Goal Oriented Image Quality Assessment Using CNN

A Final Project End Semester Report

submitted by

SRAVANTH CHOWDARY POTLURI (CS20B1006)

in partial fulfilment of requirements

for the award of the degree of

BACHELOR OF TECHNOLOGY

Department of Computer Science and Engineering
INDIAN INSTITUTE OF INFORMATION TECHNOLOGY, DESIGN AND MANUFACTURING, KANCHEEPURAM

May 2024

DECLARATION OF ORIGINALITY

I, **Sravanth Chowdary Potluri**, with Roll No: **CS20B1006** hereby declare that the material presented in the Project Report titled **Goal Oriented Image Quality Assessment Using CNN** represents original work carried out by me in the **Department of Computer Science and Engineering** at the Indian Institute of Information Technology, Design and Manufacturing, Kancheepuram.

With my signature, I certify that:

- I have not manipulated any of the data or results.
- I have not committed any plagiarism of intellectual property. I have clearly indicated and referenced the contributions of others.
- I have explicitly acknowledged all collaborative research and discussions.
- I have understood that any false claim will result in severe disciplinary action.
- I have understood that the work may be screened for any form of academic misconduct.

Sravanth Chowdary Potluri

Place: Chennai

Date: 14.05.2024

CERTIFICATE

This is to certify that the report titled **Goal Oriented Image Quality Assessment Using CNN**, submitted by **Sravanth Chowdary Potluri (CS20B1006)**, to the Indian Institute of Information Technology, Design and Manufacturing Kancheepuram, in partial fulfilment of requirements for the award of the degree of **BACHELOR OF TECHNOLOGY** is a bonafide record of the work done by him/her under my supervision. The contents of this report, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Prof. Masilamani V
Project Internal Guide
Professor
Department of Computer Science and Engineering
IIITDM Kancheepuram, Chennai - 600 127

Place: Chennai

Date: 14.05.2024

ABSTRACT

Traditional image quality assessment (IQA) methods often fall short in precision when tailored specifically for evaluating images intended for subsequent computer vision tasks. This research introduces a goal-oriented image quality assessment (GO-IQA) framework, designed specifically for depth estimation applications. Utilizing the NYU-Depth-v2 dataset, which has been augmented with blur and various types of noise, the study reflects realistic image quality challenges encountered in the field. Our approach leverages the "Depth Anything" algorithm to generate depth-related mean squared error (MSE) scores, which serve as goal-specific quality metrics. BRISQUE analysis is employed to quantify the effects of image degradation. To predict MSE scores directly from images, multiple convolutional neural network (CNN) architectures were developed and assessed. The results reveal a strong correlation between CNN-predicted scores and actual image quality in the context of depth estimation. The study also evaluates the time efficiency of this algorithm in identifying poor-quality images, thereby streamlining depth estimation processes. Ultimately, this work establishes foundational principles for the development of robust GO-IQA systems using CNNs, aimed at enhancing image quality assessment for depth estimation and similar applications.

KEYWORDS: Image Quality Assessment; Depth Estimation; Convolutional Neural Networks; BRISQUE; Data Augmentation; ResNet

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

IQA	Image Quality Assessment
NR-IQA	No Reference - Image Quality Assessment
GO-IQA	Goal Oriented - Image Quality Assessment
MSE	Mean Square Error
MAE	Mean Absolute Error
PCC	Pearson Correlation Coefficient

CHAPTER 1

Introduction

1.1 Image Quality Assessment

Image Quality Assessment (IQA) plays a fundamental role in the realm of image processing and computer vision. Its influence extends across a vast array of applications, impacting fields like security and surveillance systems, medical imaging analysis, and the entertainment industry. At its core, IQA strives to quantify the perceived visual quality of an image. Traditionally, three main approaches have been employed to achieve this goal:

- **Full-Reference IQA:** This method leverages a pristine, undistorted reference image as a benchmark. By comparing the distorted image under evaluation with the reference, the IQA algorithm calculates a quality score reflecting the degree of deviation from the original.
- **Reduced-Reference IQA:** In scenarios where a flawless reference image might not be readily available, reduced-reference IQA techniques come into play. These methods utilize partial information about the reference image, such as statistical properties, to assess the quality of the distorted image.
- **No-Reference IQA (NR-IQA):** Offering the most flexibility, NR-IQA algorithms function entirely without reference to the original image. They rely solely on the properties of the distorted image itself to estimate its quality. This makes NR-IQA particularly valuable in situations where access to a reference image is impractical or even impossible, such as with real-time image streams or captured images from unknown sources.

While these established IQA methodologies have proven valuable, a critical limitation emerges when solely evaluating visual quality.

1.2 Goal Oriented Image Quality Assessment

Imagine a scenario where an image appears visually stunning, boasting excellent clarity, color balance, and detail. However, when tasked with a specific computer vision goal,

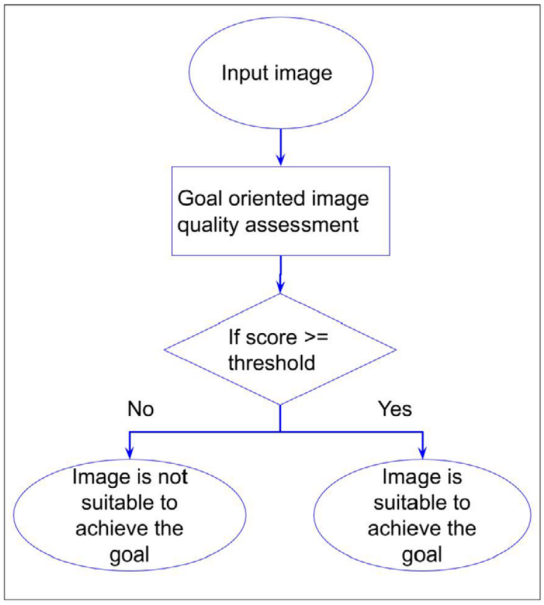

Figure 1.1: Objective of Goal Oriented Image Quality Assessment

such as accurate depth estimation, this seemingly perfect image might perform poorly. This inconsistency highlights a crucial blind spot in traditional visual-centric IQA.

