Aswan University Journal of Science and Technology Volume 4, issue 4

DeepMSE: A Lightweight Image Quality Assessment Model Based on SqueezeNet and MSE for Resource-Constrained Systems

Hossam Mady[1], Adel Agamy[1], Abdelmageed Mohamed Aly[1], Mohamed Abdel-Nasser[1]
[1]Electrical Engineering Department, Faculty of Engineering, Aswan University, Aswan 81542, Egypt

Abstract

Image Quality Assessment (IQA) is very important in many different applications. It is therefore not surprising that research into IQA has received extensive attention during the last three decades. Recent models in the field of IQA demonstrate strong performance on several standard IQA datasets. However, their reliance on computationally intensive deep learning architectures and/or complex calculations makes them unsuitable for resource-constrained systems such as embedded and mobile Systems. In this paper we propose a Full Reference (FR) IQA model, called DeepMSE, which is based on SqueezeNet for feature extraction and Mean square Error (MSE) for aggregation. Unlike existing FR-IQA models, the proposed model doesn't require training or tuning with Mean Opinion Scores (MOSs), which helps mitigate the risk of overfitting. Additionally, our model reduces computational complexity while maintaining high performance, making it well-suited for deployment on mobile or edge devices. Experimental evaluations across large standard IQA datasets demonstrate the high performance of our model and its superiority over state-of-the-art methods in aligning with human visual perception, all while maintaining simplicity, compact size, and reduced complexity.

Keywords: Image Quality Assessment (IQA), Convolution Neural Networks (CNNs), Mean Square Error (MSE)

1. Introduction

Image Quality Assessment (IQA) has received extensive attention over the past three decades, as it plays a critical role in a wide range of computer vision applications. As visual content spreads throughout media channels, IQA techniques that are automated, reliable, efficient, and perceptually relevant are becoming more and more important [1]. Full-Reference Image Quality Assessment (FR-IQA) is a category of IQA that provides a quality score for a distorted image by comparing it with a reference (pristine) image.

Knowledge-driven FR-IQA aims to mimic the Human Visual System (HVS) to accurately predict perceived image quality. Common examples of knowledge-driven FR-IQA include Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM) [2] and its variants [3][4][5], among others. Such models have become widely adopted due to their simplicity and computational efficiency. However, despite their popularity, these models often struggle to align precisely with human visual perception, especially in the presence of complex distortions.

Corresponding author E-mail: hossammady97@eng.aswu.edu.eg

Received October 20, 2024, revised November 1, 2024, accepted November 2, 2024.

In contrast, data-driven FR-IQA models utilize the power of deep learning and convolutional neural networks (CNNs) to discover and learn features that best describe image quality, rather than rely on handcrafted features. Examples such as Deep Similarity for image quality assessment (DeepSim) [6], Learned Perceptual Image Patch Similarity (LPIPS) [7] and Deep Image Structure and Texture Similarity (DISTS) [8] demonstrate strong correlations with human judgments of image quality.

However, existing deep learning models for IQA often require substantial computational resources, making them impractical for real-time and embedded applications. In response, this work presents a lightweight IQA model that employs SqueezeNet [9] which is a compact CNN architecture to extract deep features. The choice of SqueezeNet ensures that the model remains resource-efficient without compromising on the quality of the extracted features. Aggregation is achieved through MSE, a simple yet effective metric that provides a measure of perceptual similarity based on feature map differences.

The proposed model is designed with real-time applications in mind, such as mobile image processing, augmented reality, and quality control in compressed image transmission systems. By delivering a balance between efficiency and perceptual accuracy our model offers a practical solution for scenarios where computational resources are constrained but high-quality IQA is necessary.

2. Related Work

Traditional FR-IQA models have relied heavily on metrics such as PSNR and SSIM. While PSNR is simple and computationally efficient, it does not correlate well with human visual perception. SSIM [2], introduced by Wang et al. has been a widely adopted metric as it accounts for structural information by comparing luminance, contrast, and structural similarity between the reference and distorted images. Despite its popularity, SSIM has limitations, particularly when it comes to images with complex distortions or color variations, where it fails to capture perceptual quality accurately.

