levels. We used sixty percentage of the images from the database to train the model. The model is trained in a pair-wise fashion, where two patches cropped at the same spatial location from a pristine reference image and anFD image and fed into the network.

3.2 DR Baseline Model Development

The first idea of developing a DL-based DR baseline model is to use the same model structure as FR-EONSS, but with a different training method. Since DL models try to capture the features from training data, we train a DR model using FR-EONSS structure by feeding random DR images. We use random DR images for training for two reasons. On the one hand, this is the first step in building a DR model, we would like to try a model that is the simplest. On the other hand, there is no clear definition of degraded reference. To be specific, given an FD image distorted by noise and blur, the DR image can be the same image distorted by noise, the same image distorted by blur, or the same image distorted by any other distortion. What's more, the DR image can even be an image with completely different content or random noise.

The overall structure of the DR baseline model is the same as FR-EONSS, just that the output of this DR baseline model is a DR score instead of an FR score.

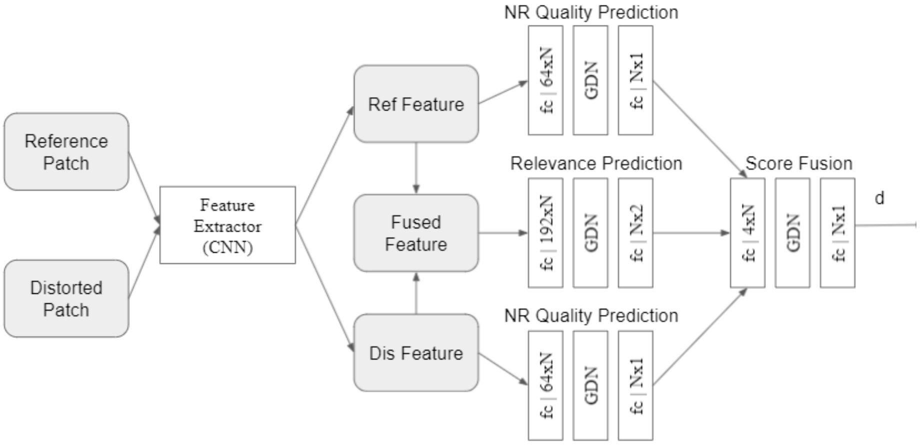


Figure 3.3: DR-Fusion Structure

To move on further in the development of DR baseline models, we introduce a relevance prediction branch into the structure. Since we are feeding random DR images as reference, it is good to let the model know whether the DR image and the FD image is of the same content. Then, the model can handle these two situations using different strategies. The first model developed is called DR-Fusion. The overall structure is shown in Figure 3.3, where there are three branches designed based on the perceptual relevance features. The NR quality prediction branch produces NR scores for the reference patches and distorted patches. The relevance prediction predicts whether the reference patch and the distorted patch have the same content. The score fusion part combines three scores into a single DR score. The general expression is given by:

$$d = f(q1, q2, r) \tag{3.2}$$

where, d denotes the DR score for a distorted patch, r denotes the relevance between a distorted patch and a reference patch, and q1 and q2 denote the NR score for a reference patch and a distorted patch, respectively.

The second model developed is called DR-Residual. The overall model structure for DR-Residual is given in Figure 3.4, where there are also three branches designed based on the perceptual relevance features, including the quality prediction branches and relevance prediction branch. There is also a residual prediction branch that predicts the residual between the NR score of the distorted image and the DR score. The expression is given by:

$$\begin{aligned} res &= f(q1, r, fm) \\ d &= res + q2 \end{aligned} \tag{3.3}$$

where, $q1$, $q2$, and r are defined similarly as in the DR-Fusion model. f here denotes the residual prediction function.

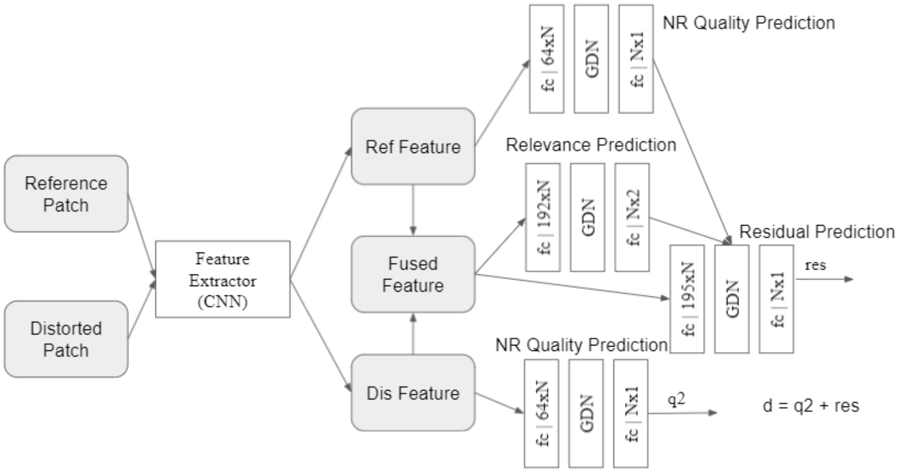


Figure 3.4: DR-Residual Structure

The third model developed is called DR-Combine. The overall model structure for DR-Combine is given in Figure 3.5, where there are four branches, each responsible for one simple task. There are two NR score prediction branches, one relevance prediction branch, and one branch producing the first-step DR score. The weight prediction branch produces weights from NR scores for the reference patch and the relevance scores. Then, these weights are used to improve the first-step DR score, resulting in a final DR score in the end. The equation for this model can be written as:

$$\begin{aligned} w &= f(q1, r) \\ d &= q2 + w * d1 \end{aligned} \tag{3.4}$$

where, $q1$, r,and $q2$ have the same meanings as in the previous models. $d1$ is the first-step DR score, w is the weight to improve the first-step DR score.

