Spring 4-22-2019

# NO REFERENCE IMAGE QUALITY ASSESSMENT

Ravi S. Ravela
*University of Texas at Tyler*

# NO REFERENCE IMAGE QUALITY ASSESSMENT

by

RAVI SHANKAR RAVELA

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Electrical Engineering
Department of Electrical Engineering

Mukul Shirvaikar, Ph.D., Committee Chair

College of Engineering and Computer Science

The University of Texas at Tyler
May 2019

The University of Texas at Tyler
Tyler, Texas

This is to certify that the Master's thesis of

RAVI SHANKAR RAVELA

has been approved for the thesis requirements on
April 3rd, 2019
for the Master of Science in Electrical Engineering

Approvals:

Thesis Chair: Dr. Mukul V. Shirvaikar, Ph.D.

Member: Dr. Premananda Indic, Ph.D.

Member: Dr. Ron J. Pieper, Ph.D.

Chair: Dr. Hassan El-Kishky, Ph.D.

FOR JK

Dr. Javier Kypuros, Ph.D.,
Dean, College of Engineering.

## ACKNOWLEDGEMENTS

I would like to express my special appreciation and thanks to my advisor Professor Dr. Mukul Shirvaikar, for being a tremendous mentor. I am grateful to him for his exceptional guidance, encouragement, support and patience to complete my thesis successfully. I would like to thank Professor Dr. Christos Grecos for his valuable suggestions and ideas in this project. I would also like to thank Dr. Melvin Robinson for allowing me to use his computer and GPU for this project, Also I would like to thank Dept. of Electrical Engineering for providing me a Computer and GPU during the latter part of the thesis for experimentation, I would also like to thank my committee members, Professor Dr. Ron J. Pieper, Professor Dr. Premananda Indic, for serving as my committee members even at hardship. I also want to thank them for letting my defense be an enjoyable moment, and for your brilliant comments and suggestions.

A special thanks to my parents and family for their constant encouragement and financial support to pursue my Master of Science in Electrical Engineering and for all of the sacrifices that they have made on my behalf. I would also like to thank all of my friends who supported me in writing and encouraged me to strive towards my goal. Also, I would like to thank Mrs. Priyanka (Ramapuram) Valluri for many valuable life lessons and making me a better person as of today.

I would like to thank each and every person for their encouragement and support to complete my thesis work successfully.

## TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

## ABSTRACT

# NO REFERENCE IMAGE QUALITY ASSESSMENT

Ravi Shankar Ravela

Thesis Chair: Mukul Shirvaikar, Ph. D.

The University of Texas at Tyler
May 2019

A no-reference image quality assessment (NR-IQA) technique can measure the visual distortion in an image without any reference image data. NR-IQA aims to predict the image quality based on the quality perceived by the Human Visual System (HVS). Image distortions can be caused through the acquisition, compression or transmission of digital images. From the several types of image distortions, JPEG and JPEG2000 compression distortions, addition of white noise, Gaussian blur and fast fading are the most common. Several approaches were proposed to tackle this problem, some were distortion specific and some were general purpose. Of these, Convolutional Neural Networks (CNN) based approaches have proven to be efficient in predicting quality of the images. Most of these models are trained and tested only for single distortion general purpose images, but in the real world the images contain more than one distortion type.

This Work mainly focusses on using deep convolutional neural networks (DCNNs) for NR-IQA, identifying the different distortion types that are present in the image using distortion type classifiers and also, find the distortion quality of each distortion types using a network of DCNNs. We name this novel approach to be multiple DCNN (MDCNN). We fine tune the networks with different activation functions, optimizers and different tunable parameters in CNNs for the better accuracy. Also, we experiment on different patch sizes

that can affect the performance of the system. This proposed model is trained on the LIVE II database and its performance is tested on the CSIQ, and TID 2008 databases which are single distortion. These models achieved high correlation coefficients and accuracy scores on these databases. We further provide the visualization of the inner layers of the DCNN for better understanding of the image quality.

# CHAPTER ONE

# INTRODUCTION

With the advent of ubiquitous mobile cameras, digital photography and powerful photo editing software's on smartphones the trend of sharing images and videos through internet applications is increasing rapidly. With the rapid growth in smartphone ownership across worldwide this trend is going to continue if not increase in the upcoming years. As of June 2015, approximately 760 million images are uploaded to the Snapchat everyday [1], which is a small player compared to the Facebook, Instagram and Google Photos. The digital images captured by the user are subjected to several distortions. These distortions include artifacts during the *capture*, *compression* which might be a lossy one, *transmission* the quality can be altered due to the insufficient bandwidth requirements, in storing the images even the *alteration* of image size according to the requirement of the user device such as from 4K image to 720P for some smartphones. The human visual system can discern some of these alterations and can judge the quality of images [2]. In order to provide good quality for the end user there is a need to detect the quality of the image with respect to the human visual system. Furthermore, this process should be automated and ideally applied in real time.

## 1.1 Image Quality Assessment

Picture quality models that can accurately predict human quality judgments can be used to greatly improve consumer satisfaction, via automatic monitoring of the qualities of massively distributed pictures and videos, and to perceptually benchmark picture processing algorithms such as compression engines, denoising algorithms, and super-resolution systems that substantially affect viewed picture quality. It is very difficult to model these algorithms that are in agreement with the human visual system as the computer stores data only as bits and pixels and is unable to sense the larger picture. Several methods were proposed in the past decade to tackle this problem ranging from subjective Image Quality Assessment (IQA) (use of human observers for IQA) to objective (use of mathematical models for IQA).

Objective Image Quality Assessment (IQA) is again divided into three categories Full Reference (FR-IQA) where the distorted image and its corresponding reference image are available for assessment, Reduced Reference (RR-IQA) where the partial information about the reference image is available, No Reference (NR-IQA) or Blind IQA where no reference image is available for assessment. Further, NR-IQA algorithms were divided into two types distortion specific algorithms, where the user has the knowledge of the specific distortion type(s) that are present in an image, and general-purpose algorithms where the user has no prior knowledge about the distortion type and has to predict the overall quality of the image.

Several algorithms were proposed for NR-IQA over space, one class of these algorithms uses handpicked features such as edge width, color, sharpness to predict the image quality. Examples include S3 [3], LPC [4], JNB [5] which can predict quality of certain distortion types. DIIVINE [6], BLIINDS [7], BRISQUE [8] are general purpose NR-IQA algorithms. The accuracy of these algorithms is acceptable but are constantly outperformed by the Convolutional Neural Network (CNN) based approaches which employ automatic learning features from the raw images. These algorithms automatically select the features that are helpful for the detection of distortions and prediction of the quality of an image. This thesis mainly focuses on general-purpose NR-IQA algorithms which can detect the multiple distortion types present in an image and can further estimate the percentage contribution of each type with respect to the total distortion in the image. The novel model developed uses a network of DNNs and is named as multiple DNN approach (MDCNN). The detailed description of the above model is given in chapter 3 and chapter 4.

## 1.2 Applications of Image Quality Assessment

There are several applications of IQA algorithms in different disciplines. IQA algorithms are used to compare and evaluate the performance of different image processing algorithms and compression techniques and select the best among the possibilities. In the process of embedding a signature into an image for authentication. IQA is used to distinguish the watermarked image and restore the original image. These algorithms are also used in image or video acquisition system to monitor and control image quality, check for artifacts and guide corrections. Similarly, these algorithms are also used in multimedia streaming services to provide better end user satisfaction. These algorithms can also be used in satellite imagery for the detection and removal of artifacts from the certain parts of high dynamic range images.

### 1.3 Challenges in Image Quality Assessment

There has been lot of research in recent years in the field of IQA and there has been some progress. There are still several issues and many new challenges still exists in this field. A survey of some of the challenges that are and current trends in IQA is provided by Wang [9].

- It is highly desired to reduce the complexity of a IQA/VQA algorithm in order to compute them in real time application or even to speed up a process.
- IQA/VQA should also consider the external factors into consideration such as viewing angels, viewing conditions along with the image while predicting the perceptual score.
- IQA algorithms should be able to evaluate the certain portion in an image, in a HDR satellite image only a certain portion of an image is distorted rather than entire image in regular images, desired IQA should be able to detect such portions.
- Desired IQA algorithms should be able to work across different type of images color and monotone images.
- An IQA/VQA algorithm should be able to evaluate the multiple distortion present in an image, as specific distortion types are caused by specific processes and find them is key to eliminate such distortions.

### 1.4 Organization of Thesis

This thesis is divided into six chapters. Chapter 2 gives a brief study about the previous work are related to the current study. Chapter 3 explains the technical terms (CNN, Max Pooling, Convolutional Layers, Activation Functions, Optimizers), and describes about the databases that are used in this experiment. Chapter 4 describes the architectures implemented and experimental procedures. Chapter 5 lists and analyzes the simulation and implementation results for the architecture using different parameters. Chapter 6 consists of the conclusions and future work.

## CHAPTER TWO

## LITERATURE SURVEY

As discussed in Chapter 1, Image Quality Assessment has many applications and is very tough to achieve because of its dependency on the align with human behavior. There are several distortion types that could affect the quality of an image. Furthermore, it is possible that there is the presence of more than one distortion type in a single image and these distortion types tend to be additive in nature and degrade the quality of an image even more. Figure 2.1 shows a few distortion types that can be present in an image. In the following section the different types of subjective IQA algorithms are briefly covered and we also delve into the objective IQA algorithms.

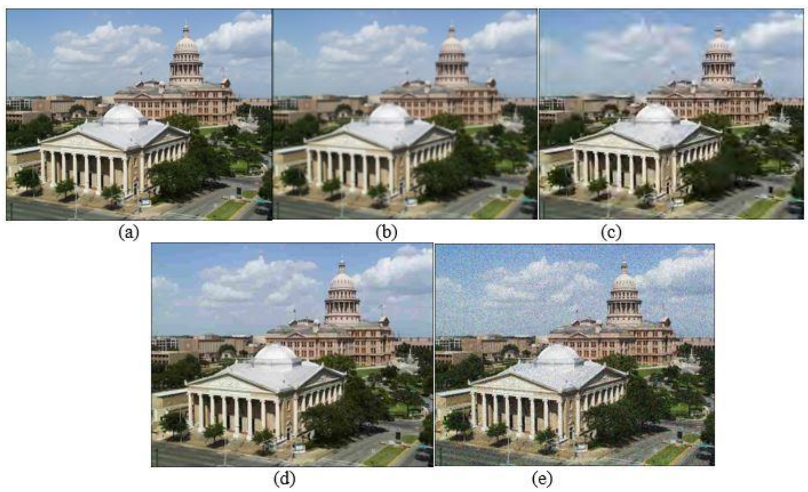



**Figure 1**. Different Distortion types. (a) reference image, (b) blurring, (c) JPEG compression, (d) JPEG2000, (e) white noise.

### 2.1 Subjective Image Quality Assessment

Subjective Image Quality assessment method uses ratings and observations from the human observers to assess the quality of an image. Subjective quality assessment typically focuses on quantifying quality as perceived by an average observer. A group of users were given certain test images to evaluate, all their opinion scores were collected and utilized for computation of the final value. It is the most accurate approach since humans are the end users in most of cases [10]. There are several subjective IQA methods, summarized as follows:

1. *Single Stimulus Rating:* Images were displayed for a short period of time and the users were asked to rate the quality of an image. Rating can be of form continuous from (0-100) or of categorical form categories ranging from good quality to poor quality.
2. *Pair-wise Similarity Judgement:* Two images are displayed, and the user has to predict the quality of an image with respect to the other image in the continuous or categorical form.
3. *Differential Mean Opinion Scores (DMOS):* DMOS rates the test images by calculating the difference between the quality score of the original versus the distorted test image and is calculated by the following equation.

$$Dmos_{i,j} = rating_{i,ref}(j) - rating_{i,j} \qquad \text{(Eq. 1)}$$

Where $rating_{i,j}$ is the raw score for $i^{th}$ subject and $j^{th}$ image and $rating_{i,ref}(j)$ is the raw score for $i^{th}$ subject and reference image which corresponding to $j^{th}$ tested image.

Also, there are several International standards proposed for performing subjective quality assessment like ITU BT 500, ITU p910 and ITUP913. Even though the Subjective assessment of image quality is most accurate and reliable, it is very impractical for real world applications due to time and resource constraints of gathering all people and collecting their opinion scores. Hence, it is practical to use objective IQA methods.

## 2.2 Objective Image Quality Assessment.

Objective image quality assessment uses mathematical models instead of human observers to predict the image quality. These models should be capable of predicting the quality of an image as perceived by the humans. These algorithms do not use humans, are fast and can be used in real time applications of image enhancement and restoration. Based on the availability of the image, IQA algorithms are divided into three categories, namely; Full Reference Image Quality Assessment (FR-IQA), Reduced Reference Image Quality Assessment (RR-IQA) and No Reference Image Quality Assessment (NR-IQA). NR-IQA is further divided into two parts distortion specific NRIQA and distortion generic or general purpose NRIQA.

