No-Reference Underwater Image Quality Assessment

CEA Internship Midterm Report

Submitted by
Sriram Dhanasekaran
(Roll No. 231035)

Department of Electrical Engineering
Indian Institute of Technology Kanpur

Under the Guidance of

External Supervisor

Prof. Balasubramanyam Appina
Assistant Professor
Department of Electrical Engineering
Indian Institute of Technology Indore

IITK Mentor

Prof. Tushar Sandhan
Assistant Professor
Department of Electrical Engineering
Indian Institute of Technology Kanpur

January 2026

Acknowledgement

I would like to express my sincere gratitude to **Prof. Balasubramanyam Appina**, Assistant Professor, Department of Electrical Engineering, Indian Institute of Technology Indore, for granting me the opportunity to undertake this research internship and for his support in facilitating this academic engagement.

I am deeply thankful to **Prof. Ajay Kumar Reddy Poreddy**, Assistant Professor, Department of Computer Science and Engineering, Sri Sivasubramaniya Nadar College of Engineering, Chennai, for recommending me for this internship and for his continuous guidance, mentorship, and encouragement throughout the course of this work.

I also wish to sincerely thank **Prof. Tushar Sandhan**, Assistant Professor, Department of Electrical Engineering, Indian Institute of Technology Kanpur, for kindly agreeing to serve as my IIT Kanpur supervisor for the Credits of External Activities (CEA), and for his valuable academic guidance and supervision.

Furthermore, I extend my sincere thanks to the **Departmental Undergraduate Committee (DUGC), Department of Electrical Engineering**, and the **Senate Undergraduate Committee (SUGC)** for approving this internship under the CEA framework and for granting approval for **18 credits**, which made this academic and research experience possible.

Finally, I am grateful to all those who directly or indirectly supported me during the course of this internship and contributed to the successful completion of this work.

Contents

List of Figures

List of Attachments

Task 1 – UIQA Fundamentals
Foundational concepts of underwater IQA, FR/RR/NR paradigms, and perceptual limitations of MSE.

Task 2 – Feature Logic and Metrics
Structural and color feature design, PLCC/SRCC interpretation, and evaluation rationale.

MATLAB: Structural & Color Feature Extraction
Implementation of gradient-domain, LBP, entropy, and HSV features.

MATLAB: Luminance Gabor Feature Generator
Multi-scale Gabor filter bank for frequency-domain perceptual modeling.

Python: Super Stack Ensemble Training
Base regressors, stacking meta-learner, and evaluation pipeline.

Complete Code & Results Archive
All scripts, plots, and spreadsheets.

1. Introduction

Underwater imaging environments introduce severe visual degradations due to wavelength-dependent light absorption and scattering phenomena. Longer wavelengths (red) attenuate rapidly, while shorter wavelengths dominate at increasing depths, resulting in strong color imbalance, contrast loss, and reduced visibility [1].

In practical deployments such as Autonomous Underwater Vehicles (AUVs), pristine reference images are unavailable, making full-reference assessment infeasible. Consequently, No-Reference (NR) methods are essential [2]. However, traditional NR-IQA models based on natural scene statistics often fail to generalize to the underwater domain [3].

To address this, this work adopts a feature-based NR-UIQA framework motivated by perceptual considerations. Handcrafted features are extracted to represent color, structural, and mid-frequency texture information, and these features are combined using an ensemble learning approach to obtain improved robustness and interpretability [4].

Objective

The primary objective of this project is to develop and validate a robust no-reference underwater image quality assessment (NR-UIQA) framework capable of predicting perceptual image quality in the absence of a pristine reference image.

At the current stage of the project, the following milestones have been **achieved**:

1. **Framework Development:** Design and implementation of a high-capacity **Super Stack Ensemble** by integrating multiple complementary regression models, resulting in strong and stable predictive performance across diverse underwater image degradations.
2. **Perceptual Validation:** Completion of systematic post-hoc analyses involving color, structural, and frequency-domain features to interpret learned model behavior and assess its consistency with human visual perception.