Goal-Oriented Image Quality Assessment (GO-IQA) [1] offers a paradigm shift by considering a specific task or goal alongside the image itself. Instead of solely focusing on visual quality, GO-IQA aims to predict how well an image would perform when processed by the best possible algorithm to achieve a defined goal. Examples of such goals include depth estimation, object detection, or image segmentation. Figure 1.1 displays the Objective of Goal Oriented Image Quality Assessment

Let's explore broader applications of GO-IQA to underscore its significance:

- **Autonomous Navigation and Robotics:** Depth estimation is a cornerstone of robotic navigation and 3D environment mapping. A GO-IQA model specialized for depth estimation could rapidly identify images that would yield inaccurate depth maps, even when processed by the most sophisticated algorithms. This could trigger adjustments, such as prompting the robot to capture additional images or reposition itself for a better vantage point.
- **Object Detection and Tracking:** Object detection in complex backgrounds remains a challenge in computer vision. GO-IQA could be trained to predict the likelihood of accurate detection, highlighting images where objects may be partially obscured, poorly lit, or inherently difficult to distinguish. This information could assist systems in adapting detection strategies or alerting users about potential detection failures.
- **Surveillance Systems:** In security surveillance, image clarity and the ability to

discern fine details are paramount. A GO-IQA model tailored to object and event detection within surveillance footage could play a vital role in quality control. Low GO-IQA scores could flag problematic footage, prompting camera adjustments or triggering alerts for human review.

- **Synthetic Image/Video Generation:** The rise of synthetic images and video generated by AI algorithms presents unique challenges in quality evaluation. A GO-IQA model could be adapted to assess the suitability of synthetic images for specific downstream tasks. For instance, in game development or virtual simulations, it could ensure generated imagery meets the requirements necessary for realistic object interaction and physics calculations.

To illustrate the critical need for Goal-Oriented Image Quality Assessment (GO-IQA), consider the triad of images presented in Figures 1.2, 1.3, and 1.4. Figure 1.2 depicts an image from the NYU-Depth-v2 dataset [2], a benchmark for depth estimation research. Figures 1.3 and 1.4 are manipulated versions of the original: a blurred image (Gaussian blur kernel, size 15x15) and a noisy image (Gaussian noise, mean 0, standard deviation 0.4).

We evaluated these images using both a state-of-the-art monocular depth estimation model, Depth Anything [3], and the popular no-reference image quality metric, BRISQUE [4]. BRISQUE scores (33.23, 67.82, 149.75) align with our intuition: the manipulated images are progressively degraded. However, the depth estimation results reveal a striking inconsistency.

Surprisingly, the Depth Anything algorithm performs reasonably well on the blurred and noisy images. In fact, the blurred image yields a lower Mean Squared Error (MSE) score (0.028) compared to the original (0.032), indicating superior depth estimation performance. This counterintuitive result exposes the fundamental shortcoming of traditional image quality metrics—they may not reliably predict performance for specific computer vision tasks, such as depth estimation.

This case study underscores the importance of GO-IQA. Instead of solely focusing on visual quality, a GO-IQA framework would assess an image's suitability for achieving accurate depth estimation results. It would analyze features directly relevant to depth algorithms, potentially identifying limitations in visually appealing images (as seen in this example) that could hinder depth estimation success.

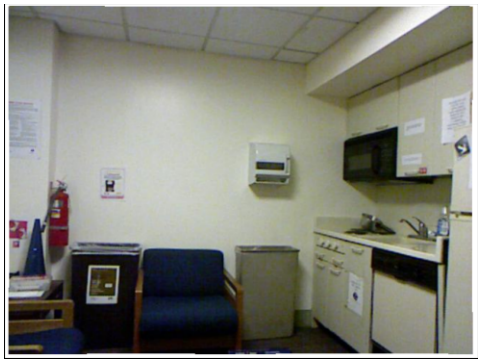

Figure 1.2: Original Image, Depth MSE:0.032, BRISQUE:33.23

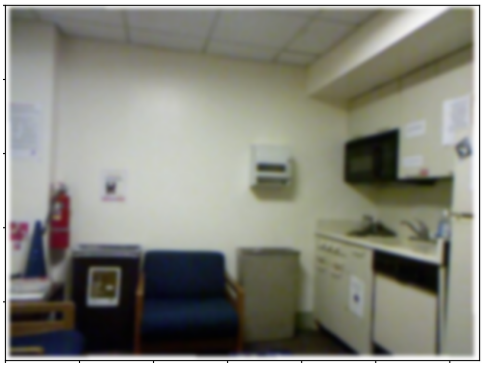

Figure 1.3: Blurred Image, Depth MSE:0.028, BRISQUE:67.82

Figure 1.4: Noisy Image, Depth MSE:0.044, BRISQUE:149.75

1.3 Motivation

1.3.1 Addressing the Challenges of Traditional GO-IQA Approaches

While existing GO-IQA approaches have proven successful, they often depend on hand-crafted features and machine learning models such as Support Vector Regression (SVR). The process of meticulous feature engineering can be a significant bottleneck, requiring extensive domain knowledge and potentially limiting adaptation to different tasks. This project aims to streamline GO-IQA, focusing specifically on the crucial task of depth

estimation, by overcoming these challenges.

1.3.2 Data-Driven Understanding of Depth Estimation Quality

We will leverage the NYU-depth-v2 [2] dataset, a comprehensive resource for building and evaluating depth estimation models. To enhance its utility, we'll strategically introduce controlled variations like noise and blur, simulating real-world conditions that impact image quality. Running a state-of-the-art depth estimation algorithm on both the original and augmented images will generate diverse depth maps. This enriched dataset will facilitate a nuanced understanding of how image quality variations affect depth estimation performance.

1.3.3 The Power of Deep Learning Automation

The heart of our approach lies in the utilization of Convolutional Neural Networks (CNNs). CNNs have a remarkable ability to automatically learn and extract complex features directly from image data. This capability eliminates the need for manual feature design, offering the potential to both streamline the GO-IQA process and improve the accuracy of depth estimation quality predictions. By integrating CNNs within our GO-IQA framework, we strive to create a more efficient and reliable quality assessment tool for depth estimation tasks.

1.4 Summary

This introduction establishes the context of your project within the field of IQA, clearly outlining the limitations of traditional approaches and the motivation behind using GO-IQA. It emphasizes your focus on depth estimation, the planned data augmentation strategy, and highlights the core innovation of using CNNs for automated and accurate quality assessment. The following sections will build upon this foundation with a literature review, in-depth methodology explanation, results along with future scope and conclusion.

CHAPTER 2

Literature Review

The goal of image quality assessment (IQA) is to evaluate the quality of images, which has become increasingly important in various applications such as image compression, augmented reality, and computer vision. One particular area of interest is the assessment of image quality for the purpose of depth estimation using convolutional neural networks (CNNs). This literature review aims to provide insights into the current state of research in goal-oriented image quality assessment for depth estimation using CNNs and identify potential future research directions.