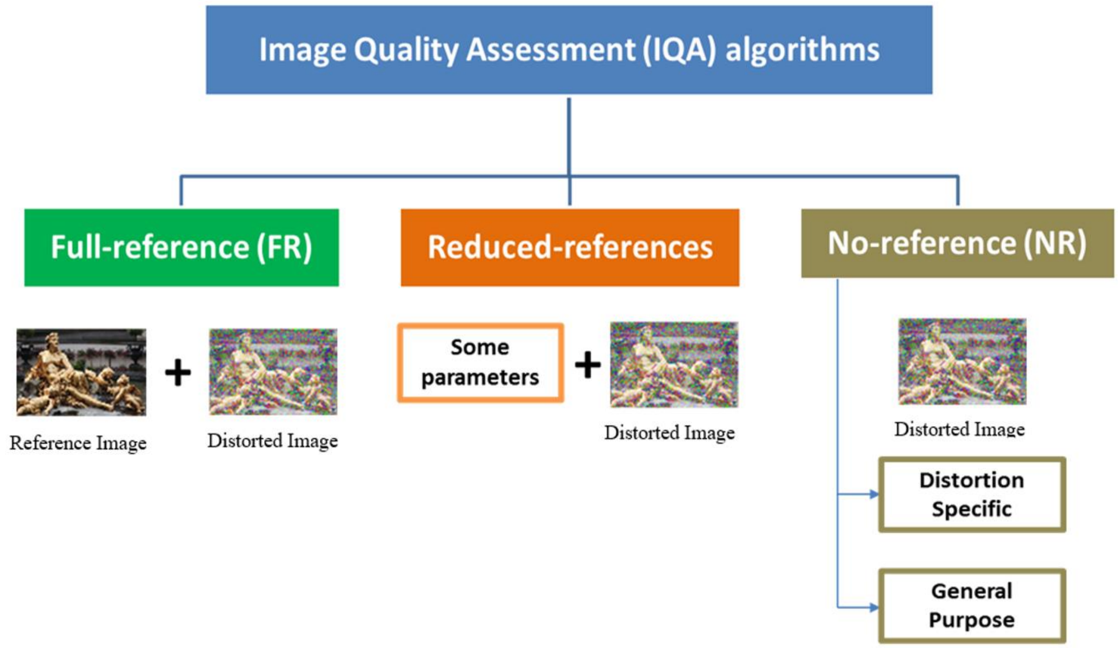



**Figure 2. Objective Image Quality Assessment Algorithms Classification [38]**

### 2.2.1 Full Reference Image Quality Assessment (FR-IQA):

As defined earlier, FR-IQA algorithms use a full reference image or the original undistorted image and a test image to predict the quality of the test image. Several FR-IQA algorithms are proposed over time. One of the algorithms calculates the quality of an image in terms of Peak Signal to Noise ratio (PSNR) which is a ratio of power of distortion and maximum possible power of a system by using Mean Square Error which is shown in the equations below.

$$MSE = \frac{1}{WH}\sum_{i=1}^{H}\sum_{j=1}^{W}\left(I_{ref}(i,j) - I_{\text{test}}(i,j)\right)^2 \quad \text{(Eq. 2)}$$

$$PSNR = 10log_{10}\frac{WH^2}{MSE} \quad \text{(Eq. 3)}$$

Where H,W are the Height and Width of an image. $I_{ref}$, $I_{test}$ are reference and test images respectively. The performance of this evaluation is not up to mark due to its ignorance of features perceived by the human visual system (HVS).

Another FR-IQA algorithm, SSIM (Structural Similarity Index) [11] focuses on the sensitivity of HVS. As HVS is highly adapted for exacting structural information from a scene, the work is based on the degradation of structural information. An image with high quality has more similar structure to the original image and a degraded image has less structural similarity. It is an improved version of the universal image quality index. It outperforms PSNR in predicting the image quality. Another FRIQA algorithm FSIM (Feature Similarity Index) [12] relies on the low-level features such as edge width and zero crossings to estimate the quality of images.

The FR-IQA algorithms by Charrier et al [13] proposes a statistics and machine learning based approach. The proposed model constructs a feature vector and then classifies the image into five quality classes Support vector regression (SVR) is performed based on the quality class to estimate a final score. The scope of FR-IQA algorithms is limited as the reference image not available in most of the cases. They are used in applications like digital watermarking and image compression where a reference image is available

### 2.2.2 Reduced Reference Image Quality Assessment (RR-IQA)

Reduced reference image quality assessment uses partial features of the reference image in estimating the quality of the test image. These partial features include frame information, edges or colors. Successful RR-IQA algorithm should satisfy the following criteria [14] : (a) they should provide an efficient summary of the reference image, (b) should be able to detect different distortion types. (c) should be in sync with human perception of image quality. These types of algorithms mainly find their application in communication systems to monitor the quality of images that are transmitted through these channels and check for distortion.

### 2.2.3 No Reference Image Quality Assessment (NR-IQA)

No Reference Image Quality Assessment predicts the quality of the image without using any reference images. It is by far the most used and most significant because it is a standalone method of predicting the quality of the image without any reference images. It is more challenging to design these algorithms compared to FR-IQA and RR-IQA algorithms. These algorithms find their use in a wide variety of applications from image acquisition systems, image processing systems to communication systems. The main goal of such algorithms is to predict the quality of an image as accurately as possible it is desired to make the algorithm computationally less intensive to allow them work to in real time and to use them in low power embedded devices. As discussed in the above sections these algorithms are classified into two categories distortion specific and distortion-generic or general-purpose NR-IQA algorithms.

Distortion Specific NR-IQA algorithms are able to function only if the distortion type is known to the user. One such algorithm is the Spectral and Spatial Sharpness measure ($s_3$) [3] it focuses on the local perceived sharpness of an image. It utilizes both spectral and spatial properties to produce the sharpness map to predict the blurriness or sharpness in an image. In order to produce the local sharpness maps this model operates on the smaller blocks within an image.

Models such as just noticeable blur (JNB) [5] and cumulative probability of blur detection (CPBD) [15], operate by detecting the edges, followed by estimating the probability of detecting blur at the detected edges. They involve calculating the density function for the obtained probabilities. Quality score is obtained by calculating the final cumulative probability for the probability density functions. These models are used to measure the quality of blur and JPEG2000 compressed images

There are also statistics and machine learning based distortion Specific NRIQA models. Pei *et al* [16] proposed a model for a sharpness measure based on large scale structures. This model uses weighted least squares filter to extract the prominent edges and then probabilities of edge widths ranging from 3 to 11 pixels are calculated. These probabilities along with contrast threshold are fed into a Support Vector Regressor to obtain a single regressor output. This kind of algorithm finds applications where there is a possibility for only a single type of distortion in an image.

General purpose NR-IQA algorithms can predict the quality of an image irrespective of the distortion type. Several algorithms were proposed, and these algorithms can be broadly classified into two categories. ***(a) Natural Scene statistic based approaches (NSS):*** the main idea is to measure the statistical changes that are present in the images that are affected due to the presence of distortions in the image. These algorithms largely depend on handcrafted features that capture the relevant information to identify the distortion levels and predict the quality of an image. ***(b) feature learning based approach:*** which instead of using handpicked features will learn features directly from the images and are derived during the training process of the algorithms. Examples of (a) include, BLIINDS II [7] by saad *et al* which operates in the Discrete Cosine Transformation (DCT) domain, these DCT coefficients are affected by the type and amount of distortion present in an image. In DIIVINE [6] the features are extracted by decomposing the image in the wavelet domain.

BRISQUE [8], by Mittal *et al* is based on spatial natural scene statistics. This model computes locally normalized luminance and uses its parameters as features. These features are fed to a regressor to get the final output as a quality score. As this model operates in the spatial domain rather than wavelet or DCT domain it substantially decreases the computational cost. Apart from BRISQUE, the other algorithms are very slow computationally but the performance is decent.

Second type of these algorithms tend to be more efficient the NSS based approaches because of their ability to extract the features from the raw images. One of such algorithms is CORNIA [17], which uses unsupervised learning to extract the codewords from raw pixel images. It then uses these codewords to learn features for predicting the quality of images. This algorithm has better performance than all the NSS based approaches stated above.

Kang *et al* [18] proposed a CNN based approach. Their model performs local contrast normalization on the images and cuts the images into non-overlapping patches. The training and quality prediction are done on these patches instead of the complete image. By this method they were able to predict the quality of certain portions of each image. Due to CNN's powerful learning capabilities the were able to achieve high accuracy scores.

Bosse *et al* [19] proposed a NR-IQA model based on Deep Convolutional Neural Networks (DNN). They used un-preprocessed image patches, instead of local contrast normalized or global contrast normalized images as input to the network to predict the quality of an image. They implemented the weighted patch aggregation method to improve the accuracy by reducing the effect of patches with minimal changes like blue sky. Hou *et al* [20] proposed a model based on a discriminative deep belief network (DBN). This model first classifies the images into 5 categories and further quality pooling converts these categories into a numerical score. Training is done using unsupervised greedy layer models.

All the above general-Purpose NR-IQA algorithms that were proposed were able to detect the quality of images with multiple distortions but were not able to quantify the amount of distortion of each type that is present in an image. Fan *et al* [21] proposed a model with a network of CNNs, each specialized in predicting the quality of each distortion type and a distortion type classifier to predict the amount of each distortion type present in an image.

# Chapter Three

# TECHNICAL BACKGROUND

The main goal of this chapter is to explain the different components in Convolution Neural Networks, operation of CNN and compare different types of popular Deep Learning (CNN) architectures and their relative advantages. We will analyze which architecture is suitable for our NR-IQA problem. We also discuss different data preprocessing algorithms that are used in this thesis. Furthermore, a brief overview is provided for different types of datasets that are present to evaluate NR-IQA models. Finally, we propose our architecture to tackle the NR-IQA problem.

Deep learning is a branch of machine learning that is specializes in automatically determining the useful features from input data such as images, it can be used to train systems that recognize set of objects in images, group of pixels or other complex patterns like IQA. The main idea behind the deep learning is to replace the handpicked feature extractors, which are difficult to design and are inefficient for complex processes like NR-IQA. Deep learning architectures include multiple stacked layers of neural networks which increases the depth of the network. This architecture helps in learning the high-level features and specific patterns which cannot be perceived by handpicked features. Convolutional Neural Networks are one of such deep learning models and the detailed description of CNN is provided below.

## 3.1 Convolutional Neural Networks (CNN)

Convolutional neural networks are deep, fully-connected feedforward neural networks that are used primarily to classify images, cluster them by similarity, and perform object recognition within scenes. They are inspired by the biology of neurons and visual system structure in animals. A CNN uses a system much like a multilayer perceptron that has been designed for reduced processing requirements. The layers of a CNN consist of an input layer, an output layer and a hidden layer that includes multiple convolutional layers, pooling layers, fully connected layers and normalization layers [22]. There are two key features of a CNN: *Local connectivity* is a concept of connectivity of neurons to a subset of pixels in an input image, and *parameter sharing* is sharing of weights by all neurons in a particular feature map. These two characters helps CNN perform supervised learning.

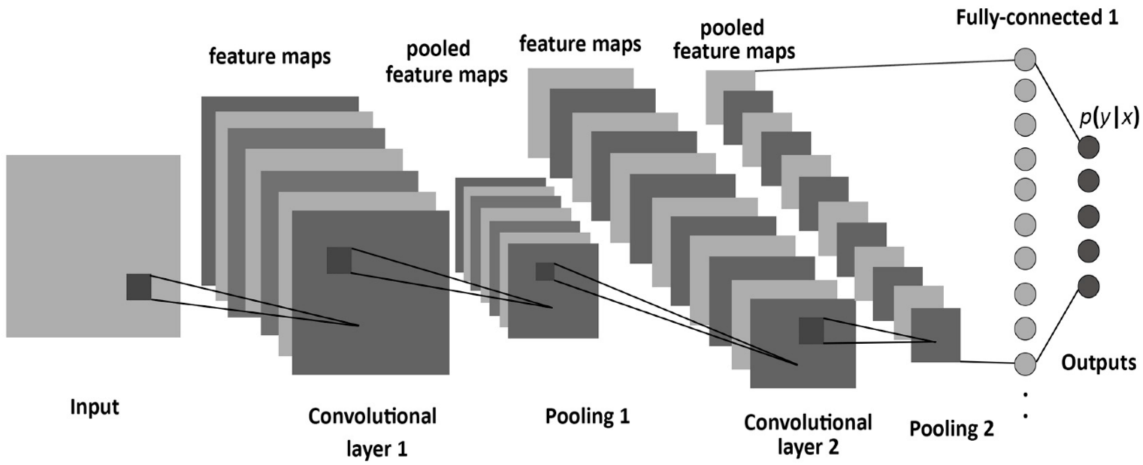



**Figure 3. Typical architecture of a convolutional neural network [22].**

As Shown in Figure 3, CNN architecture consists of several components such as convolutional layers, pooling layers, fully connected layers, input feeds and activation functions. The following sections describes further detail.

### 3.1.1 Convolutional layer

Convolutional layers are the building blocks of a CNN, and consist of a set of learnable kernels, which have a small receptive field and are extended through the full depth of input volume. Each filter takes the input of its height and width finds the convolution through the dot product operation between the input image and kernel values, producing a 2D feature map for that filter. Outputs from all filters from these kernels from one layer forms the output of convolution layer. Height, width and depth are the parameters we can control in these layers. For a convolutional layer with total C number of filters, the output of its $i^{th}$ filter, denoted by $y_i^l$, is computed by the following.

$$y_i^l = s\left(\sum_{j=1}^{C^{l-1}} f_{i,j}^l \, y_i^{l-1} + b^l\right) \qquad \text{(Eq. 4)}$$

Where $b^l$ is the bias vector, $f_{i,j}^l$ is the $i^{th}$ kernel of the convolution layer *l* that is connected to the $j^{th}$ feature map of layer *l-1*, and *s* is the activation function.

### 3.1.2 Fully Connected layer

This layer's functionality is same as a multilayer perceptron. The first fully connected (FC) layer is generally connected to all activations from its previous layers. These layers do not support parameter sharing. The function of a FC layer is to learn weight (W) and bias (b) vectors from the previous layer and forward it to next layer until the output layer. The output of FC layer is determined by the equation.

$$y^l = s(y^{l-1}.W^l + b^l) \qquad \text{(Eq. 5)}$$

Where $b^l$ is the bias vector, $W^l$ is the weight vector, $y^l$ is the current layer and $y^{l-1}$ is the layer before it, *s* is the activation function.

### 3.1.3 Pooling layers

Pooling is a form of nonlinear down sampling its main function is to reduce the dimensionality of the convolutional layers, reduce the number of calculations and avoid overfitting. There are several ways to do pooling, *max* pooling is by far the most used one and *min-max, average* pooling are the other notable algorithms. Max Pooling selects the top two values from each patch in the convolutional kernel and forwards it to the next layer.

### 3.1.4 Activation functions

Activations functions or transfer functions are the ones which supervises the transition between the two layers of a CNN. These activation functions introduce the non-linearity aspect in the implemented model. There are several activation functions and we describe a few of them that are used in this project.