These outcomes establish a solid foundation for subsequent extensions and refinements, which will be explored in the later phases of the project.

2. Background and Preliminary Studies

Before developing the proposed NR-UIQA framework, a set of focused preliminary studies was carried out to build the necessary conceptual understanding and to clarify evaluation requirements. These studies provided the foundational insights that informed the methodological choices adopted in the later stages of this work.

Task 1: Fundamentals of Underwater Image Quality Assessment

This study outlines the motivation for Underwater Image Quality Assessment (UIQA)

by discussing the limitations of conventional full-reference (FR) and reduced-reference (RR) methods when applied to practical underwater imaging scenarios. It categorizes existing image quality assessment approaches into FR, RR, and no-reference (NR) frameworks, and highlights the limitations of pixel-wise distortion metrics such as Mean Squared Error (MSE) in capturing perceptual quality degradation caused by wavelength-dependent absorption and scattering effects.
Task 1 – UIQA Fundamentals

Task 2: Feature Design Logic and Evaluation Metrics
This task defines the evaluation protocol used throughout the project. It explains the perceptual significance of correlation-based performance metrics, namely the Pearson Linear Correlation Coefficient (PLCC) and the Spearman Rank Correlation Coefficient (SRCC), and justifies their use over purely error-based measures. In addition, it describes the extraction logic for handcrafted Structural (F_S) and Color (F_C) features, with emphasis on gradient-domain statistics and HSV color-space representations suited to common underwater image distortions.
Task 2 – Feature Logic and Metrics

Recent studies on non-natural image quality assessment further support this design choice. Chen *et al.* [5] demonstrated that natural scene statistics are ineffective for cartoon images and showed that perceptually motivated gradient-domain structural features combined with HSV-based color moments and entropy provide significantly improved quality prediction. Their findings reinforce the premise that structure and color cues form a robust and transferable foundation for no-reference quality assessment beyond natural image domains.

The understanding gained from these preliminary studies influenced the design of the feature extraction framework, the selection of evaluation objectives, and the ensemble-based modeling strategy presented in the subsequent sections.

3. Methodology

3.1 Problem Formulation

The problem is formulated as a supervised regression task. Given a feature vector $\mathbf{x} \in \mathbb{R}^d$ extracted from an image I, the goal is to predict the corresponding Mean Opinion Score (MOS) y:

$$y = f(\mathbf{x}) + \epsilon$$

where $f(\cdot)$ denotes the learned regression function and ϵ represents modeling noise.

3.2 Structural Feature Extraction (F_S)

Structural degradation is analyzed in the gradient domain. The image is filtered using Sobel operators to compute the horizontal (G_x) and vertical (G_y) gradients. The gradient magnitude is then obtained as $G(x, y) = \sqrt{G_x^2 + G_y^2}$.

To describe local texture variations, **Local Binary Patterns (LBP)** [6] are applied to the gradient magnitude map:

$$\mathrm{LBP}(x, y) = \sum_{t=0}^{n-1} \psi(G_t - G_c) \cdot 2^t, \quad \psi(z) = \mathbb{I}(z \geq 0)$$

A weighted histogram of gradient magnitudes conditioned on LBP codes is used to obtain nine texture-related features. In addition, edge sharpness is quantified using the **Enhancement Measure of Entropy (EME)** [7]:

$$\mathrm{EME} = \frac{2}{k_1 k_2} \sum_{l=1}^{k_1} \sum_{k=1}^{k_2} \log \left(\frac{I_{\max,k,l}}{I_{\min,k,l}} \right)$$

The complete MATLAB implementation for structural and color feature extraction is provided in MATLAB Feature Extraction Code.