To address the limitations of SSIM, several advanced FR-IQA models have been proposed. Multi-Scale SSIM (MS-SSIM) [3] extends SSIM by incorporating multi-scale analysis, allowing it to capture distortions at different scales. Additionally, Visual Information Fidelity (VIF) [10] introduced by Sheikh and Bovik employs a natural scene statistics model to better correlate with the human visual system. Although these models achieve higher accuracy than traditional metrics, they often involve complex computations, making them less suitable for real-time applications where efficiency is paramount.

Another category of advanced FR-IQA models leverages deep learning. These models typically employ convolutional neural networks (CNNs) or other neural network architectures to extract deep features from images. By training on large-scale datasets, these models are capable of capturing intricate visual details and outperform traditional metrics in terms of accuracy. However, the reliance on large datasets and high computational costs limit the practical applicability of these methods, especially in scenarios requiring low-latency or edge-device deployment.

Recent advancements have introduced hybrid approaches that combine traditional perceptual metrics with deep learning components to enhance both accuracy and efficiency. Hybrid models typically use lightweight neural networks or leverage transfer learning to minimize computational demands. Despite these advancements, achieving a model that provides high accuracy and efficiency simultaneously remains a challenge, particularly when aiming for real-time performance on limited hardware resources.

This paper proposes a novel FR-IQA model that addresses the limitations of existing methods by introducing an efficient and effective approach that balances computational efficiency with accurate perceptual quality assessment. The proposed model utilizes SqueezeNet architecture as feature extractor and calculates the final score using MSE. By leveraging these components, the model demonstrates superior performance while maintaining low computational costs suitable for real-time applications.

3. Methodology

Our model extracts features from both the reference and distorted images, compares these features at multiple stages, and aggregates the differences to predict the overall image quality. This section describes the detailed methodology.

3.1 Feature extraction using SqueezeNet

The proposed model as shown in figure 1 utilizes five stages from SqueezeNet for feature extraction. We selected SqueezeNet as the feature extraction backbone primarily for its efficiency in resource-constrained systems. SqueezeNet achieves AlexNet-level accuracy with significantly fewer parameters (up to 50x fewer), resulting in a compact model that requires less memory and computational power. The lightweight architecture allows our model to perform real-time IQA without compromising the quality of extracted features. Using SqueezeNet allows us to balance computational efficiency with perceptual accuracy, ensuring that the model is both fast and capable of accurately aligning with human visual perception.

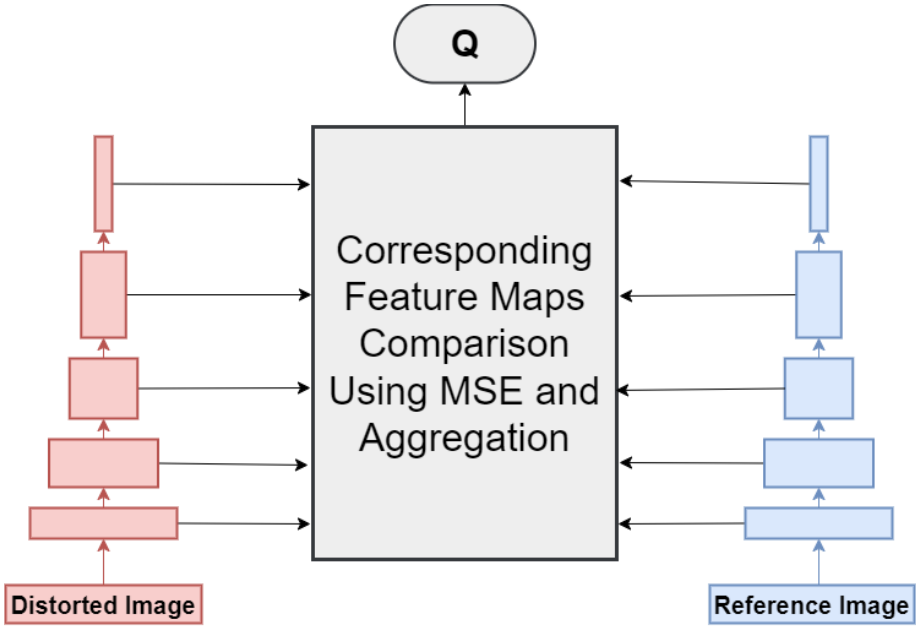
Q
Corresponding Feature Maps Comparison Using MSE and Aggregation
Distorted Image
Reference Image