Convolutional Neural Networks for Image Quality Assessment Kang, Le, Ye, and Doermann (2014) [5] explored the use of CNNs for no-reference image quality assessment. They demonstrated the effectiveness of CNNs in evaluating image quality without reference to an original image. This finding suggests that CNNs can be a powerful tool in assessing the quality of images for depth estimation.

Deep Convolutional Neural Fields for Depth Estimation Liu, Shen, and Lin (2014) [6] proposed deep convolutional neural fields for depth estimation from a single image. Their approach leveraged deep learning techniques to estimate depth from a single image, showcasing the potential of CNNs in depth estimation applications.

Unsupervised Feature Learning Framework for Image Quality Assessment Ye, Kumar, Kang, and Doermann (2012) [7] introduced an unsupervised feature learning framework for no-reference image quality assessment. Their work focused on extracting features from images without relying on reference images, which can be valuable in the context of depth estimation where reference images may not always be available.

Deep Neural Networks for Image Quality Assessment Bosse, Maniry, Müller, Wiegand, and Samek (2016) [8] investigated the use of deep neural networks for both no-reference and full-reference image quality assessment. Their findings highlighted the potential of deep learning techniques in assessing image quality, which can be particularly relevant in the context of depth estimation.

Blind Image Quality Assessment Using Convolutional Neural Networks Zhang, Ma, Yan, Deng, and Wang (2019) [9] proposed a blind image quality assessment method using a deep bilinear convolutional neural network. Their approach aimed to assess image quality without access to reference images, which aligns with the goal-oriented nature of image quality assessment for depth estimation.

Potential Future Research Directions While the existing research provides valuable insights into the use of CNNs for image quality assessment in the context of depth estimation, there are several potential future research directions to consider. Firstly, there is a need to explore the specific challenges and requirements of image quality assessment for depth estimation tasks, as traditional IQA metrics may not fully capture the quality factors relevant to depth estimation. Additionally, the integration of uncertainty-aware techniques, as proposed by Zhang, Ma, Zhai, and Yang (2020)[10], can be valuable in enhancing the robustness of image quality assessment for depth estimation, especially in real-world applications.

Furthermore, the application of 3D convolutional neural networks, as demonstrated by Ge, Liang, Yuan, and Thalmann (2017) [11], presents an intriguing avenue for assessing image quality in the context of depth estimation, considering the spatial nature of depth information. Future research can explore the adaptation of 3D CNNs for image quality assessment and depth estimation tasks. Additionally, the exploration of transfer learning techniques and domain-specific fine-tuning of CNNs for depth estimation applications can be a promising direction for future research.

In conclusion, the existing literature provides a strong foundation for utilizing convolutional neural networks in image quality assessment for the goal of depth estimation. However, there are still several knowledge gaps and opportunities for further exploration and innovation in this area, which can lead to advancements in depth estimation and related applications. This Work aims to cover the gap in the form of Goal Oriented Image Quality Assessment using CNN's for the specific task of Depth Estimation.

CHAPTER 3

Methodology

3.1 Data

3.1.1 Dataset Introduction

This research leverages the NYU-Depth-v2 dataset [2] to develop and evaluate the proposed Goal-Oriented Image Quality Assessment (GO-IQA) model for depth estimation. The NYU-Depth-v2 dataset is a widely-used benchmark in depth estimation research, providing a comprehensive collection of indoor scenes. Key characteristics of the dataset relevant to this project include:

- **RGB-D Data**: The dataset comprises around 51000 paired RGB images and corresponding depth maps (ground truth) captured with Microsoft Kinect sensors. This enables the direct relationship between image features and depth estimation performance to be investigated.
- **Diverse Scenes**: The dataset covers a variety of indoor environments, such as bedrooms, living rooms, and offices. This diversity introduces realistic challenges for depth estimation algorithms and allows the GO-IQA model to learn from a wide range of image characteristics.
- **Dataset Size**: The substantial number of images in NYU-Depth-v2 facilitates the training of data-driven deep learning models and provides a robust evaluation set for the GO-IQA model.

3.1.2 Data Preparation

To tailor the dataset for this project's GO-IQA framework, the following data preparation steps were implemented:

- **Random Selection 1:** Around 3500 images and their corresponding depth maps were randomly selected from the dataset and passed into the Data Augmentation phase.

- **Data Augmentation:** The selected images were augmented through the controlled introduction of noise (Gaussian and Salt and Pepper noise with varying standard deviations) and blur (Gaussian blur with varying kernel sizes). This augmentation expands the dataset by simulating real-world image quality variations that impact depth estimation.
- **Random Selection 2:** After the Data Augmentation which yielded around 48000 images, 5000 images were randomly selected from the dataset for the next phase which is the calculation of MSE of the images by the SOTA depth estimation algorithm.
- **MSE Calculation:** For each original and augmented image, the Mean Squared Error (MSE) score was computed using a state-of-the-art depth estimation algorithm, such as Depth Anything [3]. These MSE scores serve as ground truth labels for the GO-IQA model, reflecting the depth estimation performance achievable on each image variant.

Dataset Splits: The prepared dataset after the ground truth estimation by the used depth estimation algorithm was divided into training (e.g., 80%), validation (20%) sets. The training set is used to train the GO-IQA model and the validation set for hyperparameter tuning and evaluation of the model.

3.1.3 Data Augmentation

The NYU-Depth-v2 dataset, while comprehensive, does not fully capture the array of image quality variations commonly encountered in real-world environments. To bolster the model's robustness and enhance its generalizability across diverse imaging conditions, we refined our data augmentation strategy. We introduced controlled distortions of three types:

- **Salt and Pepper Noise:** We applied salt and pepper noise at levels of 0.01, 0.02, 0.04, and 0.08. This type of noise introduces sharp, sparse disturbances in the image, mimicking digital or transmission errors, which can significantly impact depth estimation.
- **Gaussian Blur:** Gaussian blur was applied using kernels of sizes 5x5, 9x9, 13x13, and 17x17. This simulates various degrees of out-of-focus effects caused by camera motion or focal variations, presenting challenges for depth estimation algorithms.
- **Gaussian Noise:** Gaussian noise with a zero-mean and standard deviations of 25, 50, 100, and 200 was added. This simulates sensor noise or image acquisition imperfections, further challenging the depth estimation process.

These augmentations expose the model to a broader spectrum of image quality issues, enhancing its ability to discern and adapt to noise and blur. This preparation is crucial for improving performance on real-world data that often contains such imperfections. Future iterations of this project will explore additional types of degradations to deepen our understanding of how these affect the performance of depth estimation algorithms under various conditions.