3.3 Color Feature Extraction (F_C)

Color-related distortions are analyzed in the HSV color space due to its perceptual relevance. The following features are extracted:

- **Color Moments:** Mean, standard deviation, and skewness of the H, S, and V channels, resulting in nine features.
- **Color Entropy:** To capture information loss in color channels, the entropy of Mean Subtracted Contrast Normalized (MSCN) coefficients is computed [8]:

$$I_C(x, y) = \frac{I(x, y) - \mu(x, y)}{\sigma(x, y) + 1}$$

3.4 Gabor-Based (Frequency-Domain) Features ($F_{\mathbf{Gabor}}$)

To capture mid-frequency perceptual distortions such as blur and scattering, luminance-based Gabor filters are applied at multiple scales and orientations:

$$g(x, y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2} \right) \cos\left(2\pi \frac{x'}{\lambda} \right)$$

Statistical descriptors are extracted from the resulting response maps to represent texture variations and frequency-domain degradation. The luminance Gabor filter bank and optimized feature generation procedure are implemented in MATLAB Gabor Feature Generator.

3.5 Machine Learning Strategy

3.5.1 Base Regressors (Level-0)

Multiple regression models are trained independently to capture different data characteristics:

- **Kernel-based:** Support Vector Regression (SVR) with RBF kernel [9].
- **Ensemble Trees:** Random Forest [10], XGBoost [11], LightGBM [12], and CatBoost [13].

3.5.2 Super Stack Ensemble (Level-1)

The Super Stack Ensemble combines predictions from selected high-performing base models (CatBoost, XGBoost, and SVR) using a linear meta-learner. This approach helps reduce prediction variance while retaining the ability to model non-linear relationships. The complete training and evaluation pipeline for the Super Stack Ensemble is provided in Python Super Stack Training Script.

4. Experimental Setup

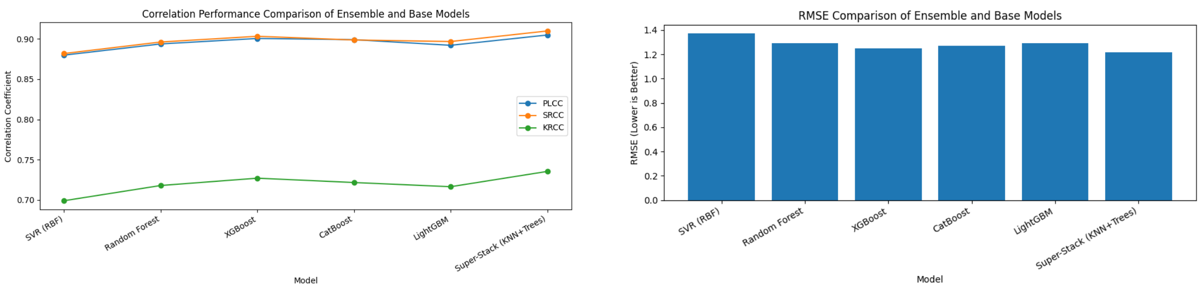
- **Datasets:** Experiments are performed on two publicly available underwater image quality datasets: **UID2021** [14], which contains 960 images with a wide range of underwater distortions, and **SAUD** [15], consisting of underwater images annotated with subjective quality scores.
- **Evaluation Protocols:**
  - *Intra-dataset evaluation:* For evaluations conducted within the same dataset (UID → UID and SAUD → SAUD), the data is split into **70%** for training, **10%** for validation, and **20%** for testing.
  - *Inter-dataset evaluation:* For cross-dataset experiments (UID → SAUD and SAUD → UID), a **90/10** split is adopted, where the source dataset is used for training and validation, and the target dataset is used exclusively for testing.
- **Repetition Strategy:** To improve the reliability of the reported results, base regression models are evaluated over **500 independent random splits**. Due to higher computational cost, the Super Stack Ensemble is evaluated over **20 independent runs**.
- **Evaluation Metrics:** Performance is assessed using the **Pearson Linear Correlation Coefficient (PLCC)** to measure prediction accuracy, the **Spearman Rank Correlation Coefficient (SRCC)** to evaluate monotonic agreement with subjective scores, and the **Root Mean Square Error (RMSE)** to quantify prediction error [2].