1. **Tanh** is a nonlinear activation function equivalent to the function hyperbolic tangent. The output range is between [-1,1]. It is one of the traditional activation functions. Saturation is one of its main disadvantages. It is slow in operation compared to other activation functions. It is calculated using the formula.

$$\tanh(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \qquad \text{(Eq. 6)}$$

2. **ReLU** Rectified Linear Unit [23] is also known as the ramp function and analogous to half-wave rectification because thresholds the activation at zero. Due to its simple mathematical operations, it is highly preferred over other conventional activation functions. It is calculated using the formula

$$Relu(x) = \max(0, x) \qquad \text{(Eq. 7)}$$

3. **Sigmoid** is also one of the traditional activation functions. It is generally used in the last or output layers of CNN. The range of sigmoid is between [0,1]. It is defined as a bounded , differentiable, real function that is defined for all real input values and has a non-negative derivative at each point. It is calculated by the equation given below.

$$S(x) = \frac{1}{1 + e^{-x}} \qquad \text{(Eq. 8)}$$

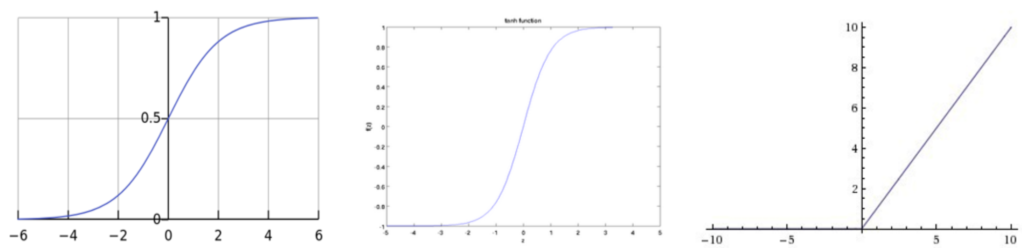



**Figure 4. graph of activation functions (a) sigmoid (b) tanh (c) ReLU**

4. **SoftMax** or normalized exponential function takes an input vector of K real numbers and normalizes it into a probability distribution consisting of K probabilities. It is generally used as the last layer in multiclass classification problem. It is calculated by the following equation.

$$\sigma(\mathbf{z})_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \quad for\ j = 1, \ldots., K\ and\ \mathbf{z} = (z_{\mathbf{1}}, \ldots, z_{\boldsymbol{k}}) \in \mathbb{R}^K \qquad \text{(Eq. 9)}$$

Apart from these, there are several activation functions such as Linear (which directly sends the output instead of amplifying it). binary crossentropy and categorical crossentropy which are again useful for classifying the data.

### 3.2 Training of Convolutional Neural Networks

There are several components that can influence the training of convolutional neural networks which include optimizers, reducing overfitting, data preprocessing methods. Different algorithms that are used in this project are explained below.

#### 3.2.1 Optimizers.

Optimization algorithms in neural network helps to minimize an error function $\boldsymbol{E}(\boldsymbol{x})$, which is dependent on learnable features by changing the values of Weights (**W**) and Bias (**b**) values of a neural network by the process of backpropagation. this project two main optimizers are used, namely Stochastic gradient descent and Adam.

*Stochastic Gradient Descent (SGD):* is an iterative method for minimizing an objective function. It changes the weights of the network for every single training iteration resulting in noisy gradient to escape local minima convergence. Noisy gradient could also make network weights difficult to converge. SGD updates the parameters $\theta$ of the objective $J(\theta)$ by the following equation. $\alpha$ is the learning rate.

$$\theta = \theta - \alpha \nabla_{\theta} E[J(\theta)] \quad \text{(Eq. 10)}$$

*Adam [24]:* Adam or Adaptive momentum estimation, computes individual adaptive learning rates for different parameters from the previous estimates of their gradients. It is computationally efficient and requires less memory. It can be calculated from the equations given below.

$$m_t = \beta_1 * m_{t-1} + (1 - \beta_1) * dp_{t-1} \quad \text{(Eq. 11)}$$

$$v_t = \beta_2 * v_{t-1} + (1 - \beta_2) * dp_{t-1}^2 \quad \text{(Eq. 12)}$$

$$\widehat{m_t} = \frac{m_t}{(1-\beta_1^t)} \quad \text{(Eq. 13)}$$

$$\widehat{v_t} = \frac{v_t}{(1-\beta_2^t)} \quad \text{(Eq. 14)}$$

$$p_t = p_{t-1} - \frac{\varepsilon * \widehat{m_t}}{(\sqrt{\widehat{v_t}} + \varepsilon)} \quad \text{(Eq. 15)}$$

$m_t$, $v_t$ are the estimates of first and second momentum gradient descent. $\widehat{m_t}$, $\widehat{v_t}$ are bias corrected second moments. $\varepsilon$ is a small constant to avoid division by zero. There are several other optimizers like Adagrad, RMSprop Adadelta their performance is subpar compared to Adam optimizer.

### 3.2.2 Controlling Overfitting

Overfitting is a condition where the network learns all the features too closely to a particular dataset including the noise present in that dataset. This results in excellent training accuracy scores, but the model fails to perform with inputs other than the trained inputs. This situation is not at all desirable. In order to reduce overfitting several methods are available. In this project dropout regularization is used, *dropout regularization* is a technique where the features learned by some randomly selected neurons are left out. This means the selected neurons are not activated in the forward and backward passes during certain epoch of the training process. All the other overfitting regularization methods are left to default while using python libraries.

## 3.3 Data Preprocessing

The performance of CNN directly depends on the data that is used for training. Uniformity of all inputs is the key, to make feature extraction process effective. It is advisable to preprocess the input data before training. In this project we experimented with three different types of data preprocessing techniques that are proven to be effective in the literature.

1. **Global Contrast Normalization (GCN):** this algorithm is proven to be an effective way of data preprocessing for object recognition tasks. In this algorithm all the pixels in an image are normalized to zero mean and unit standard deviation for each channel

(RGB) of the input image. Assuming the intensity of pixel at position $(i,j)$ is $I(i,j)$, its normalized intensity value $\hat{I}(i,j)$ can be calculated as.

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(z)}{\sigma(i,j) + C} \quad \text{(Eq. 16)}$$

$$\mu(z) = \frac{1}{P * Q} \sum_{p=1}^{P} \sum_{q=1}^{Q} I(i,j) \quad \text{(Eq. 17)}$$

$$\sigma(i,j) = \sqrt{\frac{1}{P * Q} \sum_{p=1}^{P} \sum_{q=1}^{Q} \left(I(i,j) - \mu(z)\right)^2} \quad \text{(Eq. 18)}$$

where *P, Q, z* are the dimensions of the image $\mu(z)$ is the mean and $\sigma(i,j)$ is the standard deviation.

2. **Local Contrast Normalization (LCN):** It is the most used and proven to method of data preprocessing for NR-IQA tasks. In this algorithm, the local patch data is normalized to zero mean and unit standard deviation for each channel of the input image. It was previously used in NR-IQA algorithms such as BRISQUE [8], IQA-CNN [18], IQA-MCNN [21]. Assuming the intensity of pixel at position $(i,j)$ is $I(i,j)$, its normalized intensity value $\hat{I}(i,j)$ can be calculated as.

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + C} \quad \text{(Eq. 19)}$$

$$\mu(i,j) = \sum_{p=-p}^{p=P} \sum_{q=-Q}^{q=Q} I(i+p, j+q) \quad \text{(Eq. 20)}$$

$$\sigma(i,j) = \sqrt{\sum_{p=-P}^{p=P} \sum_{q=-Q}^{q=Q} \left(I(i+p, j+q) - \mu(i,j)\right)^2} \quad \text{(Eq. 21)}$$

where *P, Q* are the height and width of the image patch, $\mu(i,j)$ is the mean and $\sigma(i,j)$ is the standard deviation of the image patch where the LCN is performed.

3. ***RGBtoGray:*** It is a simple yet effective way in improving the accuracy. This method converts the 3D RGB images to 2D grayscale images. using the equation given below.

$$gray = 0.29 * R + 0.59 * G + 0.11 * B \qquad \text{(Eq. 22)}$$

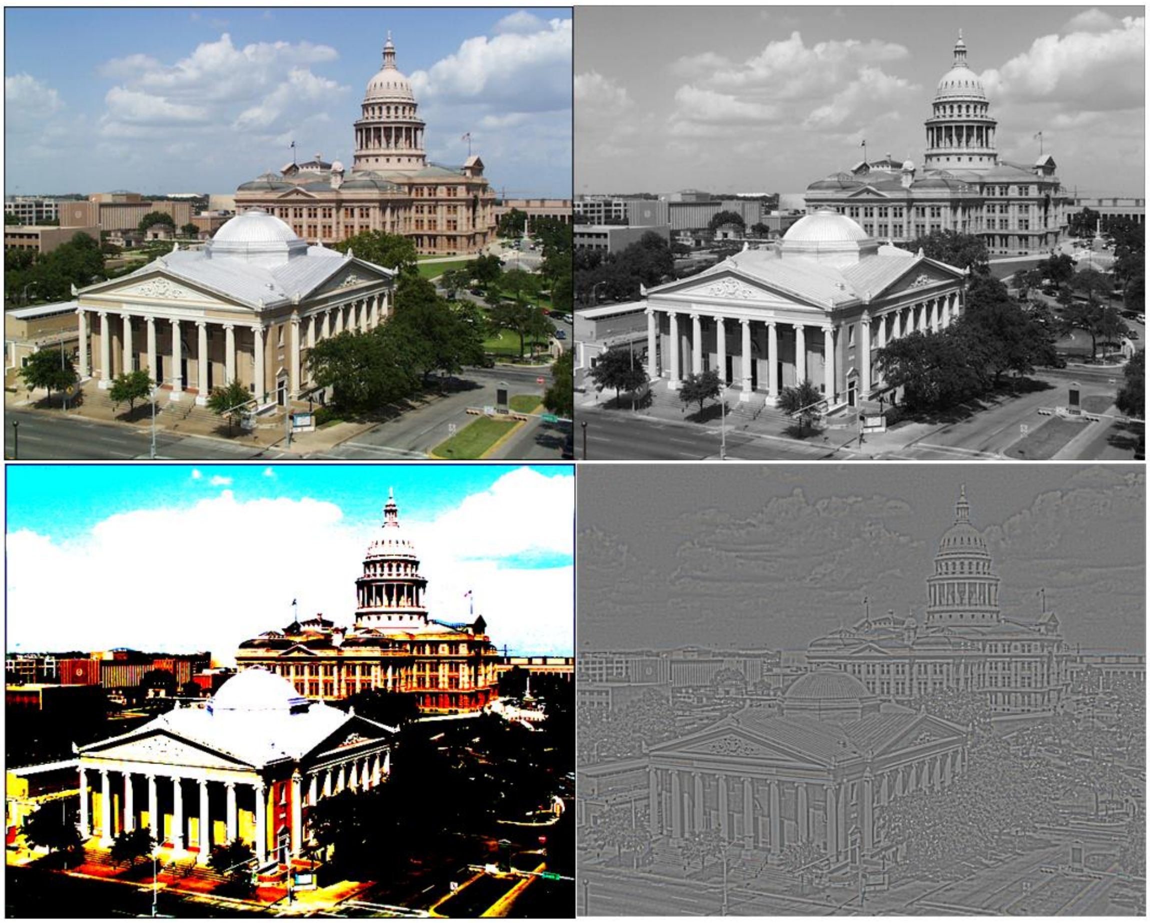


**Figure 5. Image after different preprocessing algorithms (a)original image (b)gray scale conversion (c)global contrast normalization (d)local contrast normalization**

### 3.4 Image Quality Assessment Databases

In this project we use three major databases that are most commonly used to evaluate IQA algorithms, these databases provide us with the reference images, their induced distortions and the DMOS scores which were evaluated across a wide variety of observers from different countries these databases are summarized below.

(1) *LIVE Image Quality database* [25]: This database is developed by LIVE lab at University of Texas at Austin. It contains of 29 reference images of different sizes and 779 distorted images with 5 distortion types namely: JPEG compression, JPEG2000

compression, Gaussian blur, white noise and fast fading. All the images are distorted with only single distortion type. DMOS values ranges from 0-100, with higher values being the most degraded image. DMOS ratings were determined with 15 human subjects.

(2) *CSIQ database* [26]: This database is developed by Computational Perception and Image Quality lab at Oklahoma State University. It contains a total of 30 reference images. All images are of same size and 866 distorted images with six distortion types namely: JPEG compression, JPEG2000 compression, white noise, pink noise, Gaussian blur and contrast stretching. Each distortion type has five distortion levels. All the images are distorted with single distortion type. DMOS values ranges from 0.0-1.0, higher the value lesser the degradation of the image. DMOS ratings were determined with 25 human subjects.

(3) *TID2008 database* [27]: this database was created by Signal Processing Laboratory at Tampere University of Technology. This database has 25 reference images and all images are of the same size, there are 1700 distorted images with 17 distortion types. Each distortion type has five distortion levels. DMOS values ranges from 0.0-9.0 with highest being the least degraded image. DMOS ratings were determined with 838 human subjects.

# CHAPTER FOUR

# METHODS AND EXPERIMENTAL PROCEDURES

In the previous chapter we studied the key components and operation of CNN. In this chapter we analyze a few successful CNN architectures. These architectures along with the discussed pre-processing techniques proved successful in solving problems such as image classification and object detection. We build a few models based on the described architectures to tackle the problem of NRIQA.

## 4.1 CNN Architectures

### 4.1.1 Shallow CNN architecture

There are several shallow CNN architectures proposed over time such as LeNet [28]. Some of the architectures that proved successful for the problem of NR-IQA were IQA CNN [18] and IQA CNN++ [20]. These architectures generally have one or two convolution layers with a small number of convolutional filters of size ranging around 50-100. These are computationally less intensive with around 60,000 tunable parameters. While the accuracy of these networks is descent, feature extraction capabilities of these networks are limited due to the lack of depth in the layers.