3.2 Leveraging Depth Anything for Depth Estimation

In this work, we have chose our goal as depth estimation and now we apply a State-Of-The-Art Algorithm in the field of depth estimation to act as an expert system to enable our model to learn about the different features required for depth estimation. Our proposed approach relies on the Depth Anything [3] framework for predicting depth maps from the augmented and original images. Depth Anything prioritizes a practical approach to robust monocular depth estimation. It achieves this by:

- **Large-Scale Unlabeled Data:** The approach utilizes a data engine to collect and automatically annotate a massive dataset (around 62 million images) of unlabeled data. This significantly broadens data coverage and reduces generalization errors.
- **Data Augmentation for Model Robustness:** Depth Anything leverages data augmentation techniques to create a more challenging learning scenario. This pushes the model to actively seek additional visual knowledge and develop robust feature representations, making it more resilient to variations in image quality.
- **Auxiliary Supervision with Pre-trained Encoders:** An auxiliary supervision mechanism is employed to transfer rich semantic information from pre-trained encoders to the Depth Anything model. This injects valuable knowledge about the underlying scene structure and relationships between image elements, aiding in accurate depth estimation.

The Depth Anything model demonstrates impressive generalization abilities across various datasets and real-world scenarios, making it a suitable foundation for our proposed framework. We leverage its depth estimation capabilities within our model's pipeline to generate the preliminary depth estimation scores used for our model training and evaluation.

3.3 Evaluating Image Quality with BRISQUE: A Case for Goal-Oriented Assessment

Beyond traditional image quality metrics, our work explores the limitations of relying solely on such measures for tasks like depth estimation. To address this, we incorporate the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) algorithm[4]. This section details how BRISQUE is used in our methodology and how its outcomes highlight the need for a goal-oriented image quality assessment (GO-IQA) approach.

3.3.1 BRISQUE for No-Reference Image Quality Assessment

BRISQUE is a no-reference image quality assessment algorithm. It predicts the perceived quality of an image without requiring a pristine reference image for comparison. This characteristic makes it suitable for our scenario where we aim to assess image quality without access to ground truth depth information.

BRISQUE operates by analyzing the spatial statistics of natural scenes within an image. It compares the distribution of pixel intensities in the image to a model of natural scenes with similar distortions. Deviations from this model, such as the presence of noise or blur, indicate a decrease in perceived image quality. BRISQUE outputs a score, where lower scores represent higher perceived quality.

3.3.2 BRISQUE Scores and the Need for GO-IQA

We evaluate both the original and augmented images (noise and blur variations) using BRISQUE. This provides us with scores reflecting the perceived quality of each image variant based on visual appearance.

Here's where the key insight emerges: as discussed in the introduction, we anticipate a correlation between BRISQUE scores and the performance of the Depth Anything model on these images. In simpler terms, images with lower BRISQUE scores (indicating worse visual quality) should translate to poorer depth estimation results (higher MSE).

However, our hypothesis is that this correlation might not always hold true. This potential discrepancy underscores the need for a GO-IQA system. Let's delve deeper:

- **Visual Quality vs. Depth Estimation:** BRISQUE assesses visual quality based on human perception, which might not directly translate to optimal conditions for depth estimation algorithms.
- **Focus on Features Relevant to Depth:** A GO-IQA system, in contrast, would specifically analyze features within the image that are crucial for accurate depth estimation. These features might be visually subtle and not necessarily reflected in BRISQUE scores.

By comparing BRISQUE scores with the actual depth estimation performance (MSE) on the original and augmented images, we aim to demonstrate these potential discrepancies. This analysis, already introduced in the introduction chapter, will further be discussed in the results section, will solidify the argument for a GO-IQA system that goes beyond traditional, visually-oriented metrics.

3.4 CNN Model Development

This project aimed to develop robust convolutional neural network (CNN) models that can predict the MSE score from visual input data of depth estimation images and their associated MSE scores. In this iteration, we have streamlined our approach by focusing on two advanced models based on the ResNet architecture. The models employ the MSE scores generated by the Depth Anything model as labels, facilitating a goal-oriented image quality assessment specifically tailored for depth estimation tasks.

3.4.1 Model Development Using ResNet50

Model Architecture

- **Base Model:** We employed the ResNet50 architecture[12] shown in 3.1, pre-loaded with ImageNet weights, with all layers set to be trainable. This allowed for extensive feature learning relevant to our specific task.
- **Additional Layers:**
  - Global Average Pooling 2D: Condenses the feature maps to a single vector per map, reducing model complexity and computational cost.

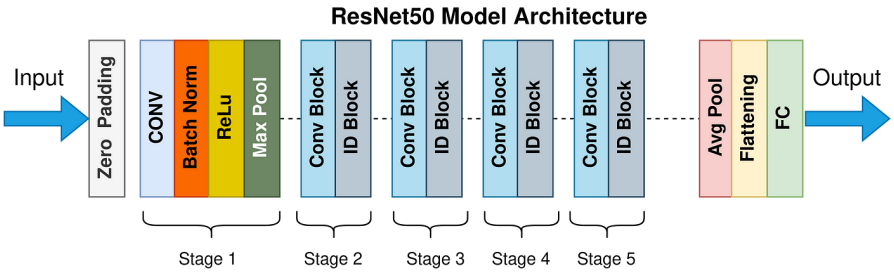

Figure 3.1: ResNet50 Architecture

- Dense Layer: 512 units, 'relu' activation for high-level feature learning.
- Dropout: 0.5 to mitigate overfitting by randomly dropping units during training.
- Dense Layer: 256 units, 'relu' activation to further refine the features.
- Output Layer: 1 unit with 'linear' activation to predict the MSE score directly.

Training Configuration

- Optimizer: Adam Optimizer was used and it is known for its efficient gradient handling and adaptive learning rate.
- Loss Function: Mean Squared Error (MSE), which directly corresponds to our regression goal.
- Early Stopping: Implemented with a patience of 20 epochs and restoration of the best weights to prevent overfitting and ensure model generalization.

3.4.2 Model Development Using ResNet152

Model Architecture

- **Base Model:** ResNet152, initialized with ImageNet weights and customized for depth image assessment by excluding the top layer and adapting input dimensions to (224, 224, 3).
- **Additional Layers:**
  - Global Average Pooling 2D: Focuses on the most essential features by averaging out the spatial dimensions.
  - Dense Layers: Sequentially arranged with 1024, 512, and 256 units, each with 'relu' activation, enhance the network's capacity to learn complex patterns and relationships.
  - Output Layer: Single unit with 'linear' activation, tailored for MSE prediction.