5. Results and Analysis

All numerical results, plots, and intermediate experiment logs referenced in this section are provided in the Complete Code and Results Archive.

5.1 Super Stack Ensemble Performance

As shown in Fig. 1, the Super Stack Ensemble achieves the highest correlation scores across evaluation metrics. Boosting-based models consistently outperform the traditional SVR baseline, and their aggregation through stacking further reduces RMSE.

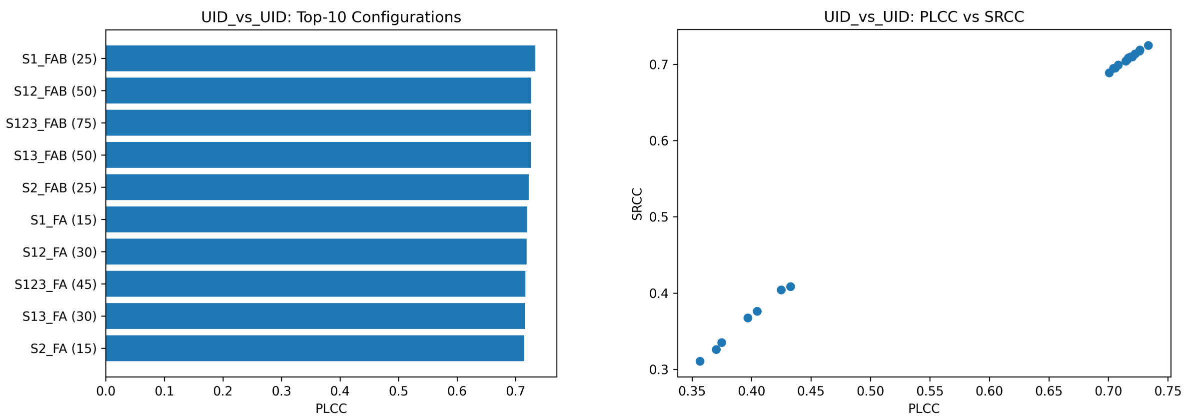


(a) PLCC/SRCC Comparison (b) RMSE Comparison

Figure 1: Performance comparison between base models and the Super Stack Ensemble.

5.2 Feature Configuration Analysis

Using a subsequent feature analysis on the UID2021 dataset (Fig. 2), it is observed that joint **Color + Structure** feature configurations consistently outperform single-domain features. This suggests that underwater image quality perception depends on multiple complementary cues rather than a single distortion factor.

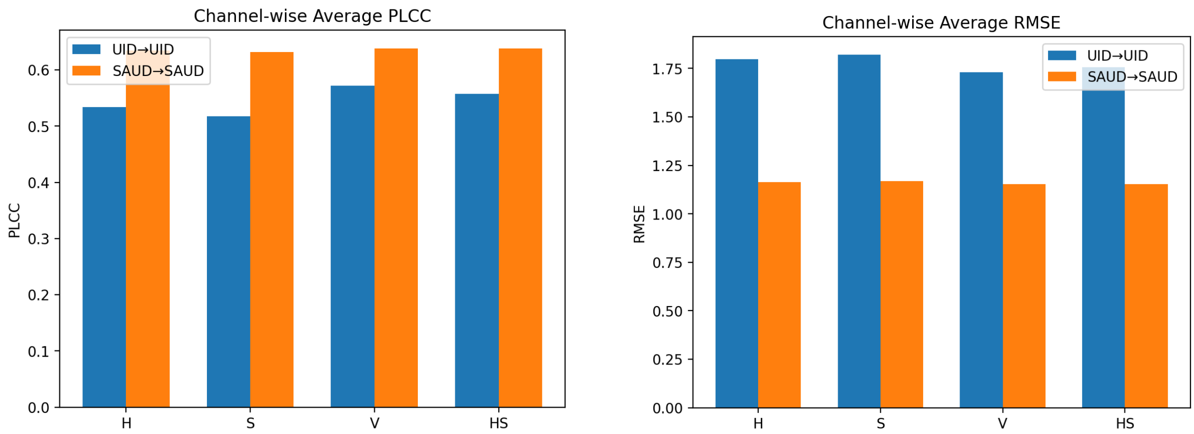


(a) Feature Ranking (UID) (b) Accuracy vs. Monotonicity

Figure 2: Intra-dataset feature analysis on UID2021.