Figure 1. DeepMSE Architecture

Both the reference image (I_{ref}) and the distorted image (I_{dist}) are fed into the model simultaneously. Both images produce five feature maps. The feature maps at each stage represent the extracted image features at different levels of abstraction, where $F_{\mathrm{ref},i}$ and $F_{\mathrm{dist},i}$ represent the feature maps for I_{ref} and I_{dist} at stage i, respectively.

3.2 MSE Calculation Between Corresponding Feature Maps and aggregation

For each corresponding pair of feature maps ($F_{\mathrm{ref},i}$, $F_{\mathrm{dist},i}$) obtained from the reference and distorted images at stage i, the model computes a feature similarity score. The similarity score D_i is calculated as the MSE between the two feature maps:

$$D_i = MSE\left(F_{ref,i}, F_{dist,i}\right) = \frac{1}{N}\sum_{j=1}^{N}\left(F_{ref,i,j} - F_{dist,i,j}\right)^2 \quad (1)$$

where N is the number of elements in the feature maps.

The distortion scores D_i from each of the five stages are then aggregated to produce an overall quality score. Aggregation is performed by taking the average of the five MSE values, representing the overall perceived difference between the reference and distorted images:

$$Q = \frac{1}{5}\sum_{i=1}^{5} D_i \qquad (2)$$

where Q is the final predicted quality score, indicating the level of distortion present in the distorted image relative to the reference image. Lower Q value suggests higher similarity and, thus better quality.

4. Experiments

4.1 Implementation Details

We employed SqueezeNet architecture as the feature extractor, initialized using pre-trained weights from ImageNet. During evaluation, each image is resized to 256×256 pixels. A computer with an Intel(R) core(TM) i7-9700K processor (3.60 GHz) and 32 GB of RAM was used for this experiment. Using PyTorch version 2.2.0, we ran all computations on an NVIDIA GeForce GTX 1650 GPU. CUDA version 12.1 was installed on our Windows 10 64-bit system in order to facilitate GPU acceleration.

4.2 Results

To demonstrate the effectiveness of our model, we test it on two large IQA datasets which contain a large number of distorted images and subjective human evaluations of each. The datasets are TID2013 [11] and KADID-10K [12]. TID2013 consists of 25 original images distorted with 24 types of distortions —such as Gaussian noise, image compression artifacts, and color saturation changes— at five levels of severity for each type of distortion, so that the total number of distorted images is 3000 images. Each image is accompanied by mean opinion scores (MOS) gathered from human assessments. On the other hand, KADID-10k consists of 81 original images distorted with 25 types of distortions at five levels of intensity, resulting in 10125 distorted images, each with an associated difference mean opinion score (DMOS). The distortions include types such as white noise, contrast change, and chromatic aberration. Together, these datasets offer a diverse range of distortion types and levels, enabling robust testing of our model's accuracy and generalizability in image quality assessment.

Table 1 shows the SRCC and KRCC values for our model compared to some state-of-the-art models. It is obvious that our model outperforms other traditional models (PSNR, SSIM [2], MS-SSIM [3], FSIMc [13], VIF [10], and NLPD [14]) and data-driven models (PieAPP [15], LPIPS [7], DISTS [8], DeepDC [16], and SWLGV [17]) achieving SRCC =0.877 and KRCC = 0.697 on TID2013 and SRCC = 0.901 and KRCC = 0.723 on KADID-10k dataset.

Table 1. Performance comparison between our model and various state-of-the-art IQA methods on the TID2013 and KADID-10k datasets. The best two results are highlighted in **bold**.