  - Dropout: 0.5 included after the first dense layer to reduce overfitting.

Training Configuration

- Optimizer: Adam Optimizer was used and it is known for its efficient gradient handling and adaptive learning rate.
- Loss and Metric: MSE was used as both the loss function and the performance metric to directly evaluate the prediction accuracy.
- Early Stopping: Configured with a patience of 10 epochs, optimizing training duration and preventing overfitting by restoring the best performing model weights.

These updated models, leveraging the advanced capabilities of the ResNet architectures, are better suited to capture the intricate details required for precise MSE prediction, underpinning the development of more accurate and reliable GO-IQA systems for depth estimation.

3.5 Evaluation Metrics

The following comprehensive set of metrics was used to benchmark the models:

MSE (Mean Squared Error): This provided a central measure of error by quantifying the average squared discrepancy between predicted and ground-truth MSE scores.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3.1}$$

n is the number of samples

y_i is the true ground-truth value for sample i

$\hat{y}_i$ is the predicted value for sample i

MAE (Mean Absolute Error): The average magnitude of prediction errors was captured by the MAE, offering a complementary perspective on model performance.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3.2}$$

n is the number of samples

y_i is the true ground-truth value for sample i

$\hat{y}_i$ is the predicted value for sample i

Pearson Correlation Coefficient: This metric revealed the degree of linear association between the predicted and actual MSE scores as well as the BRISQUE scores, shedding light on the models' ability to capture trends within the data.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{3.3}$$

n is the number of samples

x_i and y_i are the paired data points for sample i

$\bar{x}$ and $\bar{y}$ are the means of the x and y variables, respectively

R2 Score (Coefficient of Determination): This statistical measure represents the proportion of variance in the dependent variable that is predictable from the independent variable(s). It provides insight into how well the observed outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3.4}$$

n is the number of samples

y_i is the true ground-truth value for each sample i

$\hat{y}_i$ is the model's prediction for sample i

$\bar{y}$ is the mean value of all y_i

3.6 Summary

This methodology outlined a comprehensive approach to investigating the prediction of image quality scores for depth estimation. The foundation was built upon a carefully prepared and augmented dataset, ensuring a robust basis for model development. The Depth Anything model was strategically employed to generate ground-truth scores. To assess image quality, the goal-oriented BRISQUE metric was applied, aligning evaluation with the task of depth estimation.

In this iteration, we focused on a streamlined suite of advanced CNN architectures based on the ResNet50 and ResNet152 models. This refinement was aimed at harnessing their enhanced capabilities for deep feature extraction and learning, which are critical for accurately predicting image quality in the context of depth estimation. These models were meticulously trained with Adam optimization, incorporating strategies like fully trainable layers and extensive fine-tuning. The training process also featured early stopping to ensure optimal model performance without overfitting, and model efficacy was rigorously evaluated using metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), Pearson Correlation Coefficient and The R2 Score.

This methodology provides a strong framework for investigating the relationship between CNN-predicted image quality scores and their efficacy in depth estimation. The insights gained from employing sophisticated ResNet architectures will be further discussed in the Results section.

CHAPTER 4

Results

This section delves into the key findings and quantitative outcomes of the investigation into predicting image quality scores for depth estimation. Recapping from the Methodology section, we explored the Depth Anything model on the NYU Depth V2 dataset, which was strategically augmented with blur and noise to enhance model robustness. The NR-IQA BRISQUE metric was employed to assess image quality, directly aligning with the objective of depth estimation.

In this iteration, we focused on two advanced CNN architectures based on the ResNet50 and ResNet152 models, both configured for extensive learning of depth-related features. Each model underwent rigorous training with Adam optimization and early stopping to achieve optimal performance. Their effectiveness was comprehensively measured using Mean Squared Error (MSE), Mean Absolute Error (MAE), and Pearson Correlation Coefficient.

This section unveils the following key aspects:

- The impact of blur and noise variations on image quality scores derived from the Depth Anything model.
- BRISQUE scores and their correlation with the introduced image quality variations.
- The training performance of the ResNet-based models, including loss curves and convergence behaviors, indicating significant improvements in learning depth-related features.
- The evaluation results of the trained models using MSE, MAE, and Pearson Correlation Coefficient, highlighting their enhanced predictive accuracy and reliability in depth estimation tasks.

By dissecting these findings, we aim to establish a refined understanding of how CNN-predicted image quality scores, derived from sophisticated ResNet architectures, correlate with the success of depth estimation tasks on the NYU Depth V2 dataset with introduced variations.

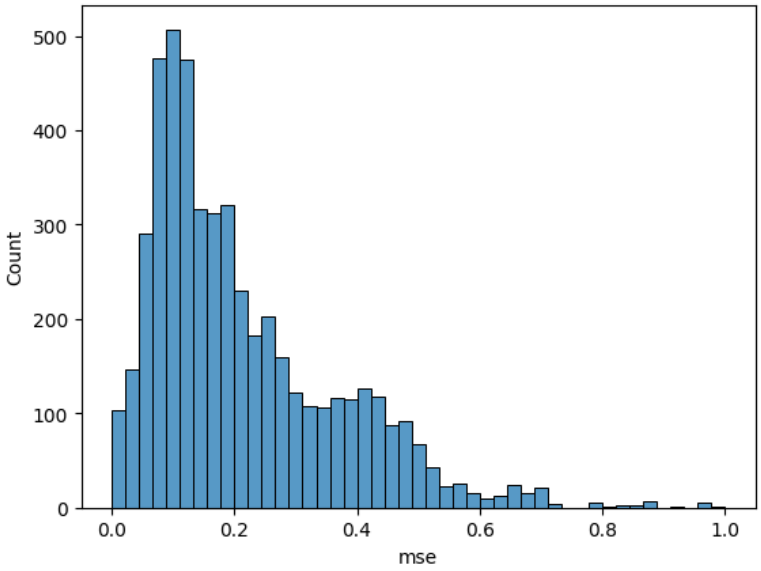

Figure 4.1: Min-Max Normalized Distribution of MSE Values

4.1 Depth Estimation Using Depth Anything

We applied the Depth Anything model to the NYU-Depth-v2 dataset to generate depth maps and assess the impact of various image quality degradations. Figure 4.1 displays the distribution of MSE values from the depth anything model after Min-Max Normalization. Figures 1.2, 1.3, and 1.4 illustrate a visual comparison of the ground truth depth map (4.2) alongside Depth Anything predictions for the original image (4.3), a blurred version (4.4), and two noisy versions (4.5).