5.3 Channel-wise Importance

The multi-scale HSV analysis shown in Fig. 3 indicates that the **Value (V) channel** and combined **HS** representations contribute most strongly to prediction performance. This observation is consistent with the characteristics of underwater imaging, where luminance degradation due to scattering and contrast loss is often more dominant than absolute hue variations.

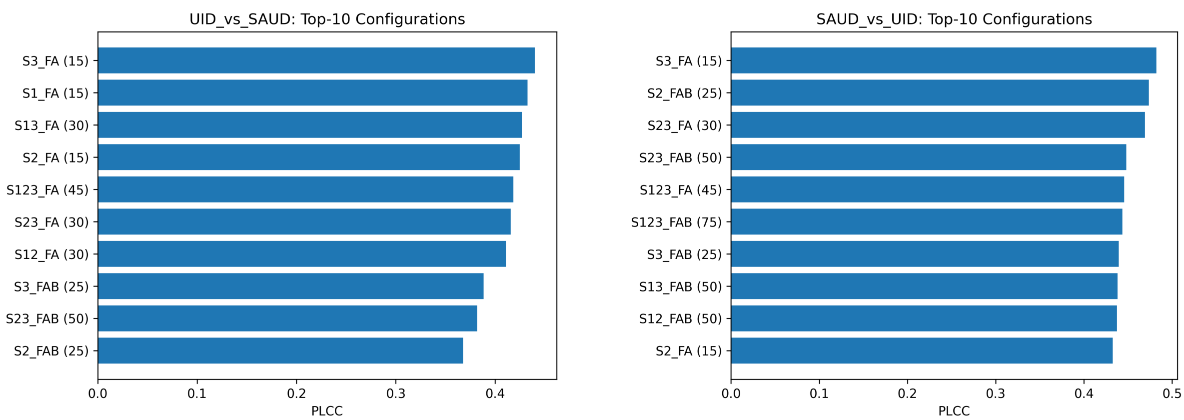


(a) Channel-wise PLCC (b) Channel-wise RMSE

Figure 3: Effect of individual color channels on prediction performance.

5.4 Cross-Dataset Generalization

Cross-dataset evaluation (UID → SAUD) results in a noticeable performance drop, as expected due to domain differences (Fig. 4). However, models using joint feature representations maintain higher monotonic correlation compared to single-feature baselines, indicating improved generalization behavior.



(a) UID → SAUD Generalization (b) SAUD → UID Generalization

Figure 4: Cross-dataset evaluation illustrating generalization trends.

5.5 Luminance Gabor Feature Evaluation: Cross-Domain Validation

While the main NR-UIQA framework in this work is based on perceptually motivated color and structural features designed for underwater imagery, an auxiliary study is conducted to examine whether the observed performance trends extend to alternative feature representations. Specifically, frequency-domain luminance features derived from Gabor filter responses are evaluated.

These features are motivated by prior studies in perceptual quality assessment for medical and laparoscopic imaging [16], where luminance-driven frequency responses have shown strong alignment with human perception. This experiment, therefore, investigates whether such representations can provide meaningful cues for underwater image quality assessment.

5.5.1 Motivation and Cross-Domain Perspective

Although underwater image degradation and laparoscopic video degradation originate from different physical processes, both domains commonly exhibit reductions in contrast, edge clarity, and mid-frequency texture content. Gabor filters, which model localized spatial-frequency and orientation-selective responses, offer a suitable framework for examining these shared perceptual characteristics.

The following questions are explored:

- Can luminance-based frequency features provide useful quality predictions for underwater images?
- Are the dominance of luminance cues and feature-combination trends consistent across different feature families?