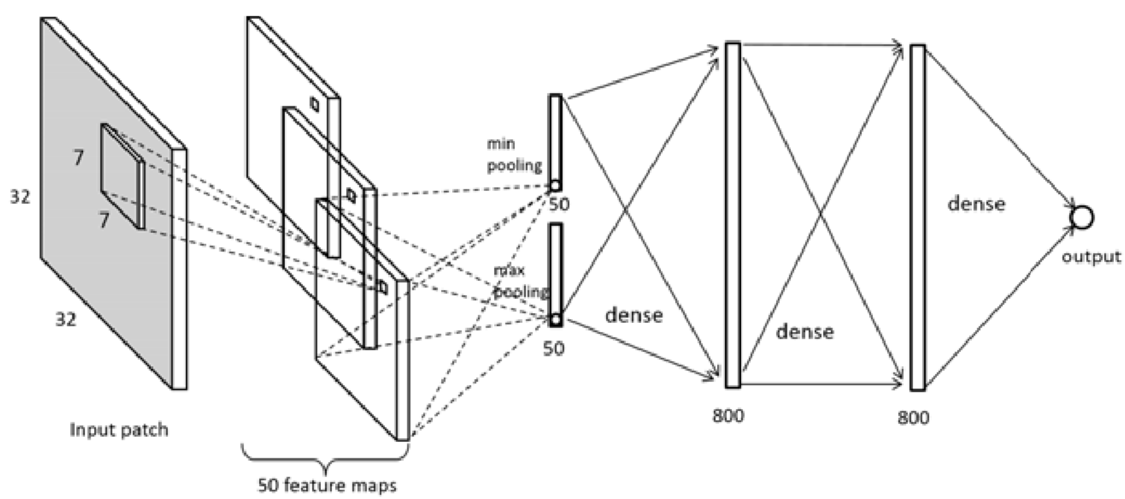


**Figure 6. Architecture for IQA CNN [18]**

### 4.1.2 VGG-16

VGG stands for Visual Geometric Group, it was developed by Simonyan and Zisserman [29]. It consists of 16 convolution layers with two fully connected layers. All the layers are stacked over each other increasing the depth of the CNN. This model only uses $3 \times 3$ convolutional kernels with lots of filters. This is the most preferred model for extracting the features from the images. Because of its huge number of 138 million tunable parameters training and computing it is often a tedious task. With increasing depth it creates vanishing and exploding gradient issues which are also not recommended. Bosse *et al* [19] used an architecture similar to VGG net for NR-IQA and achieved high correlation scores.

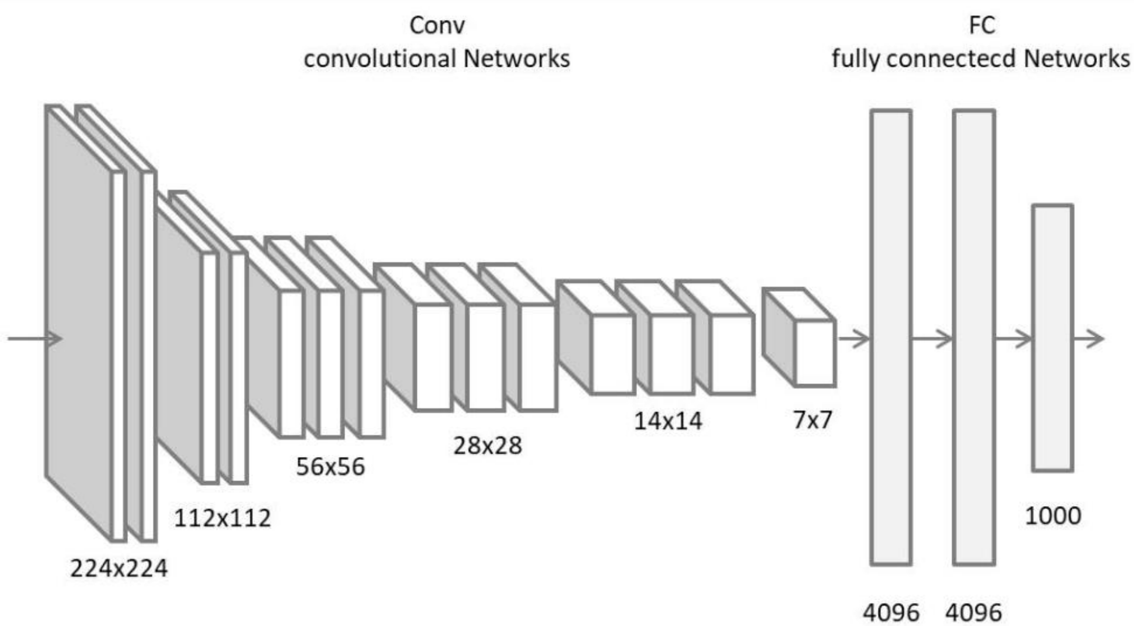



**Figure 7. VGG-16 network architecture.**

### 4.1.3 RESNET

RESNET stands for Residual Neural Network [30]. Unlike the other Deep Neural networks all the layers are not sequential. These are built on micro-architecture modules which are also called as network in networks. Each of these blocks consists of a set of convolutional, pooling layers and normalization. The core idea is to identify the shortcut connections that skip modules, connect few of these micro block modules. Using this technique, they were able to reduce the complexity of the network even with the increase in depth and reduce the number of tunable parameters. This also solved the vanishing and exploding gradient problems. Computation and training of RESNETs is also not tedious compared to VGG net. Hongyu *et al* [31] used architecture like RESNET for NR-IQA.

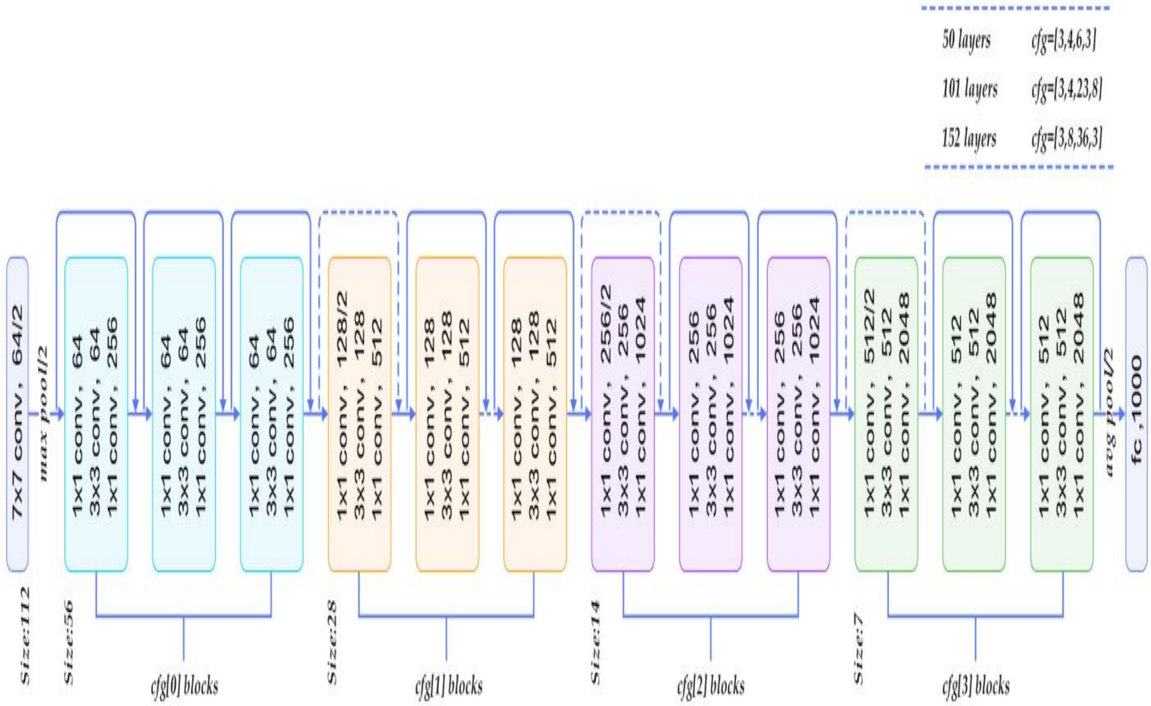



**Figure 8. RESNET architecture.**

### 4.2 Proposed Architecture

With all the problems stated in section 1.4, the proposed model should have to solve the following problems that are present in NR-IQA:

1. It is important for the NR-IQA algorithm to be general purpose and able to predict the quality of images with multiple distortions in it.
2. It is important for the NR-IQA algorithm to be able to find the distortion levels in certain parts of an image, as the distortions in an image are not uniform, and in HDR images the chances of distortion in a specific part of an image is highly likely.
3. It is important to identify the number of distortion types in an image and their probability in the total distortion level of an image.

The ideal model should at least solve the above quoted problems to be a successful general-purpose NR-IQA algorithm. Based on the successful architectures in the literature survey we found that the performance of the model is optimum if we know the distortion types that are present in an image and the percentage of each distortion type. Furthermore, it is desired that every distortion type should have a novel expert IQA network for more accurate results rather than a single combined expert IQA for all distortion types. It is important to evaluate the image for all the distortion types to predict more distortion aware features. For this propose a distortion type classifier is designed. Then finally it is ideal to fuse all the probabilities together to get the single regressor output for evaluation purposes. This model is similar to the one that is proposed by the Fan *et al* [21]. The real differences between their approach and our model is the use of deep convolutional neural networks and RESNETs for the evaluation of the images and specialized expert IQA's for the image quality assessment mainly to improve the performance of the model.

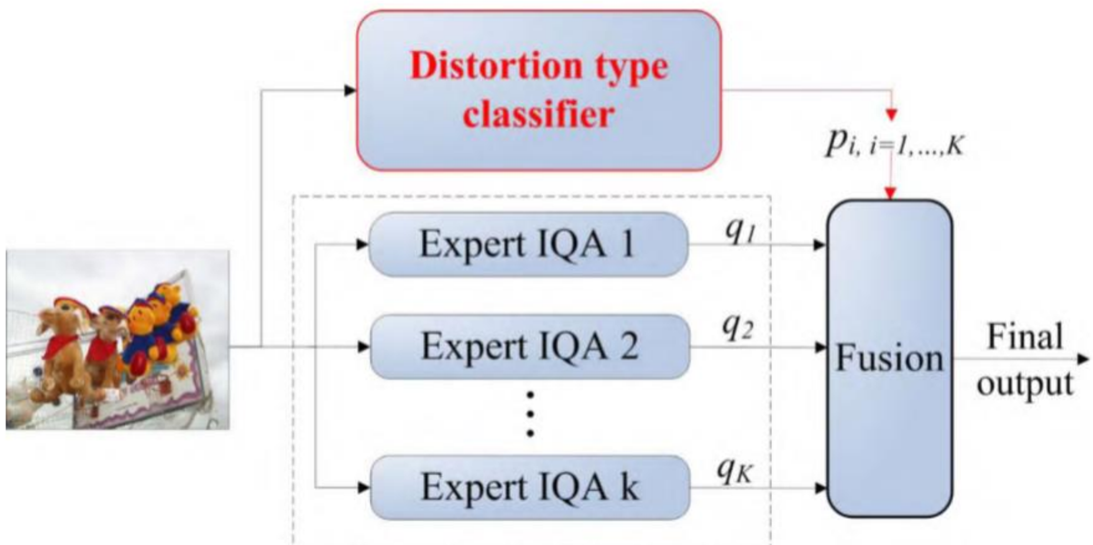



**Figure 9. Proposed MCNN model by Fan *et al* [21]**

Our proposed model architecture consists of three major components Distortion type classifier (DTC) to classify the distortion types and the number of distortion types along with their percentages present in the image, a network of Expert IQA's one for each distortion type, to predict the image quality for that particular distortion type and finally a fusion algorithm to combine all the outputs from the networks to make it a single which is relatable to the DMOS score or human evaluation score. The fusion algorithm proposed is based on weighted average pooling, as inspired from the literature [21]. Such a design is proposed since the distortion noise tends to be additive. It is represented as

$$E(q) = \sum_{q_i \in Q} q_i . p(q_i) \quad \text{(Eq. 23)}$$

where $q_i$ is the distortion level predicted by the expert IQA and the $p(q_i)$ is the probability of that distortion type predicted by the distortion type classifier.

Based on this architecture three such models were built to evaluate the performance of the system. Each model has same architecture shown in the figure 9 but has CNN architectures: shallow network, deep neural network and residual network.

The shallow model is based on IQA CNN [18] architecture. Each of the expert IQAs and DTC. Each of shallow nets networks consists of a *conv7-50, maxpool, FC512, FC256,* and *output(FC-1)* layers. All the layers except output layer are activated by the ReLU activation function. As mentioned above distortion type classifier contains *n* outputs one for each distortion type and is activated by the SoftMax activation function. Expert IQA contains one regressor output and the final layer for this network is activated by the sigmoid activation function.

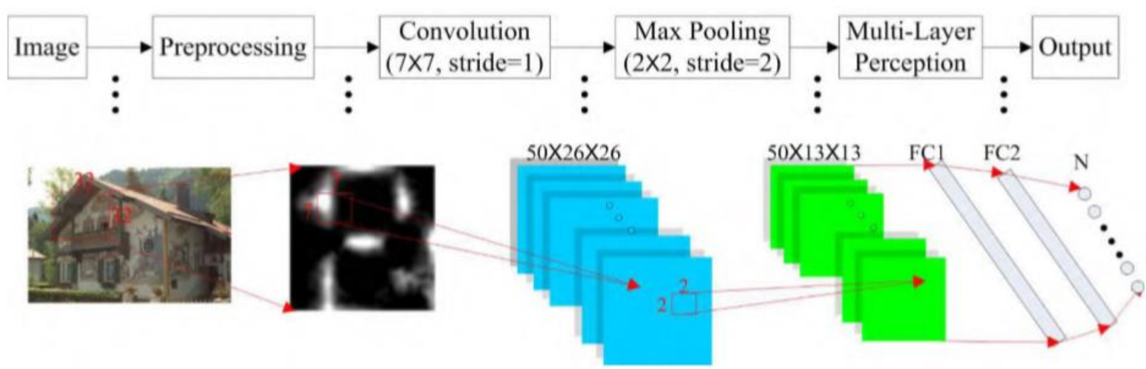



**Figure 10. Shallow CNN architecture flow diagram for DTC and Expert IQA**

The deep neural network model architecture is based on the VGG-16 [29] architecture for each of the expert IQAs and DTC. Each VGG-16 network is used consists of a *a conv3-32, conv3-32, maxpool, conv3-64, conv3-64, maxpool, conv3-128, conv3-128, maxpool, conv3-256, conv3-256, maxpool, conv3-512, conv3-512, maxpool, FC512* and *output(FC-1)* layers. The layer depth is reduced from 16 to 12 in order to reduce the complexity for faster realtime processing. All the layers the except output layer are activated by the ReLU activation function. As mentioned above, distortion type classifier contains *n* outputs one for each distortion type and is activated by the SoftMax activation function. Expert IQA contains one regressor output and the final layer for this network is activated by the Sigmoid activation function. Zero padding is added before all convolutional layers to avoid shrinking of the image pixels as we use image patch as $32 \times 32$ pixel size.

