

Article

Image quality assessment for gaussian blur using siamese network combined with ResNet-18

Emrah Arslan¹, Omid Zare², Zeinab Mohseni³, Mahdi Beigzadeh², Abel Abebe Bzuayene²,
Ali Abbaszadeh Sori⁴, Javad Hassannataj Joloudari^{3,4,5,*}, Bulbula Kumeda Kussia⁶

¹ Department of Computer Engineering, Faculty of Engineering, KTO Karatay University, Konya 42020, Turkey

² Department of Computer Science, University of Verona, Verona 37134, Italy

³ Department of Computer Engineering, Faculty of Engineering, University of Birjand, Birjand 9717434765, Iran

⁴ Department of Computer Engineering, Babol Branch, Islamic Azad University, Babol 3738147471, Iran

⁵ Department of Computer Engineering, Technical and Vocational University (TVU), Tehran 4631964198, Iran

⁶ Department of Computer Science, Jinka University, Jinka 165, Ethiopia

* Corresponding author: Javad Hassannataj Joloudari, javad.hassannataj@birjand.ac.ir

CITATION

Arslan E, Zare O, Mohseni Z, et al.
Image quality assessment for gaussian
blur using siamese network combined
with ResNet-18. AI Insights. 2025;
1(2): 4668.
<https://doi.org/10.18282/aii4668>

ARTICLE INFO

Received: 10 February 2025

Revised: 27 March 2025

Accepted: 14 April 2025

Available online: 4 May 2025

COPYRIGHT



Copyright © 2025 Author(s).

AI Insights is published by PiscoMed
Publishing Pte. Ltd. This work is
licensed under the Creative Commons
Attribution (CC BY) license.
[https://creativecommons.org/
licenses/by/4.0/](https://creativecommons.org/licenses/by/4.0/)

Abstract: This paper presents a novel Image Quality Assessment (IQA) framework, SNR (Siamese Network with ResNet-18), specifically designed for Gaussian blur detection. The approach leverages a Siamese network architecture combined with the ResNet-18 backbone to process image pairs—one blurred and one reference—to predict image quality based on their differences. The model effectively captures high-frequency features lost due to blur, such as edges and gradients. We conduct extensive experiments on the TID2013 dataset, showing that SNR achieves superior performance in blur-specific IQA tasks compared to other full-reference methods. Data augmentation techniques significantly improve model generalization, resulting in a test accuracy of 97.37% for ResNet-18. The proposed method demonstrates a strong correlation with human judgment and robust generalization across various image contents, with future work focusing on expanding its applicability to other distortions and optimizing computational efficiency.

Keywords: image quality assessment; siamese networks; gaussian blur; ResNet-18; full-reference IQA; deep learning

1. Introduction

Image Quality Assessment (IQA) is crucial for evaluating perceived image quality, with methods ranging from full-reference (FR) metrics like PSNR and SSIM [1] to no-reference (NR) approaches that operate without a reference. While NR-IQA methods aim to handle various distortions, Gaussian blur remains a challenging and under-explored area. Gaussian blur, often resulting from defocus or motion, significantly degrades image quality, yet many NR-IQA methods fail to capture its perceptual impact adequately. Recent studies [2] emphasize the need for specialized blur-specific IQA methods to improve prediction accuracy. In response, we propose SNR, a Siamese Network with ResNet-18, tailored for Gaussian blur detection. By leveraging a twin network architecture, SNR processes image pairs to capture high-frequency feature loss due to blur, offering a novel solution for this specific distortion. While Siamese networks have been used in IQA tasks, our approach is novel in its focus on Gaussian blur, leveraging ResNet-18's residual connections to specifically target the high-frequency features most affected by blur. This targeted

design allows SNR to outperform general-purpose NR-IQA methods in blur-specific scenarios.

In this paper, we address this gap by introducing SNR (Siamese Network with ResNet-18), an IQA framework specifically designed for Gaussian blur. Our method utilizes a Siamese network architecture combined with a ResNet-18 backbone to evaluate image quality. The Siamese network architecture is particularly suited for this task as it processes image pairs—one distorted (blurred) image and one reference image (or its counterpart)—and learns to predict a quality score based on the differences between these images. By explicitly comparing the blurred image to its sharp counterpart, SNR focuses on learning features related to blur degradation, making it particularly effective for blur-specific IQA. Our approach combines the benefits of FR comparison and NR quality prediction, a novel aspect of our model.