5.5.2 Gabor Feature Construction

A bank of Gabor filters is applied to the luminance channel of each image across multiple scales and orientations. Statistical descriptors are extracted from the resulting response maps to form compact feature representations.

The following configurations are evaluated:

- **Single-feature configurations:** Individual Gabor responses.
- **Two-feature concatenations:** Pairwise combinations of distinct responses.

In total, 66 distinct feature configurations are analyzed.

5.5.3 Experimental Protocol

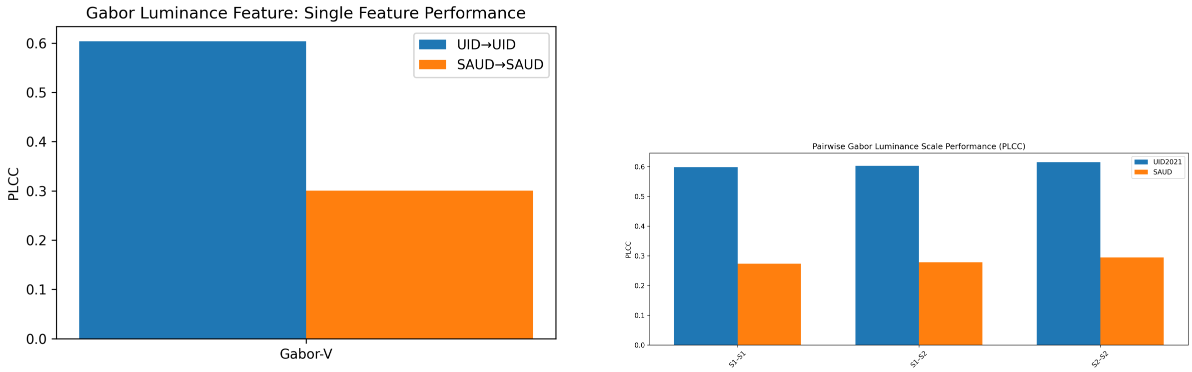
All Gabor-based experiments are conducted using a Support Vector Regression (SVR) baseline to isolate the effect of the feature representation. Evaluations are performed independently on:

- UID2021 (intra-dataset evaluation)
- SAUD (intra-dataset evaluation)

Performance is measured using PLCC, SRCC, and RMSE, consistent with earlier experiments.

5.5.4 Results: Single vs. Concatenated Gabor Features

Figure 5 compares the performance of individual and concatenated Gabor-based luminance features evaluated on the UID2021 and SAUD datasets. Single Gabor responses achieve moderate correlation with subjective quality scores, indicating that localized frequency information captures certain perceptual degradation characteristics. However, pairwise Gabor concatenations consistently improve PLCC and SRCC, while reducing RMSE, demonstrates that complementary frequency responses provide additional perceptual cues.



(a) Single Gabor-based luminance features.

(b) Pairwise Gabor feature concatenations.

Figure 5: Comparison of single and concatenated Gabor-based luminance features on UID2021 and SAUD datasets.

Overall, these results confirm that while individual Gabor features are informative, Combining responses across multiple scales and orientations yields more robust quality prediction. This trend is consistent with earlier findings in this report, where joint feature representations consistently outperformed single-domain configurations.

5.5.5 Discussion and Relation to Main Framework

Despite their cross-domain origin, Gabor-based luminance features show stable behavior on underwater datasets. Importantly, the observed trends align with earlier findings:

- Luminance-related cues play a dominant role in quality prediction.
- Single-feature representations are insufficient for complex distortions.
- Combining complementary features improves robustness.

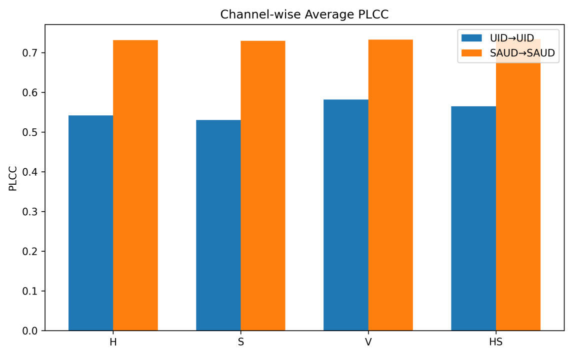
Nevertheless, even the strongest Gabor-based configurations do not exceed the performance of the proposed color–structure feature set or the Super Stack Ensemble. This indicates that while luminance frequency information is important, it is not sufficient on its own for comprehensive underwater image quality assessment.