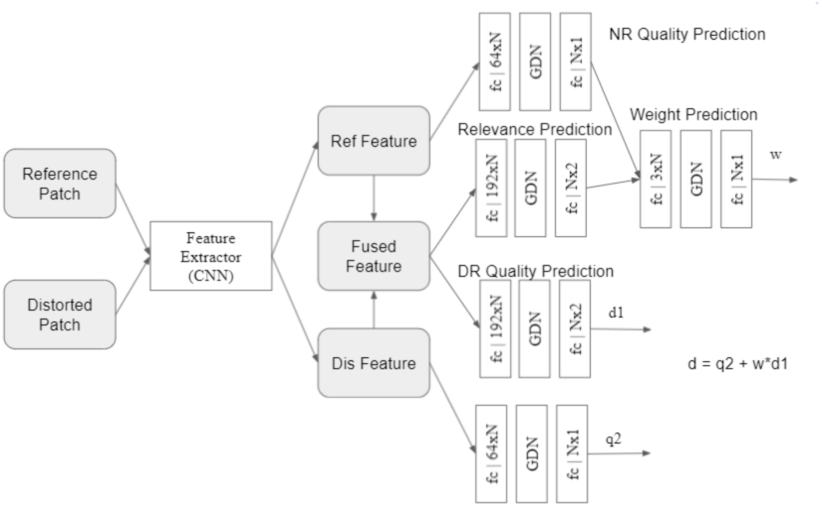


Figure 3.5: DR-Combine Structure

The last model developed in this series is called DR-Multitask. The overall model structure for DR-Multitask is given in Figure 3.6, where the DR-Multitask model contains four branches as well. We take the output from the third branch as the DR score predicted for the distorted image. No specific assumption is made in this model, the model is designed to figure out information for prediction DR score itself all through the entd-to-end learning process.

For training these DR models, we also utilized the Waterloo-Exploation2 database. Instead of pairing each FD image with one reference image, we pair each distorted image with a few random selected DR images.

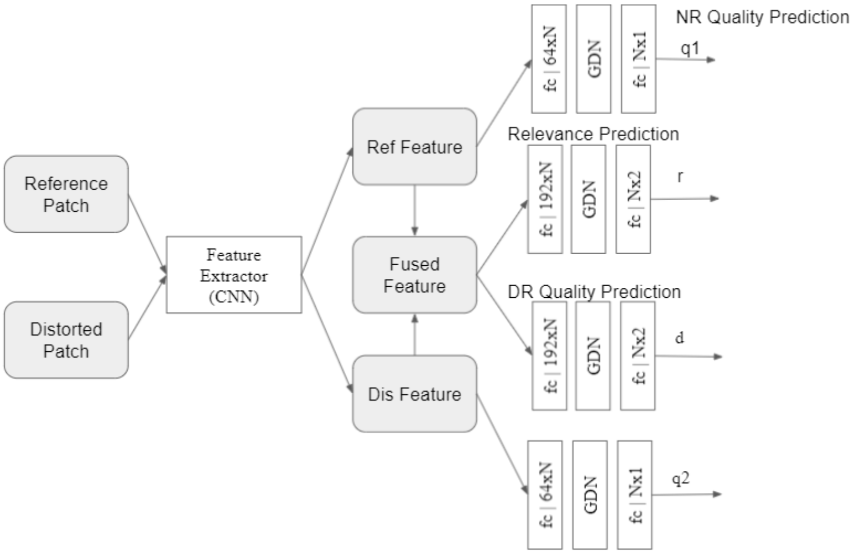


Figure 3.6: DR-Multitask Structure

3.3 Two-Step DR Model Development

As mention above, 2stepQA framework is used to predict the quality of a compressed-after-distorted image using the distorted image as reference. Here we propose to extend this model for a more general DR task.

Firstly, in [56], the NR algorithm and the FR algorithm used in the implementation of the framework are some traditional methods. For example, BRISQUE or CORNIA are used as NR metrics in the framework, SSIM or MS-SSIM are used as FR metrics in the framework. This idea has some drawbacks. One most significant one is that, traditional FR metrics are designed to handle one stage distortion only. However, in our application senario the compressed-after-distorted image has gone through at least two stages of distortions. As such, the traditional FR methods might not handle this type of distorted images well. Also, the performance of traditional NR and FR methods is limited. With the development of the IQA field, more and more large scale IQA databases have been made available. One recently designed multiply distorted IQA database has more than 3 million images[1]. This makes it feasible to use DL technologies to improve the IQA model performance because sufficient labelled data for training is available. Finally, the

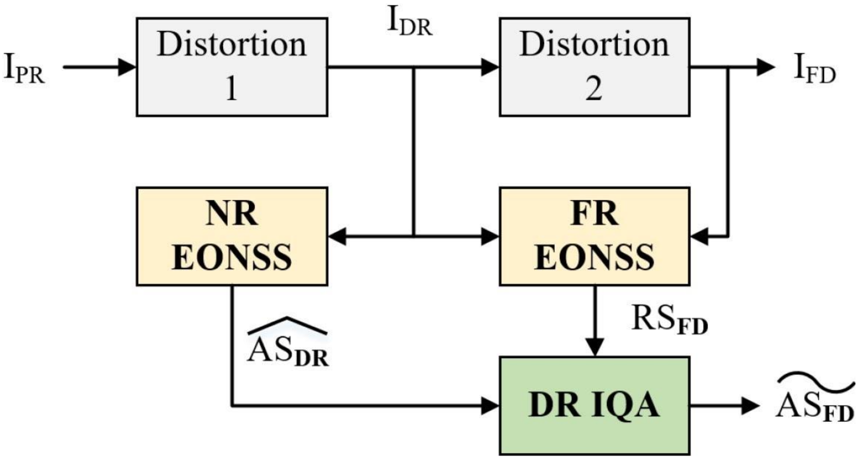


Figure 3.7: DRIQA Model Structure

framework in [56] can only handle compressed-after-distorted images. We aim to extend the framework to handle more types of multiply distorted images.

To address the first problem, we have designed the FR-EONSS model which is able to handle multiply distorted images. We adopt FR-EONSS into the 2stepQA framework, so that the FR module in the framework has the ability to handle multiply distorted images. The second improvement is that, we use EONSS as the NR module inside the 2stepQA framework. Both EONSS and FR-EONSS are developed using DL technology. They are developed using a large IQA database containing more than 3 million images. The large database significantly boosts the performance of EONSS and FR-EONSS. For the last problem, EONSS and FR-EONSS are developed to handle more types of multiply distorted images. More specifically, EONSS is trained to handle three type of singly distorted images (Noise, Blur, Joint Photographic Experts Group (JPEG)) and five types of multiply distorted image (Blur-JPEG, Blur-Noise, JPEG-JPEG, Noise-JPEG, Noise-Joint Photographic Experts Group 2000 (JP2)). FR-EONSS is able to handle the above mentioned nine types of distorted images. Using EONSS and FR-EONSS in 2stepQA framework makes it possible to handle five types of multiply distorted images using the corresponding DR images.

The overall structure of the proposed DR IQA model is shown in Figure 3.7. The distortion types in Distortion 1 include Blur, Noise and JPEG, while in Distortion 2 contains

Noise, JPEG and JP2. The NR module and FR module are pre-trained. The DR IQA module is constructed using FC layers. To train the DR IQA model, we pair each FD image with its corresponding DR image. For example, for a given FD image distorted by blur firstly and noise secondly, the DR image has the same content as the given FD image. At the same time, the DR image should be distorted by blur at the same distorted level as the blur distortion applied in the given FD image.