IQA Model	TID2013 [11]		KADID-10k [12]	
	SRCC	KRCC	SRCC	KRCC
PSNR	0.687	0.496	0.676	0.488
SSIM [2]	0.720	0.527	0.724	0.537
MS-SSIM [3]	0.786	0.605	0.826	0.635
FSIMc [13]	0.851	0.666	0.854	0.665
VIF [10]	0.677	0.518	0.679	0.507
NLPD [14]	0.800	0.625	0.812	0.623
PieAPP [15]	**0.876**	**0.683**	0.836	0.647

LPIPS [7]	0.670	0.497	0.843	0.653
DISTS [8]	0.830	0.639	0.887	0.709
DeepDC [16]	0.844	0.651	**0.905**	**0.733**
SWLGV [17]	**0.804**	**0.637**	**0.840**	**0.655**
DeepMSE (Ours)	**0.877**	**0.697**	**0.901**	**0.723**

Figure 2 and Figure 3 compare our model with several existing FR-IQA methods in terms of frames per second (FPS) and floating point operations (FLOPS). The bar chart in Figure 2 highlights our model's FPS of 248.44, the highest among the compared methods, while Figure 3 shows it achieves the lowest FLOPS at 0.694 billion, indicating both high processing speed and low complexity. This proves the efficiency of our method, which makes it not only more accurate, as shown in Table 1, but also much faster and more lightweight. The combination of high FPS and low FLOPS of our model indicates its suitability for real-time applications, where both speed and accuracy are crucial factors.

Figures 4 and 5 present heatmaps displaying SRCC and KRCC values for different distortion types in the CSIQ [18] and LIVE [19] datasets, respectively. The consistently high SRCC and KRCC values across all distortion types demonstrate the superiority and generalizability of our model. Notably, even the lowest results observed—SRCC = 0.94 and KRCC = 0.80 for Contrast Decrement in CSIQ, and SRCC = 0.93 and KRCC = 0.77 for Fastfading in LIVE—are exceptionally high. These results indicate that the model performs robustly across various distortion types without noticeable limitations, underscoring its reliability and effectiveness in diverse quality assessment scenarios.

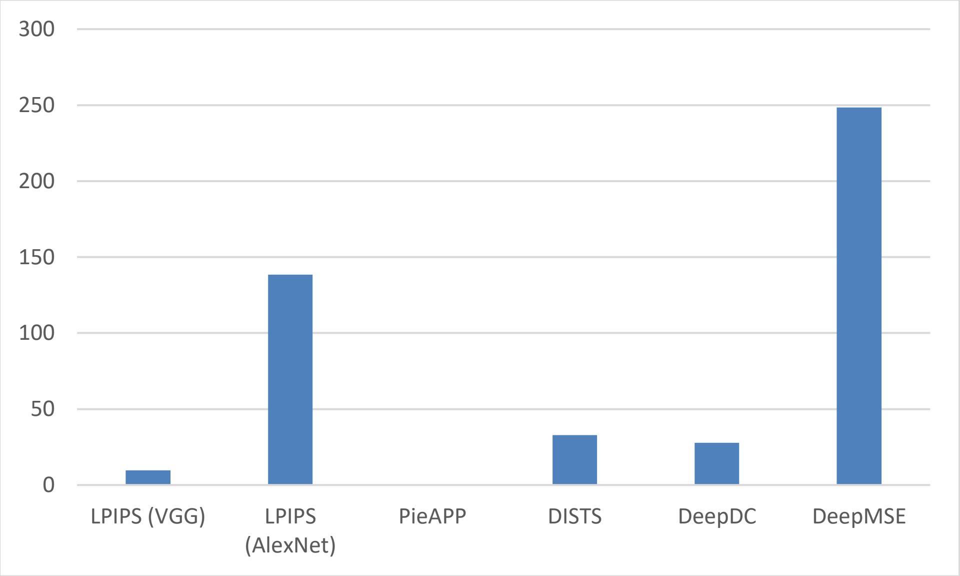

Figure 2. Comparison of FPS performance between the proposed DeepMSE model and various state-of-the-art IQA methods.

