Master in Computational Colour and Spectral Imaging (COSI)

Full Reference Image Quality Assessment using Siamese Neural Network

Master Thesis Report

Presented by

Akib Jayed Islam

and defended at the

Norwegian University of Science and Technology

September 2023

Academic Supervisor(s): Professor Marius Pedersen, NTNU, Gjøvik

Jury Committee:

1. Dr. Eva Valero, University of Granada, Spain
2. Dr. Joni Hyttinen, University of Eastern Finland, Finland

Submission of the thesis: 10th August 2023
Day of the oral defense: 5th September 2023

Abstract

This study aims to explore the application of Siamese Neural Networks (SNN) in Full-Reference Image Quality Assessment (FR-IQA). The goal is to develop an efficient and accurate model to predict perceived image quality and correlate it with human scores. While deep learning has gained popularity in FR-IQA, the potential of Siamese Neural Networks still needs to be explored. The research addresses the challenges associated with capturing complex and nonlinear distortions and accurately assessing perceptual image quality in line with human judgment.

The proposed method represents a significant advancement in image quality assessment (IQA) by harnessing the capabilities of SNNs. It introduces an approach to accurately measuring and evaluating image quality with the publicly available KADID-10k dataset. The FR-IQA model proposed in this research has undergone thorough evaluation and comparison with existing techniques. This evaluation provides valuable insights into the strengths and weaknesses of different approaches, ultimately highlighting the effectiveness of the SNN-based model. We achieved a Pearson Linear Correlation Coefficient (PLCC) of 0.781, surpassing the Structural Similarity Index Measure (SSIM) with a PLCC of 0.671. Furthermore, the Spearman Rank Order Correlation Coefficient (SROCC) and Kendall Rank Order Correlation Coefficient (KROCC) of our proposed model exhibit higher values of 0.823 and 0.804, respectively, compared to other image quality metrics. The study showcases the versatility and capacity to enhance the SNN-based model through fine-tuning. The drawbacks of the suggested model have been addressed, and potential strategies to overcome these limitations and achieve superior performance have been highlighted.

Acknowledgment

I would like to extend my utmost gratitude to my supervisor, Professor Marius Pedersen, for his invaluable guidance and support. His guidance, kindness, and understanding have played a crucial role in helping me successfully complete my thesis. I am truly grateful and deeply honoured to have had the privilege of working under his thoughtful mentorship. With his support, I have embarked on exciting endeavours and gained valuable knowledge.

I would also like to express my sincere gratitude to my family for their unwavering support throughout my master's degree and the process of writing my thesis. Their unwavering belief in me has kept me motivated and positive throughout this journey. I would not have achieved this significant milestone without their assistance and love.

I would like to extend my heartfelt appreciation to my friends Aleef, Jishan, Nasim, Sadman for their unwavering support and invaluable assistance in sharing their knowledge with me.

Furthermore, I want to express my profound gratitude to my exceptional classmates in the COSI program. Their unwavering support and constant presence have been a source of strength for me. I want to give Lakshay, Tawsin, Borhan, and Ehsan special recognition for their support during challenging times.

Lastly, I would like to express my gratitude to all of the COSI coordinators for providing me with opportunities to grow and succeed.

Acronyms

HVS - Human Visual System

CNN - Convolutional Neural Network

IQA - Image Quality Assessment

NR-IQA - No-Reference Image Quality Assessment

FR-IQA - Full Reference Image Quality Assessment

RR-IQA - Reduced-Reference Image Quality Assessment

SNN - Siamese neural network

GSN - Gradient Siamese Network

LRSN - Learning-to-Rank Siamese Network

NIMA - Neural Image Assessment

RADN - Region Adaptive Deformable Network

PIPAL - Perceptual Image Processing ALgorithms IQA Dataset

PLCC - Pearson Linear Correlation Coefficient

RMSE - Root Mean Squared Error

KROCC - Kendall Rank Order Correlation Coeffcient

SROCC - Spearman Rank-Ordered Correlation Coeffcient

MSE - Mean Squared Error

MAE - Mean Absolute Error

VIF - Visual Information Fidelity

PSNR - Peak Signal-to-Noise Ratio

SSIM - Structural Similarity Index

FSIM - Feature Similarity Index

CDC - Central Difference Convolution

DMOS - Difference Mean Opinion Scores

MS-SSIM - Mean Structural Similarity

MDSI -Mean Deviation Similarity Index

Contents

1 Introduction

1.1 Background and Motivation

Image Quality Assessment (IQA) is an integral part of numerous image processing applications including compression, restoration, watermarking, and visualization. Traditionally, subjective IQA metrics, such as Mean Opinion Scores, have been used to judge the perceptual quality of images. However, this approach has inherent limitations, such as inconsistency and subjectivity. To overcome these issues, objective IQA techniques have been developed that attempt to emulate the Human Visual System (HVS) using computational models. Although substantial progress has been made, a considerable gap still exists between the performance of current IQA models and human judgment Talebi and Milanfar (2018); Niu et al. (2019); Cong et al. (2022)..

Full Reference Image Quality Assessment (FR-IQA) Pedersen et al. (2012); Varga (2022) provides a way to evaluate the quality of distorted images by comparing them with their original high-quality versions. However, the current FR-IQA models face challenges in capturing high-level semantic features and handling complex, nonlinear distortions. Recently, deep learning methods, such as Convolutional Neural Networks (CNNs), have shown potential in improving IQA performance Cheon et al. (2021); Cong et al. (2022); Varga (2020); Liu et al. (2019). Among them, Siamese Neural Networks (SNN), known for their strength in comparing and contrasting features between input pairs, present an attractive opportunity for FR-IQA.

1.2 Problem Statement

This research will investigate the application of Siamese Neural Networks in FR-IQA, with the goal of developing an efficient and accurate model. Despite the increasing use of deep learning in FR-IQA, the potential of Siamese Neural Networks has not been fully explored. In particular, the research will address the difficulty in capturing complex and nonlinear distortions and the challenges of accurately assessing the perceptual image quality in line with human judgment Ahmed et al. (2022); Ayyoubzadeh and Royat (2021); Cong et al. (2022).

1.3 Research Objectives

The primary objectives of this research are:

- To explore the application and effectiveness of Siamese Neural Networks for FR-IQA.
- To develop an SNN-based FR-IQA model that accurately reflects human judgment of image quality.
- To assess the proposed model's performance against existing FR-IQA techniques using a standard IQA database.
- To explore how the SNN's unique properties can be leveraged to improve FR-IQA performance.

1.4 Contribution of the Thesis

This thesis aims to make several significant contributions:

- The proposal and development of a new SNN-based FR-IQA model.
- An extensive evaluation and comparison of the proposed model with existing FR-IQA techniques.
- A comprehensive study on the application of Siamese Neural Networks in the field of image quality assessment.
- Insights on how Siamese Neural Networks can be fine-tuned to enhance FR-IQA performance.

1.5 Thesis Organization

The subsequent sections of the thesis are structured as follows:

Chapter II offers a comprehensive review of image quality assessment, specifically emphasizing Full Reference (FR) methods and the utilization of deep learning techniques.

Chapter III elucidates the methodology behind SNNs and outlines the development process of the proposed FR-IQA model.

Chapter IV presents the experimental setup, encompassing details regarding the datasets used, evaluation metrics employed, and comparison methods employed.

Chapter V engages in a thorough discussion of the experimental results, providing an in-depth analysis of the performance exhibited by the proposed model.

Chapter VI serves as the concluding section of the thesis, summarizing the key findings and contributions while also suggesting potential avenues for future research.

2 Background and Literature Review

The purpose of this chapter is to offer a broad understanding of the theoretical foundation and perform an extensive review of existing literature. It will be divided into two main sections.

Background Theory on Image Quality and Siamese Network: In this section, the chapter will delve into the theoretical foundations related to image quality and the concept of a Siamese network. It will explore the principles, techniques, and methodologies of assessing and enhancing image quality. Additionally, it will explain the concept of a Siamese network, a type of neural network architecture commonly used in various applications, including image comparison and similarity measurement.

State-of-the-Art Literature Review: The second part of the chapter will focus on reviewing the current state-of-the-art literature in the field. It will provide an in-depth analysis and evaluation of the most recent and relevant research papers, articles, and studies related to the topic of interest. This literature review aims to identify the existing gaps, trends, and advancements in the field, and it will serve as a foundation for the subsequent chapters of the research or study.

2.1 Background Theory on Image Quality and Siamese Network

2.1.1 Image Quality Assessment (IQA)

Image Quality Assessment (IQA) is a field of study that focuses on developing algorithms and methodologies to assess subjective and objective IQA. IQA aims to model human perception and provide quality scores that correlate with how humans perceive the visual quality of images Talebi and Milanfar (2018); Niu et al. (2019); Cong et al. (2022). IQA is of utmost importance in various domains and applications. In image processing and restoration, IQA helps evaluate the effectiveness of techniques by comparing the quality of the output with the original image. IQA ensures that transmitted or compressed images meet acceptable quality standards in multimedia communications. In computer vision and machine learning, IQA is a pre-processing step to ensure input images are of good quality, free from distortions that could affect algorithm performance. In medical imaging, IQA is crucial for accurate and reliable diagnostic results. In human-computer interaction, IQA plays a role in applications relying on visual information, such as virtual reality and gaming Niu et al. (2019); Cong et al. (2022).

Several factors influence image quality. Compression artifacts introduced by techniques like JPEG degrade perceived quality through blocking, blurring, and ringing. Image noise from sensor limitations, transmission errors, or environmental factors reduces clarity and increases distortions. Blur caused by camera shake, defocus, or motion during capture results in reduced sharpness and loss of details. Inaccurate color reproduction or shifts can significantly affect visual quality. Low-resolution images lack fine details, leading to a perceived decrease in quality. Improper contrast and brightness adjustments can make images appear too dark or bright. In lossy compression, higher compression ratios lead to more significant quality loss. Additionally, the subjective content of an image, including the presence of meaningful objects or elements, can influence perceived quality. Understanding these factors and their impact on image quality is essential for developing effective image processing techniques, ensuring accurate diagnostic results, and enhancing user satisfaction in various applicationsTalebi and Milanfar (2018); Shi et al. (2021).

IQA can be divided into subjective IQA and objective IQA. Talebi and Milanfar (2018); Niu et al. (2019).

Subjective Image Quality Assessment (IQA) involves human observers providing their subjective judgments on the perceived quality of images. Human participants are presented with pairs of reference and distorted images and asked to rate the quality of the distorted images based on their visual perception and subjective experience. These subjective quality scores serve as the ground truth for evaluating objective IQA algorithms and other image-processing techniques. Subjective IQA is based on the principles of human perception, considering cognitive and psychological factors that influence how humans perceive and evaluate visual information. It captures the real-world context in which images are perceived by actual users, accounting for diverse preferences and expectations regarding image quality. The evaluation process involves collecting subjective quality scores from multiple human observers to create a database of human ratings, which serves as a reference for developing and validating objective IQA algorithmsTalebi and Milanfar (2018); Niu et al. (2019).

Subjective IQA has applications in validating and benchmarking objective IQA algorithms, providing a benchmark for comparing different image quality assessment methods. It is also crucial in developing image and video coding standards, helping to set quality targets and determine acceptable levels of quality degradation for various applications. By incorporating human perception and preferences, subjective IQA enhances the accuracy and relevance of image quality assessment Mohammadi et al. (2014); Pedersen et al. (2012).

Objective Image Quality Assessment (Objective IQA) is a computational approach that quantitatively assesses the visual quality of images without human observers. Objective IQA algorithms analyze visual features of reference and distorted images to generate objective quality scores. This approach is used in computer vision, image processing, multimedia communications, and quality control.

Objective IQA principles include feature extraction, where relevant features capturing perceptual differences are extracted from images. Psychophysics modeling simulates human perception efficiently. Performance metrics like PLCC, SROCC, and KROCC Sheikh et al. (2006) quantify the correlation between predicted quality scores and subjective ratings. Objective IQA methods include Full Reference (FR-IQA), comparing distorted images with high-quality references; No-Reference (NR-IQA), assessing quality using only the distorted image; and Reduced-Reference (RR-IQA), using a reduced set of features from both images Talebi and Milanfar (2018); Niu et al. (2019); Mohammadi et al. (2014).

Applications of Objective IQA include automation, quality control, and guiding image enhancement and processing techniques. Challenges in Objective IQA include diverse distortions, subjectivity variability, and dataset bias. Objective IQA is a valuable tool for automated quality assessment, benefiting industries and optimizing image processing techniques. Researchers continue to address challenges, improving the accuracy and reliability of automatic image quality assessmentMohammadi et al. (2014).

2.1.2 Full Reference Image Quality Assessment (FR-IQA)

Full Reference Image Quality Assessment (FR-IQA) Larson and Chandler (2010); Pedersen et al. (2012); Zhang et al. (2012); Sun et al. (2018); Madhusudana et al. (2022) is a method used to assess the quality of distorted images by comparing them with a known high-quality reference image. Here are some key points about FR-IQA:

- Comparison with a Reference Image: FR-IQA algorithms assess the quality degradation brought about by distortions by contrasting the distorted image with a reference image. The reference image is typically an original, undistorted version of the same image.

- Image Similarity Metrics: FR-IQA methods often use image similarity metrics to compare the features extracted from the reference and distorted images. These metrics evaluate the differences in pixel values, local texture patterns, color histograms, and edge information between the two images.

- Perceptual Differences: FR-IQA algorithms aim to capture the perceptual differences between the reference and distorted images. They consider factors such as loss of details, color shifts, artifacts, and overall visual degradation caused by the distortion Pedersen et al. (2012); Larson and Chandler (2010); Sun et al. (2018).

- Psychophysics-Based Models: Many FR-IQA algorithms are designed based on psychophysical studies, where human perception of image quality is quantified and used as a reference for developing computational models Pedersen et al. (2012).

- Evaluation Metrics for Full Reference IQA: The performance of FR-IQA Larson and Chandler (2010); Sun et al. (2018); Madhusudana et al. (2022) algorithms is evaluated by comparing the predicted quality scores with subjective ratings provided by human observers. Several evaluation metrics are used to assess the performance of Full Reference IQA algorithms. These metrics compare the quality scores generated by the IQA model against human judgments or subjective quality scores provided by human observers. Some commonly used metrics include: Pearson Correlation Coefficient (PCC), Kendall Rank-Order Correlation Coefficient (KRCC), Mean Squared Error (MSE), Spearman Rank-Order Correlation Coefficient (SRCC), Root Mean Squared Error (RMSE) Sheikh et al. (2006).