Chapter 4

Performance Evaluation

In this chapter, we conduct extensive experiments to evaluate the performance of the proposed models. We first evaluate the performance of the FR-EONSS, and compare it with state-of-the-art FR methods on commonly used subject-rated IQA databases. Then, we evaluate the performance of DR models based on different DL architectures, and we compare the performance of these methods with state-of-the-art FR and NR methods. Finally, we evaluate the performance of the proposed two-step DR methods, and compare their performance with some state-of-the-art FR and NR methods.

4.1 FR Model Performance Evaluation

In this section, we evaluate the performance of FR-EONSS which is an FR IQA model we developed for multiply distorted images. We evaluate the model on Waterloo-Exploration2 database test set and other commonly used singly distorted or multiply distorted databases in the IQA field. We report the model performance mainly using Spearman's Rank Correlation Coefficient (SRCC):

$$\rho = 1 - \frac{6\sum d_i^2}{n\left(n^2-1\right)} \tag{4.1}$$

where ρ is the SRCC coefficient. d_i is the difference between two ranks of each observation. n is the total number of observations. Similar results are obtained when other evaluation criteria such as Pearson Linear Correlation Coefficient (PLCC) are used. The results show that the proposed FR model outperforms several state-of-the-art FR models.

4.1.1 Databases, Methods and Criteria used for Performance Evaluation

The Waterloo Exploration2 database is used for training. Waterloo Exploration2 database is the largest database till now in the field of IQA. There are over 3 million images, 3570 reference images in the database. It is a multiply distorted database. For the first distortion process, each of the reference images is distorted by Blur, Noise or JPEG distortion. To cover the whole range of visual quality, eleven distortion levels are used in this process. To produce multiply distorted image, five distortion combinations are used, which are Blur-JPEG, Blur-Noise, JPEG-JPEG, Noise-JPEG, and Noise-JP2. There are eleven levels in the first distortion type and seventeen levels in the second distortion type. The database is mainly used in the training phase. Since the images in the database have multiple resolutions, we extract 235*235 patches from the images for training.

To labeling this database, since there are millions of images, it is impossible to employ humans and conduct subjective test. Therefore, a novel data annotation mechanism, called Synthetic Quality Benchmarky (SQB)[1] are developed, where the perceptual relevance scores are create by using fused FR methods.

The databases used for testing and comparing the models are discussed here. These include Waterloo-Exploration2 database test set, Colourlab Image Database Image Quality (CIDIQ) database, Categorical Image Quality (CSIQ) database, LIVE Release2 (LIVE R2)) database, Tampere Image Database 2013 (TID2013) database, Video Communications Laboratory @ FER (VCLFER) database, LIVE Multiply Distorted (LIVEMD) database, Multiply Distorted Image Database (MDID) database, Multiply Distorted Image Database2013 (MDID2013) database, and Multiple Distorted IVL (MDIVL) database. These databases are best known IQA datasets available in the public domain, widely used for performance comparisons. They can be categorized into singly distorted databases and multiply distorted databases.

CIDIQ is a singly distorted database containing 23 reference images[27]. Each reference image is distorted by six distortion types and 5 distortion levels, producing 690 singly distorted images. All images have the resolution of 800*800. To label the distorted images, each image is given a Mean Opinion Score (MOS) from 1 to 9. There are five levels of scores, denoting Bad, Poor, Fair, Good, Excellent quality, respectively. The subjective tests are carried out at two different viewing distances, thus there are two sets of scores named CIDIQ50 and CIDIQ100, respectively.

CSIQ is a singly distorted database containing 30 reference images[21]. Each reference image is distorted by six distortion types and four to five distortion levels. There are 5000 subjective ratings from 35 observers. The ratings are reported using the form of Difference Mean Opinion Score (DMOS).

LIVE R2 is a singly distorted database developed by the Laboratory for Image and Video Engineering at University of Texas at Austin, and is widely used when testing obejective IQA algorithms[41]. There are 29 reference images. Each reference image is distorted by five distortion types and up to five distortion levels. The total number of distorted images is 779. Scores are given in the range of 1 to 100. Observers are asked to use a slider to determine the quality of a given image. The subjective data is reported in the form of DMOS.

TID2013 is another singly distorted database. It consists of 25 reference images and 3000 distorted images created from these 25 reference images. For each reference image, there are 24 distortion types and 5 distortion levels. The resolution of images is 512*384. The subjective test is carried out by 971 people from five different countries. The subjective test is carried out using a tristimulus methodology [34], where two distorted images along with their reference image are given to the observer at the same time. The observer is asked to select the better one between the two distorted images. Each distorted image is presented nine times and the winning image get one point. The sum of the winning points is reported as the final MOS.

VCLFER is also a singly distorted database[57]. There are 23 reference images, each distorted by four distortion types and six ditortion levels. The subjective rating uses single stimulus methodology[4]. The scores are in the range from 0 to 100. The subjective data is reported in the form of MOS.

LIVEMD is a multiply distorted database[16], and is the first database designed for assessing the quality of multiply distorted images. There are 15 reference images in the database. The reference images are distorted by 3 distortion types (Gaussian blur, JPEG compression, white Gaussian noise) at 3 distortion levels, producing 135 singly distorted images. There are two multiple distortion combinations which are Gaussian blur followed by JPEG compression and Gaussian noise followed by Gaussian noise. This results in 270 multiply distorted images. For subjective rating, the single stimulus strategy is used. The score ranges from 0 to 100. Subjective scores for this database are provided in the form of DMOS.

MDID database has 20 reference images. There are five distortion types: Gaussian noise, contrast change, JPEG, JP2 and Gaussian blur. An image in the database is distorted in three steps. Firstly, the image acquisition process is simulated, for which Gaussian blur or contrast change is added to the original image. Secondly, the transmission process is simulated. Starting with the image from the first step, JPEG or JP2 is added to the image. In the final step, in order to simulate the display process, Gaussian noise is added to the image from the second step. In each step, the distortion is added randomly. 80 distorted images are created from one reference image. Pair comparison is used to conduct subjective ratings, for which two distorted images and one reference image are presented to the observer, and the observer is asked to tell which distorted image has better quality. The subjective socres are presented in the form of MOS.

MDID2013 database has 12 reference images and 324 multiply distorted images. It contains three distortion types which are Gaussian blur, JPEG and white Gaussian noise, and there are three distortion levels for a given distortion type. There is only one type of multiple distortion combination (Gaussian blur followed by JPEG compression). Subjective experiment uses the single stimulus strategy, and the subjective ratings range from 0 to 1. The subjective results are reported using DMOS.