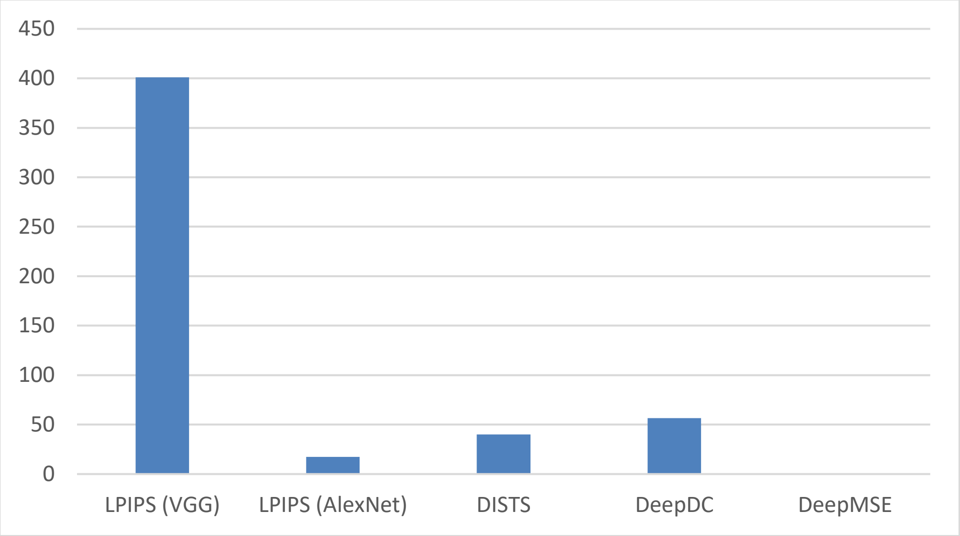

Figure 3. Comparison of FLOPS performance between the proposed DeepMSE model and various state-of-the-art IQA methods.

5. Conclusion

In this paper, we proposed a new lightweight FR-IQA model based on SqueezeNet and mean squared error (MSE) which optimizes both accuracy and computational efficiency for resource-constrained systems. Our model was tested on two widely used datasets, TID2013 and KADID-10k, and demonstrated excellent performance when compared to state-of-the-art IQA methods. Specifically, the proposed model achieved high SRCC and KRCC, which indicates its effectiveness in evaluating image quality. Additionally, the model's processing speed, measured in FPS, outperformed other methods, proving its suitability for real-time applications.