- Applications: FR-IQA is used in various applications, including image and video compression, image restoration, and quality control. It helps optimize algorithms and techniques by providing quantitative measures of the quality degradation caused by different distortions Zhang et al. (2012); Madhusudana et al. (2022).

- Limitations: FR-IQA requires access to a high-quality reference image, which may not always be available. It also assumes that the reference image perfectly represents the original, which may not be accurate in some scenarios Sun et al. (2018); Zhang et al. (2012).

Overall, FR-IQA plays a crucial role in objectively assessing the quality of distorted images by comparing them with a reference image, providing valuable insights for image processing and quality evaluation tasks Zhang et al. (2012); Larson and Chandler (2010); Sun et al. (2018); Madhusudana et al. (2022).

FR IQA methods rely on feature extraction techniques. Despite advancements in image processing for IQA, many approaches still employ low-complexity metrics such as mean square error (MSE) and peak signal-to-noise ratio (PSNR). However, these metrics do not fully account for the characteristics of the human visual system (HVS) Wang et al. (2004), are sensitive to pixel locations, and often do not align well with subjective judgments of image quality.

To address these limitations, Wang et al. (2004) introduced the structural similarity index (SSIM), which assumes that the HVS is highly adaptable to structural information. Building upon SSIM, subsequent methodologies such as multi-scale SSIM (MS-SSIM) and information-weighted SSIM (IW-SSIM) have been proposed to incorporate image details across multiple resolutions and viewing conditions for IQA. Larson argues that the HVS employs various methodologies to analyze image quality, including the examination of local statistics related to local brightness, contrast masking, and spatial frequency components, in order to identify distortions Larson and Chandler (2010); Wang and Li (2010).

2.1.3 FR-IQA with Deep Learning

The performance and speed of the FR-IQA algorithm are improved by analyzing multiple perspectives of the Human Visual System (HVS). Deep learning-based approaches for Image Quality Assessment (IQA) have shown significant advancements in various image-processing applications. Gao proposed Gao et al. (2017) introduced a method called DeepSim, which estimates the local similarities of features between distorted and reference images. This approach predicts the overall quality by aggregating these small similarities. In their study titled "DeepSim," Gao et al. Gao et al. (2017) demonstrate the extraction of feature maps from different layers of a pre-trained deep Convolutional Neural Network (CNN) used for image recognition. The research reveals that these features, originally learned for image recognition, also play a significant role in assessing perceived quality.

Bosse et al. (2017) presented a neural network architecture in their study that achieves simultaneous learning of local quality and local weights. This approach integrates feature learning and regression within a single end-to-end framework. The effectiveness of deep-learning-based IQA approaches has been validated through the NTIRE 2021 Perceptual Image Quality Assessment Challenge Gu et al. (2022). The LIPT team, as reported by Cheon et al. (2021), emerged as the winners of the NTIRE 2021 IQA challenge. Their approach utilized a transformer architecture to address the perceptual IQA task. The MT-GTD team secured the second place in the NTIRE 2021 Jinjin Gu (2020) IQA challenges. MT-GTD proposed a novel bilateral-branch multi-scale Image Quality Estimation (IQMA) network Guo et al. (2021) to effectively extract multi-scale features from patches of both the reference image and the distorted image individually.

Utilizing a DOG-based channel decomposition, the Gaussian (DOG)-SSIM Pei and Chen (2015) technique emulates the frequency bands of the contrast sensitivity function. Through channel-wise quality calculations, SSIM generates quality values that are subsequently aggregated by a trained regression model to provide an overall quality estimate. The MAD Larson and Chandler (2010) metric effectively distinguishes between supra- and near-threshold aberrations, thereby accommodating the distinct domains of human quality perception. It has been observed that the combination of multiple manually crafted Image Quality Metrics (IQMs) Lukin et al. (2015) can enhance performance Lukin et al. (2015). Feature learning Gao et al. (2017) and regression Pei and Chen (2015) are employed in FR IQA.

2.1.4 Loss Functions

Triplet Loss Funciton:

Triplet loss is a loss function commonly used in image quality assessment tasks, particularly in the context of siamese networks. The goal of triplet loss is to learn a feature representation that can effectively distinguish between images of different qualities. In the triplet loss formulation, each training sample consists of an anchor image, a positive image, and a negative image. The anchor image represents the reference image, while the positive image is a similar image to the anchor (e.g., a high-quality version of the same image), and the negative image is a dissimilar image (e.g., a low-quality version of the same image or a different image altogether) Niu et al. (2019).

$$\mathcal{L} = \max(d(a,p) - d(a,n) + \mathrm{mar}\,gin, 0)$$

Contrastive loss: Contrastive loss is a loss function commonly used in Image Quality Assessment (IQA) tasks. Its purpose is to learn a similarity metric that can distinguish between high-quality and low-quality images. The goal is to minimize the distance between similar image pairs and maximize the distance between dissimilar image pairs. In IQA, contrastive loss is typically applied in a Siamese network architecture, where two identical networks share weights and process two input images simultaneously. The output of each network is a feature vector that represents the image. The contrastive loss function takes into account the similarity between pairs of images and encourages the network to learn embeddings that preserve this similarity.

$$\text{loss}(d, Y) = \frac{1}{2} * Y * d^2 + (1 - Y) * \frac{1}{2} * \max(0, m - d)^2$$

Where: - d is the distance between the outputs of the encoder; - Y is the label of the model inputs (1 if similar, 0 if dissimilar); - m, the margin parameter

Soft-max cross-entropy: Softmax cross-entropy Levina and Bickel (2001) is a commonly used loss function in the field of Image Quality Assessment (IQA). It is used to measure the discrepancy between predicted quality scores and ground truth quality scores. Here are some key points about softmax cross-entropy for IQA: Loss Function: Softmax cross-entropy combines the softmax activation function with the cross-entropy loss. It is suitable for multi-class classification problems, where the goal is to assign quality scores to different image classes.

Softmax Activation: Softmax is used to convert the output of a neural network into a probability distribution over the different image classes. It ensures that the predicted scores sum up to 1 and are within the range of 0 to 1.

Cross-Entropy Loss: Cross-entropy measures the dissimilarity between the predicted probability distribution and the true probability distribution. In IQA, the true probability distribution represents the ground truth quality scores of the images Levina and Bickel (2001).

2.1.5 Siamese Networks

An SNN is a type of neural network architecture that contains two or more identical sub-networks. The sub-networks have the same configuration, weights, and parameters, and the updating of parameters is mirrored across both sub-networks. SNNs are beneficial when limited data is available for training, making them suitable for tasks like image processing, facial recognition, and signature verification Cong et al. (2022); Koch et al. (2015).

SNNs can determine which image is higher quality in a pair of images. They are skilled at capturing gradient features and can make near-precise predictions with just a few images. By comparing the features of the two images, the Siamese network can learn a similarity function and predict multiple classes based on this learned function. Siamese neural networks offer several benefits, including the ability to learn from minimal data, making them popular in data science. They also allow for classifying new data classes without retraining the network. However, it's important to note that Siamese neural networks also have some drawbacks, including increased computational complexity and the need for careful design and training Koch et al. (2015).

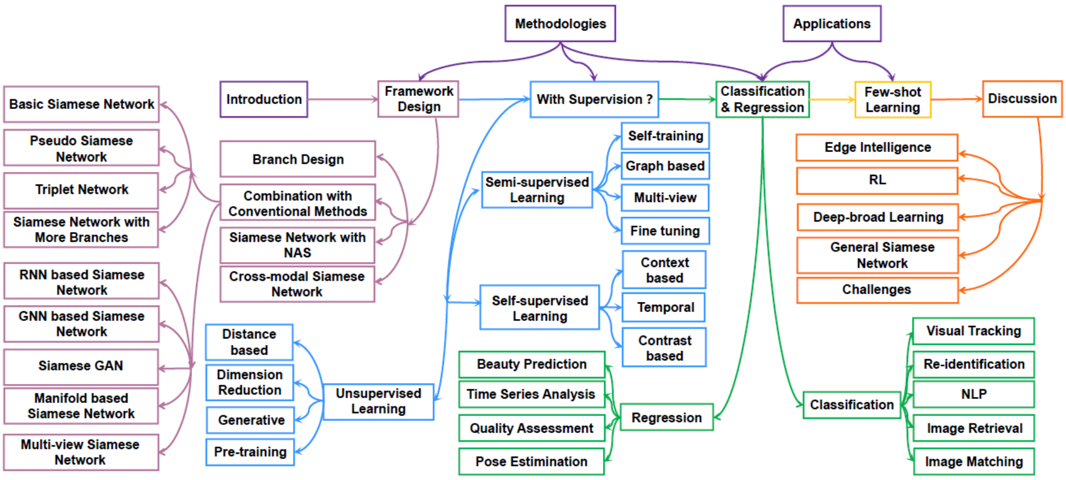

Figure 2.1: *Overview of Siamese network [Image credit: Li et al. (2022)]*

Overall, SNNs provide a powerful architecture for image quality assessment and other tasks with limited data. They offer the potential for accurate predictions and can be valuable in machine learning and computer vision. The overview of the applications of the Siamese network is shown in Fig 2.1

Siamese Networks are a class of artificial neural networks designed to learn the similarity or dissimilarity between two inputs. The name "Siamese" comes from the famous "Siamese Twins" who were conjoined twins. Siamese Networks (shown in Fig 2.2 consist of two or more identical subnetworks (commonly referred to as twins or branches) with shared weights. These twins process the two input samples independently but with the same architecture and parameters Li et al. (2022); Niu et al. (2019); Varga (2020); Ayyoubzadeh and Royat (2021).

The primary objective of Siamese Networks is to learn a similarity metric or distance measure between the inputs. They are commonly used in tasks like image verification, face recognition, and signature verification. Siamese Networks excel at handling tasks where there is a need to compare and measure the similarity between pairs of data points Li et al. (2022); Niu et al. (2019); Varga (2020); Ayyoubzadeh and Royat (2021).

Siamese Networks are a type of neural network architecture used for tasks such as similarity comparison, face recognition, and signature verification. They consist of twin networks, each processing one input sample and producing an embedding or feature vector. The key characteristic is that the weights of the twins are shared, enabling them to learn similar representations for similar inputs. The similarity between the embeddings is calculated using a distance metric, such as Euclidean distance or cosine similarity. This distance is then fed into an output layer to make

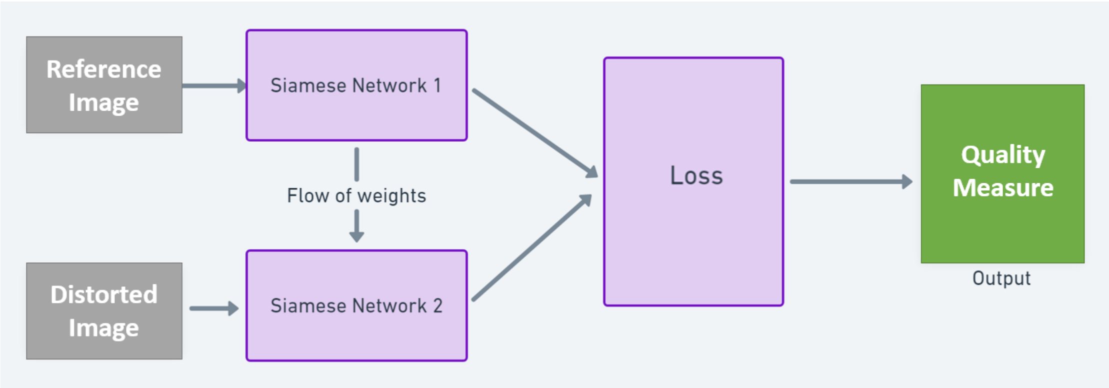

Figure 2.2: *General Model of Siamese network, [Image credit (inspired from): Dey et al. (2017)]*

a final decision based on a threshold Niu et al. (2019); Varga (2020); Ayyoubzadeh and Royat (2021).

During the training process, input samples are paired and labeled based on their similarity or dissimilarity. The primary goal is to reduce the distance between similar pairs and increase the distance between dissimilar pairs. This is accomplished by utilizing loss functions such as Contrastive Loss, which penalizes the network when the distance between similar pairs exceeds a specified margin, and Triplet Loss, which aims to minimize the distance between an anchor and a positive sample while simultaneously maximizing the distance between the anchor and a negative sample Li et al. (2022); Niu et al. (2019); Ayyoubzadeh and Royat (2021).

To optimize Siamese Networks, common techniques like Stochastic Gradient Descent (SGD) and its variants (Adam, RMSprop) are used, with the choice depending on the specific task and dataset.

Siamese Networks have shown promising results in various applications that require similarity learning and distance-based comparisons. By sharing weights between twins and using specialized loss functions, Siamese Networks can effectively learn powerful embeddings that facilitate similarity-based decision-making. The flexibility and versatility of Siamese Networks make them a valuable tool in various machine learning tasks Li et al. (2022); Niu et al. (2019); Varga (2020); Ayyoubzadeh and Royat (2021).

2.2 State-of-the-Art Literature Review

- **Advantages of Siamese Networks in IQA:**

Fine-Grained Image Comparison: Siamese Networks excel at capturing fine-grained differences between images. In IQA, this ability is crucial, as small variations in image quality need to be accurately detected and quantified Li et al. (2022); Ayyoubzadeh and Royat (2021).

Distortion-Specific Quality Assessment: Siamese Networks can be designed to focus on specific types of distortions. By training on pairs of reference and distorted images, the network can learn to specialize in assessing the quality degradation caused by particular distortions, such as compression artifacts, blur, or noise.

Fewer Training Samples: Siamese Networks require fewer training samples compared to other deep learning approaches since they learn the similarity metric rather than directly predicting the quality score. This makes them particularly useful for tasks with limited annotated data Koch et al. (2015).

Handling No-Reference IQA: Siamese Networks can be adapted to No-Reference IQA scenarios, where only the distorted image is available for quality assessment. Training the network to compare the distorted image against itself as a reference can still provide quality estimates without access to an explicit reference image Niu et al. (2019); Varga (2020).

Several Siamese Network architectures have been proposed for different applications. Some of the notable ones include:

- **Siamese Networks with CNNs:** Using Convolutional Neural Networks (CNNs) as the subnetworks in Siamese Networks has been a popular choice. CNNs are effective in feature extraction, and their hierarchical nature allows them to learn meaningful representations for image quality assessment.