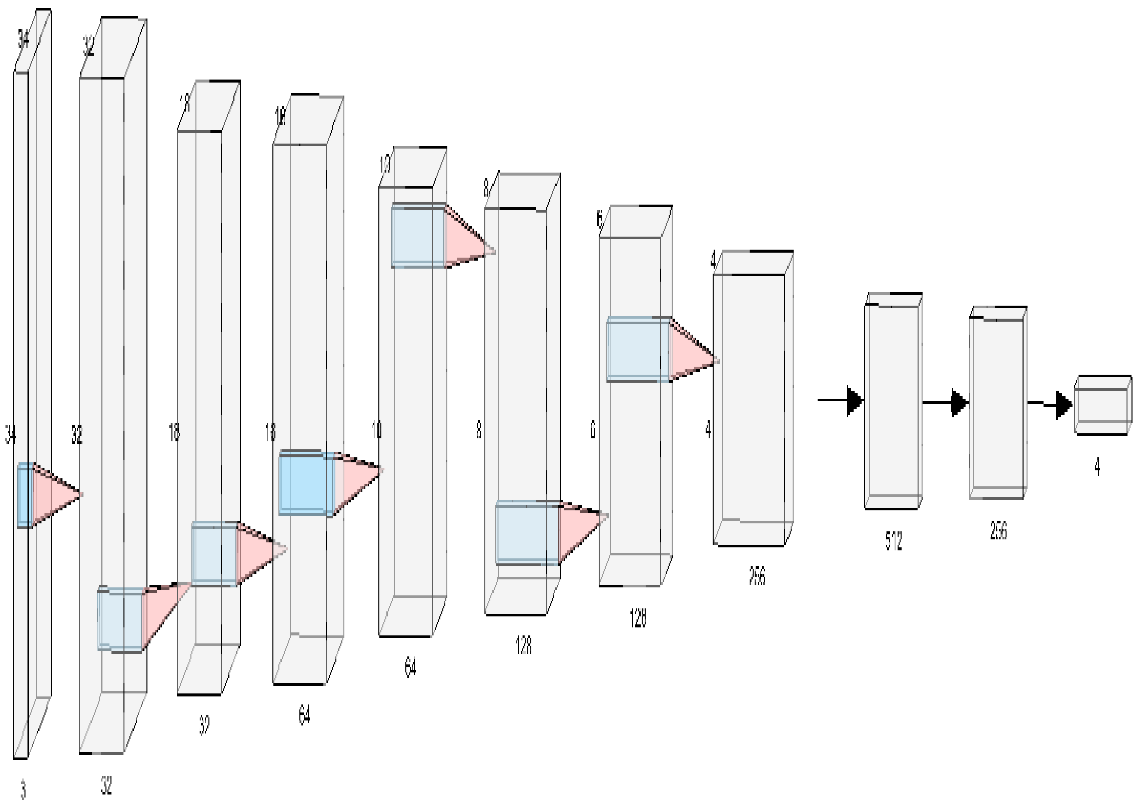


**Figure 11. shortened VGG network for IQA**

Residual network model is based on a 15-layer residual network architecture with 2 residual blocks [30] for each of the expert IQA and DTC. Each network consists of *a conv3-16, batch normalization (BN), L1-conv3-16, L1-BN, L1-conv3-16, L1-BN2, res (merge), conv3-32, BN, L2-conv3-32, L2-BN, L2-conv3-32, L2-BN, res2 (merge), FC-256, FC-128,* and *output* layers. All the layers except the output layer are activated by the ReLU activation function. As mentioned above, distortion type classifier contains *n* outputs one for each distortion type and is activated by the SoftMax activation function. Expert IQA contains one regressor output and the final layer for this network is activated by the

Sigmoid activation function. Zero padding is added before all convolutional layers to avoid shrinking of the image pixels as use image patch of $32 \times 32$ pixel size.

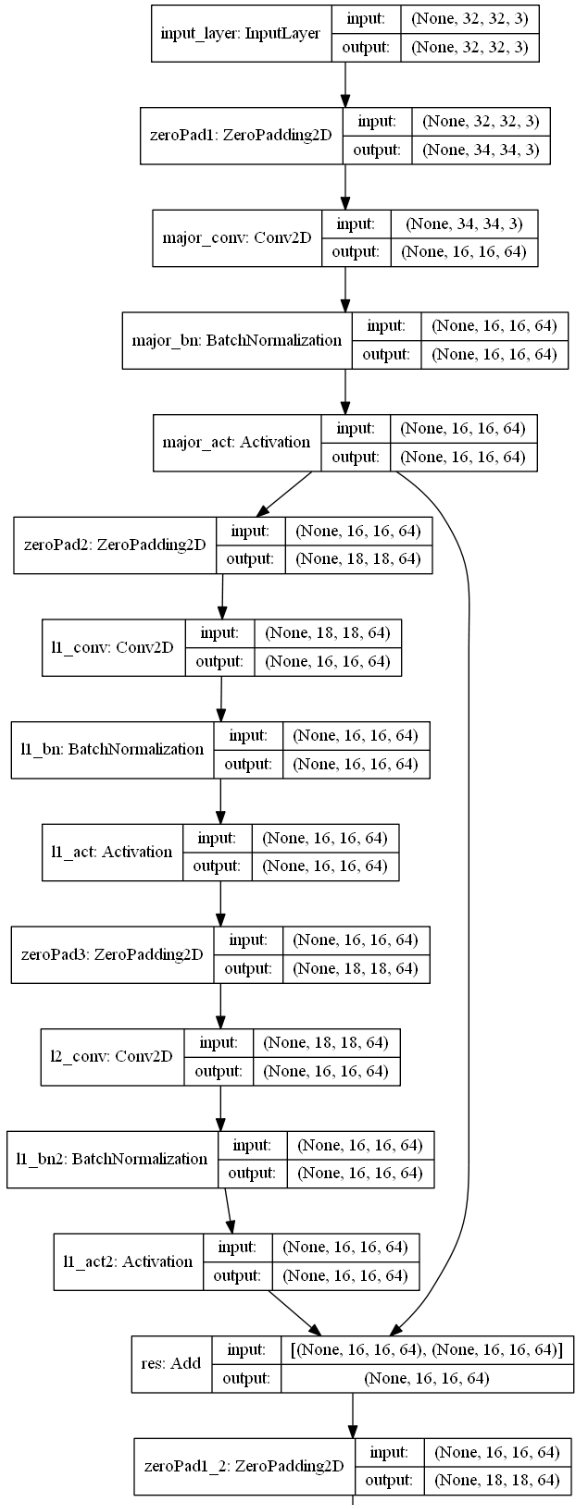


**Figure 12. Architecture of a residual block in RESNET for IQA.**

### 4.3 Evaluation process

We implement correlation coefficients as a method of measurement to measure the prediction statistics and compare their performance with the DMOS values or the reference scores. Two types of such coefficients namely Pearson linear correlation coefficient (PLCC) and Spearman rank order correlation coefficients (SROCC).

*Pearson linear correlation coefficient (PLCC):* is a measure of the linear correlation between two variables X and Y. Its range is between -1 and 1 where -1 shows the negative correlation, 0 shows no correlation at all and 1 shows the positive correlation. It is calculated by the equation given below.

$$\rho X, Y = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \qquad \text{(Eq. 24)}$$

where $\sigma_X$ and $\sigma_Y$ are the standard deviations of X and Y respectively. $\mu_X$ is the mean of X and $\mu_Y$ is the mean of Y and E is the expectation.

*Spearman's rank order correlation coefficients (SROCC):* is a nonparametric measure of rank correlation. It accesses how well the relationship between two variables can be described as a monotonic function. Its range is between -1 and 1 where -1 shows the negative correlation, 0 shows no correlation at all and 1 shows the positive correlation. It is computed by the equation given below.

$$r_\delta = \rho_{rgx,rgy} = \frac{cov(rg_X, rg_Y)}{\sigma_{rg_x}\sigma_{rg_Y}} \qquad \text{(Eq. 25)}$$

where $\rho$ denotes the PLCC, but applied to the rank variables, $cov(rg_X, rg_Y)$ is the covariance of the rank variables, and $\sigma_{rg_x}\sigma_{rg_Y}$ are the standard deviations of the rank variables.

The networks that are developed above needs to be fine-tuned with all the tunable parameters like number of layers in a network, use of different activation functions, use of different preprocessing methods, different optimizers and other initialization techniques, that can improve the accuracy of the system and reduce the overfitting. this process will be covered in the next chapter .

## CHAPTER FIVE

# DESIGN IMPLEMENTATION AND RESULTS

In this chapter, we implement the three models described in the previous chapter with different parameters to evaluate and optimize their performance.

### 5.1 Evaluation with different parameters

The evaluation process follows the standard protocol of splitting 60% of total available images with all their distortion types and levels as the training data, 20% of the available images as the validation data and remaining 20% of the images as the test data. Only the LIVE II database is used for the fine-tuning process and cross database validation is done on the fine-tuned network architecture when comparing our model with other state-of-the-art-models.

#### 5.1.1 Image Preprocessing

Different preprocessing techniques were tested on the three distortion type classifiers. Three expert IQAs were randomly chosen from each of the three architectures to check the performance of the different models with different preprocessing methods. Table 1 shows the normalized accuracy scores for the test image patches that were tested using different preprocessing techniques (tested for 20 epochs and standard architecture).

**Table 1: DTC evaluation using different preprocessing methods.**

| **Accuracy score** | **No preprocessing** | **Gray Scale** | **GCN** | **LCN** |
|---|---|---|---|---|
| **Shallow net** | 0.2783 | 0.3183 | 0.8029 | 0.9487 |
| **Deep net** | 0.9418 | 0.9465 | 0.9471 | 0.9657 |
| **RESNET** | 0.8702 | 0.9102 | 0.8826 | 0.9460 |

Table 1 shows the evaluation of the Distortion type classifiers with 4 different preprocessing algorithms. Local contrast normalization shows better results in all three networks. Also, it can be summarized that shallow CNNs are not able to perceive high level features if the proper preprocessing applied, due to the lack of depth. In case of the deep net and RESNET the models can learn the features even without the proper preprocessing algorithms. Since distortion type classification is a simple task compared to the distortion quality prediction, preprocessing algorithms still needed to be applied.

**Table 2.** PLCC evaluation of expert IQA's using different preprocessing methods.

| **PLCC** | **No Preprocessing** | **Gray scale** | **GCN** | **LCN** |
|---|---|---|---|---|
| **Shallow net** | 0.8657 | 0.9080 | 0.9306 | 0.9628 |
| **Deep net** | 0.9085 | 0.9432 | 0.9476 | 0.9832 |
| **RESNET** | 0.9057 | 0.9335 | 0.9445 | 0.9739 |

**Table 3. SROCC evaluation of expert IQA's using different preprocessing methods.**

| **SROCC** | **No Preprocessing** | **Gray scale** | **GCN** | **LCN** |
|---|---|---|---|---|
| **Shallow net** | 0.8169 | 0.8769 | 0.9238 | 0.9576 |
| **Deep net** | 0.9069 | 0.9392 | 0.9369 | 0.9800 |
| **RESNET** | 0.90 | 0.9376 | 0.9361 | 0.9646 |

Table 2 and Table 3 show the PLCC and SROCC coefficients to measure the quality of the Expert IQA. This test of preprocessing algorithms is done on Gaussian blur expert IQA. Training and testing were done on the LIVE II database. Even for the regression problem for one distortion type, the LCN algorithm outperformed all other preprocessing steps. LCN is chosen as the preprocessing algorithm for the rest of the experiments.

### 5.1.2 Image patch size

Tests were conducted with different patch sizes. We start with a patch size of $32 \times 32$ and increasing the sizes to $64 \times 64$, $96 \times 96$ and $128 \times 128$ respectively.

**Table 4. PLCC Evaluation of impact of patch size on performance of models.**

| **PLCC** | **32** | **64** | **96** | **128** |
|---|---|---|---|---|
| **Shallow net** | 0.9788 | **0.9789** | 0.9646 | 0.9212 |
| **Deep net** | 0.9717 | **0.9758** | 0.9728 | 0.9642 |
| **RESNET** | **0.9793** | 0.9757 | 0.9723 | 0.9759 |

**Table 5.** SROCC Evaluation of impact of patch size on performance of models.

| SROCC | 32 | 64 | 96 | 128 |
|---|---|---|---|---|
| **Shallow net** | **0.9444** | 0.9359 | 0.8879 | 0.8779 |
| **Deep net** | 0.9041 | 0.8944 | **0.9400** | 0.8948 |
| **RESNET** | **0.9186** | 0.9158 | 0.8984 | 0.8951 |

Table 4 and Table 5 show the PLCC and SROCC evaluation of the JPEG expert IQA, trained and tested on JPEG compression from LIVE II database. It is seen that the optimum patch size is $32 \times 32$ and the accuracy decreases with the increase in patch size. Also, with increase in the patch size requires a lot of memory resources to process the input which is not recommended from a computational and real time application point of view. The patch size of $32 \times 32$ is chosen for all upcoming experimentation

### 5.1.3 Optimizers and learning rate regularization.

We experimented with different activation functions at different learning rates and dropouts. First, we try the system with SGD optimizer and then with the Adam optimizer.

**Table 6.** PLCC Evaluation of different optimizers with different parameters.

| PLCC | description | Shallow net | Deep net | RESNET |
|---|---|---|---|---|
| **SGD** | Lr= 0.01, Dcy= 1e-6, mntm= 0.9 | Nan | Nan | Nan |
| **Adam** | Lr=0.001, dcy=0, b1=0.9, b2=0.999 | **0.9788** | **0.9703** | 0.9709 |
| **Adam** | Lr=0.01, dcy=1e-6, b1=0.9, b2=0.999 | 0.9763 | Nan | 0.9773 |
| **Adam** | Lr=0.1, dcy=1e-4, b1=0.9,b2=0.999 | 0.6689 | -0.6200 | Nan |
| **Adam** | Lr=0.001,dcy=1e-4,b1=0.9, b2=0.999 | 0.9780 | 0.9680 | **0.9741** |

In Table 6, $Lr$ represents learning rate, $Dcy$ represents decay of learning rate, mntm represents momentum and $b1$ and $b2$ are the internal parameters of the Adam optimizer. The table shows the importance of learning rates and optimizers in the training of a neural network. The experiment shows that the Optimizer Adam with (learning rate =0.001, decay =1e-6, n1=0.9,b2=0.999) are the better optimization parameters for this problem. All the experiments performed here after this will use these parameters in the models.