MDIVL is another multiply distorted database. There are 10 reference images and two multiple distortion combinations in the database, which are Blur-JPEG and Noise-JPEG. For Blur-JPEG, a reference image is firstly distorted by Blur at seven levels, then JPEG compression is applied to the image at five different levels. For Noise-JPEG, a reference image is distorted by Noise at ten levels and then four levels of JPEG compression is applied to the noisy image. The above mentioned process creates 750 multiply distorted images. Subjective scores range from 0 to 100 rated using the single stimulus strategy. The subjective scores are reported in the form of MOS.

4.1.2 Performance of FR Methods

Datasets/FR Models	FR-EONSS	WaDIQaM-FR	IW-SSIM
LIVEMD	0.8812	**0.8839**	0.8836
MDID	0.9132	**0.9158**	0.8911
MDIVL	**0.9090**	0.8877	0.8588
WA AVG	**0.9069**	0.9031	0.8812

Table 4.1: SRCC Comparison of FR Models on Three Multiply Distorted Databases

To evaluate the performance of FR model, we use two state-of-the-art FR IQA algorithms in the field: IW-SSIM and WaDIQaM-FR. We mainly use SRCC as the criteria for evaluation. The comparison result on 3 multiply distorted database is given in Table 4.1. WA AVG denotes Weighted Average, where the weights for each database is the number of images in the database. From the Table, we can see that FR-EONSS has the best performance compared to IW-SSIM and WaDIQaM-FR on three commonly used multiply distorted databases.

The comparison result on all ten IQA databases is given in Table 4.2, where we can see that FR-EONSS achieves competitive performance on both singly distorted databases and multiply distorted databases. For some large IQA databases such as TID2013 and CSIQ, FR-EONSS has better performance than other FR IQA algorithms. Although FR-EONSS is developed mainly for multiply distorted images, some singly distorted images are added into the training set. Eventually, FR-EONSS produces a balanced performance on both singly distorted and multiply distorted databases.

Datasets/FR Models	FR-EONSS	WaDIQaM-FR	IW-SSIM
CIDIQ50	0.8714	**0.8808**	0.8484
CIDIQ100	0.8157	**0.8409**	0.8564
CSIQ	**0.9459**	0.9215	0.9212
LIVEMD	0.8812	**0.8839**	0.8836
LIVE2	0.9612	**0.9643**	0.9567
MDID	0.9132	**0.9158**	0.8911
MDID2013	0.7773	0.8357	**0.8551**
MDIVL	**0.909**	0.8877	0.8588
TID2013	**0.8052**	0.8016	0.7779
VCLFER	**0.9541**	0.9468	0.9163
WA AVG	**0.8726**	0.8725	0.8559

Table 4.2: SRCC Comparison FR Models on Ten IQA Databases

4.2 DR Baseline Performance Evaluation

In this section, details of the performance evaluation results of DR of different DL architectures are provided. The models are evaluated on the DR database: DR-LIVEMD. We compare the models with FR and NR models when pristine reference images or degraded reference images are given. The results show that these models have good performance when pristine reference is given, but the performance drops significantly when the reference is degraded.

4.2.1 Databases, Methods and Criteria used for Comparison

The DR-LIVEMD database contains 270 final distorted images. There are two distortion combinations. The first one is Blur-Noise, and the second one is Blur-JPEG. For each final distorted image, its degraded reference and pristine reference are both available.

The models being tested include the one with the same structure as FR-EONSS, DR-Fusion model, DR-Residual model, DR-Combine model and DR-Multitask model. Two models offer meaningful reference points in assessing the relative performances of the models under testing: FR-EONSS as the upper bound, and EONSS as the lower bound. Ideally, the DR models should perform better than EONSS but worse than FR-EONSS. The same as the criterion used for FR model comparison, we compare the SRCC results on the DR-LIVEMD database.

4.2.2 Performance of DR Baseline Methods

The results of FR-EONSS, DR of different network architectures, and EONSS on DR-LIVEMD database with pristine reference are provided in Table 4.3. From the table we can see that DR-Residual and DR-Combine have performance comparable to FR-EONSS when the pristine reference is given, all DR baseline models have better performance than EONSS. Overall, DR models' performance is good on DR-LIVEMD database when pristine reference images are given. DR-Residual and DR-Combine models have SRCC around 0.73 which is not too far from FR-EONSS's 0.76.

Algorithms/Data	LIVEMD BN	LIVEMD BJPG	LIVMD ALL
FR-EONSS	**0.7607**	**0.7717**	**0.7639**
DR-Residual	0.7211	0.7425	0.7309
DR-Combine	0.7199	0.7646	0.7366
DR-Fusion	0.6691	0.7627	0.7132
DR-Multitask	0.5474	0.6580	0.5862
DR-Baseline	0.5355	0.5356	0.5213
EONSS	0.4641	0.4283	0.4124