  The literature Liu et al. (2019) presents a Siamese Convolutional Neural Network (CNN) for remote sensing scene classification. The model combines identification and verification models of CNNs and enforces robustness through metric learning regularization. The paper describes the framework of Siamese CNNs and the role of convolutional layers in feature extraction. Experimental results demonstrate that the proposed method outperforms existing remote sensing scene classification methods. The authors conducted experiments on three widely used remote sensing datasets and compared their method with state-of-the-art approaches, showing the effectiveness of the proposed Siamese CNN model.

- **Siamese VGG:** The VGG architecture, known for its depth, has been employed in Siamese Networks for IQA. VGG-based Siamese Networks have shown competitive performance in various image quality benchmarks.

The paper Yuan et al. (2019) introduces an improved method called Siamese-VGG for optimizing face tracking tasks. This method addresses the limitations of traditional face tracking methods by enhancing the feature extraction template. It exhibits strong adaptability and processing capabilities, particularly in handling light changes, motion blur, and occlusion. The algorithm is designed to be faster and more suitable for practical applications. Experimental results demonstrate that the face tracking algorithm using Siamese-VGG achieves an average overlap of up to 80

- **Siamese ResNet:** Residual Networks (ResNets) have demonstrated superior performance in various computer vision tasks. Siamese ResNet architectures leverage the skip connections in ResNets to facilitate learning rich and discriminative features for image quality comparisons.

  The paper Qiu et al. (2018) introduces a loop closure detection method for visual SLAM systems using a Siamese-ResNet network. The network is trained through supervised learning to extract features from images and determine if they belong to the same scene. Experimental results on the FabMap dataset demonstrate that the Siamese-ResNet outperforms the traditional FabMap2.0 method in terms of loop closure detection accuracy and computational efficiency. The Siamese-ResNet network improves feature extraction by replacing the smaller LeNet with the larger ResNet, enabling better feature extraction in complex environments.

- **Siamese GANs:** Some studies have explored the use of Generative Adversarial Networks (GANs) in Siamese architectures. These networks learn to generate image representations that are indistinguishable from the real representations, helping in generating high-quality embeddings for IQA.

  The literature Yan et al. (2021) discusses a research paper on using neural networks to remove snow from images. The paper introduces a snow-free model and a pseudo-siamese generative adversarial network. The proposed method incorporates techniques like autoencoder structure, multi-scale perception structure, and skip-connect operation. The paper evaluates the model's effectiveness in recovering snow-free images and enhancing target detection algorithms.

It is worth noting that the field of Siamese Networks in Image Quality Assessment is an active area of research. New architectures and techniques are continuously being explored to enhance the accuracy and robustness of IQA models. By addressing the challenges and leveraging the advantages, Siamese Networks have the potential to significantly contribute to the advancement of image quality assessment and related applications.

2.2.1 Siamese Network Image Quality Metrics

- **Challenges of Siamese Networks in IQA:**

  Siamese Networks in Image Quality Assessment (IQA) face several challenges. Firstly, they require a large training dataset to generalize well, which can be time-consuming and resource-intensive. Gathering high-quality references and distorted image pairs, along with subjective quality scores, is necessary Niu et al. (2019); Varga (2020).

  Secondly, designing an appropriate distance metric is crucial. The distance metric should align well with human perception, but finding the right one may require careful experimentation and fine-tuning Li et al. (2022); Niu et al. (2019).

  Thirdly, imbalanced data distribution can affect model performance. The number of similar and dissimilar pairs in real-world datasets may differ significantly, leading to biased results. Special attention is needed during training to address this class imbalance Li et al. (2022); Ayyoubzadeh and Royat (2021).

  Lastly, ensuring generalization across different types of distortions is challenging. Siamese Networks should be robust enough to handle various distortions commonly encountered in real-world scenarios.

2.2.1.1 Applications of Siamese Networks

Several Siamese Network architectures have been proposed for IQA. Some notable ones include using Convolutional Neural Networks (CNNs) as subnetworks, employing the VGG architecture, leveraging Residual Networks (ResNets), and exploring the use of Generative Adversarial Networks (GANs) in Siamese architectures Li et al. (2022); Niu et al. (2019); Varga (2020); Ayyoubzadeh and Royat (2021)..

Siamese Networks with CNNs are popular due to CNNs' effectiveness in feature extraction. Siamese VGG, known for its depth, has shown competitive performance. Siamese ResNet architectures utilize skip connections in ResNets to learn rich and discriminative features. Siamese GANs generate high-quality embeddings for IQA Li et al. (2021).

It's important to note that Siamese Networks in IQA is an active area of research, with continuous exploration of new architectures and techniques to enhance accuracy and robustness. By addressing challenges and leveraging advantages, Siamese Networks can significantly contribute to image quality assessment and related applications Li et al. (2022); Niu et al. (2019); Varga (2020); Ayyoubzadeh and Royat (2021).

The distance metric, typically Euclidean distance or cosine similarity, is then calculated between the embeddings. This distance value represents the similarity between the reference and distorted images. A smaller distance indicates higher similarity, implying that the distorted image retains a high level of quality similar to the reference image. Conversely, a larger distance indicates lower similarity, implying that the distortion has significantly affected the perceived qualityLi et al. (2022); Niu et al. (2019); Varga (2020).

2.2.1.2 IQA method with Siamese by Varga (2020)

This study proposes a technique for evaluating image quality using a Siamese architecture of pre-trained convolutional neural networks (CNNs) (illustrated in Fig 2.3) . The method captures the detailed characteristics of images without any alterations, such as scaling or cropping. The proposed model is trained on the KADID-10k dataset, currently, the most extensive image-quality database, consisting of 10,125 digital images. Experimental results show that the proposed methodology outperforms other contemporary algorithms. The findings are further validated through cross-database examinations of additional publicly accessible datasets. The proposed approach effectively captures images' nuanced, quality-aware features by utilizing pre-trained CNNs at a higher resolution than previous studies. The system combines feature pooling and a neural network to process input photographs and evaluate their quality. Overall, this study presents a novel IQA technique demonstrating superior image quality evaluation performance Varga (2020).

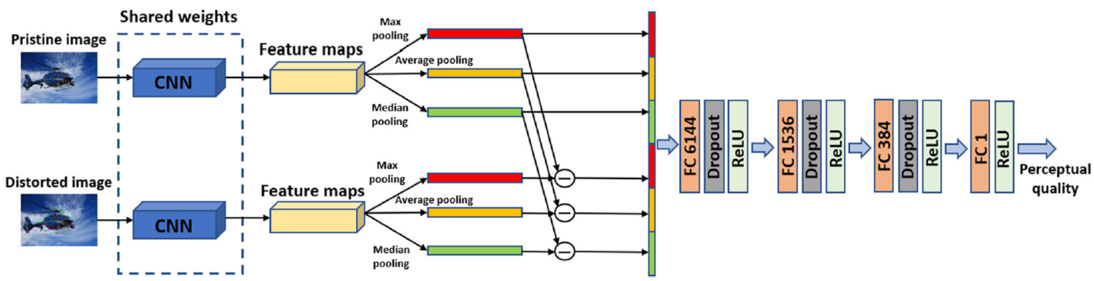

Figure 2.3: *IQA method with Siamese by Varga. [Image credit: Varga (2020)]*

2.2.1.3 Gradient Siamese Network(GSN)

The GSN architecture, as described by Cong et al. (2022) and shown in Fig. 2.4, consists of three main components: a weight-sharing Siamese Network, a multi-level feature fusion module, and a regression head. To handle the reference picture and distorted image, the authors utilized a Siamese network that incorporates CDC and spatial attention mechanisms. This network is composed of hierarchical blocks, each extracting features that are uniformly down-sampled to a predetermined size. Additionally, the authors proposed a multi-level feature fusion module, which involves combining features at the same level and then fusing features from different levels. Finally, to provide the expected quality ratings, a regression head is employed, which consists of a fully connected layer. In summary, the GSN architecture includes a multi-level feature fusion module, a weight-sharing Siamese Network, and a regression head, all working together to handle image distortion and provide quality ratings.

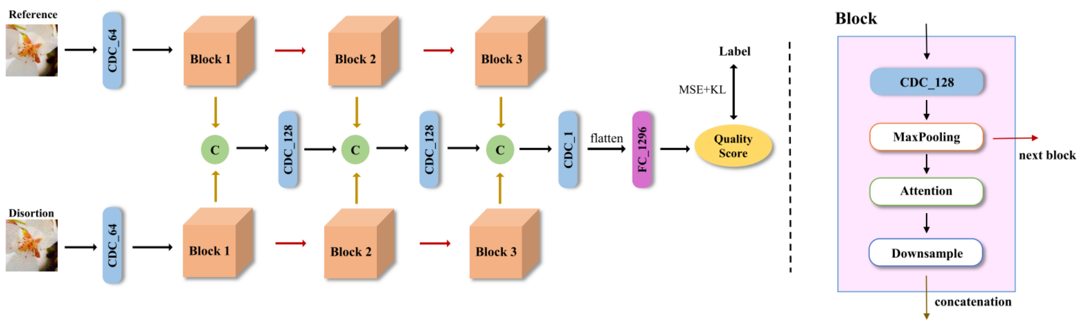

Figure 2.4: *GSN model architecture. [Image credit: Cong et al. (2022)]*

2.2.1.4 Attention-based Siamese-Difference Neural Network (ASNA)

This research Ayyoubzadeh and Royat (2021) presents a novel approach that introduces a convolutional neural network with a Siamese-Difference neural network architecture (refer to Figure 2.5). The study utilizes two variations of Attention-based Siamese-Difference Neural Network (ASNA) designs, each trained independently, to accurately estimate Mean Opinion Score (MOS). The Siamese-Difference network design effectively captures subtle variations between reference and distorted images.

The proposed architecture incorporates spatial and channel-wise attention mechanisms to enhance the overall performance of our method. Additionally, our model is trained using an additional loss function. This recommended supplementary cost

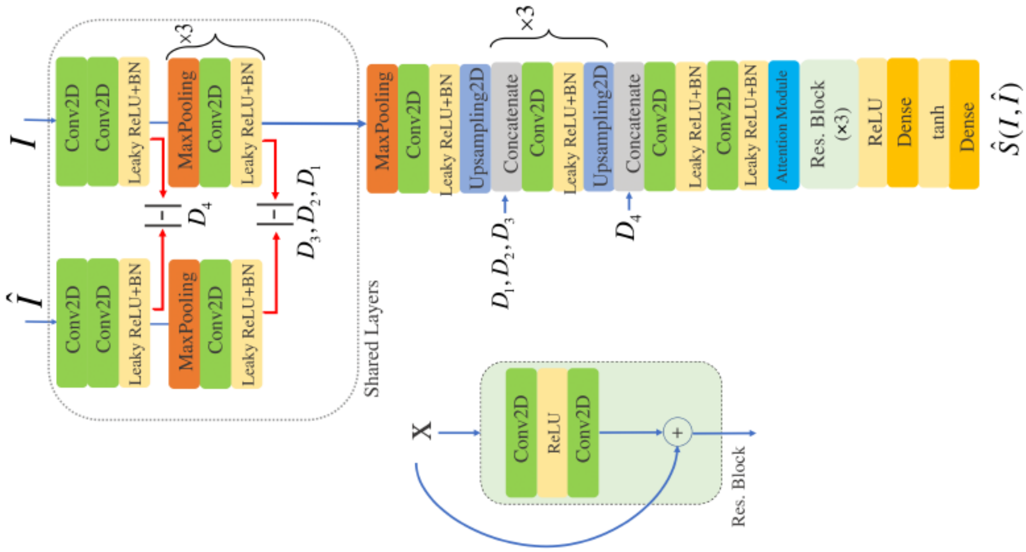

Figure 2.5: *Model architecture of ASNA. [Image credit: Ayyoubzadeh and Royat (2021)]*

function serves as a surrogate for ranking loss, effectively boosting Spearman's rank correlation coefficient while maintaining the differentiability of the neural network parameters. Notably, this technique emerged as the winning solution in the NTIRE 2021 Perceptual Image Quality Assessment Challenge Jinjin Gu (2020).

2.2.1.5 Learning-to-Rank Siamese Network (LRSN)

The LRSN Niu et al. (2019) is a type of Siamese network that is utilized to rank the quality scores of two image patches. LRSN employs Siamese CNN for the purpose of ranking IQA. In order to address the issue of limited samples, the researchers introduced no-reference 2D and 3D IQA indexes. These indexes propose a transformation of IQA into image quality comparison. By generating image pairings, the training data is effectively doubled, thereby mitigating the scarcity of samples. The LRSN model, which consists of two branches sharing weights and taking two picture patches as input, is designed to learn the task of quality comparison.

The ranking of image patch quality scores is performed by LRSN. To determine the relative quality score of a test image, its image patches are compared to several image patches from other images. The number of times the test image's patches

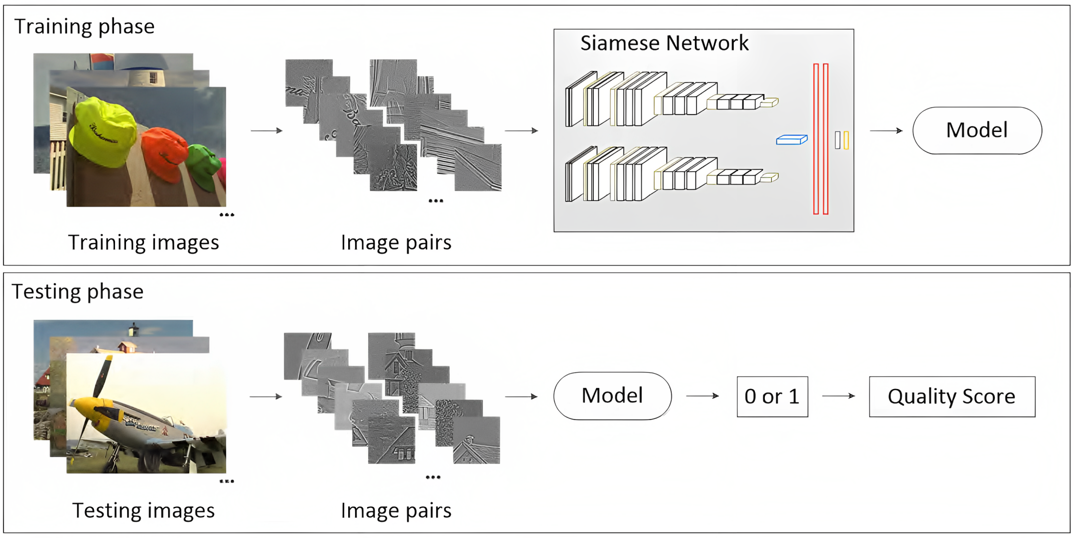

Figure 2.6: *Model architecture of LRSN. [Image credit: Niu et al. (2019)]*

are judged superior is counted, and this count is used to calculate the relative quality score. The proposed LRSN model has demonstrated superior performance compared to the state-of-the-art no-reference 2D and 3D IQA metrics on three 2D IQA databases (LIVE, CSIQ, and TID2013) and three 3D IQA databases (LIVE 3D Phase-I, Phase-II, and NBU) Niu et al. (2019); Li et al. (2021).