Key Observations:

- **Depth Anything Robustness to Blur:** The blurred image shows only minor deviations in the predicted depth map compared to the original, indicating a degree of robustness to blur in the Depth Anything model.
- **Noise Sensitivity:** The images with salt and pepper noise, as well as Gaussian noise, exhibit significantly deteriorated depth maps with a loss of detail and precision, underlining the impact of noise on Depth Anything's performance.

Table 4.1 shows the relationship between image quality degradations and the accuracy of depth masks predicted by Depth Anything, using min-max normalized median scores:

Key Insights

Figure 4.2: Ground Truth Of The Image

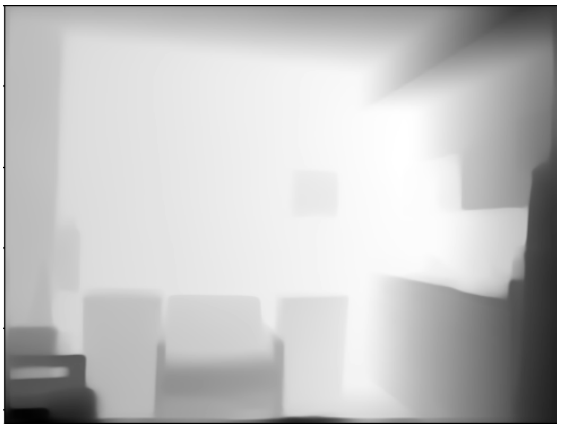

Figure 4.3: Depth Anything Prediction

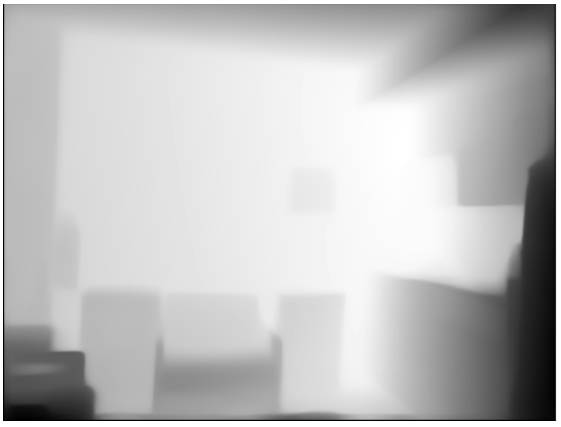

Figure 4.4: Depth Anything Prediction For Blurred Image

Figure 4.5: Depth Anything Predictions For Noisy Images

Image Type	Normal Images	Blurred Images	Salt and Pepper Noise	Gaussian Noise
Median Scores	0.170	0.165	0.171	0.179

Table 4.1: Min-Max Normalized Median Scores of MSE of Predicted Depth Masks

- **Noise Sensitivity:** The increased median scores for images with salt and pepper and Gaussian noise reaffirm the vulnerability of the Depth Anything model to noise, highlighting the importance of noise reduction techniques or the development of models that are more robust to these types of distortions.
- **Unexpected Blur Resilience:** The slight improvement in median scores for blurred images invites further investigation. This suggests that the blurring process may suppress some image features that otherwise interfere with accurate depth prediction.

4.2 BRISQUE Score Analysis

The application of the BRISQUE No-Reference Image Quality Assessment (NR-IQA) metric reveals distinct trends in perceived image quality across the normal, blurred, and two types of noisy image sets. Table 4.2 summarizes these findings:

- **Normal Images:** The baseline BRISQUE score of 35.07 falls within a range typically associated with images of reasonable quality.
- **Blurred Images:** A BRISQUE score of 59.70 reflects the expected reduction in image quality due to blurring, indicating moderate degradation.
- **Noisy Images:**
  - Salt and Pepper Noise: A BRISQUE score of 88.83 confirms quality degradation more severe than blurring but less than expected for other noise types.

Image Type	Normal Images	Blurred Images	Salt and Pepper Noise	Gaussian Noise
BRISQUE Scores	35.07	59.70	88.83	71.93

Table 4.2: BRISQUE Scores For Images

- Gaussian Noise: The BRISQUE score of 71.93, although lower than salt and pepper noise, still signifies significant quality reduction, indicative of substantial image degradation.

4.2.1 Connecting BRISQUE to Depth Estimation

These BRISQUE scores provide a quantitative basis for understanding the impact of image quality on Depth Anything predictions:

Noise Correlation: The elevated BRISQUE scores for noisy images correlates with their respective increased median scores from the Depth Anything predictions. This reinforces the notion that noise significantly hinders the Depth Anything model's accuracy. However These scores are taken on an aggregate level and the actual scenario of Brisque score's Pearson Correlation Coefficient is quite different as explained in further sections

Blur Intrigue: Despite the higher BRISQUE score for blurred images, their similar median scores to normal images suggest an intriguing resilience in depth map predictions against blur. This points to the potential need for refining image quality metrics like GO-IQA, as discussed previously, to better capture the nuances affecting depth estimation.

4.3 Results of The CNN Models

Table 4.3 presents the performance metrics of the evaluated CNN models based on the ResNet architectures. With the focus on establishing a threshold for the GO-IQA framework, the Pearson Correlation Coefficient (PCC) and R-squared (R2) values emerge as critical metrics, providing insights into the linear relationship and variance explanation of the models respectively.

Model	MSE	MAE	PCC	R2
ResNet50	0.13	0.029	0.89	0.60
ResNet152	0.15	0.041	0.85	0.72

Table 4.3: MSE, MAE, PCC, and R2 scores of CNN Models

4.3.1 PCC-Focused and R2 Analysis

- **Threshold Potential:** ResNet50 demonstrates the strongest PCC (0.89) and a solid R2 value (0.60), indicating a robust linear relationship and substantial explanation of variance with the ground-truth values. This supports its use in setting a quality threshold within the GO-IQA system.
- **ResNet152 Performance:** While ResNet152 shows a slightly lower PCC (0.85), it excels with a higher R2 value (0.72), suggesting better overall fit and predictive accuracy, albeit with slightly less linearity compared to ResNet50.

MSE (Mean Squared Error) and MAE (Mean Absolute Error):

- The ResNet50 model shows slightly better performance in terms of MSE (0.13) compared to ResNet152 (0.15), suggesting smaller average squared deviations from the ground-truth MSE scores.
- In terms of MAE, ResNet50 also performs better (0.029) than ResNet152 (0.041), indicating smaller average magnitude of errors.

4.3.2 Overall Trends

Best PCC and R2 Scores: The ResNet50 model not only provides the best PCC score but also shows a commendable R2, indicating it is particularly effective in capturing and interpreting the relationship between the features in the images and the MSE scores predicted by the Depth Anything algorithm.