Table 4.3: SRCC Comparison on DR-LIVEMD with Pristine Reference

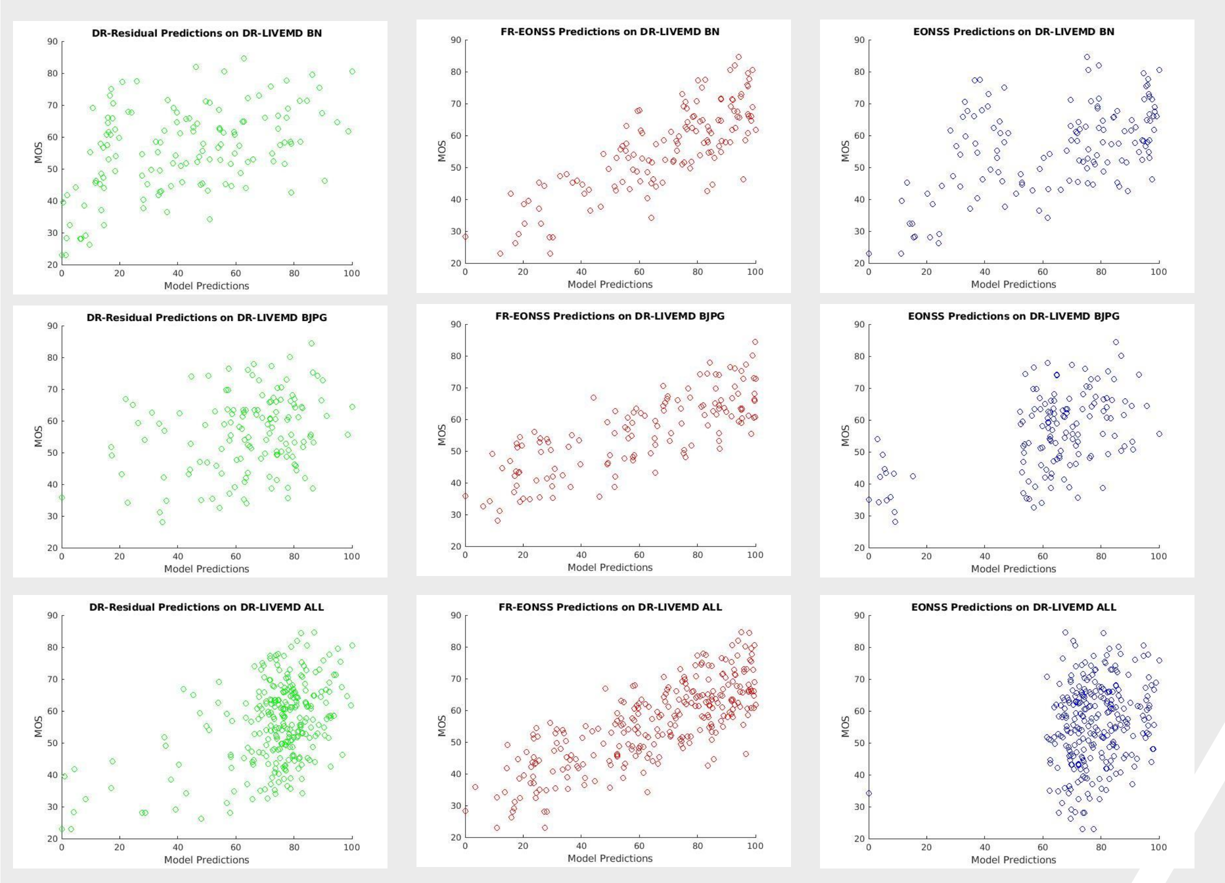
To further verify whether the DR of different network architectures have good performance when only the degraded reference is available, we test DR baseline models on DR-LIVEMD database given degraded reference, the test results are provided in Table 4.4. and Figure 4.1. From the Table, we notice that the performance of the DR models is worse than NR models. This is true for all models including DR-Baseline, DR-Residual, DR-Combine, DR-Fusion, DR-Multitask. Although, DR-Residual improves upon DR-Baseline, the improvement is not sufficient to outperform EONSS. Figure 4.1 gives the scatter plots of DR-Residual, FR-EONSS and EONSS model predictions versus MOS. We can see from the figure that FR-EONSS has good correlation with MOS in all score ranges, DR-Residual does not correlate well with MOS in all score ranges, EONSS correlates well with MOS only in higher score range. These results may due to the defect in the model structure, or problems in the dataset or training process, and are worth deep investigations in the future. To find an alternative, we propose to develop a new DR model based on the two-step framework[56]. The earlier two-step method was proposed for handling the compressed-after-distorted images using degraded reference. On the one hand, the application is limited to compressed-after-distorted images. On the other hand, old fashioned FR/NR IQA algorithms tuned to singly distorted images are adopted. In the next section, we provide evaluation results of the proposed two-step DR model.

Algorithms/Data	LIVEMD BN	LIVEMD BJPG	LIVMD ALL
FR-EONSS	**0.7607**	**0.7717**	**0.7639**
DR-Residual	0.4239	0.2447	0.3237
DR-Combine	0.3885	0.2554	0.2976
DR-Fusion	0.3809	0.2873	0.3058
DR-Multitask	0.4095	0.258	0.2988
DR-Baseline	0.3982	0.0439	0.1934
EONSS	0.4641	0.4283	0.4124

Table 4.4: SRCC Comparison on DR-LIVEMD with Degraded Reference

4.3 Two-Step DR Model Performance Evaluation

Two databases are used for evaluating the DR model performance. The first is DR-LIVEMD which includes 270 multiply distorted images. Blur-JPEG and Blur-Noise are two distortion combinations in this database. The second database is DR-MDIVL. All the distorted images contain multiple distortions. The multiple distortion combinations are Blur-JPEG and Noise-JPEG.



DR-Residual Predictions on DR-LIVEMD BN

FR-EONSS Predictions on DR-LIVEMD BN

EONSS Predictions on DR-LIVEMD BN

DR-Residual Predictions on DR-LIVEMD BJPG

FR-EONSS Predictions on DR-LIVEMD BJPG

EONSS Predictions on DR-LIVEMD BJPG

DR-Residual Predictions on DR-LIVEMD ALL

FR-EONSS Predictions on DR-LIVEMD ALL

EONSS Predictions on DR-LIVEMD ALL

Figure 4.1: Scatter Plots of DR-Residual, FR-EONSS, EONSS Model Predictions Versus MOS on DR-LIVEMD with Degraded Reference

We evaluate the performance of the proposed two-step DR model on the above mentioned databases. State-of-the-art FR model FR-EONSS is considered as the upper bound of the performance, while state-of-the-art NR model EONSS as the lower bound. The evaluation results are given in the Table 4.5.

Data/Models	Two-step DR	FR-EONSS	EONSS
LIVEMD ALL	0.7301	**0.7639**	0.4124
LIVEMD BJPG	0.7402	**0.7717**	0.4283
LIVEMD BN	0.7494	**0.7607**	0.4641
MDIVL ALL	**0.9034**	0.8868	0.8666
MDIVL BJPG	**0.9033**	0.9078	0.853
MDIVL NJPG	**0.9274**	0.9	0.8966

Table 4.5: SRCC Comparison on DR-LIVEMD and DR-MDIVL with Degraded Reference

To have a close comparison of the two-step DR model, FR-EONSS and EONSS. We provide scatter plots of model predictions versus MOS on DR-LIVEMD and DR-MDIVL. As shown in Figure 4.2, the proposed two-step DR model and FR-EONSS have good correlation with MOS, while EONSS correlates well in the higher score range only. The scatter plots for DR-MDIVL database are shown in Figure 4.3, where all three models correlate well with MOS

Details of the testing procedure are provided as follows. All the images evaluated are multiply distorted. When evaluating FR-EONSS, a pair of pristine reference and multiply distorted images are fed into the model. When evaluating the two-step DR model, the multiply distorted image is paired with its degraded reference. For evaluating EONSS, only multiply distorted images are fed into the model. The results from Table 4.5 suggests that, the proposed two-step DR model is successful. Even when only degraded reference images are provided, the performance reaches the upper bound. In summary, the proposed two-step DR IQA model meets all the objectives as laid out in Section 1.2, as it is able to handle multiple distortion combinations and achieve highly competitive performance even when the pristine reference images are not available.