2.2.1.6 Neural Image Assessment (NIMA)

The study discusses a deep learning-based approach called NIMA Talebi and Milanfar (2018) that automatically predicts the aesthetics of images. NIMA utilizes a Convolutional Neural Network (CNN) that has been trained on a large dataset of photos with human-rated aesthetic scores. Unlike traditional methods that only provide mean scores, NIMA predicts quality rating distributions, which improves the accuracy of quality prediction and its correlation with ground truth. The researchers also used two models, one for aesthetic assessment and another for technical assessment, to control parameters of image-enhancing operators. Through tests, these models were found to guide denoising and tone enhancement processes, resulting in visually improved outcomes.

The authors Talebi and Milanfar (2018) suggest that the trained models can be utilized in various applications for further improvement of pictures. However, the experimental setup used in the study involved evaluating numerous enhancement operators, which makes it unsuitable for real-time applications. To address this limitation, the authors propose that NIMA could serve as a more efficient loss

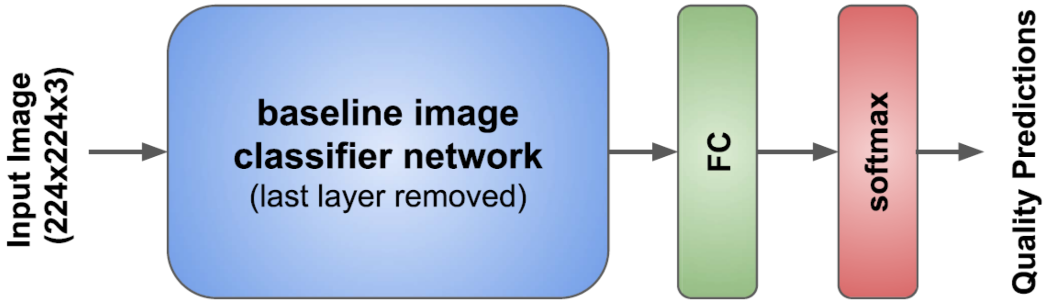

Figure 2.7: *Model architecture of NIMA. [Image credit: Talebi and Milanfar (2018)]*

function for an enhancement operator that has well-defined derivatives.

2.2.1.7 Region Adaptive Deformable Network (RADN)

A recent study aimed to enhance IQA by utilizing GAN-generated images Shi et al. (2021). GAN-generated images often pose challenges for existing IQA methods due to spatial changes and texture noise. To address this, the study introduced a reference-oriented deformable convolution approach that effectively handles spatial misalignment, resulting in improved IQA performance for GAN-based distortions. Additionally, a patch-level attention module was developed to enhance the interaction between patch regions, which were previously processed separately in patch-based approaches.

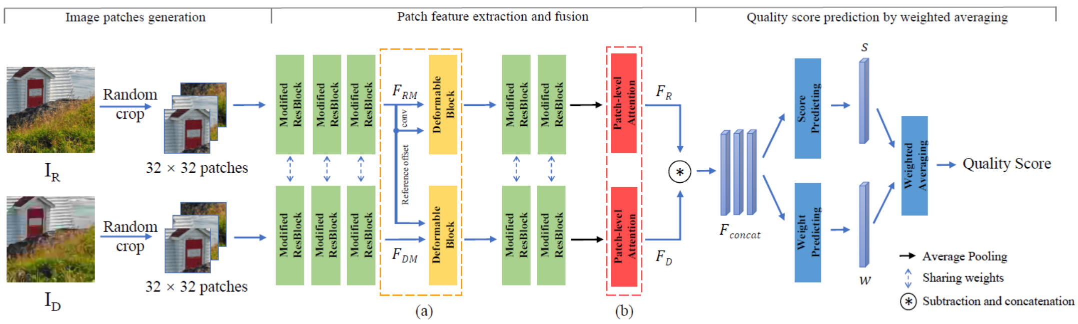

Figure 2.8: *Model architecture of RADN. Image credit: Shi et al. (2021)*

The study also introduced WResNet, a patch-region-based baseline, which was derived from the remaining block. These proposed modules were combined to form the Region-Adaptive Deformable Network (RADN). RADN demonstrated strong performance on the NTIRE 2021 Perceptual Image Quality Assessment Challenge dataset Jinjin Gu (2020), and when incorporated into an ensemble strategy, it achieved fourth place in the final testing phase.

2.2.2 Databases

The LIVE Image Quality Assessment Database (LIVE) Sheikh (2005), the Categorical Subjective Image Quality (CSIQ) database Larson and Chandler (2010), and the Tampere image database 2013 (TID2013) Ponomarenko et al. (2013), Ponomarenko et al. (2015) are publicly accessible.

2.2.2.1 LIVE

The dataset presented by the authors Sheikh et al. (2006) conducted a comprehensive subjective quality evaluation research, wherein a total of 779 distorted photos were assessed by nearly a dozen human subjects. The authors Sheikh et al. (2006) evaluated the performance of several well-known full-reference image quality evaluation algorithms using the "ground truth" image quality data, which was derived from approximately 25,000 unique human quality judgments. The research provided is one of the largest subjective picture quality studies published, considering the number of images, types of distortion, and the number of human judgments for each image. Furthermore, the authors have made all the collected data openly accessible to the scholarly community.

2.2.2.2 CSIQ

Categorical Image Quality (CSIQ) Larson and Chandler (2010) is a database developed by the Image Coding and Analysis Lab at Oklahoma State University. This database contains 30 original images that have been distorted using six distinct varieties and levels of distortion, including JPEG compression, JPEG-2000 compression, global contrast decrements, additive pink Gaussian noise, and Gaussian blurring. As a result, there are a total of 866 distorted images in the database. The subjective quality of these images was evaluated by 35 different observers using a linear displacement method on four calibrated LCD monitors. The DMOS ratings are derived from 5,000 subjective ratings.

2.2.2.3 TID2013

The TID2013 dataset is an extension of the TID2008 dataset. It is designed to evaluate the visual quality of images using a full-reference approach. TID2013 provides an estimation of human perception based on a specific metric. The dataset consists of 25 reference images and 3000 distorted images, resulting from the combination of 25 reference images, 24 types of distortions, and 5 levels of distortion. The reference images are cropped from the Kodak Lossless True Colour Image Suite Ponomarenko et al. (2013), Ponomarenko et al. (2015).

2.2.2.4 KADID-10k

The advancement of deep learning techniques for IQA can be facilitated by the creation of extensive IQA databases encompassing diverse types of data. In this regard, the authors of this study have introduced two notable databases, namely the Konstanz Artificially Distorted Image quality Database (KADID-10k) Lin et al. (2019) and the Konstanz Artificially Distorted Image quality Set (KADIS-700k), as a result of their research efforts. KADID-10k comprises 81 photographs that have been intentionally distorted to varying degrees, offering a total of 25 distinct distortion options. On the other hand, KADIS-700k consists of 140,000 pristine images, each accompanied by five additional copies that have been subjected to random distortions. Leveraging the KADID-10k database, the authors conducted a subjective IQA crowdsourcing investigation, obtaining 30 deterioration category ratings (DCRs) for each image. By combining the annotated KADID-10k dataset with the unlabeled KADIS-700k dataset, the full potential of deep learning-based IQA approaches can be harnessed through the utilization of weakly-supervised learning techniques.

2.2.2.5 CID:IQ

In the conducted research, a novel IQ database named Colourlab Image Database: Image Quality (CID:IQ) Liu et al. (2014) was introduced. The authors of this research proposed several approaches to construct reference images and employed various forms of distortions within the database. Notably, an additional advancement incorporated into the CID:IQ database is the inclusion of perceptual trials conducted at two distinct viewing distances Liu et al. (2014).

The CID:IQ database encompasses three significant features. Firstly, it incorporates state-of-the-art methodologies for creating reference pictures and evaluation procedures. Secondly, it introduces new color-related distortions to enhance the comprehensiveness of the database. Lastly, the CID:IQ database conducts its studies under controlled viewing conditions, utilizing two different viewing distances during the trials Liu et al. (2014).

2.2.2.6 PIPAL

The PIPAL Jinjin Gu (2020) dataset, also known as the Perceptual Image Processing ALgorithms IQA Dataset, stands out from other Image Quality Assessment (IQA) datasets in several ways. Firstly, it is the largest IQA dataset to date, containing a vast number of distorted images with varying levels of visual quality that have been annotated by humans. This comprehensive subjective assessment sets it apart. Secondly, PIPAL employs the Elo rating system to provide subjective scores. Compared to other rating systems, the Elo rating system generates more reliable findings as it is based on probabilities. This enhances the credibility of the dataset. Lastly, PIPAL exhibits high extensibility, allowing users to update the dataset by directly adding additional types of distortions using the extendable characteristic of the Elo system. This flexibility is facilitated by the dataset's good extensibility.

The PIPAL training set comprises over one million human ratings, more than 23,000 distorted images, and over 200 reference images. It encompasses 40 different types of distortions, including the incorporation of results from GAN-based algorithms as a new form of distortion. The Elo rating system is utilized to issue Mean Opinion Scores (MOS) in this context.

The table 2.1 summarizes information about different databases. It includes columns such as Dataset Name, No of Reference and Resolution of Images, Distorted Images, Distortion Types and Levels, Judgement Type, and Subjective Test Method. It summarizes the information from the six databases.

Table 2.1: *Summary of datasets*

Dataset	Ref. & Res. of Img.	Dist. Img.	Dist. Types & (Level)	Judgement Type	Subjective Test Method
LIVE Sheikh et al. (2006)	29 (512x786)	779	5 (5 or 4)	DMOS	Single Stimulus
TID2013 Ponomarenko et al. (2013, 2015)	25 (384x521)	3000	25 (5)	MOS	Pair-wise Comparison
KADID-10k Lin et al. (2019)	81 (512x512)	10125	25 (5)	DMOS	NA
CID:IQ Liu et al. (2014)	23 (800x800)	690	6 (5)	MOS	Double Stimulus
CSIQ Larson and Chandler (2010)	30 (384x521)	866	6 (5 or 4)	DMOS	Simultaneous Comparison
PIPAL Jinjin Gu (2020)	250 (288x288)	29000	40 (116)	MOS	NA

2.2.3 Performance measures

2.2.3.1 PLCC(Pearson Linear Correlation Coefficient)

Pearson's Linear Correlation Coefficient (PLCC) is a statistical measure that quantifies the linear relationship between two variables. In image quality assessment, the PLCC is frequently employed to assess the similarity between subjective image quality ratings and objective quality metrics Sheikh et al. (2006); Varga (2022).

To calculate the PLCC for image quality, a set of subjective ratings provided by human observers and corresponding objective quality scores obtained from an image quality metric are typically required. The following steps outline the process of calculating the PLCC:

Subjective rating collection: A dataset of images is gathered, and a group of human observers subjectively rates the quality of each image using a numerical scale, such as 1 to 5.

Objective quality scores computation: The same set of images is subjected to an image quality metric, such as the Structural Similarity Index (SSIM), Peak Signal-to-Noise Ratio (PSNR), or Mean Opinion Score (MOS), to obtain objective quality scores.

The PLCC value ranges from -1 to 1. A value close to 1 indicates a strong positive linear correlation, signifying that the objective quality scores align well with the subjective ratings. Conversely, a value close to -1 suggests a strong negative linear correlation, indicating a disagreement between the subjective and objective measures. A PLCC value near 0 suggests a weak or no linear correlation Sheikh et al. (2006); Varga (2022).

2.2.3.2 SROCC(Spearman Rank-Ordered Correlation Coefficient)

The Spearman Rank-Ordered Correlation Coefficient (SROCC) is used to evaluate the correlation or similarity between two sets of ranked data. In the realm of image quality assessment, SROCC is frequently employed to assess the effectiveness of objective image quality metrics Sheikh et al. (2006); Varga (2022). Objective image quality metrics aim to automatically quantify the perceived quality of an image, without relying on subjective human judgments. These metrics typically compare the quality scores predicted by the metric with the subjective quality ratings provided by human observers.

To calculate SROCC for image quality assessment, the following steps are typically undertaken:

A dataset of images, along with their corresponding subjective quality ratings provided by human observers, is collected. The objective image quality metric is applied to the images in the dataset, and the predicted quality scores are obtained. Both the subjective quality ratings and the predicted quality scores are independently ranked.

The Spearman's rank correlation coefficient is calculated between the two sets of ranks. This coefficient measures the strength and direction of the monotonic relationship between the subjective ratings and the predicted scores.

A strong positive correlation, indicated by a high positive SROCC value (close to +1), suggests that the objective metric's predictions align well with the subjective ratings. Conversely, a strong negative correlation, indicated by a high negative

SROCC value (close to -1), suggests an inverse relationship between the metric's predictions and the subjective ratings. A value close to 0 indicates a weak or no correlation.

It is important to note that SROCC is just one of several statistical measures used to evaluate the performance of image quality metrics. Other commonly used measures include Pearson's correlation coefficient (PCC) and Kendall's rank correlation coefficient (KROCC). These measures offer different perspectives on the correlation between subjective and objective quality assessments Sheikh et al. (2006); Varga (2022).

Overall, SROCC serves as a valuable tool for assessing the performance of objective image quality metrics, enabling researchers and practitioners to understand how well these metrics align with human perception of image quality.

2.2.3.3 KROCC(Kendall Rank Order Correlation Coefficient)

The Kendall rank order correlation coefficient (KROCC) is a statistical measure that is utilized to assess the strength and direction of the relationship between two ranked variables. Its application is commonly observed in the evaluation of agreement or similarity between two rankings or ordinal data sets Sheikh et al. (2006); Varga (2022).