### 5.1.4 Activation functions.

In this section all the experiments are performed using LCN preprocessing technique, images with the patch size of $32 \times 32$, optimizer Adam with above mentioned parameters. We experimented with two different activation functions in the convolutional and hidden layers, namely: ReLU and tanh. We also experimented with two different activation functions in the output layer of the expert IQA, namely: sigmoid and linear. In order to use the sigmoid all the label values are scaled between 0.00 and 1.00.

**Table 7.** PLCC Evaluation of different optimizers with different parameters.

| **PLCC** | **Shallow net** | **Deep net** | **RESNET** |
|---|---|---|---|
| **ReLU with Linear** | **0.9780** | 0.9717 | **0.9793** |
| **ReLU with Sigmoid** | 0.9747 | 0.8942 | Nan |
| **Tanh with sigmoid** | 0.9756 | **0.9808** | 0.9788 |
| **Tanh with linear** | Nan | 0.9736 | 0.9708 |

**Table 8.** SROCC Evaluation of different optimizers with different parameters.

| **SROCC** | **Shallow net** | **Deep net** | **RESNET** |
|---|---|---|---|
| **ReLU with Linear** | **0.9472** | 0.9041 | 0.9186 |
| **ReLU with Sigmoid** | 0.9004 | 0.8567 | Nan |
| **Tanh with sigmoid** | 0.8892 | **0.9198** | 0.9178 |
| **Tanh with linear** | Nan | 0.9226 | **0.9331** |

Table 7 and Table 8 shows the performance of the models with different combinations of activation functions. Correct combination of activation functions can increase the performance of the system. From the above tables we can see the use of two exponential activation (tanh and sigmoid) or the two linear activation functions (ReLU and linear) together yields to the better performance for the system rather than using them otherwise (ReLU + sigmoid) or the (tanh + linear). All these experiments are for the expert IQA's. For distortion type classifier ReLU is used for all hidden layers and SoftMax for the output layer which shows the highest accuracy for the classification problems.

### 5.2 Comparison with other State of the art models

Based on the previous observations, the three models were built with these chosen parameters. Each network (Distortion type classifier and expert IQAs) are trained on the LIVE II database and tested on the LIVE II, CSIQ, TIB2008 databases. All these results will be compared with other state of the art models from the literature described in Chapter2.

**Table 9.** PLCC comparison of different state of the art models trained and tested on LIVE II database

| **PLCC** | **JPEG** | **JPEG2000** | **GBLUR** | **AWGN** | **ALL** |
|---|---|---|---|---|---|
| **PSNR** | 0.9463 | 0.9542 | 0.9932 | 0.9211 | 0.9292 |
| **SSIM** | 0.9849 | 0.9805 | 0.967 | 0.9428 | 0.9647 |
| **BLINDS II** | 0.9426 | 0.9386 | 0.8994 | 0.9635 | 0.930 |
| ***BRISQUE*** | 0.9734 | 0.9229 | 0.9506 | 0.9851 | 0.942 |
| ***Kang's CNN*** | **0.981** | 0.953 | 0.953 | 0.984 | 0.953 |
| ***Fan's CNN*** | 0.9570 | **0.9643** | 0.9459 | 0.9869 | 0.9572 |
| ***Shallow net*** | 0.9549 | 0.9122 | 0.9580 | 0.9825 | 0.9227 |
| ***Deep net*** | 0.9428 | 0.9504 | **0.9613** | **0.9888** | 0.9555 |
| ***RESNET*** | 0.9459 | 0.9197 | 0.9472 | 0.9815 | 0.8808 |

**Table 10.** SROCC comparison of different state of the art models trained and tested on LIVE II databases

| **SROCC** | **JPEG** | **JPEG2000** | **GBLUR** | **AWGN** | **ALL** |
|---|---|---|---|---|---|
| **PSNR** | 0.9463 | 0.9542 | 0.9932 | 0.9211 | 0.9020 |
| **SSIM** | 0.9849 | 0.9805 | 0.967 | 0.9428 | 0.9582 |
| **BLINDS II** | 0.9426 | 0.9386 | 0.8994 | 0.9635 | 0.931 |
| ***BRISQUE*** | 0.9734 | 0.9229 | 0.9506 | 0.9851 | 0.942 |
| ***Kang's CNN*** | **0.981** | 0.953 | 0.953 | 0.984 | 0.951 |
| ***Fan's CNN*** | 0.9570 | **0.9643** | 0.9459 | 0.9869 | 0.9531 |
| ***Shallow net*** | 0.9384 | 0.9397 | 0.9515 | 0.9869 | 0.9264 |
| ***Deep net*** | 0.8541 | 0.9622 | **0.9469** | 0.9870 | 0.9535 |
| ***RESNET*** | 0.9094 | 0.9534 | 0.9415 | **0.9900** | 0.8783 |

Table 9 and table 10 shows the PLCC and SROCC comparison of different state of the art models with the proposed three networks, it is shown that deep neural networks and residual networks with multiple expert IQA's will have the better performance compared to the single DNN's or single RESNETs or even with the shallow networks with multiple expert IQA's multiple convolutional neural networks. Also, in this case the performance of deep neural networks is slightly better than the RESNET, but it should be noted that the number of convolutional kernels for each layer (breadth of the CNN) is very high for the DNN compared with the RESNET.

### 5.2.1 Cross Dataset Evaluation.

In this Section the models that are trained with the LIVE II database are tested with the CSIQ and TID 2008 databases to check the independence of the model and its real-world performance. The four common distortion types across the 4 databases namely JPEG, JPEG2000, White Noise and Gaussian Blurring were selected for testing.

**Table 11.** PLCC comparison of different models on CSIQ database

| PLCC | JPEG | JPEG2000 | GBLUR | AWGN | ALL |
|---|---|---|---|---|---|
| **PSNR** | 0.8907 | 0.9468 | 0.9252 | 0.9532 | 0.9218 |
| **SSIM** | 0.9786 | 0.9694 | 0.9496 | 0.8983 | 0.9269 |
| ***Kang's CNN*** | 0.9330 | 0.8106 | 0.9105 | 0.7613 | 0.8110 |
| ***Fan's CNN*** | 0.9654 | 0.9151 | 0.8882 | 0.8590 | 0.8935 |
| **Shallow net** | **0.9656** | 0.8982 | 0.9626 | 0.9304 | 0.9071 |
| **Deep net** | 0.9600 | **0.9223** | **0.9722** | **0.9690** | 0.9147 |
| **RESNET** | 0.9423 | 0.8900 | 0.9321 | 0.9126 | 0.8824 |

**Table 12.** PLCC comparison of different models on TID 2008 database

| PLCC | JPEG | JPEG2000 | GBLUR | AWGN | ALL |
|---|---|---|---|---|---|
| **Shallow net** | 0.6204 | 0.7809 | 0.8107 | 0.8350 | 0.6490 |
| **Deep net** | 0.7630 | 0.7299 | 0.8795 | 0.9150 | 0.8218 |
| **RESNET** | 0.7785 | 0.7338 | 0.8406 | 0.9099 | 0.6797 |

**Table 13.** SROCC comparison of different models on CSIQ database

| SROCC | JPEG | JPEG2000 | GBLUR | AWGN | ALL |
|---|---|---|---|---|---|
| **PSNR** | 0.8879 | 0.9363 | 0.9291 | 0.9361 | 0.9218 |
| **SSIM** | 0.9543 | 0.9605 | 0.9608 | 0.8974 | 0.9325 |
| ***Kang's CNN*** | 0.9114 | 0.7953 | 0.8759 | 0.7534 | 0.7909 |
| ***Fan's CNN*** | 0.9309 | 0.8925 | 0.8167 | 0.8538 | 0.8766 |
| **Shallow net** | **0.9350** | 0.9042 | 0.9223 | 0.9678 | 0.8851 |
| **Deep net** | 0.8727 | **0.9381** | **0.9506** | **0.9715** | 0.8962 |
| **RESNET** | 0.9194 | 0.8812 | 0.9096 | 0.9375 | 0.8692 |

**Table 14.** SROCC comparison of different models tested on TID2008

| SROCC | JPEG | JPEG2000 | GBLUR | AWGN | ALL |
|---|---|---|---|---|---|
| **Shallow net** | 0.6438 | 0.7891 | 0.8790 | 0.8151 | 0.6054 |
| **Deep net** | 0.7904 | 0.7547 | 0.8691 | 0.9139 | 0.8020 |
| **RESNET** | 0.8191 | 0.7351 | 0.8255 | 0.9205 | 0.6206 |

Table 11 and Table 13 are the CSIQ database evaluation comparison and Table 12 and 14 are the TID2008 evaluation comparison between different models. From these results it is evident that the deep models can do better than shallow models, performance of RESNETs even with less convolutional filters is on par with the deep model with a slight reduction in accuracy but the with a lot lesser computational requirement. The accuracy of the shallow net is good for the LIVE II database, but it does not have a par performance on CSIQ and TID2008 databases. It can be observed that all models perform well and showed near accurate results on the LIVE II database. The deep nets struggle with JPEG and JPEG2000 distortions due to overfitting and due to perceiving the over compressed images as blur and noise distortions.

### 5.3 Correlation Scatter Plots

In this section the scatter plots of test images from all the three datasets are shown. These test images were randomly selected twenty percent of images with all their distortion types and distortion levels from all the datasets. As all models are trained on LIVE II database, it expected to have better results on LIVE II databases. Robustness of the models depend on how they perform on the other two databases. In all the scatter plots the X- axis represents the DMOS values and Y-axis represents predicted values.