Balanced Performance: While ResNet152 does not top the PCC charts, its higher R2 value indicates it might be more suited for applications where understanding the variance is more critical than the linear correlation alone.

Model Complexity vs. Performance: The comparison highlights that while both models are complex, the nuances in their architectural differences and training might influence their specific applicability and performance in depth estimation tasks.

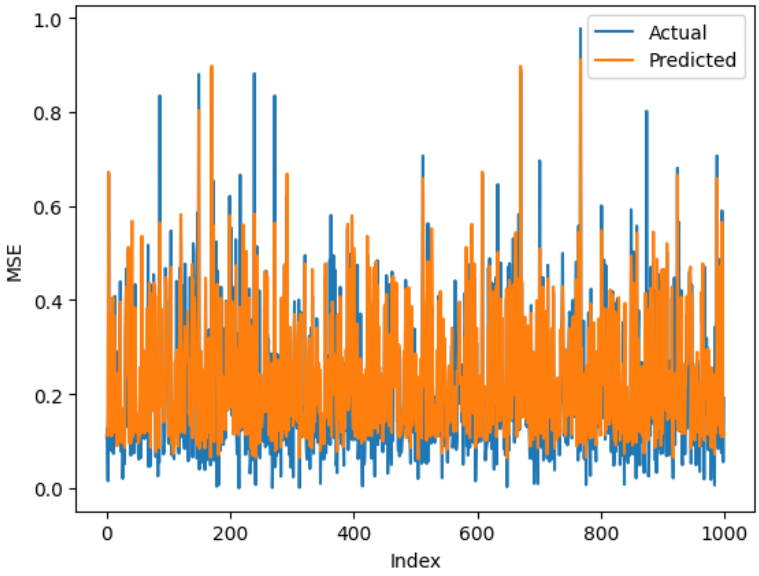

Figure 4.6: Expected Vs Predicted Labels Scatter Plot

4.3.3 Expected Vs Predicted Labels Scatter Plot

4.6 displays the scatter plot between the expected values of MSE given and image and the predicted values for our ResNet150 Model on the validation data. From this scatter plot and the zoomed scatter plot show in 4.7 we can gather that our model is performing satisfactorily in capturing the variances in the MSE-Depth scores across different kinds of images having different types of degradations.

4.4 Correlation Analysis in Results

The correlation analysis plays a pivotal role in assessing the effectiveness of the GO-IQA framework. The Pearson Correlation Coefficients (PCC) between different metrics provide insights into how well the assessed image quality aligns with actual performance metrics. The table below shows the PCC between BRISQUE scores and MSE, and the output scores from the ResNet50 model against MSE.

Metric	Pearson Correlation Coefficient
BRISQUE vs. MSE	-0.035
ResNet50 vs. MSE	0.89

Table 4.4: Correlation between image quality scores and MSE

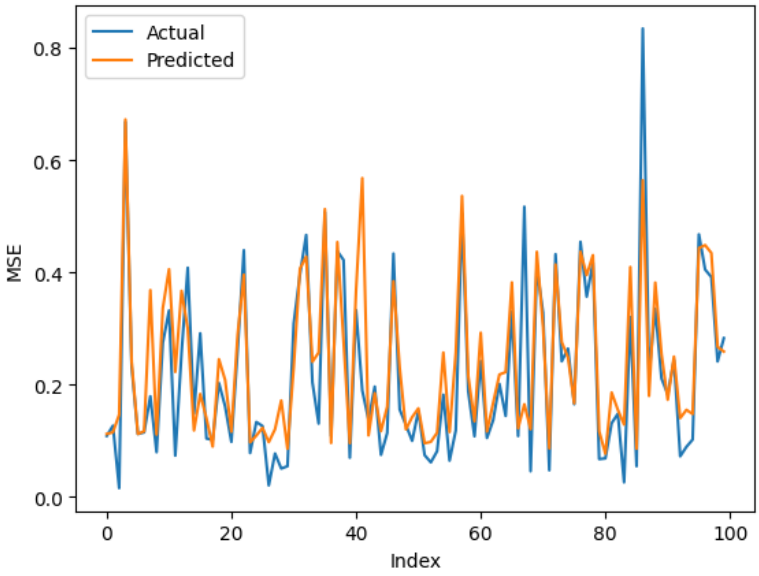

Figure 4.7: Expected Vs Predicted Labels Scatter Plot Zoomed

The BRISQUE metric, traditionally used for generic image quality assessment, shows a negligible correlation with MSE (*PCC = -0.035*). This suggests that standard IQA metrics might not adequately reflect the nuances required for specific applications like depth estimation. Conversely, the high correlation (*PCC = 0.89*) between ResNet50's output and MSE underscores the potential of GO-IQA systems. This robust correlation demonstrates that tailored models like ResNet50, which are aligned with specific end goals of the imaging process, are far more effective in predicting practical image quality metrics that influence system performance.

This analysis highlights the critical need for developing GO-IQA frameworks that are closely integrated with the specific requirements of the application, ensuring that the quality assessment is not only theoretical but practically relevant to the system's operational success.

4.5 Efficiency Analysis of GO-IQA System

The implementation of a GO-IQA system using ResNet architectures prior to depth estimation demonstrates significant efficiency improvements in terms of processing speed and resource utilization. The following table outlines the running times per image and the number of parameters for each system:

System	Running Time (per image)	No. of Parameters
Depth Anything SOTA	13 ms	335 Million
GO-IQA ResNet50	0.7 ms	24 Million
GO-IQA ResNet150	2.1 ms	61 Million

Table 4.5: Comparison of running times and parameter counts for depth estimation and GO-IQA systems

The results clearly show that the GO-IQA models, particularly ResNet50, are markedly faster and leaner in terms of computational requirements compared to the state-of-the-art (SOTA) Depth Anything model. With a running time of only 0.7 milliseconds per image and requiring far fewer parameters (24 million compared to 335 million), ResNet50 in the GO-IQA framework provides a swift and efficient preliminary quality assessment. This allows for rapid screening of images to exclude those of insufficient quality before engaging the more resource-intensive depth estimation process.

Furthermore, even the more complex GO-IQA ResNet150 model, which provides additional robustness and potentially higher accuracy at a slightly increased computational cost (2.1 milliseconds per image and 61 million parameters), still represents a substantial efficiency gain over the traditional Depth Anything approach.

This efficiency is critical in real-world applications where processing speed and resource management are key constraints. By incorporating a GO-IQA system to filter out unsuitable samples, organizations can optimize their processing pipelines, ensuring that only images with sufficient quality are subjected to depth estimation. This not only speeds up the overall process but also reduces computational waste, leading to faster, more cost-effective operations.