In the realm of image quality assessment, the correlation between subjective rankings of image quality and objective quality metrics can be measured using KROCC. Objective quality metrics are mathematical algorithms that aim to quantify the perceived quality of an image automatically.

To calculate KROCC in the context of image quality assessment, the following steps are typically followed:

A set of images is collected, and subjective rankings of their quality are obtained from human observers. These rankings are usually acquired through subjective experiments where participants rate or rank the images based on their perceived quality.

Objective quality metrics are applied to the same set of images. These metrics generate numerical scores that represent the objective quality of each image.

The images are ranked based on their objective quality scores.

The Kendall rank order correlation coefficient is calculated between subjective and objective rankings. This calculation can be performed using statistical software or programming languages that support statistical calculations Sheikh et al. (2006); Varga (2022).

2.2.3.4 RMSE(Root Mean Squared Error)

The difference between two images regarding image quality is commonly evaluated using the Root Mean Squared Error (RMSE) metric. The average pixel-wise difference between the predicted image and the ground truth image is measured by RMSE Sheikh et al. (2006); MATLAB (2023a).

To calculate the RMSE between two images, the following steps can be followed:

- If the images are in color, both images should be converted to grayscale.
- If the images have different sizes, they should be resized to the same dimensions.
- The pixel-wise difference between the corresponding pixels in the two images should be computed.
- Each difference value should be squared.
- The mean of the squared differences should be calculated.
- The square root of the mean should be taken to obtain the RMSE value.

2.2.3.5 Alternative Performance Metrics forFR-IQA

- Mean Absolute Error (MAE): It measures the average absolute difference between the predicted quality scores and the ground truth scores. It is calculated by taking the absolute difference between each predicted and ground truth score, and then averaging themSheikh et al. (2006); Varga (2022); MATLAB (2023a).

- Mean Squared Error (MSE): It measures the average squared difference between the predicted quality scores and the ground truth scores. It is calculated by taking the squared difference between each predicted and ground truth score, and then averaging themSheikh et al. (2006); MATLAB (2023a).

2.2.3.6 IQ Metrics

- Structural Similarity Index (SSIM): It measures the similarity between two images based on luminance, contrast, and structural information. SSIM ranges from -1 to 1, with 1 indicating perfect similaritySheikh et al. (2006); MATLAB (2023a).

- Peak Signal-to-Noise Ratio (PSNR): It measures the ratio between the maximum possible power of a signal and the power of corrupting noise. PSNR is commonly used to evaluate image or video quality, with higher values indicating better qualitySheikh et al. (2006); MATLAB (2023a).

- Visual Information Fidelity (VIF): It measures the amount of visual information preserved in the distorted image compared to the reference image. VIF ranges from 0 to 1, with 1 indicating perfect fidelitySheikh et al. (2006); MATLAB (2023a).

- Feature Similarity Index (FSIM): It measures the similarity between two images based on luminance, contrast, and structural information. FSIM ranges from 0 to 1, with 1 indicating perfect similaritySheikh et al. (2006); MATLAB (2023a). These metrics provide different perspectives on image quality assessment and can be used depending on the application's specific requirements.

3 | Methodology

This chapter provides a comprehensive overview of the methodology employed for the thesis. In our pursuit to assess image quality using a Siamese Neural Network, the methodology employed forms a critical part of our research. Here, we outline various steps to design, develop, and evaluate our model.

Figure 3.1 presents a system for our proposed method, FR-IQA using Siamese Neural Networks. The system consists of a Siamese network that learns a similarity metric between pairs of images. We utilize existing Image Quality Metrics (IQM) and databases such as LIVE Sheikh et al. (2006), TID2013 Ponomarenko et al. (2013, 2015), KADID-10k Lin et al. (2019), and CID:IQ Liu et al. (2014) for training and evaluation. The evaluation stage involves comparing the similarity scores generated by the Siamese network with the quality scores in the databases to predict the perceived quality of input images.

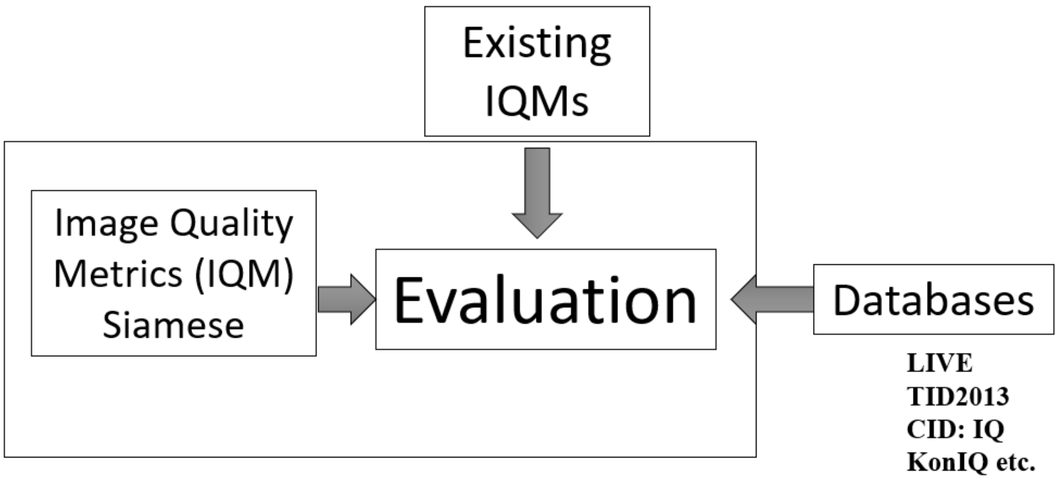

Figure 3.1: *Block diagram of proposed method*

3.1 Problem Formulation

The primary aim of our study revolves around FR-IQA. FR-IQA means the original, undistorted image is available for comparison, serving as the reference. The goal is to assess the quality of a distorted image by comparing it with this reference image.

The problem we are addressing can be broken down into the following components:

Image Distortion Various factors can lead to image distortions, such as noise, compression, transmission errors, or blurring. Our goal is to evaluate the degree of these distortions and measure the perceptual quality of the distorted image. A 'pristine' or undistorted version of the image acts as a reference for comparison. The key is quantifying deviations from this reference image's distorted image.

Quality Assessment involves creating a quality score that accurately reflects human perceptual judgment. The challenge here is to build a model that can consistently predict how humans would rate the quality of the distorted image.

To tackle the above problem, we propose using a Siamese Neural Network. Siamese Neural Networks are a class of neural network architectures that contain two or more identical subnetworks. They are used to calculate the similarity between inputs. In the context of FR-IQA, one subnetwork will intake the reference image and the other the distorted image. The network is trained to minimize the difference between the output of the reference image and the distorted image if they are of the same identity (i.e., essentially the same image, barring the distortions) and maximize the difference if they are not of the same identity.

The problem can thus be formulated as a supervised learning task, where the goal is to learn a function that maps a pair of images to a quality score. The Siamese network can be trained using a loss function, which penalizes large differences in the output for similar images and small differences for dissimilar images.

In the following sections, we will delve into the design and development of the Siamese Neural Network, the choice of loss functions, and the performance metrics to evaluate the model's efficacy.

3.2 Dataset Collection and Preprocessing

3.2.1 Image Dataset Selection

Selecting an appropriate image dataset is a fundamental part of our methodology. Given the nature of Full Reference Image Quality Assessment (FR-IQA), the dataset must contain pairs of reference and distorted images along with human-rated subjective quality scores for these images. These scores serve as the ground truth during the training phase.

A commonly used dataset for FR-IQA tasks is the LIVE Sheikh et al. (2006), IQA Database. The LIVE IQA dataset provides a rich variety of distortions, including JPEG and JPEG2000 compression, additive white Gaussian noise, Gaussian blur, and transmission errors, all common in real-world applications. It also provides subjective quality scores for these images obtained through human studies. Another notable dataset is the TID2013 Ponomarenko et al. (2013, 2015) dataset, which offers a broader variety of distortions and is more challenging due to the diversity of reference images and distortions.

The KADID-10K Lin et al. (2019) dataset is widely used for FR-IQA research. It comprises 10,125 distorted images with various types and severities of distortions, such as compression artifacts, noise, and blur. Each image is associated with a subjective quality score. This dataset provides a diverse and challenging set of images for training and evaluating FR-IQA algorithms, and it has been used in many studies to develop and benchmark state-of-the-art algorithms.

We used the KADID-10k Lin et al. (2019) dataset to train and evaluate our proposed architecture. To be more specific, this dataset was randomly split into three distinct sets: a training set (60%), a validation set (20%), and a test set (20%), all based on the pristine images. Consequently, there was no overlap in content between these sets. Figure 3.2 illustrates resference images from KADID-10K Lin et al. (2019) dataset.

All the models and loss functions were implemented using MATLAB Deep Learning Toolbox MATLAB (2023d). The MATLAB Deep Learning Toolbox is powerful for developing and implementing deep learning models. It supports popular frameworks like TensorFlow, Keras, and PyTorch MATLAB (2023c). It offers prebuilt neural network architectures, training and optimization functions, data preprocessing tools, visualization and interpretation functions, deployment options, and seamless integration with other MATLAB toolboxes. It simplifies the process of building and training models, making it accessible to beginners and experienced researchers.

3.2.2 Image Preprocessing Techniques

Before the data is fed into the Siamese Neural Network, it is crucial to preprocess the images to ensure optimal model performance. Some of the preprocessing steps include:

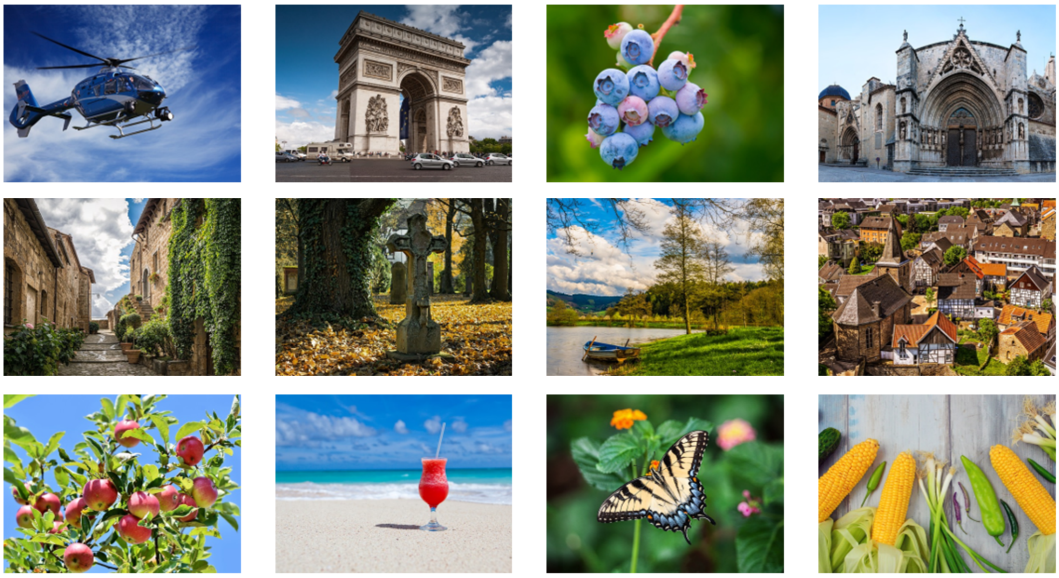

Figure 3.2: *Image samples from KADID-10k dataset. [Image credit: Lin et al. (2019)]*

3.2.2.1 Resizing

All the images in the dataset are resized to a standard size to maintain consistency. This is essential as neural networks require input data of a fixed size.

3.2.2.2 Normalization

The image's pixel intensities are often normalized to fall within a specific range, usually 0-1. This step helps speed up the learning process and improve the model's performance by ensuring all input features (pixel intensities) have similar scales.

3.2.2.3 Augmentation

Data augmentation techniques such as rotation, flipping, scaling, or adding noise can be used to increase the dataset size artificially. This helps improve the model's generalization capability by providing more varied examples during training.

3.2.2.4 Pairwise preparation

Given the nature of Siamese Networks, the data must be prepared in pairs. Each pair contains a reference image and a distorted version of the same image. These pre-processing steps play a crucial role in determining the performance of the model and can be adjusted based on the specific requirements of the task at hand.

ImageDatastore MATLAB (2023a) is a MATLAB feature that simplifies managing and processing large image collections by efficiently storing image data, supporting various file formats, providing automatic transformations, customizable read options, and seamless integration with other MATLAB functions and toolboxes. It also supports labeling, making tasks like image classification or object detection convenient.

3.3 Siamese Network Architecture Design

3.3.1 Network Components

A typical Siamese Network Architecture for Full Reference Image Quality Assessment consists of two identical sub-networks, each taking one image of the pair (reference and distorted) as input. Each sub-network usually consists of the following components:

- Convolutional Layers: These are the primary building blocks of our Siamese Network, aimed at learning spatial hierarchies from the images. They can capture complex patterns, shapes, and structures in the image data that can differentiate between good and poor quality images.
- Pooling Layers: Positioned between the convolutional layers, these layers are employed to reduce the spatial dimensions of the output volume, which in turn reduces computational complexity and helps the network generalize better.
- Activation Functions: The Rectified Linear Unit (ReLU) is commonly used as the activation function in the convolutional layers due to its efficiency and ability to mitigate the vanishing gradient problem.
- Fully Connected Layers: These layers come after several convolutional and pooling layers and are used to perform high-level reasoning on the extracted features. They can help draw complex inferences about the data.
- Output Layer: Each sub-network's output layer produces a feature vector representing the input image's features.

  The final stage of the Siamese Network involves a distance metric or a learned function that takes as input the two feature vectors and outputs a quality score representing the perceived image quality.

3.3.2 Proposed Architecture Details

In our FR-IQA task, we begin by employing a basic Siamese neural network (SNN) in MATLAB.

The Siamese Network architecture consists of several layers that process the input images. The feature extraction part of the Siamese Network is defined by a set of layers. It comprises convolutional layers, activation layers (ReLU), and max pooling layers. The input size of the images (128x128x1) is specified by the *imageInputLayer*, which also disables normalization. The subsequent layers perform feature extraction through the application of convolutions, ReLU activations, and max pooling operations. Finally, a fully connected layer with 4096 neurons is included.