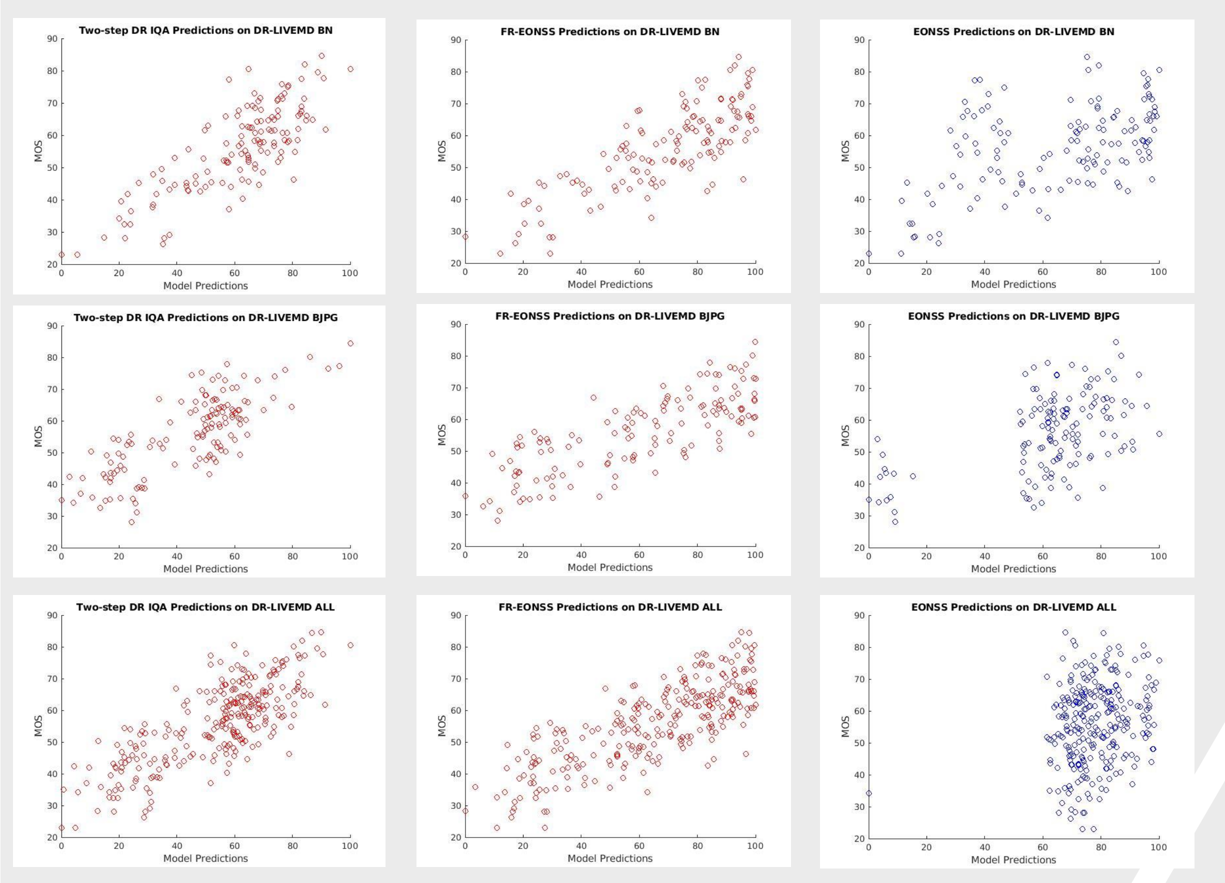


Figure 4.2: Scatter Plots of Two-Step DR Model, FR-EONSS, EONSS Versus MOS on DR-LIVEMD with Degraded Reference

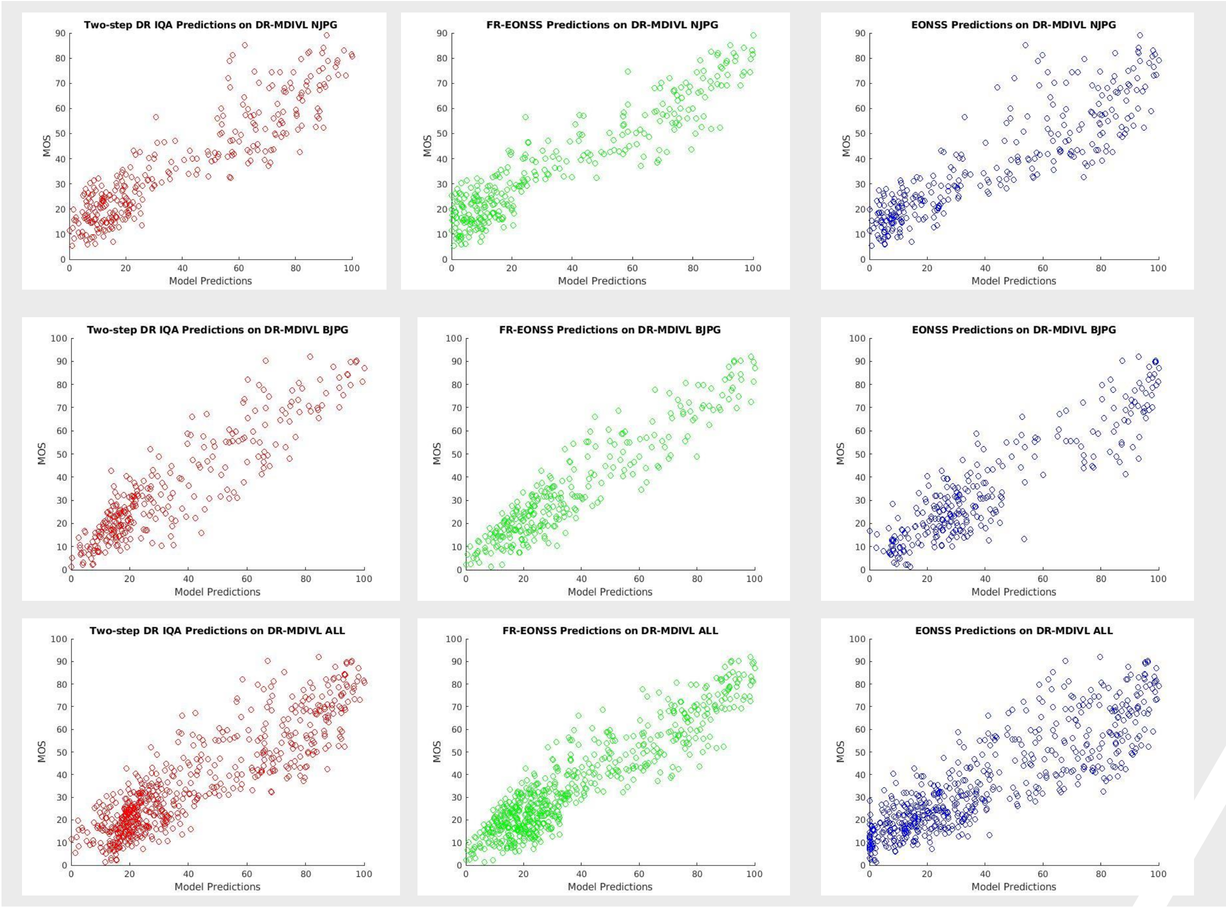


Figure 4.3: Scatter Plots of Two-Step DR Model, FR-EONSS, EONSS Versus MOS on DR-MDIVL with Degraded Reference

Chapter 5

Conclusion and Future Work

5.1 Conclusion

In this thesis we have focused on the new paradigm of DR IQA, which overcomes the limitation of FR IQA approaches that require full access to the pristine reference image, and meanwhile outperforms NR IQA in quality prediction performance. To reach this goal, firstly, we have developed an FR-EONSS model that can handle multiply distorted images with multiple distortion combinations. We train our FR-EONSS model on one of the largest database in the IQA field. The experiment results have demonstrated the competitive performance of FR-EONSS on commonly used multiply distorted databases.