4.6 Summary

This chapter presented the results of our exploration into predicting the MSE scores of depth estimation images using advanced CNN models within the GO-IQA framework. The findings highlight the efficacy and efficiency of the approach, showcasing substantial improvements over traditional methods.

Model Performance:

The ResNet-based models, particularly ResNet50 and ResNet152, demonstrated outstanding performance, with high Pearson Correlation Coefficients and R-squared values, indicating their robust ability to predict MSE scores accurately. This underscores the potential of these models to serve as effective tools in GO-IQA systems for depth estimation tasks.

4.6.1 Efficiency of GO-IQA:

The GO-IQA models drastically reduce the computational load and processing time compared to traditional depth estimation systems. For instance, ResNet50 in the GO-IQA setup runs over 18 times faster than the state-of-the-art Depth Anything model and requires significantly fewer parameters. This efficiency enables quicker assessments and potential cost savings in real-world applications by filtering out unsuitable images before depth estimation.

4.6.2 Opportunities for Refinement

- **Algorithmic Enhancement:** Although the current models perform well, there is always room for improvement, particularly in reducing the slight lag in performance seen with ResNet152 in terms of MSE and MAE scores compared to ResNet50. Integrating different state-of-the-art algorithms to refine ground truth MSE estimations can enhance the robustness and reliability of our GO-IQA system, ensuring that it adapts to varying complexities in image quality assessment.
- **Hybrid Architectures:** Further exploration into hybrid models that combine the strengths of different CNN architectures could lead to even more robust GO-IQA systems. Such architectures might leverage the unique capabilities of each model to better capture and analyze the nuances of image quality relevant to depth estimation.
- **Expanded Validation:** Extending validation on larger and more varied datasets could help confirm the models' effectiveness across different scenarios and conditions. This broader testing will also allow us to evaluate the practicality of our refined MSE estimation methods under diverse operational environments.

4.6.3 Future Work

The next phase of this research will focus on further refining the efficiency and accuracy of the GO-IQA framework. Efforts will be directed towards enhancing algorithmic performance, exploring hybrid architectures, and expanding dataset diversity. This will ensure that the GO-IQA system remains adaptable and effective for real-world applications in depth estimation, ultimately contributing to the development of an optimized and highly accurate image quality assessment system tailored to specific operational needs.

CHAPTER 5

Conclusion

This report explored the realm of goal-oriented image quality assessment (GO-IQA), developing a tailored framework for depth estimation. Our initial steps included a review of existing image quality assessment methods, underscoring the necessity for an approach directly aligned with specific operational goals. To address this, the NYU-Depth-v2 dataset was enhanced with blur and noise to better simulate the challenges found in real-world scenarios. The BRISQUE metric was employed to effectively measure the impact of these degradations on image quality.

Utilizing the Depth Anything algorithm, we generated MSE scores to act as a specific image quality metric for depth estimation tasks. Inspired by a comprehensive review of related literature, we then crafted various CNN architectures designed to predict these MSE scores directly from images. These models, particularly those based on advanced ResNet architectures, showed promising results and also highlighted the potential advantages of using GO-IQA algorithms along with Depth Estimation and in other situations for improved time efficiency.

5.1 Key Findings

- **GO-IQA for Depth Estimation:** The study confirmed the viability of using CNN-predicted MSE scores as a measure of image quality specifically for depth estimation, illustrating the effectiveness of the GO-IQA approach.
- **Depth Anything and BRISQUE:** Utilizing Depth Anything to generate depth-specific scores alongside BRISQUE for traditional quality assessment provided a comprehensive toolset for analyzing how image quality affects depth estimation performance.
- **Model Potential and Refinement:** The implementation of ResNet50 and ResNet152 models revealed strong potential, though there is also scope for further improvement.

5.2 Future Directions

- **Robust Datasets Ground Truths:** Developing larger, more diverse datasets that include a broader spectrum of image degradations will be critical for training more resilient models. Employing advanced data augmentation techniques will play a key role in this expansion. Integrating diverse state-of-the-art algorithms for more accurate ground truth MSE estimation will further enhance the robustness of the models, ensuring they are well-tuned to real-world variability in image quality.
- **Architectural Innovation:** Exploring state-of-the-art CNN architectures that incorporate elements such as attention mechanisms might enable models to capture more complex image features and relationships effectively. This can also include hybrid architectures that blend multiple advanced techniques to optimize performance.
- **Thresholding Strategies:** Implementing Pearson Correlation Coefficient (PCC)-based thresholding strategies, driven by in-depth analysis of model performance and dataset characteristics, will be vital for establishing dependable image quality thresholds for depth estimation tasks. This includes refining thresholding approaches as new data and models evolve, ensuring optimal decision-making in GO-IQA applications.

This research has established a solid foundation for developing a specialized GO-IQA system for depth estimation. The insights garnered here will guide further enhancements, aiming to boost the precision and reliability of applications dependent on accurate depth information.

REFERENCES

[1] S. Kiruthika and V. Masilamani, "Goal oriented image quality assessment," *IET Image Processing*, vol. 16, no. 4, pp. 1054–1066, 2021.

[2] P. K. Nathan Silberman, Derek Hoiem and R. Fergus, "Indoor segmentation and support inference from rgbd images," in *ECCV*, 2012.

[3] L. Yang, B. Kang, Z. Huang, X. Xu, J. Feng, and H. Zhao, "Depth anything: Unleashing the power of large-scale unlabeled data," 2024.

[4] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on image processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[5] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1733–1740.

[6] F. Liu, C. Shen, and G. Lin, "Deep convolutional neural fields for depth estimation from a single image," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 5162–5170.

[7] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 1098–1105.

[8] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Transactions on image processing*, vol. 27, no. 1, pp. 206–219, 2017.

[9] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, "Blind image quality assessment using a deep bilinear convolutional neural network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 1, pp. 36–47, 2018.

[10] W. Zhang, K. Ma, G. Zhai, and X. Yang, "Uncertainty-aware blind image quality assessment in the laboratory and wild," *IEEE Transactions on Image Processing*, vol. 30, pp. 3474–3486, 2021.

[11] L. Ge, H. Liang, J. Yuan, and D. Thalmann, "3d convolutional neural networks for efficient and robust hand pose estimation from single depth images," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1991–2000.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.

Weekly Review Report

Insert the image of the Weekly review report

Plagiarism Report

Insert the 1st page of the plagiarism report (containing the similarity index). Make sure that it is duly signed by you as well as your supervisor.