The second branch of the Siamese Network is defined by a set of layers, which is identical to the first branch. It follows a similar structure as the first set of layers but with different filter sizes and additional dropout and batch normalization layers.

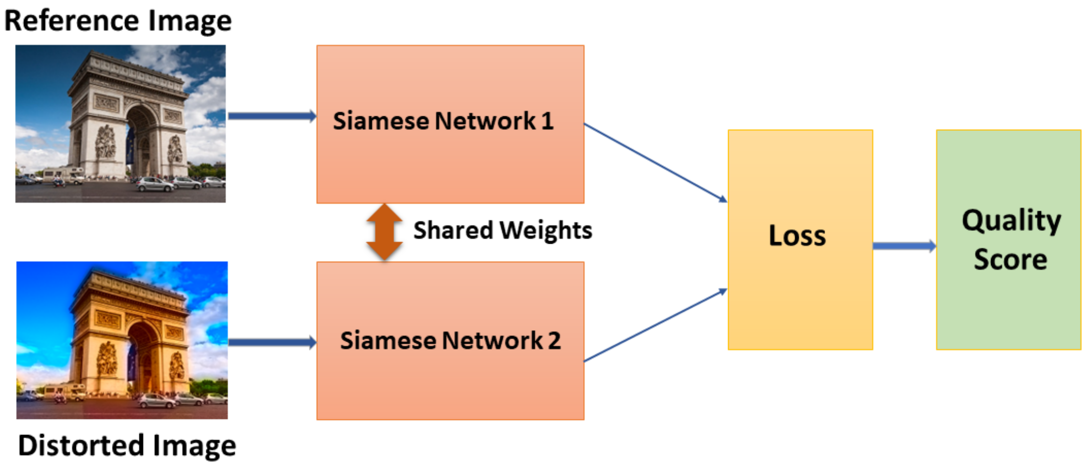

Figure 3.3: *Block diagram of the proposed FR-IQA with Siamese network*

To combine the two branches of the Siamese Network, the layers from the second set can be added to the existing layers from the first set using the *addLayers* function. The outputs of the two branches are combined using the *additionLayer*, and the final regression task is accomplished by adding the *fullyConnectedLayer* and *regressionLayer*. A loss function assesses the predicted image scores and improves the proposed model.

Figure 3.4 shows the proposed FR-IQA system, which uses a Siamese network architecture. This architecture comprises three main components: a multi-level feature fusion module, a weight-sharing Siamese Network, and a regression head.

Firstly, we use a Siamese network that incorporates CDC (central difference convolution) Cong et al. (2022) and spatial attention to process both the reference image and the deformed image. The Siamese network is made up of multiple blocks, and the features obtained from each block are down-sampled to the same size to ensure consistency.

Secondly, we introduce a multi-level feature fusion module. This module combines features at the same level by concatenating them and then fuses features from different levels. This fusion process helps enhance the image representation and improve the quality assessment.

Finally, to obtain the desired quality scores, we employ a fully connected layer as a regression head at the end of the process. This regression head uses fused features to predict the quality scores of the images.

Overall, the proposed FR-IQA system utilizes a Siamese network architecture with a multi-level feature fusion module and a regression head to assess the quality of images accurately.

Block 1 and Block 2 have identical layers but differ in the size of their attention modules. Block 1, which uses a 7x7 convolution kernel, focuses on extracting low-level features. On the other hand, Block 2, which employs a 3x3 convolution kernel, is designed to extract high-level features.

In Fig. 3.4, the letter "C" represents concatenation, and "CDC 64" indicates that there are a total of 64 output channels. The final features are flattened into a vector with a length of 1296 to calculate a quality score. Then, a fully connected layer called "FC1296" is added. It's worth noting that crossing the red line signifies transitioning to the next block while crossing the golden line indicates moving to the multi-level feature fusion structure.

3.3.3 Training Strategy

The training strategy for the Siamese Network involves the use of a suitable loss function and an optimizer.

- Loss Function: The loss function utilized in our study, as described in the literature Cong et al. (2022), serves as a means to assess and optimize the model under consideration. To align the predicted scores with the subjective evaluation scores provided, the IQ evaluation scores are normalized.

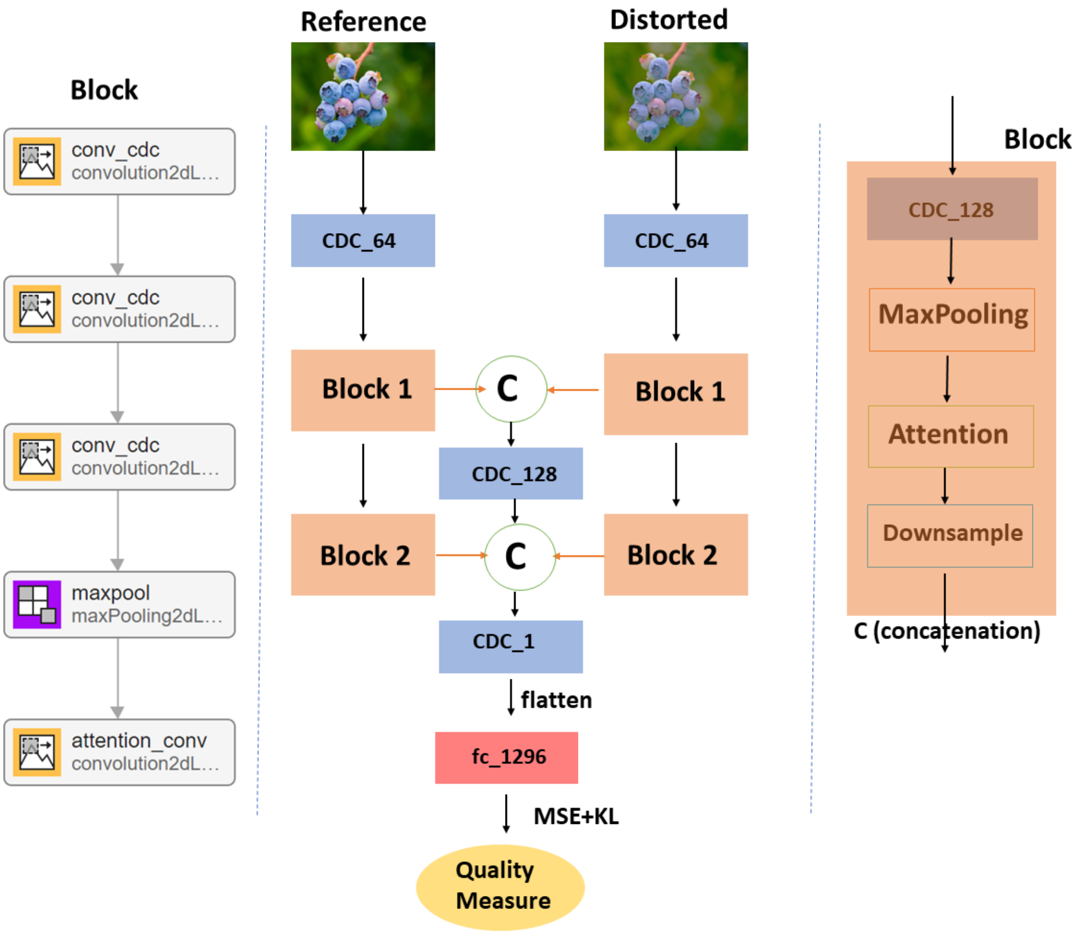

Figure 3.4: *Model Architecture of the proposed FR-IQA with Siamese network*

$$\hat{S}_i = \frac{\left(\sum_{i=1}^{N} \left|\hat{M}_i - \frac{1}{N}\sum_{i=1}^{N}\hat{M}_i\right|^q\right)^{\frac{1}{q}}}{\hat{M}_i - \frac{1}{N}\sum_{i=1}^{N}\hat{M}_i} (i = 1, \cdots, N),$$

$$S_i = \frac{\left(\sum_{i=1}^{N} \left|M_i - \frac{1}{N}\sum_{i=1}^{N}M_i\right|^q\right)^{\frac{1}{q}}}{M_i - \frac{1}{N}\sum_{i=1}^{N}M_i} (i = 1, \cdots, N),$$

where M and $\hat{M}$ are the subjective evaluation scores and predicted quality scores, respectively. $q \geq 1$ is a hyperparameter, here, we set q as 2.$\hat{S}$ and S are the normalized predicted quality scores and the normalized subjected evaluation scores, respectively. The mean square error (MSE) is appiled and the function is defined as

$$L_{\text{pair}}\,(M, \hat{M}) = \text{MSE}\left(\hat{S}_i - S_i\right).$$

Additional, we apply Softmax regression for another set of normalized quality scores (W and $\hat{W}$), given by

$$\hat{W}_i = \frac{\exp \hat{M}_i}{\sum_{i=1}^{N} \exp \hat{M}_i}(i = 1, \cdots, N),$$
$$W_i = \frac{\exp M_i}{\sum_{i=1}^{N} \exp M_i}(i = 1, \cdots, N).$$

The KL divergence | between M and $\hat{M}$ is calculated by

$$L_{list} = D_{KL}(M, \hat{M}) = \sum_{i=1}^{N} \hat{W}_i \times \log \frac{\hat{W}_i}{W_i}.$$

The total loss function can be defined as

$$L(M, \hat{M}) = \alpha L_{\text{pair}}\,(M, \hat{M}) + \beta L_{\text{list}}\,(M, \hat{M}),$$

where α and β are both hyper-parameter. Here, $\alpha = 0.1$ and $\beta = 0.1$.

- Optimizer: Optimizers reduce losses by changing the neural network's weights and learning rate. The choice of optimizer can vary, but Stochastic Gradient Descent (SGD) or Adam are commonly used due to their efficiency in managing large datasets and converging on the optimal weights for the network. The learning rate and other hyperparameters can be tuned based on the specifics of the task. Adam (derived from adaptive moment estimation) Kingma and Ba (2014) employs RMSProp-like parameter updating with a momentum term

  The *InitialLearnRate* training option can be used to specify the learning rate a for all optimization algorithms. The effect of the learning rate differs for different optimization algorithms, so the optimal learning rates are also generally different. Specifying learning rates that differ by layer and parameter is also possible. For more information, please see the section on setting up parameters in Convolutional and Fully Connected Layers Kingma and Ba (2014).

- Regularization Techniques: Overfitting is a common problem in machine learning, where a model performs well on training data but poorly on unseen data. To prevent overfitting, regularization techniques are used in the training of Siamese Networks.

One common method is dropout Srivastava et al. (2014), where random neurons are 'dropped out' or temporarily deactivated during training, which helps to prevent the model from becoming too dependent on any single neuron. To prevent overfitting in the given layers, dropout regularization can be applied by adding dropout layers after the fully connected layers. A dropout layer with a dropout rate of 0.5 (50%) is added after the first fully connected layer. The dropout rate can be adjusted to find the right balance between preventing overfitting and maintaining model performance.

3.4 Training Process

3.4.1 Hyperparameter Selection and Tuning

3.4.1.1 Hyperparameters Selected for the Network

Hyperparameters are the variables that govern the training process and the structure of a machine-learning model. For Siamese Networks, common hyperparameters include:

- Learning Rate: The learning rate controls the size of the updates to the model's weights during training. The *learningRate* is 1e-5 for our model.
- Mini batch Size: The choice of minibatch size depends on various factors, including the available computational resources, the size of the dataset, and the specific IQA algorithm or model being used. We used minibatch size of 128.
- Number of iteration: An "iteration" refers to a single update of the model's parameters using a batch of training data. In each iteration, the model processes a subset of the training data, calculates the loss, and updates the parameters based on the loss gradient. The number of iterations is determined by the size of the training dataset and the batch size used during training. We used 10k *numIterations* MATLAB (2023b).

3.5 Model Evaluation and Validation

- **Metrics Used to Evaluate the Model:**

Objective IQA measures evaluate how well the predicted scores align with the actual scores. These scores serve as a basis for assessing the quality of images. In the majority of articles, you will find three commonly used correlation coefficients: Pearson's rank order correlation coefficient (PLCC), Spearman's rank order correlation coefficient (SROCC), and Kendall's rank order correlation coefficient (KROCC). PLCC is particularly useful when there is a linear relationship between the variables being compared. It quantifies the strength of this linear relationship. However, if the relationship between the variables is not linear, PLCC may not accurately capture the strength of the link. This is because PLCC does not consider nonlinear relationships, limiting its effectiveness in such cases Sheikh et al. (2006); Varga (2022).

To address this limitation, both SROCC and KROCC are provided in the relevant literature. SROCC serves as the nonparametric counterpart to PLCC. It analyzes the strength of a monotonic relationship between the variables, which means it assesses whether the relationship consistently increases or decreases without necessarily being linear. Additionally, SROCC determines the direction of this monotonic relationship Sheikh et al. (2006).

It's important to note that the assumptions for a monotonic relationship are less restrictive than those for a linear relationship. In a monotonic relationship, the variables may not change at a constant rate, but the overall trend is consistently increasing or decreasing. In contrast to PLCC and SROCC, KROCC takes pairs of observations into account. It assesses the strength of the link based on the concordance (agreement) and discordance (disagreement) between these pairs of observations. By considering the agreement and disagreement between pairs, KROCC provides a different perspective on the relationship between the variables Zhang et al. (2012); Varga (2022).

In summary, while PLCC is suitable for linear relationships, SROCC and KROCC offer alternative approaches for assessing the strength and direction of relationships, especially when the relationships are nonlinear or when pairs of observations are involved.

- **Model Validation:** Model validation is an essential step in assessing the performance of a machine learning model. It involves testing the model on a separate set of data that was not used during training. This allows us to evaluate how well the model can handle new, unseen data, which is crucial for its effectiveness.

  In our case, we utilized the KADID-10k dataset, as described in the study by Lin et al. (2019), to both train and evaluate our proposed architecture. To ensure a fair evaluation, we divided this dataset into three distinct sets: a training set comprising 60% of the data, a validation set comprising 20%, and a test set also comprising 20%. Importantly, these sets were created in such a way that there was no overlap in the content of the pristine images between them. This ensures that the model's performance is assessed on diverse and independent data samples.

4 | Experimental Results and Analysis

In this chapter, we will explore experimental results and analysis of the proposed FR-IQA. We will cover important aspects of evaluating and comparing IQA techniques. First, we will talk about the evaluation metrics used to measure the performance of these techniques. Then, we will discuss the experimental setup, including the methods and tools used in our experiments. Additionally, we will compare traditional IQA techniques with the newer Deep Learning IQA techniques. This comparison will help us understand the advancements and potential of deep learning in image quality assessment.