Based on the general architecture of FR-EONSS, we have developed five deep neural network architectures for DR IQA, which we call DR-Baseline, DR-Residual, DR-Combine, DR-Fusion, DR-Multitask, respectively. According to the experimental results, these DR models have good performance when pristine reference images are available, but the performance drops significantly when the pristine reference images are not available.

We extend the two-step IQA framework by employing the EONSS and FR-EONSS models. Our experiments have demonstrated the superior performance of the proposed two-step DR model, which successfully fullfills our objectives.

5.2 Future Work

The work presented in this thesis can be served as a foundation of the following future research directions.

5.2.1 Fully Scalable DR IQA

It is desirable to develop a fully scalable DR IQA model. In the current work, the two-step DR model is able to take degraded reference as input, but its ability is restricted by the Waterloo Exploration2 database as the training dataset. Specifically, the number of multiple distortion combinations is limited (Blur-JPEG, Blur-Noise, JPEG-JPEG, Noise-JPEG, and Noise-JP2). Moreover, the distortion levels in the database is limited (eleven distortion levels for distortion process one and seventeen distortion levels for distortion process two). To achieve the goal of a fully scalable DR IQA model, the database can be extended in two directions. The first is to cover more multiple distortion combinations, and the second is to design a larger database with more distortion levels.

5.2.2 DR Video Quality Assessment (VQA)

The two-step DR IQA model proposed in this thesis can served as a basic step for the development of a DR Video Quality Assessment (VQA) models. In this thesis, only spatial relations between the reference and distorted images are explored. To design a DR VQA model, firstly, a large scale multiply distortion VQA database is needed. This database should be large enough to cover adequate distortion combinations, distortion levels, and video contents. Apart from a large database, a model that can handle temporal relations is needed. A direct extension of the current work is to add on top of our successful two-step approach with an extra component that takes into account the temporal relations. More sophisticated machine learning approaches and deep learning architectures may also be investigated to achieve end-to-end DR VQA.

References

[1] Shahrukh Athar. Image quality assessment: Addressing the data shortage and multi-stage distortion challenges. 2020.

[2] Chanda Bhabatosh et al. *Digital image processing and analysis.* PHI Learning Pvt. Ltd., 1977.

[3] 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, 2017.

[4] RECOMMENDATION ITU-R BT. Methodology for the subjective assessment of the quality of television pictures. *International Telecommunication Union*, 2002.

[5] Mathieu Carnec, Patrick Le Callet, and Dominique Barba. An image quality assessment method based on perception of structural information. In *Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429)*, volume 3, pages III–185. IEEE, 2003.

[6] Mathieu Carnec, Patrick Le Callet, and Dominique Barba. Visual features for image quality assessment with reduced reference. In *IEEE International Conference on Image Processing 2005*, volume 1, pages I–421. IEEE, 2005.

[7] Wu Cheng and Keigo Hirakawa. Corrupted reference image quality assessment. In *2012 19th IEEE International Conference on Image Processing*, pages 1485–1488. IEEE, 2012.

[8] Niranjan Damera-Venkata, Thomas D Kite, Wilson S Geisler, Brian L Evans, and Alan C Bovik. Image quality assessment based on a degradation model. *IEEE transactions on image processing*, 9(4):636–650, 2000.

[9] Keyan Ding, Kede Ma, Shiqi Wang, and Eero P Simoncelli. Image quality assessment: Unifying structure and texture similarity. *arXiv preprint arXiv:2004.07728*, 2020.

[10] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[11] Karen Egiazarian, Jaakko Astola, Nikolay Ponomarenko, Vladimir Lukin, Federica Battisti, and Marco Carli. New full-reference quality metrics based on hvs. In *Proceedings of the Second International Workshop on Video Processing and Quality Metrics*, volume 4, 2006.

[12] Rony Ferzli and Lina J Karam. A no-reference objective image sharpness metric based on the notion of just noticeable blur (jnb). *IEEE transactions on image processing*, 18(4):717–728, 2009.

[13] Leon Gatys, Alexander S Ecker, and Matthias Bethge. Texture synthesis using convolutional neural networks. *Advances in neural information processing systems*, 28:262–270, 2015.

[14] Irwan Prasetya Gunawan and Mohammed Ghanbari. Reduced-reference picture quality estimation by using local harmonic amplitude information. In *London Communications Symposium*, volume 2003, pages 353–358, 2003.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[16] Dinesh Jayaraman, Anish Mittal, Anush K Moorthy, and Alan C Bovik. Objective quality assessment of multiply distorted images. In *2012 Conference record of the forty*

sixth asilomar conference on signals, systems and computers (ASILOMAR), pages 1693–1697. IEEE, 2012.

[17] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5148–5157, 2021.

[18] Fatma Kerouh and Amina Serir. A perceptual blind blur image quality metric. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2784–2788. IEEE, 2014.

[19] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.

[20] Tubagus-Maulana Kusuma and H-J Zepernick. A reduced-reference perceptual quality metric for in-service image quality assessment. In *SympoTIC'03. Joint 1st Workshop on Mobile Future and Symposium on Trends in Communications*, pages 71–74. IEEE, 2003.

[21] Eric Cooper Larson and Damon Michael Chandler. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of electronic imaging*, 19(1):011006, 2010.

[22] Patrick Le Callet, Christian Viard-Gaudin, and Dominique Barba. Continuous quality assessment of mpeg2 video with reduced reference. In *First International Workshop on Video Processing and Quality Metrics for Consumer electronics, Phoenix*, 2005.

[23] Qiang Li and Zhou Wang. Reduced-reference image quality assessment using divisive normalization-based image representation. *IEEE journal of selected topics in signal processing*, 3(2):202–211, 2009.