5.5.6 Key Takeaway

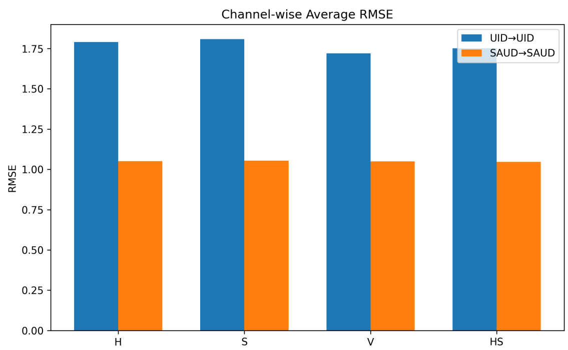
This cross-domain evaluation suggests that the proposed NR-UIQA framework captures broader perceptual trends rather than being narrowly optimized for a single feature type. The improved performance of the full model can be attributed to effective feature fusion across complementary perceptual cues.

5.6 XGBoost Channel-wise Performance Check

To further examine channel-level behavior, a focused analysis is conducted using the XGBoost regressor, which performs strongly as a single-model baseline. This experiment evaluates prediction performance when the model is restricted to specific channel subsets (H, S, V, and HS).



(a) Channel-wise PLCC using XGBoost.



(b) Channel-wise RMSE using XGBoost.

Figure 6: XGBoost-based channel-wise performance comparison on UID2021 and SAUD.

As shown in Fig. 6, the **Value (V) channel** consistently achieves the highest PLCC and lowest RMSE across both datasets, reinforcing the importance of luminance-driven contrast information. In contrast, the **Hue (H) channel** shows weaker performance, likely due to its sensitivity to wavelength-dependent absorption and non-linear color shifts in underwater environments.

Combined channel configurations, particularly **HS**, demonstrate improved robustness compared to individual chromatic channels. The consistency of these observations with earlier SVR-based and ensemble-level analyses indicates that the channel prioritization observed in the Super Stack Ensemble reflects underlying perceptual relevance rather than model-specific effects.

6. Conclusion

This midterm report presented a no-reference underwater image quality assessment (NR-UIQA) framework based on perceptually motivated feature extraction and ensemble learning. The study examined how handcrafted features and supervised regression models can be combined to predict subjective image quality without access to reference images.

Experimental results show that ensemble-based learning, particularly the proposed Super Stack Ensemble, consistently achieves stronger correlation with subjective opinion scores than individual regression models. This indicates that aggregating complementary learners helps capture the diverse distortion characteristics commonly present in underwater imagery.

Feature analysis further demonstrates that no single feature family is sufficient to model perceptual degradation. Structural features in the gradient domain, color features in the HSV space, and frequency-domain luminance features each capture complementary aspects of underwater image quality. Joint feature configurations, especially those combining color and structure, consistently outperform single-domain representations, reflecting the multimodal nature of underwater visual perception.

Channel-wise and cross-dataset evaluations highlight the dominant role of luminance-driven information. Features associated with the Value (V) channel and mid-frequency texture content are more informative than hue-only representations, which are more sensitive to wavelength-dependent absorption effects. Although cross-dataset evaluation exhibits an expected performance drop due to domain shift, the proposed feature combinations retain reasonable monotonic agreement with human opinion scores, indicating stable generalization behavior.

An auxiliary cross-domain study using Gabor-based luminance features supports these observations by reinforcing the importance of luminance and frequency information, while also confirming that such features alone are insufficient to surpass the full framework.