4.1 Evaluation Metrics

Different statistical measures are often used to evaluate how well the predicted scores from a Full-Reference IQA (FR-IQA) metric match up with the actual quality scores in an Image Quality Assessment (IQA) benchmark database. These measures include Pearson's linear correlation coefficient (PLCC), Spearman's rank-order correlation coefficient (SROCC), Kendall's rank order correlation coefficient (KROCC), and mean square error (MSE) Sheikh et al. (2006); Varga (2020); Ayyoubzadeh and Royat (2021); Liu et al. (2019); Cong et al. (2022).

PLCC is useful when there is a linear relationship between the variables being compared. It quantifies the strength of this linear relationship. However, PLCC may not accurately capture the strength of the link if the relationship between the variables is not linear. This limitation led to the development of SROCC and KROCC. SROCC serves as the nonparametric counterpart to PLCC. It assesses the strength of a monotonic relationship between the variables, which means it evaluates whether the relationship consistently increases or decreases without necessarily being linear. SROCC also determines the direction of this monotonic relationship. The assumptions for a monotonic relationship are less restrictive than those for a linear relationship. KROCC takes pairs of observations into account and assesses the strength of the link based on the concordance (agreement) and discordance (disagreement) between these pairs. By considering the agreement and disagreement between pairs, KROCC provides a different perspective on the relationship between the variables Cong et al. (2022); Sheikh et al. (2006); Varga (2020); Ayyoubzadeh and Royat (2021); Liu et al. (2019).

4.2 Experimental Setup

The KADID-10k dataset, described in the study by Lin et al. (2019), was used in our research to train and evaluate our proposed architecture. To ensure a fair evaluation, we divided this dataset into three separate sets: a training set containing 60% of the data, a validation set containing 20%, and a test set also containing 20%. It is worth noting that these sets were carefully created to ensure that there was no overlap in the content of the pristine images between them. This approach guarantees that the model's performance is assessed using diverse and independent data samples.

We utilized an Intel Core i7 processor, along with 32GB of RAM and a NVIDIA GeForce GTX 1660Ti graphics card boasting 6GB of memory. All the models and loss functions were implemented using the deep learning toolbox available in MATLAB MATLAB (2023c). This toolbox provides advanced functionalities and tools for developing and training deep learning models. The toolbox offers a comprehensive set of functions for tasks such as data preprocessing, network architecture design, training, and evaluation.

4.3 Performance Comparison with Baseline Methods

This table presents correlation scores among various state-of-the-art metrics. The metrics are evaluated using different algorithms and compared using three correlation measures: PLCC, SROCC, and KROCC.

The table provides a comparison of the correlation scores for each metric-algorithm combination. Higher correlation scores indicate a stronger relationship between the metric and the algorithm's performance.

In the **SSIM**, the PLCC, SROCC, and KROCC scores are 0.579, 0.671, and 0.489, respectively. This suggests that the SSIM metric, evaluated using the algorithm described in the reference Wang et al. (2004), has moderate positive correlations with the performance measures. Similarly, the table shows correlation scores for other metrics and algorithms, including Mean Structural Similarity (MS-SSIM) Wang et al. (2003), Mean Deviation Similarity Index (MDSI) Nafchi et al. (2016).

The Siamese network presented in Fig 3.3 was modified by adding AlexNet Krizhevsky et al. (2012), VGG16 Simonyan and Zisserman (2014), MobileNet V2 Sandler et al. (2018) as a predefined network. AlexNet Krizhevsky et al. (2012), VGG16 Simonyan and Zisserman (2014), MobileNet_V2 Sandler et al. (2018) as the subnetworks in our Siamese code, we replaced the existing feature extraction layers with the corresponding layers from these networks. These networks are commonly available in deep learning frameworks such as TensorFlow or PyTorch, making it easier to integrate them into the code. The pretrained AlexNet, VGG16, or MobileNet V2 model can be loaded using the appropriate deep-learning framework in MATLAB. The fully connected layers should be removed from the loaded model, as they are typically used for classification tasks and are not necessary for FR-IQA. The remaining layers can be used as the feature extraction layers in your Siamese network. The reference and distorted images can be passed through the Siamese network to obtain their respective feature representations.

The PLCC values between Siam+AlexNet, Siam+MobileNet V2, Siam+VGG16 (0.663, 0.716, 0.707) are shown in the Table 4.1. Based on this comparison, it is evident that our approach achieves the highest (0.781) PLCC values.
The SROCC values between Siam+AlexNet, Siam+MobileNet V2, Siam+VGG16 (0.659, 0.727, 0.679) are shown in the Table 4.1. Based on this comparison, it is evident that our approach achieves the highest (0.823) SROCC values.
The KROCC values between Siam+AlexNet, Siam+MobileNet V2, Siam+VGG16 (0.529, 0.683, 0.624) are shown in the Table 4.1. Based on this comparison, it is evident that our approach achieves the highest (0.804) KROCC values.

In summary, when comparing the performance of these three approaches, our proposed method consistently achieves the highest correlation values across all three metrics: PLCC, SROCC, and KROCC.

When discussing the reasons behind the subpar performance of our proposed FR-IQA model using a Siamese Network, there are several factors to consider. Let us collectively examine these hypotheses.

Firstly, Siamese networks thrive on a large and diverse training dataset to learn meaningful representations. We only used KADID-10k dataset Lin et al. (2019) which is small or lacks diversity, the network may struggle to generalize effectively to new images. This could be one reason for the model's underperformance.

Secondly, the architecture we provided consists of convolutional and fully connected layers. However, it's possible that the network architecture is not deep or complex enough to capture the necessary features for accurate quality assessment.

Table 4.1: *Analyzing Correlation Scores among State-of-the-Art Metrics using KADID-10K Lin et al. (2019)*

	PLCC	SROCC	KROCC
SSIM Wang et al. (2004)	0.579	0.671	0.489
MS-SSIM Wang et al. (2003)	0.625	0.712	0.534
MDSI Nafchi et al. (2016)	-0.844	-0.885	-0.701
Siam+AlexNet	0.663	0.659	0.529
Siam+MobileNet_V2	0.716	0.727	0.683
Siam+VGG16	0.707	0.679	0.624
Ours	0.781	0.823	0.804

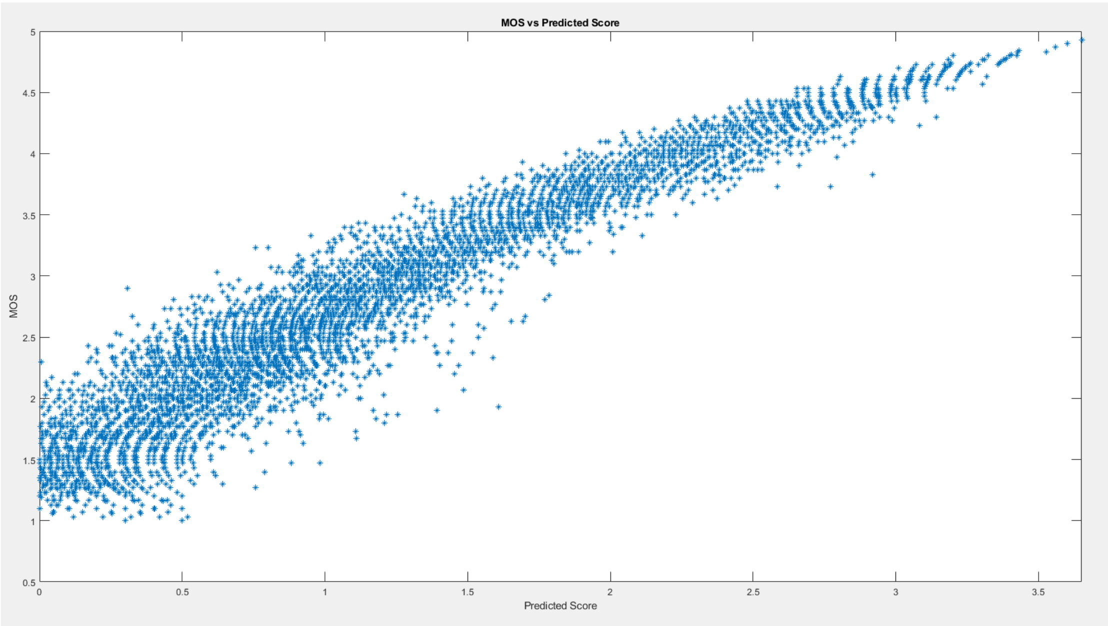

Figure 4.1: *Comparison of Predicted Scores and MOS Values through Scatter Plot*

To address this, we can consider experimenting with different architectures or increasing the depth and complexity of the current network.

Additionally, hyperparameters play a crucial role in the network's performance. Parameters such as learning rate, weight decay, and dropout rate can significantly impact the model's effectiveness. It's important to carefully tune these hyperparameters to find the optimal values for the specific task at hand. We may need to experiment with different values to improve the model's performance.

Lastly, preprocessing techniques can also contribute to enhancing the network's performance. Applying data normalization, augmentation, or resizing can help the network learn relevant features more effectively. It's essential to ensure that the data is properly preprocessed and augmented to maximize the network's ability to assess image quality accurately.

By considering these factors and exploring potential solutions, we can work towards improving the performance of our FR-IQA model using a Siamese Network.

5 | Discussions and Limitations

5.1 Limitations of the Proposed Method

Several limitations exist in the proposed FR-IQA using a Siamese Network. Here are some of them:

- **Computational limitations:** In our proposed method, we acknowledge that computation resources can be a limitation when using Siamese networks. Siamese networks can be computationally expensive, especially with large-scale datasets or high-resolution images. The training and evaluation processes often require substantial computational resources, including memory and processing power. Considering these limitations when implementing Siamese networks in practical applications is important to ensure that the available resources can effectively handle computational demands.

- **Lack of dataset:** The performance of the proposed method heavily relies on the availability and quality of the dataset used for training. If the dataset is limited in size or lacks diversity, the network may not generalize well to unseen images, resulting in reduced prediction accuracy.

- **Cross-validation dataset:** Obtaining a suitable cross-validation dataset can pose challenges for the proposed method, especially when the available dataset is small or needs representative samples. Cross-validation is an essential technique for evaluating the performance of machine learning models.

- **Improving result/prediction accuracy:** Although the proposed method may yield reasonable results, there is always room for improvement in prediction accuracy. Enhancing the accuracy of the predictions could involve fine-tuning the network architecture, optimizing hyperparameters, or exploring advanced training techniques such as data augmentation and transfer learning.

  To address these limitations, the following ideas can be considered:

- **Computational optimization:** Techniques like model compression, quantization, or distributed training can be explored to reduce the computational requirements of the Siamese network. This would make training and evaluating the network on resource-constrained systems more feasible.

- **Transfer learning:** Leveraging pre-trained models on large-scale image datasets, such as ImageNet Deng et al. (2009), and fine-tuning them for the task of image quality assessment can potentially enhance the prediction accuracy of the Siamese network Iman et al. (2023).

By implementing these ideas, the proposed method can overcome some limitations and achieve better performance in Full Reference Image Quality Assessment.

5.2 Future Directions and Improvements

This work has the potential to expand into the following techniques in the future:

- **Comparison to more IQMs:** It would be highly beneficial to conduct a comparative analysis between our Siamese Network approach and the existing IQMs. Such an evaluation would offer a comprehensive assessment of our method and contribute to establishing its efficacy in comparison to other cutting-edge techniques.

- **Cross-dataset validation:** Explore the performance of Siamese Networks for IQA across different datasets to assess their generalizability and robustness. This would involve training the network on one dataset and evaluating it on another, ensuring that the model can effectively handle image characteristics and quality variations Ahmed et al. (2022); Yang et al. (2022); Cong et al. (2022); Varga (2022); Ayyoubzadeh and Royat (2021).

- **Using patches of the images:** Investigate the effectiveness of using image patches instead of full images for IQA. This approach can provide a more fine-grained quality assessment by analyzing local regions within an image, allowing for better detection of specific distortions or artifacts Yang et al. (2022); Cong et al. (2022); Varga (2022); Ayyoubzadeh and Royat (2021).

- **Ensembling different Siamese Networks:** Combine multiple Siamese Networks with different architectures or training strategies to create an ensemble model. This ensemble approach could improve the overall IQA performance by leveraging the strengths of each network Nanni et al. (2021).

- **Siamese Network with Transformer:** Explore the integration of Transformer models into Siamese Networks for IQA Tang et al. (2023). Transformers have shown great success in various computer vision tasks, and their attention mechanisms can capture long-range dependencies in images, potentially enhancing the performance of Siamese Networks for IQA.

- **Multi-Resolution Siamese Networks:** Develop Siamese Networks that can handle images at multiple resolutions. This approach can account for the importance of different image regions at different resolutions, allowing for more accurate and comprehensive quality assessment Yin et al. (2023).

- **Combining Multiple Quality Metrics:** Investigate the fusion of multiple quality metrics within Siamese Networks. The network can capture a broader range of image quality aspects by combining different metrics such as structural similarity index (SSIM), peak signal-to-noise ratio (PSNR), and perceptual metrics like deep features.

- **Explainable Quality Assessment:** Enhance the interpretability of Siamese Networks for IQA by incorporating explainability techniques. This would involve developing methods to visualize and understand the network's decision-making process, providing insights into the specific image regions or features contributing to the quality assessment Jalaboi et al. (2023).

- **Cross-Modality IQA:** Extend Siamese Networks to handle cross-modality IQA, where the reference and distorted images belong to different modalities (e.g., images and videos, images and 3D models). This would require adapting the Siamese Network architecture to compare and assess quality across different modalities effectively.

- **Consider computational complexity, generalizability, interpretability, and scalability:** The challenges related to computational complexity should be addressed, ensuring that the Siamese Network models are efficient and scalable. Furthermore, attention should be given to improving generalizability across diverse datasets, enhancing the interpretability of the network's decisions, and ensuring scalability to handle large-scale IQA tasks.

- **User-Centric Quality Assessment:** Incorporate user-centric factors into Siamese Networks for IQA. This could involve integrating user feedback or preferences during training to create personalized quality assessment models that align with individual user preferences and expectations.