[24] Xin Li. Blind image quality assessment. In *Proceedings. International Conference on Image Processing*, volume 1, pages I–I. IEEE, 2002.

[25] Anmin Liu, Weisi Lin, and Manish Narwaria. Image quality assessment based on gradient similarity. *IEEE Transactions on Image Processing*, 21(4):1500–1512, 2011.

[26] Lixiong Liu, Bao Liu, Hua Huang, and Alan Conrad Bovik. No-reference image quality assessment based on spatial and spectral entropies. *Signal processing: Image communication*, 29(8):856–863, 2014.

[27] Xinwei Liu, Marius Pedersen, and Jon Yngve Hardeberg. Cid: Iq–a new image quality database. In *International Conference on Image and Signal Processing*, pages 193–202. Springer, 2014.

[28] Kede Ma, Wentao Liu, Kai Zhang, Zhengfang Duanmu, Zhou Wang, and Wangmeng Zuo. End-to-end blind image quality assessment using deep neural networks. *IEEE Transactions on Image Processing*, 27(3):1202–1213, 2017.

[29] Anish Mittal, Anush K Moorthy, and Alan C Bovik. Blind/referenceless image spatial quality evaluator. In *2011 conference record of the forty fifth asilomar conference on signals, systems and computers (ASILOMAR)*, pages 723–727. IEEE, 2011.

[30] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing*, 21(12):4695–4708, 2012.

[31] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012.

[32] Anush Krishna Moorthy and Alan Conrad Bovik. A two-step framework for constructing blind image quality indices. *IEEE Signal processing letters*, 17(5):513–516, 2010.

[33] Anush Krishna Moorthy and Alan Conrad Bovik. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE transactions on Image Processing*, 20(12):3350–3364, 2011.

[34] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al.

Image database tid2013: Peculiarities, results and perspectives. *Signal processing: Image communication*, 30:57–77, 2015.

[35] Ekta Prashnani, Hong Cai, Yasamin Mostofi, and Pradeep Sen. Pieapp: Perceptual image-error assessment through pairwise preference. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1808–1817, 2018.

[36] Klaus Rank, Markus Lendl, and Rolf Unbehauen. Estimation of image noise variance. *IEE Proceedings-Vision, Image and Signal Processing*, 146(2):80–84, 1999.

[37] Soroosh Rezazadeh and Stéphane Coulombe. A novel approach for computing and pooling structural similarity index in the discrete wavelet domain. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 2209–2212. IEEE, 2009.

[38] Michele A Saad, Alan C Bovik, and Christophe Charrier. A dct statistics-based blind image quality index. *IEEE Signal Processing Letters*, 17(6):583–586, 2010.

[39] Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*, 2013.

[40] Hamid R Sheikh and Alan C Bovik. A visual information fidelity approach to video quality assessment. In *The First International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, volume 7. sn, 2005.

[41] Hamid R Sheikh, Muhammad F Sabir, and Alan C Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing*, 15(11):3440–3451, 2006.

[42] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[43] Jing Tian and Li Chen. Image noise estimation using a variation-adaptive evolutionary approach. *IEEE Signal Processing Letters*, 19(7):395–398, 2012.

[44] Ci Wang, Minmin Shen, and Chen Yao. No-reference quality assessment for dct-based compressed image. *Journal of Visual Communication and Image Representation*, 28:53–59, 2015.

[45] Zhongling Wang, Shahrukh Athar, and Zhou Wang. Blind quality assessment of multiply distorted images using deep neural networks. In *International Conference on Image Analysis and Recognition*, pages 89–101. Springer, 2019.

[46] Zhou Wang and Alan C Bovik. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE signal processing magazine*, 26(1):98–117, 2009.

[47] Zhou Wang and Alan C Bovik. Reduced-and no-reference image quality assessment. *IEEE Signal Processing Magazine*, 28(6):29–40, 2011.

[48] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[49] Zhou Wang and Qiang Li. Information content weighting for perceptual image quality assessment. *IEEE Transactions on image processing*, 20(5):1185–1198, 2010.

[50] Zhou Wang and Eero P Simoncelli. Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In *Human vision and electronic imaging X*, volume 5666, pages 149–159. International Society for Optics and Photonics, 2005.

[51] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pages 1398–1402. Ieee, 2003.

[52] Arthur A Webster, Coleen T Jones, Margaret H Pinson, Stephen D Voran, and Stephen Wolf. Objective video quality assessment system based on human perception. In *Human vision, visual processing, and digital display IV*, volume 1913, pages 15–26. International Society for Optics and Photonics, 1993.

[53] Shaoping Xu, Shunliang Jiang, and Weidong Min. No-reference/blind image quality assessment: a survey. *IETE Technical Review*, 34(3):223–245, 2017.

[54] Wufeng Xue, Lei Zhang, Xuanqin Mou, and Alan C Bovik. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Transactions on Image Processing*, 23(2):684–695, 2013.

[55] Peng Ye, Jayant Kumar, Le Kang, and David Doermann. Unsupervised feature learning framework for no-reference image quality assessment. In *2012 IEEE conference on computer vision and pattern recognition*, pages 1098–1105. IEEE, 2012.

[56] Xiangxu Yu, Christos G Bampis, Praful Gupta, and Alan Conrad Bovik. Predicting the quality of images compressed after distortion in two steps. *IEEE Transactions on Image Processing*, 28(12):5757–5770, 2019.

[57] Anela Zarić, Nenad Tatalović, Nikolina Brajković, Hrvoje Hlevnjak, Matej Lončarić, Emil Dumić, and Sonja Grgić. Vcl@ fer image quality assessment database. *AUTOMATIKA: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije*, 53(4):344–354, 2012.

[58] Chen Zhang, Wu Cheng, and Keigo Hirakawa. Corrupted reference image quality assessment of denoised images. *IEEE Transactions on Image Processing*, 28(4):1732–1747, 2018.

[59] 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.

[60] Weixia Zhang, Kede Ma, Jia Yan, Dexiang Deng, and Zhou Wang. Blind image quality assessment using a deep bilinear convolutional neural network. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(1):36–47, 2018.

[61] Heliang Zheng, Huan Yang, Jianlong Fu, Zheng-Jun Zha, and Jiebo Luo. Learning conditional knowledge distillation for degraded-reference image quality assessment. *arXiv preprint arXiv:2108.07948*, 2021.

[62] Luo-yu Zhou and Zheng-bing Zhang. No-reference image quality assessment based on noise, blurring and blocking effect. *Optik*, 125(19):5677–5680, 2014.

[63] Hancheng Zhu, Leida Li, Jinjian Wu, Weisheng Dong, and Guangming Shi. Metaiqa: Deep meta-learning for no-reference image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14143–14152, 2020.