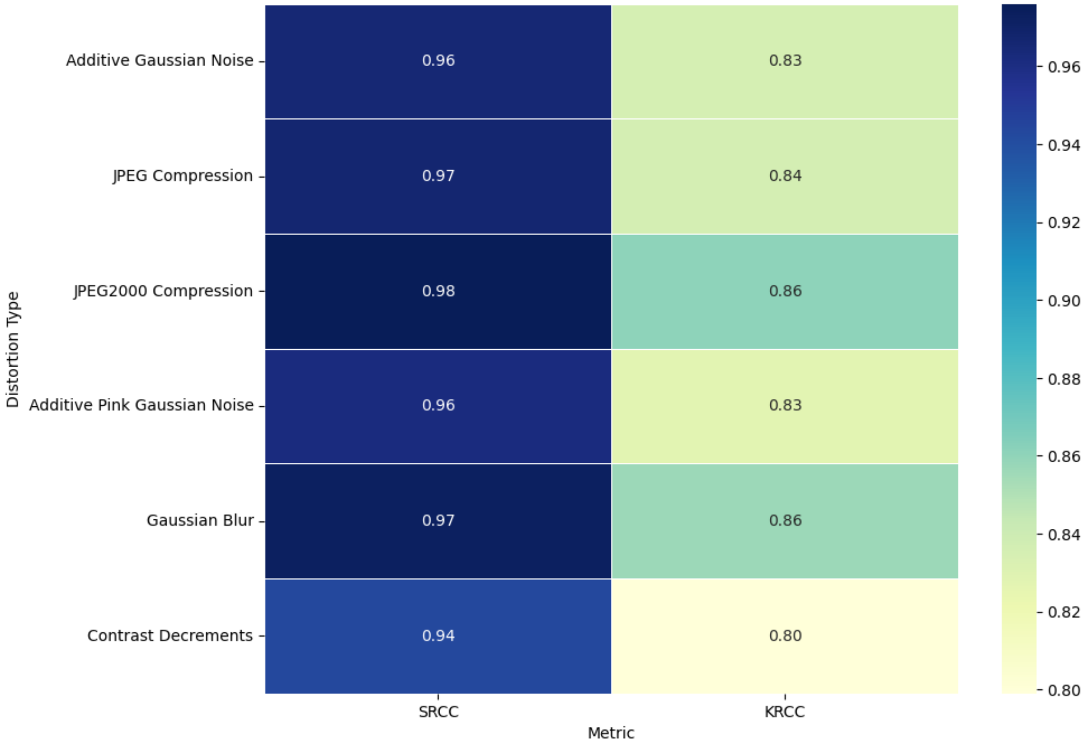

Figure 4. Model performance on different distortion types in the CSIQ [18] dataset.

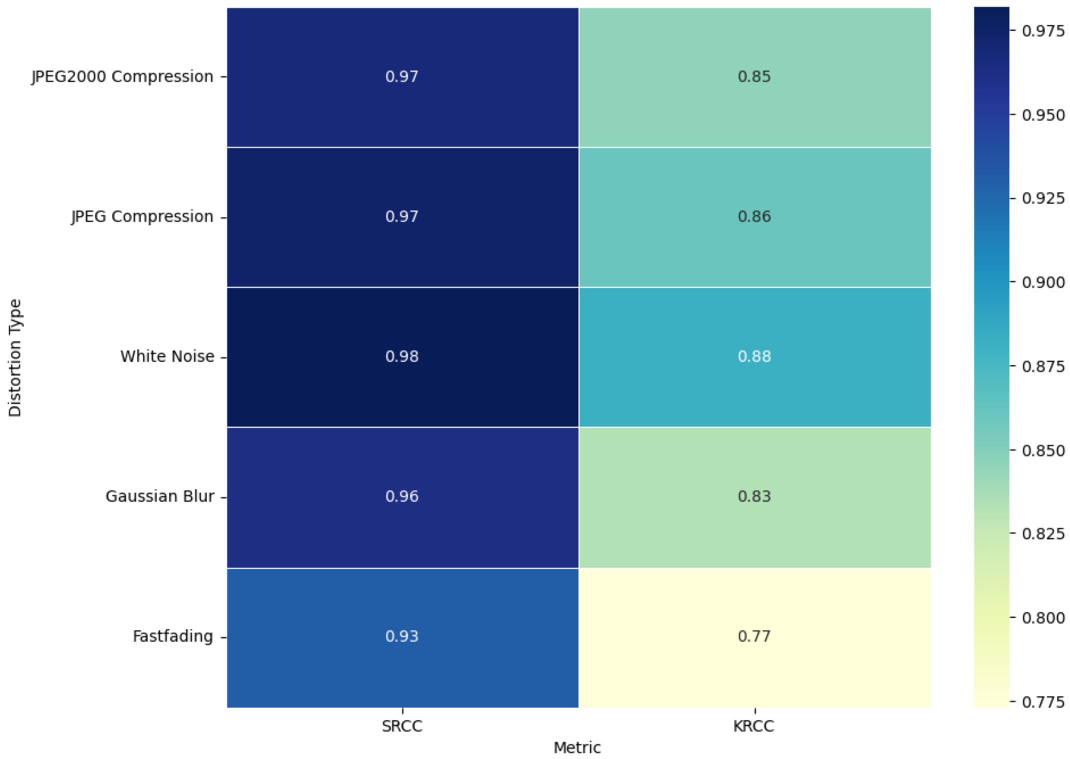

Figure 5. Model performance on different distortion types in the LIVE [19] dataset.

The success of this model can be attributed to the integration of the lightweight SqueezeNet architecture and the simplicity of MSE. Our results suggest that this approach is promising for practical deployment in various image quality evaluation tasks, especially in scenarios where computational resources are limited. Future work could further optimize the model by exploring other loss functions to enhance both accuracy and computational efficiency even more.