6 | Conclusion

Image Quality Assessment (IQA) is crucial in various image processing applications. While traditional subjective metrics have limitations in terms of consistency and subjectivity, objective IQA techniques have been developed to address these issues. These techniques aim to replicate the functioning of the Human Visual System (HVS) through computational models. However, there is still a gap between the performance of current IQA models and human judgment Niu et al. (2019); Bosse et al. (2017).

Existing FR-IQA models face challenges in capturing high-level semantic features and effectively handling complex, nonlinear distortions. Deep learning methods, particularly Convolutional Neural Networks (CNNs), have shown promising results in improving IQA performance. Siamese Neural Networks, known for their ability to compare and contrast features between pairs of inputs, offer a potential avenue for enhancing FR-IQA Gao et al. (2017).

This research aims to investigate the use of Siamese Neural Networks in FR-IQA and create a model that is both efficient and precise. The goal is to overcome the difficulties associated with capturing intricate and nonlinear distortions and accurately evaluate the perceived quality of images, aligning with human judgment.

6.1 Summary of Contributions

In summary, I would like to highlight the contributions as follows:

- Proposed and developed a Siamese Neural Network-based FR-IQA model: This model represents an advancement in image quality assessment by leveraging the power of Siamese Neural Networks. It offers a new approach to measuring and evaluating image quality accurately.

- Conducted extensive evaluation and comparison: The proposed FR-IQA model was thoroughly evaluated and compared against existing techniques. This evaluation provides valuable insights into the strengths and weaknesses of different approaches, highlighting the effectiveness and superiority of the Siamese Neural Network-based model.

- Explored the application of Siamese Neural Networks in image quality assessment: Through a comprehensive study, we have investigated the potential of Siamese Neural Networks in image quality assessment. This exploration sheds light on the capabilities and limitations of this approach, contributing to the overall understanding of image quality assessment techniques.

- Demonstrated the fine-tuning of Siamese Neural Networks for enhanced FR-IQA performance: By fine-tuning the Siamese Neural Network-based model, its adaptability and potential are shown for further improvement. This insight opens possibilities for future research and development in image quality assessment.

Overall, significant contributions have been made by introducing an FR-IQA model, evaluating its performance, exploring its application in image quality assessment, and providing insights on fine-tuning.

6.2 Revisiting Research Objectives

The application and effectiveness of Siamese Neural Networks for FR-IQA were explored. A Siamese Neural Network-based FR-IQA model was developed to reflect human judgment of image quality. The proposed model's performance was assessed against existing FR-IQA techniques using a standard IQA database. Additionally, the unique properties of the Siamese Neural Network were investigated to leverage their potential for improving FR-IQA performance.

Bibliography

Ahmed, N., Shahzad Asif, H., Bhatti, A. R., and Khan, A. (2022). Deep ensembling for perceptual image quality assessment. *Soft Computing*, 26(16):7601–7622. (cited on pages 2 and 48)

Ayyoubzadeh, S. M. and Royat, A. (2021). (asna) an attention-based siamese-difference neural network with surrogate ranking loss function for perceptual image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 388–397. (cited on pages 2, 12, 13, 14, 16, 17, 18, 19, 41, 42, 48, and 59)

Bosse, S., Maniry, D., Müller, K.-R., Wiegand, T., and Samek, W. (2017). Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Transactions on image processing*, 27(1):206–219. (cited on pages 9 and 51)

Cheon, M., Yoon, S.-J., Kang, B., and Lee, J. (2021). Perceptual image quality assessment with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 433–442. (cited on page 1)

Cong, H., Fu, L., Zhang, R., Zhang, Y., Wang, H., He, J., and Gao, J. (2022). Image quality assessment with gradient siamese network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1201–1210. (cited on pages 1, 2, 6, 11, 18, 35, 41, 42, 48, and 59)

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee. (cited on page 48)

Dey, S., Dutta, A., Toledo, J. I., Ghosh, S. K., Lladós, J., and Pal, U. (2017). Signet: Convolutional siamese network for writer independent offline signature verification. *arXiv preprint arXiv:1707.02131*. (cited on pages 13 and 59)

Gao, F., Wang, Y., Li, P., Tan, M., Yu, J., and Zhu, Y. (2017). Deepsim: Deep similarity for image quality assessment. *Neurocomputing*, 257:104–114. (cited on pages 9, 10, and 51)

Gu, J., Cai, H., Dong, C., Ren, J. S., Timofte, R., Gong, Y., Lao, S., Shi, S., Wang, J., Yang, S., et al. (2022). Ntire 2022 challenge on perceptual image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 951–967. (cited on page 9)

Guo, H., Bin, Y., Hou, Y., Zhang, Q., and Luo, H. (2021). Iqma network: Image quality multi-scale assessment network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 443–452. (cited on page 10)

Iman, M., Arabnia, H. R., and Rasheed, K. (2023). A review of deep transfer learning and recent advancements. *Technologies*, 11(2):40. (cited on page 48)

Jalaboi, R., Winther, O., and Galimzianova, A. (2023). Explainable image quality assessments in teledermatological photography. *Telemedicine and e-Health.* (cited on page 49)

Jinjin Gu, Haoming Cai, H. C. X. Y. J. R. C. D. (2020). Pipal: a large-scale image quality assessment dataset for perceptual image restoration. In *European Conference on Computer Vision (ECCV) 2020*, pages 633–651, Cham. Springer International Publishing. (cited on pages 10, 19, 22, and 24)

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980.* (cited on page 37)

Koch, G., Zemel, R., Salakhutdinov, R., et al. (2015). Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, volume 2. Lille. (cited on pages 11 and 14)

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25. (cited on page 43)

Larson, E. C. and Chandler, D. M. (2010). Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of electronic imaging*, 19(1):011006–011006. (cited on pages 7, 8, 9, 10, 22, and 24)

Levina, E. and Bickel, P. (2001). The earth mover's distance is the mallows distance: Some insights from statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 251–256. IEEE. (cited on page 11)

Li, Y., Chen, C. P., and Zhang, T. (2022). A survey on siamese network: Methodologies, applications and opportunities. *IEEE Transactions on Artificial Intelligence.* (cited on pages 12, 13, 14, 16, 17, and 59)

Li, Y., Wang, S., Zhang, X., Wang, S., Ma, S., and Wang, Y. (2021). Quality assessment of end-to-end learned image compression: The benchmark and objective measure. In *Proceedings of the 29th ACM international conference on multimedia*, pages 4297–4305. (cited on pages 16 and 20)

Lin, H., Hosu, V., and Saupe, D. (2019). Kadid-10k: A large-scale artificially distorted iqa database. In *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–3. IEEE. (cited on pages 23, 24, 29, 31, 32, 39, 42, 43, 44, 59, and 61)

Liu, X., Pedersen, M., and Hardeberg, J. Y. (2014). Cid: Iq–a new image quality database. In *International Conference on Image and Signal Processing*, pages 193–202. Springer. (cited on pages 23, 24, and 29)

Liu, X., Zhou, Y., Zhao, J., Yao, R., Liu, B., and Zheng, Y. (2019). Siamese convolutional neural networks for remote sensing scene classification. *IEEE Geoscience and Remote Sensing Letters*, 16(8):1200–1204. (cited on pages 1, 14, 41, and 42)

Lukin, V. V., Ponomarenko, N. N., Ieremeiev, O. I., Egiazarian, K. O., and Astola, J. (2015). Combining full-reference image visual quality metrics by neural network. In *Human Vision and Electronic Imaging XX*, volume 9394, pages 172–183. SPIE. (cited on page 10)

Madhusudana, P. C., Birkbeck, N., Wang, Y., Adsumilli, B., and Bovik, A. C. (2022). Image quality assessment using contrastive learning. *IEEE Transactions on Image Processing*, 31:4149–4161. (cited on pages 7, 8, and 9)

MATLAB (Accessed on March, 2023a). Matlab. *https://www.mathworks.com/*. (cited on pages 27, 28, and 33)

MATLAB, S. (Accessed on January, 2023b). Train a siamese network to compare images. *https://www.mathworks.com/help/deeplearning/ug/train-a-siamese-network-to-compare-images.html.* (cited on page 38)

MATLAB, T. (Accessed on April, 2023c). Tensorflow with matlab. *https://blog.tensorflow.org/2023/03/tensorflow-with-matlab.html.* (cited on pages 31 and 42)

MATLAB, T. (Accessed on January, 2023d). Toolbox deep learning matlab. *https://www.mathworks.com/products/deep-learning.html.* (cited on page 31)

Mohammadi, P., Ebrahimi-Moghadam, A., and Shirani, S. (2014). Subjective and objective quality assessment of image: A survey. *arXiv preprint arXiv:1406.7799.* (cited on page 7)

Nafchi, H. Z., Shahkolaei, A., Hedjam, R., and Cheriet, M. (2016). Mean deviation similarity index: Efficient and reliable full-reference image quality evaluator. *Ieee Access*, 4:5579–5590. (cited on pages 43 and 44)

Nanni, L., Minchio, G., Brahnam, S., Sarraggiotto, D., and Lumini, A. (2021). Closing the performance gap between siamese networks for dissimilarity image classification and convolutional neural networks. *Sensors*, 21(17):5809. (cited on page 48)

Niu, Y., Huang, D., Shi, Y., and Ke, X. (2019). Siamese-network-based learning to rank for no-reference 2d and 3d image quality assessment. *IEEE Access*, 7:101583–101595. (cited on pages 1, 6, 7, 10, 12, 13, 14, 16, 17, 19, 20, 51, and 59)

Pedersen, M., Hardeberg, J. Y., et al. (2012). Full-reference image quality metrics: Classification and evaluation. *Foundations and Trends® in Computer Graphics and Vision*, 7(1):1–80. (cited on pages 1, 7, and 8)

Pei, S.-C. and Chen, L.-H. (2015). Image quality assessment using human visual dog model fused with random forest. *IEEE Transactions on Image Processing*, 24(11):3282–3292. (cited on page 10)

Ponomarenko, N., Ieremeiev, O., Lukin, V., Egiazarian, K., Jin, L., Astola, J., Vozel, B., Chehdi, K., Carli, M., Battisti, F., et al. (2013). Color image database tid2013: Peculiarities and preliminary results. In *European workshop on visual information processing (EUVIP)*, pages 106–111. IEEE. (cited on pages 22, 23, 24, 29, and 31)

Ponomarenko, N., Jin, L., Ieremeiev, O., Lukin, V., Egiazarian, K., Astola, J., Vozel, B., Chehdi, K., Carli, M., Battisti, F., et al. (2015). Image database tid2013: Peculiarities, results and perspectives. *Signal processing: Image communication*, 30:57–77. (cited on pages 22, 23, 24, 29, and 31)

Qiu, K., Ai, Y., Tian, B., Wang, B., and Cao, D. (2018). Siamese-resnet: Implementing loop closure detection based on siamese network. In *2018 IEEE Intelligent Vehicles Symposium (IV)*, pages 716–721. IEEE. (cited on page 15)

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520. (cited on page 43)

Sheikh, H. (2005). Live image quality assessment database release 2. *http://live. ece. utexas. edu/research/quality*. (cited on page 22)

Sheikh, H. R., Sabir, M. F., and Bovik, A. C. (2006). A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing*, 15(11):3440–3451. (cited on pages 7, 8, 22, 24, 25, 26, 27, 28, 29, 31, 39, 41, and 42)

Shi, S., Bai, Q., Cao, M., Xia, W., Wang, J., Chen, Y., and Yang, Y. (2021). Region-adaptive deformable network for image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 324–333. (cited on pages 6, 21, and 59)

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. (cited on page 43)

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958. (cited on page 38)

Sun, W., Liao, Q., Xue, J.-H., and Zhou, F. (2018). Spsim: A superpixel-based similarity index for full-reference image quality assessment. *IEEE Transactions on Image Processing*, 27(9):4232–4244. (cited on pages 7, 8, and 9)

Talebi, H. and Milanfar, P. (2018). Nima: Neural image assessment. *IEEE transactions on image processing*, 27(8):3998–4011. (cited on pages 1, 6, 7, 20, 21, and 59)

Tang, Z., Chen, Z., Li, Z., Zhong, B., Zhang, X., and Zhang, X. (2023). Unifying dual-attention and siamese transformer network for full-reference image quality assessment. *ACM Transactions on Multimedia Computing, Communications and Applications*. (cited on page 49)

Varga, D. (2020). Composition-preserving deep approach to full-reference image quality assessment. *Signal, Image and Video Processing*, 14(6):1265–1272. (cited on pages 1, 12, 13, 14, 16, 17, 41, 42, and 59)

Varga, D. (2022). Full-reference image quality assessment based on an optimal linear combination of quality measures selected by simulated annealing. *Journal of Imaging*, 8(8):224. (cited on pages 1, 24, 25, 26, 27, 39, and 48)

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612. (cited on pages 9, 42, and 44)

Wang, Z. and Li, Q. (2010). Information content weighting for perceptual image quality assessment. *IEEE Transactions on image processing*, 20(5):1185–1198. (cited on page 9)

Wang, Z., Simoncelli, E. P., and Bovik, A. C. (2003). Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pages 1398–1402. IEEE. (cited on pages 43 and 44)

Yan, S., Liang, S., Li, X., Zhang, L., Wang, S., and Dong, C. (2021). Ps-gan: A single image snow removal framework using pseudo-siamese gans. In *International Conference on Image Processing and Intelligent Control (IPIC 2021)*, volume 11928, pages 151–159. SPIE. (cited on page 15)

Yang, S., Wu, T., Shi, S., Lao, S., Gong, Y., Cao, M., Wang, J., and Yang, Y. (2022). Maniqa: Multi-dimension attention network for no-reference image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1191–1200. (cited on page 48)

Yin, H., Weng, L., Li, Y., Xia, M., Hu, K., Lin, H., and Qian, M. (2023). Attention-guided siamese networks for change detection in high resolution remote sensing images. *International Journal of Applied Earth Observation and Geoinformation*, 117:103206. (cited on page 49)

Yuan, S., Yu, X., and Majid, A. (2019). Robust face tracking using siamese-vgg with pre-training and fine-tuning. In *2019 4th International Conference on Control and Robotics Engineering (ICCRE)*, pages 170–174. IEEE. (cited on page 15)

Zhang, L., Zhang, L., Mou, X., and Zhang, D. (2012). A comprehensive evaluation of full reference image quality assessment algorithms. In *2012 19th IEEE International Conference on Image Processing*, pages 1477–1480. IEEE. (cited on pages 7, 8, 9, and 39)

List of Figures

List of Tables