Overall, this work demonstrates that a feature-based NR-UIQA pipeline, when combined with ensemble learning, offers an effective and interpretable approach to underwater image quality assessment. The findings from this midterm phase provide a solid foundation for further refinement and extension in the final stage of the project.

7. Future Scope & Expansion

Based on the observations from the current experiments and the available computational resources, the following directions are identified for extension during the end-term phase:

1. **Feature Selection and Interpretability Analysis:** Building on the existing feature analysis, systematic feature selection techniques (e.g., correlation-based filtering or model-driven importance ranking) can be applied to identify the most informative subset of handcrafted features. This would help reduce redundancy, improve computational efficiency, and enhance interpretability by clarifying which perceptual cues contribute most strongly to quality prediction.

2. **Gabor Feature Refinement:** Additional ablation studies can be conducted to analyze the influence of Gabor filter parameters, such as scale and orientation, and to identify configurations that are most sensitive to underwater-specific degradations like turbidity and blur.

3. **Hybrid Feature Modeling:** The existing handcrafted feature set can be extended by incorporating semantic features extracted from a pre-trained convolutional neural network (e.g., ResNet-50) adapted for underwater imagery. Such a hybrid representation would allow the model to combine low-level perceptual statistics with higher-level semantic information.

4. **Domain Adaptation Strategies:** The observed reduction in cross-dataset performance suggests the presence of distributional differences between datasets. Exploring unsupervised domain adaptation techniques, such as feature distribution alignment between UID and SAUD, may help improve generalization without requiring additional subjective annotations.

5. **Expanded Ensemble Evaluation:** With access to increased computational resources, the number of Super Stack Ensemble iterations can be scaled beyond the current setting to obtain more statistically stable performance estimates comparable to those used for base regressors.

References

[1] J. S. Jaffe, "Computer modeling and the design of optimal underwater imaging systems," *IEEE J. Ocean. Eng.*, vol. 15, no. 2, pp. 101–111, 1990.

[2] S. Dhanasekaran, "Research Internship Task 1: UIQA basics and statistical metrics," task report, Department of Electrical Engineering, IIT Kanpur, December 2025. Internal Report.

[3] S. Dhanasekaran, "Research Internship Task 2: Feature extraction logic and metric differences," task report, Department of Electrical Engineering, IIT Kanpur, December 2025. Internal Report.

[4] D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992.

[5] H. Chen, X. Chai, F. Shao, X. Wang, Q. Jiang, X. Meng, and Y.-S. Ho, "Perceptual quality assessment of cartoon images," *IEEE Trans. Multimedia*, vol. 25, pp. 140–153, 2023.

[6] T. Ojala *et al.*, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, 2002.

[7] K. Panetta, A. Samani, and S. Agaian, "Choosing the optimal spatial domain measure of enhancement for mammogram images," *Int. J. Biomed. Imaging*, vol. 2014, 2014.

[8] A. Mittal *et al.*, “No-reference image quality assessment in the spatial domain,” *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, 2012.

[9] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[10] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[11] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proc. ACM SIGKDD*, pp. 785–794, 2016.

[12] G. Ke *et al.*, “LightGBM: A highly efficient gradient boosting decision tree,” in *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[13] L. Prokhorenkova *et al.*, “CatBoost: Unbiased boosting with categorical features,” in *Adv. Neural Inf. Process. Syst.*, vol. 31, 2018.

[14] G. Hou *et al.*, “UID2021: An underwater image dataset for evaluation of no-reference quality assessment metrics,” *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 19, no. 4, pp. 1–24, 2023.

[15] Q. Jiang *et al.*, “Underwater image enhancement quality evaluation: Benchmark dataset and objective metric,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 9, pp. 5944–5958, 2022.

[16] A. K. R. Poreddy, P. Kokil, and B. Appina, “Enhancing surgical laparoscopic video quality assessment with integrated feature fusion accounting for sensor and transmission distortions,” *IEEE Sensors Letters*, vol. 9, no. 5, pp. 1–4, 2025. Art. no. 6005004.