References

[1] 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, Jan. 2018, doi: 10.1109/TIP.2017.2760518.

[2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.819861.

[3] Z. Wang, E. P. Simoncelli, and A. C. Bovik, “MULTI-SCALE STRUCTURAL SIMILARITY FOR IMAGE QUALITY ASSESSMENT,” *IEEE Asilomar Conference on Signals, System and Computers*, pp. 1398–1402, 2003.

[4] Z. Wang and Q. Li, “Information content weighting for perceptual image quality assessment,” *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185–1198, May 2011, doi: 10.1109/TIP.2010.2092435.

[5] Z. Wang and E. P. Simoncelli, “Translation insensitive image similarity in complex wavelet domain,” *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2, 2005, doi: 10.1109/ICASSP.2005.1415469.

[6] F. Gao, Y. Wang, P. Li, M. Tan, J. Yu, and Y. Zhu, “DeepSim: Deep similarity for image quality assessment,” *Neurocomputing*, vol. 257, pp. 104–114, Sep. 2017, doi: 10.1016/J.NEUCOM.2017.01.054.

[7] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The Unreasonable Effectiveness of Deep Features as a Perceptual Metric,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018.

[8] K. Ding, K. Ma, S. Wang, and E. P. Simoncelli, “Image Quality Assessment: Unifying Structure and Texture Similarity,” *IEEE Trans Pattern Anal Mach Intell*, vol. 44, no. 5, pp. 2567–2581, Apr. 2020, doi: 10.1109/TPAMI.2020.3045810.

[9] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, “SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size,” *arXiv preprint arXiv:1602.07360*, Feb. 2016, [Online]. Available: http://arxiv.org/abs/1602.07360

[10] H. R. Sheikh and A. C. Bovik, “A VISUAL INFORMATION FIDELITY APPROACH TO VIDEO QUALITY ASSESSMENT,” *The First International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, vol. 7, 2005.

[11] N. Ponomarenko *et al.*, “Image database TID2013: Peculiarities, results and perspectives,” *Signal Process Image Commun*, vol. 30, pp. 57–77, Jan. 2015, doi: 10.1016/j.image.2014.10.009.

[12] H. Lin, V. Hosu, and D. Saupe, “KADID-10k: A Large-scale Artificially Distorted IQA Database,” *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–3, Mar. 2018, doi: 10.1109/TIP.2020.2967829.

[13] L. Zhang, L. Zhang, X. Mou, and D. Zhang, “FSIM: A Feature Similarity Index for Image Quality Assessment,” *IEEE transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.

[14] V. Laparra, J. Ballé, A. Berardino, and E. P. Simoncelli, “Perceptual image quality assessment using a normalized Laplacian pyramid,” *Electronic Imaging*, vol. 2016, no. 16, pp. 1–6, 2016.

[15] E. Prashnani, H. Cai, Y. Mostofi, and P. Sen, “PieAPP: Perceptual Image-Error Assessment through Pairwise Preference,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1808–1817, Jun. 2018, [Online]. Available: http://arxiv.org/abs/1806.02067

[16] H. Zhu, B. Chen, L. Zhu, S. Wang, and W. Lin, “DeepDC: Deep Distance Correlation as a Perceptual Image Quality Evaluator,” *arXiv preprint, arXiv:2211.04927v2*, Nov. 2023, [Online]. Available: http://arxiv.org/abs/2211.04927

[17] D. Varga, “Full-Reference Image Quality Assessment Based on Grünwald–Letnikov Derivative, Image Gradients, and Visual Saliency,” *Electronics (Switzerland)*, vol. 11, no. 4, Feb. 2022, doi: 10.3390/electronics11040559.

[18] D. M. Chandler and E. C. Larson, “Most apparent distortion: full-reference image quality assessment and the role of strategy,” *J Electron Imaging*, vol. 19, no. 1, pp. 1–21, Jan. 2010, doi: 10.1117/1.3267105.

[19] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, “Image and video quality assessment research at LIVE.” [Online]. Available: http://live.ece.utexas.edu/research/quality/.