### 5.3.1 Shallow nets.

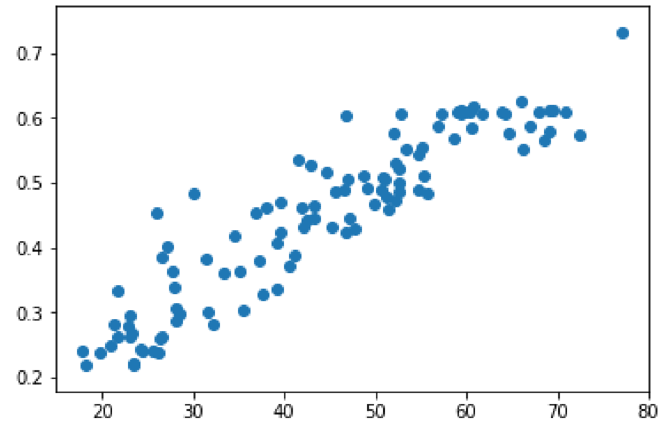



**Figure 13. Scatter plot of LIVE II test images for shallow nets.**

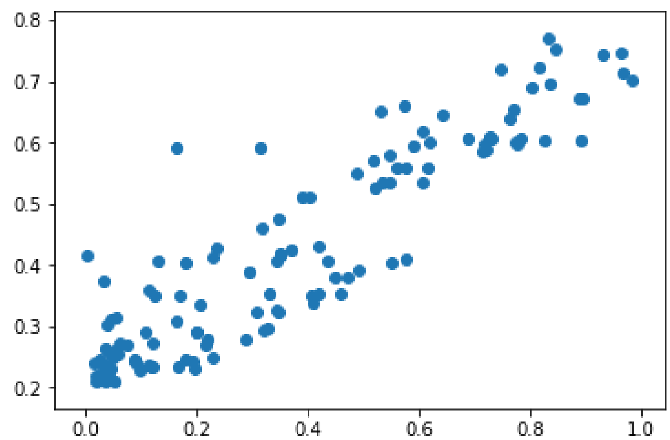


**Figure 14. Scatter plot of CSIQ test images for shallow nets.**

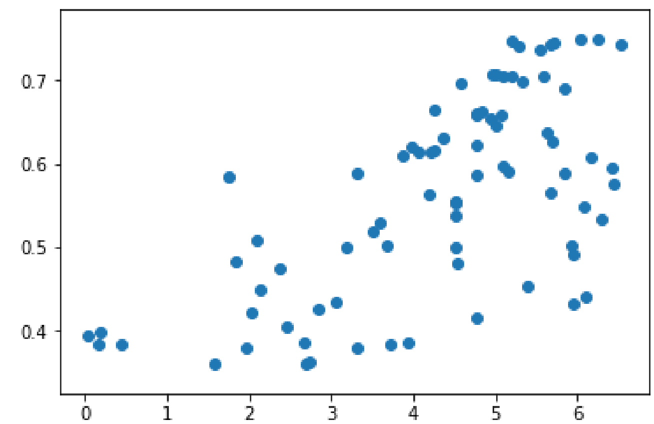



**Figure 15. Scatter plot of TID2008 test images for shallow nets.**

### 5.3.2 Deep nets.

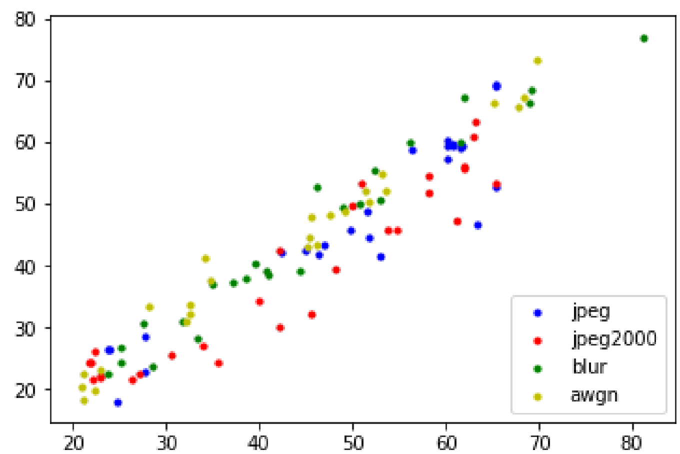



**Figure 16. Scatter plot of LIVE II test images for deep nets.**

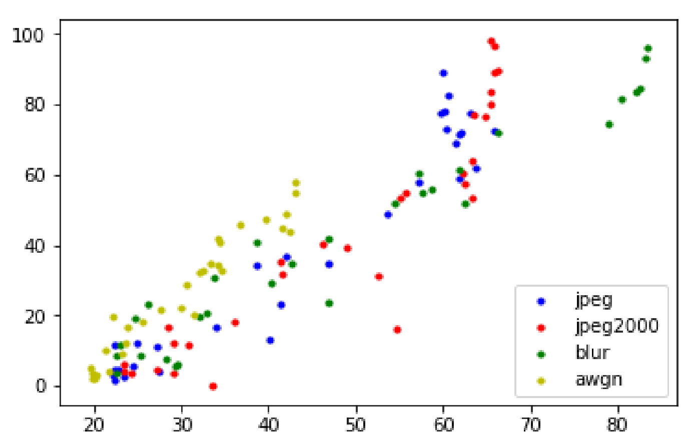


**Figure 17. Scatter plot of CSIQ test images for deep nets.**

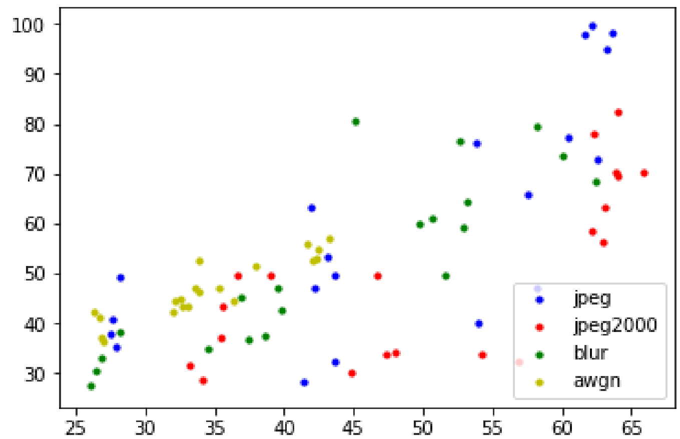



**Figure 18. Scatter plot of TID 2008 test images for deep nets.**

### 5.3.3 RESNETs.

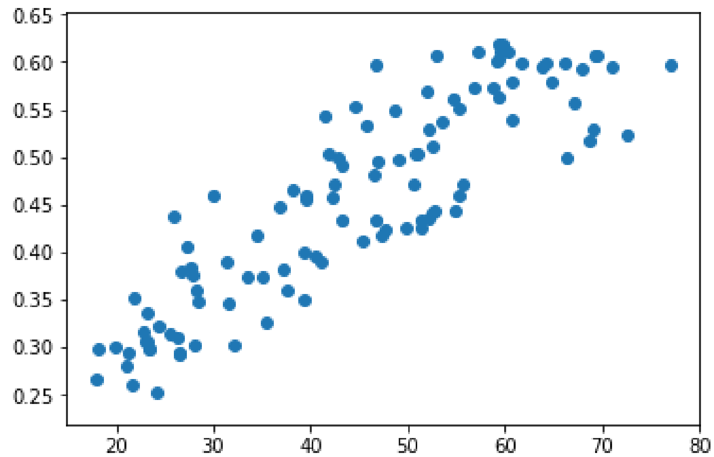


**Figure 19. Scatter plot of LIVE II test images for RESNETs.**

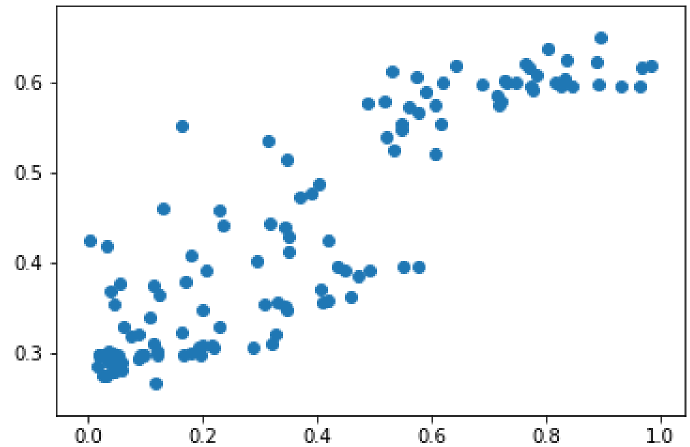



**Figure 20. Scatter plot of CSIQ test images for RESNETs.**

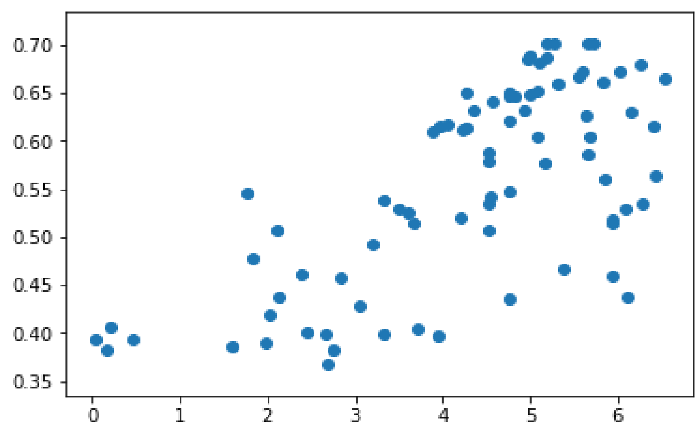



**Figure 21. Scatter plot of TID 2008 test images for RESNETs.**

## 5.4 Visualization of Convolutional Kernels

Convolutional kernels are a major part of the convolutional neural networks. All the features that are learned are stored in those kernels. They are responsible for the prediction and most of the performance accuracy depends on them. It is difficult to understand and assess their internal working because of their huge numbers and their deep architectures.

Lately, visualization techniques have been developed, and with these techniques we can plot the work of convolutional kernels on the image patches. This gives us insights into the working of the CNN kernels the features that are learnt and also the features that are transferred to further layers in a deep network.

In this section we analyze the first convolutional layer in each Expert IQA and DTC networks for all the 3 models to examine how the learning of features occurs in the layers.

From Figures 22, 23 and 24, we can see that the features learned by the kernels of JPEG compression expert IQA are in the shape of blocks, Features learned by the  JPEG2000 are more rounded and of spread shapes,  Blur expert IQA are looking for more spread out edges in the image and AWGN expert IQA are activated by the more isolated points in an image. Also, the Distortion type classifier has learned a mixture of every feature.

### 5.4.1 Shallow nets.

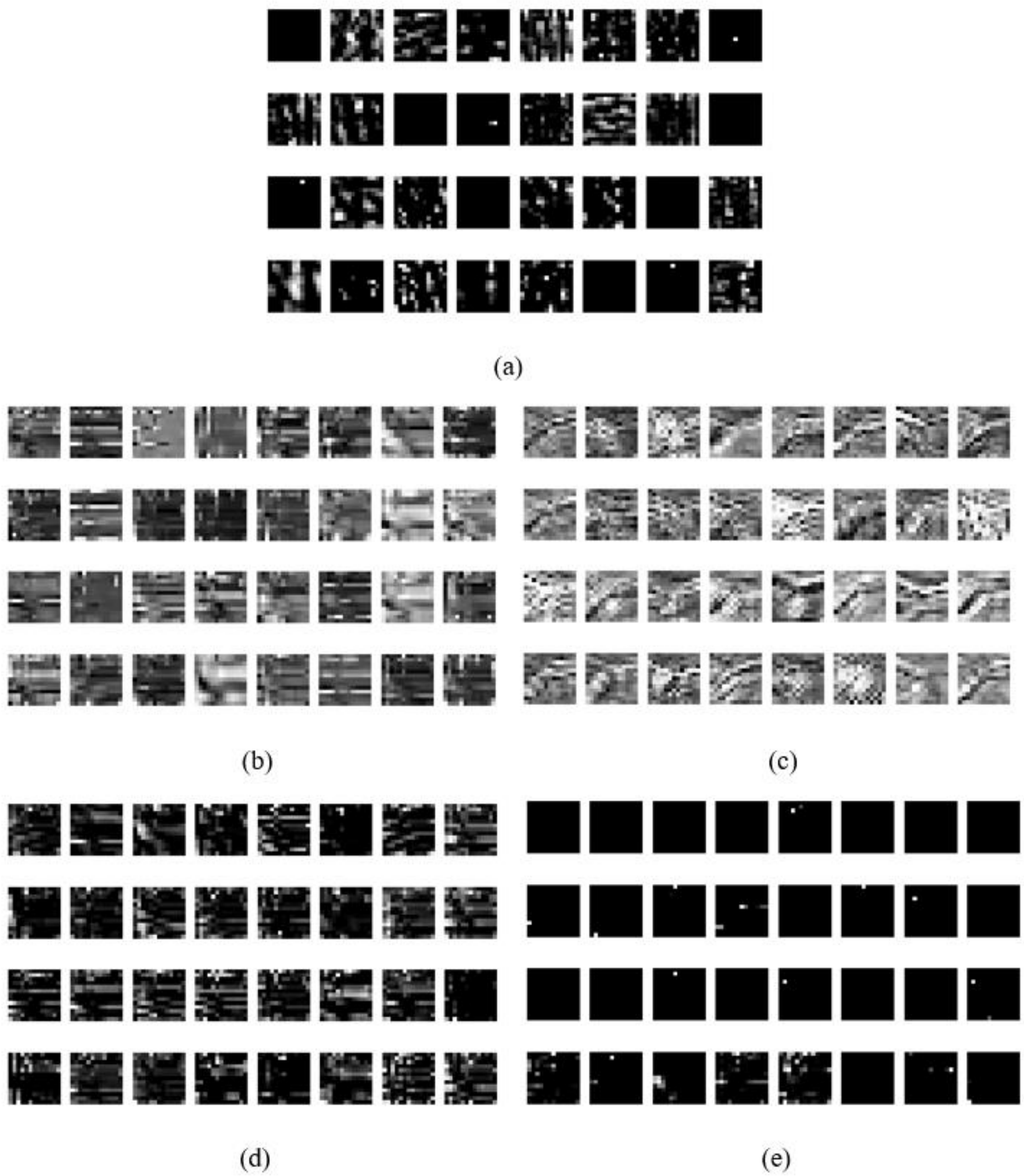
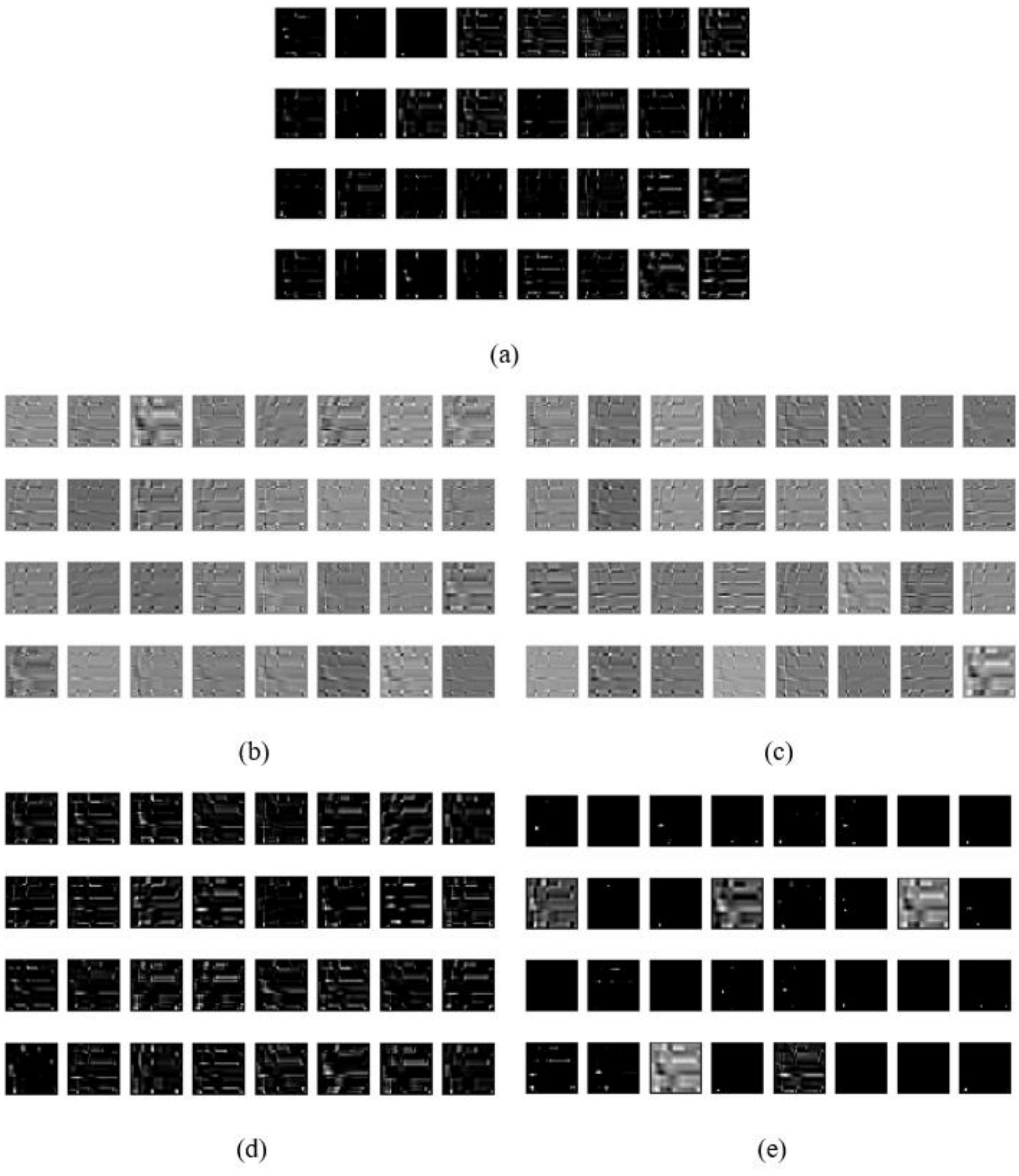


**Figure 22.** Visualization of first 32 filters of first convolutional layer in (a)Distortion type classifier, (b)JPEG expert IQA, (c)JPEG2000 expert IQA (d)Gaussian blur expert IQA (e)WGN expert IQA for shallow nets

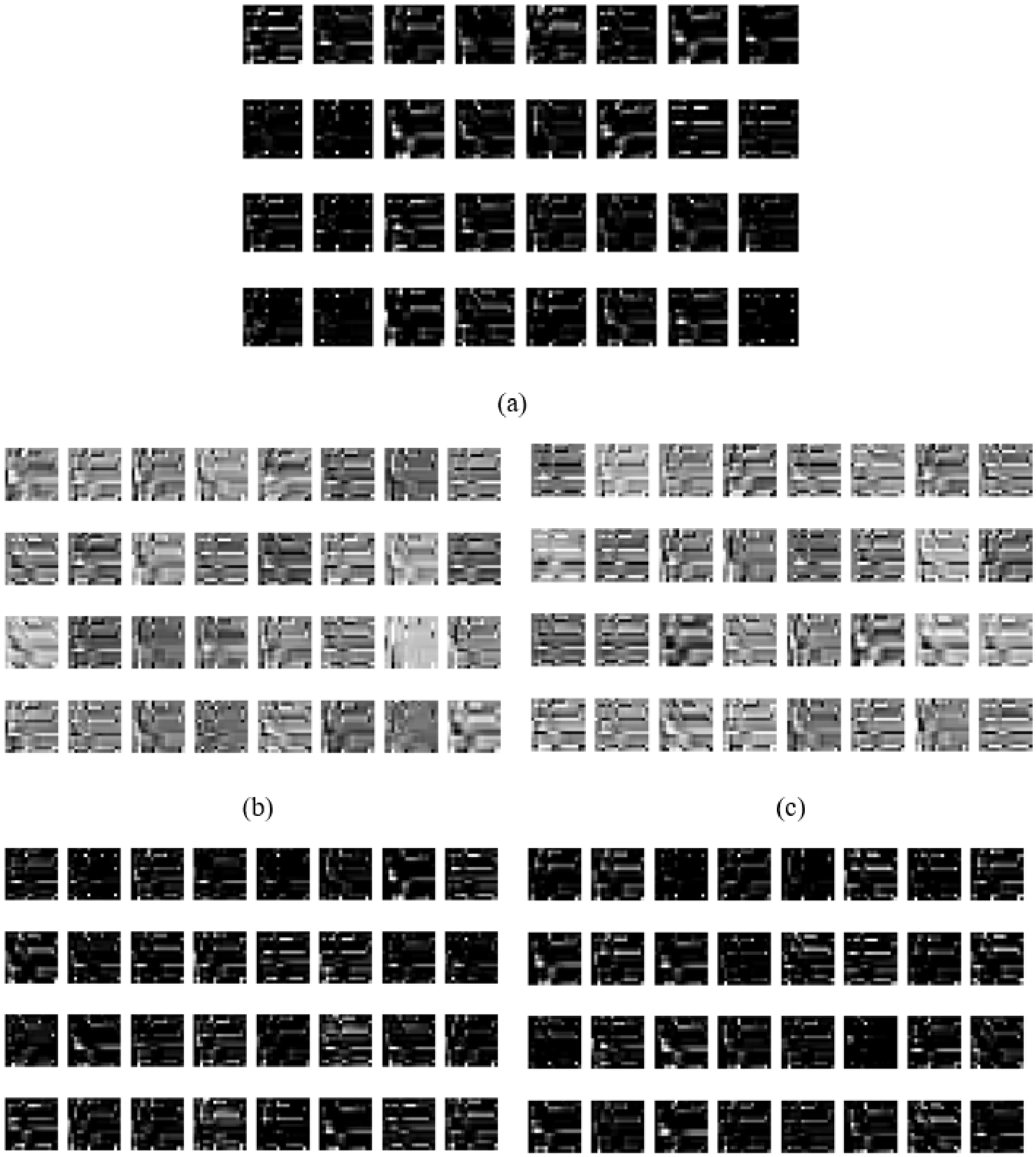
### 5.4.2 Deep nets



(a) (b) (c) (d) (e)

**Figure 23.** Visualization of first 32 filters of first convolutional layer in (a)Distortion type classifier, (b)JPEG expert IQA, (c)JPEG2000 expert IQA (d)gaussian blur expert IQA (e)AWGN expert IQA for deep nets.

### 5.4.3 RESNETs



(a) (b) (c) (d) (e)

**Figure 24.** Visualization of first 32 filters of first convolutional layer in (a)Distortion type classifier,(b) JPEG expert IQA, (c)JPEG2000 expert IQA (d)gaussian blur expert IQA (e)AWGN expert IQA for RESNETs.

## 5.5 Conclusion

This concludes all the experimentation and  results section in this project. In summary, a novel model for NR-IQA was proposed, with multiple expert IQAs to analyze the different distortion types present in an image and predict the quality of an image with respect to human visual system. These models were able to outperform most of the existing techniques with different distortions and mixtures. Each of the expert IQAs is associated with one distortion type, enabling it to learn completely all of its features. Training is a supervised learning process and different parameters were fine-tuned to improve the performance of the system. All the training is done on the LIVE II database and cross database evaluation is done on the CSIQ and TID 2008 databases. Results were tabulated and finally to get insights into the CNN learning, convolutional kernels were visualized.

# CHAPTER SIX

## DISCUSSION AND CONCLUSION

In this study we have demonstrated the use of multiple neural networks for the problem of NR-IQA. Our approach not only improved the accuracy of the system but also able to detect the multiple distortion present in an image. The major contributions of this thesis to the research in image quality assessment is:

- Development of a distortion type classifier that can classify and predict the probabilities of different distortion types in an image.
- Use of deep neural networks such as VGG-16 and RESNET for the problem of NR-IQA and improvement of the networks for better accuracy.
- Use of multiple expert IQAs for image quality assessment. With each one specialized in assessing a specific distortion type.
- Visualization of the convolutional kernels for the better insights into the functioning of convolutional neural networks.

Though this thesis is successful in improving the accuracy of NR-IQA, there are lot of unexplored areas such as increasing the number of convolution kernels in a network, adding more residual layers in RESNETs (this project experimented with only two residual layers). The use of higher dimensional convolutional filters $5 \times 5$ or above can also be explored (we have only $3 \times 3$ convolutional filters for deep networks). Further, the performance of model can be evaluated on multiple distortion databases such as LIVE MD [32], TID2013 [33], CID 2013 [34] databases.

For future work, in order to further improve the performance, the distortion type classifier should be enhanced since system accuracy depends on it. It is recommended to explore other deep neural network architectures such as GoogleNet [35], generative adversarial networks (GANs) [36] for this problem. Visualizing different layers other than convolutional layers such as activation layers for better insights can be done. In addition this algorithm can be expanded to video quality assessment.

## REFERENCES


[1] K. Morrison, “How Many Photos Are Uploaded to Snapchat Every Second?" https://www.adweek.com/digital/how-many-photos-are-uploaded-to-snapchat-every-second/, June 2015.

[2] D. M. Chandler “Seven Challenges in Image Quality Assessment: Past, Present, and Future Research,” ISRN Signal Processing. 2013. 10.1155/2013/905685..

[3] C. T. Vu and D. M. Chandler, "S3: A Spectral and Spatial Sharpness Measure," 2009 First International Conference on Advances in Multimedia, Colmar, 2009, pp. 37-43. doi: 10.1109/MMEDIA.2009.15

[4] R. Hassen, Z. Wang and M. M. A. Salama, "Image Sharpness Assessment Based on Local Phase Coherence," in IEEE Transactions on Image Processing, vol. 22, no. 7, pp. 2798-2810, July 2013. doi: 10.1109/TIP.2013.2251643.

[5] Ferzli R, Karam L, “A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)”. IEEE Trans Image Process 18(4):717–728.

[6] A. K. Moorthy and A. C. Bovik, `` Blind Image Quality Assessment: From Scene Statistics to Perceptual Quality ", IEEE Transactions Image Processing, pp. 3350-3364, vol. 20, no. 12, 2011.

[7] M. A. Saad, A. C. Bovik and C. Charrier, `` Model-Based Blind Image Quality Assessment: A natural scene statistics approach in the DCT domain' ", IEEE Transactions Image Processing, pp. 3339-3352, vol. 21, no. 8, 2012.

[8] A. Mittal, A. K. Moorthy and A. C. Bovik, “ Referenceless Image Spatial Quality Evaluation Engine ". 45th Asilomar Conference on Signals, Systems and Computers. November 2011.

[9] Z. Wang, “Objective Image Quality Assessment: Facing The Real-World Challenges”. Electronic Imaging, Image Quality and System Performance XIII, pp. 1-6(6)

[10] P. Mohammadi, A. Ebrahimi-Moghadam, and S. Shirani, “Subjective and objective quality assessment of image: A survey”. arXiv preprint arXiv:1406.7799, 2014.

[11] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, Apr. 2004.

[12] L. Zhang, L. Zhang, X. Mou and D. Zhang, "FSIM: A Feature Similarity Index for Image Quality Assessment," in IEEE Transactions on Image Processing, vol. 20, no. 8, pp. 2378-2386, Aug. 2011.

[13] C. Charrier, O. Lézoray, and G. Lebrun, "Machine learning to design full reference image quality assessment algorithm," Signal Processing: Image Communication, vol. 27, no. 3, pp. 209–219, 2012.

[14] Z. Wang and E. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," Electronic Imaging 2005, 2005.

[15] N. D. Narvekar and L. J. Karam, "A No-Reference Image Blur Metric Based on the Cumulative Probability of Blur Detection (CPBD)," in IEEE Transactions on Image Processing, vol. 20, no. 9, pp. 2678-2683, Sept. 2011.

[16] P. Bingzhi, L. Xiaolin and F. Zicheng. "A No-Reference Image Sharpness Metric Based on Large-Scale Structure". Journal of Physics: Conference Series. 960. 012018. 10.1088/1742-6596/960/1/012018.

[17] P. Ye, J. Kumar, Le Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment. In Computer Vision and Pattern Recognition (CVPR)", 2012 IEEE Conference on, pages 1098–1105. IEEE, 2012.

[18] L. Kang, P. Ye, Y. Li and D. Doermann, "Convolutional Neural Networks for No-Reference Image Quality Assessment," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 1733-1740.

[19] S. Bosse, D. Maniry, T. Wiegand and W. Samek, "A deep neural network for image quality assessment," 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, 2016, pp. 3773-3777. doi: 10.1109/ICIP.2016.7533065.

[20] W. Hou, X. Gao, D. Tao and X. Li, "Blind Image Quality Assessment via Deep Learning," in IEEE Transactions on Neural Networks and Learning Systems, vol. 26, no. 6, pp. 1275-1286, June 2015. doi: 10.1109/TNNLS.2014.2336852.

[21] C. Fan, Y. Zhang, L. Feng and Q. Jiang, "No Reference Image Quality Assessment based on Multi-Expert Convolutional Neural Networks," in IEEE Access, vol. 6, pp. 8934-8943, 2018. doi: 10.1109/ACCESS.2018.2802498.

[22] LeCun, Yann. "LeNet-5, convolutional neural networks". Retrieved 16 November 2013.

[23] V. Nair and G. Hinton. "Rectified Linear Units Improve Restricted Boltzmann Machines", ICML'10 Proceedings of the 27th International Conference on International Conference on Machine Learning.

[24] D. P. Kingma, J. Ba, "Adam: A Method for Stochastic Optimization", 3rd International Conference for Learning Representations, San Diego, 2015.

[25] H. R. Sheikh, Z. Wang, L. Cormack and A. C. Bovik, "LIVE Image Quality Assessment Database Release 2", http://live.ece.utexas.edu/research/quality

[26] E. C. Larson and D. M. Chandler, "Most Apparent Distortion: Full-Reference Image Quality Assessment and the Role of Strategy," Journal of Electronic Imaging, 19 (1), March 2010.

[27] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and Jay Kuo, Image database TID2013: Peculiarities, results and perspectives, Signal Processing: Image Communication, vol. 30, Jan. 2015, pp. 57-77.

[28] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. "Gradient-Based Learning Applied to Document Recognition". Proceedings of the IEEE, November 1998.

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," CoRR, vol. abs/1409.1556, 2014

[30] He, Kaiming, X. Zhang, S. Ren and J. Sun. “Deep Residual Learning for Image Recognition.” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016): 770-778.

[31] Ren, Hongyu, D. Chen and Y. Wang. “RAN4IQA: Restorative Adversarial Nets for No-Reference Image Quality Assessment.” AAAI (2018).

[32] D. Jayaraman, A. Mittal, AK. Moorthy and A.C. Bovik, “Objective Quality Assessment of Multiply Distorted Images”, Proceedings of Asilomar Conference on Signals, Systems and Computers, 2012.

[33] E. C. Larson and D. M. Chandler, "Most Apparent Distortion: Full-Reference Image Quality Assessment and the Role of Strategy," Journal of Electronic Imaging, 19 (1), March 2010.

[34] Virtanen T, Nuutinen M, Vaahteranoksa M, Oittinen P and Hakkinen J. “CID2013: A database for evaluating no-reference image quality assessment algorithms”. IEEE Transactions on Image Processing. 2015 Jan; 24(1): 390±402. https://doi.org/10.1109/TIP.2014.2378061 PMID: 25494511

[35] C. Szegedy and W. Liu, “Going deeper with convolutions”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[36] Goodfellow, Ian and Pouget-Abadie, “Generative Adversarial Nets”, Advances in Neural Information Processing Systems 27, 2672—2680, 2014.

[37] Amritpal Singh Gill, "No-Reference Image Quality Assessment using Deep Convolutional Neural Networks", M.Sc Thesis Delft University of Technology, 2016.

[38] Omar Alaql, "General Purpose Approaches for No-Reference Image Quality Assessment ", Ph.D. Thesis Kent State University, 2017.

# APPENDIX

All the experiments in this project are performed in python programming language the following software libraries were used:

1. **openCV2:** is a python library that works mainly with the images. In this project it is used mainly in executing the image preprocessing applications such as Local Contrast Normalization, GCN and Gray scale conversion.
2. **TensorFlow:** is a python deep learning library provides great support for coding deep learning algorithms. It is efficient evaluating and optimizing multi-dimensional arrays. Keras wrapper is used over the TensorFlow in this project.
3. **Matplotlib:** is a python visualization library. In this project we used it to plot scatterplots and visualize the convolutional kernels of the networks.
4. **cuDNN:** is the NVIDIA library provides an optimized version of some mathematical operations like convolution.

Hardware specifications for the computer used in this project is divided into two parts we used an Intel i7 2.40 GHz CPU and NVIDIA GTX1070 + NVIDIA TITAN GPUs for the initial stages in this project and Intel i7-3.20 GHz CPU and Nvidia GTX 1080 GPU. With GPU we were able to speedup the training process by many folds.

Python codes developed in this project were made available to the public for evaluation in the following weblink https://github.com/alien2rv/NR-IQA