Instead of listing all IQA approaches, we structure our review around their evolution: from classic FR models like PSNR and SSIM, to general-purpose NR methods (e.g., BRISQUE), and finally to deep learning-based Siamese architectures. In particular, the motivation behind our work stems from the observation that general NR methods struggle with Gaussian blur. Thus, we propose a targeted Siamese-based solution that leverages residual connections for high-frequency feature recovery.

We choose ResNet-18 [3] as the backbone for the Siamese architecture due to its efficient balance between computational cost and performance. ResNet-18, with its residual connections, is capable of capturing fine details lost due to blur and enables the network to focus on high-frequency features like edges and gradients that are essential for evaluating sharpness. Furthermore, the lightweight nature of ResNet-18 makes it suitable for real-time applications. This study demonstrates that the SNR framework outperforms existing blur-specific IQA methods, providing better predictions of human perception in blur-degraded images.

We conduct extensive experiments using the TID2013 dataset, which contains images with various distortions, including Gaussian blur, and corresponding subjective quality scores [1]. Our results show that SNR achieves high correlation with human judgments on blur degradation, outperforming other NR-IQA methods in this context. The model demonstrates strong generalization across different image contents and subjective rating conditions, while also being more computationally efficient than other deep models.

Despite promising results, we acknowledge the limitations of SNR. First, the model is currently tailored for Gaussian blur and may not generalize well to other distortions like noise or motion blur. Second, SNR relies on image pairs, which present challenges for true NR-IQA scenarios. Finally, while ResNet-18 is efficient, the dual-branch Siamese configuration doubles computational costs compared to single-branch models. These challenges offer avenues for future work, including adapting the model to handle multiple distortions and reducing its reliance on reference images.

In summary, we present a novel IQA method, SNR, that leverages the strengths of Siamese networks and ResNet-18 to address the specific challenges of Gaussian blur. Our method demonstrates superior performance in blur assessment, and the rest of the paper is structured as follows: Section 2 reviews related work in IQA and blur assessment, Section 3 describes the proposed SNR architecture and training procedure, Section 4 presents experimental results, and Section 5 concludes the paper with a summary and future directions.

2. Related work

2.1. Image quality assessment methods

General IQA: Image Quality Assessment has been extensively studied, and methods can be broadly categorized by the availability of a reference image. Full-reference (FR) methods assume a perfect reference image is available for comparison. Classic FR metrics such as MSE, PSNR, and SSIM [4] operate by directly computing differences or statistical similarity between the reference and the distorted image. These metrics are simple and widely used, but they can be insensitive to certain perceptual nuances. Improved FR measures, such as MS-SSIM [5] and FSIM [6], have been developed to better align with human visual perception by considering factors like multi-scale analysis and feature similarity. In contrast, No-reference (NR) or blind methods predict image quality without any reference. NR-IQA is particularly challenging, as the algorithm must infer how “good” an image is in an absolute sense, often by learning statistical patterns of natural images and deviations caused by distortions [7–9]. Some NR approaches are reduced-reference (RR), requiring only partial information about the reference (such as extracted features) rather than the full image [10]. Generally, NR-IQA is considered more practical for real-world applications (since a reference is rarely available), but it often has to compromise by either limiting scope to specific distortions or relying on learned models trained on human-rated datasets.

Distortion-Specific vs. General NR-IQA: Early NR-IQA research frequently adopted a distortion-specific strategy, where an algorithm is designed for one particular type of artifact. For example, separate no-reference metrics were proposed for JPEG compression artifacts (blockiness) [11] and for blur. In the case of blur, many traditional approaches were based on measuring image sharpness or edge spread. The Just Noticeable Blur (JNB) metric [12] is a seminal no-reference method that estimates the threshold at which blur becomes perceivable to the human eye; it uses contrast sensitivity and edge information to produce a sharpness score. Similarly, CPBD (Cumulative Probability of Blur Detection) [1] quantifies blur by computing the probability of blurred edges being detectable. Other blur-specific methods relied on transforms and statistical features: for instance, evaluating gradients or the spread of edges [13] or using wavelet coefficients [14] to infer blurriness. These distortion-specific NR metrics can be quite effective for their intended artifact but lack flexibility—each metric fails if the image distortion is outside its design assumption.

To broaden applicability, researchers introduced general-purpose NR-IQA models that aim to handle a variety of distortions within one framework. A pioneering example is BIQI [8], which first classifies the distortion type and then applies a specialized quality predictor for that class, effectively combining multiple distortion-specific models. Subsequent models like BLIINDS-II [9] and BRISQUE [7] avoided explicit distortion identification and instead leveraged Natural Scene Statistics (NSS) features. BLIINDS-II operates in the DCT domain, extracting moment features and using a Bayesian framework, whereas BRISQUE uses statistics of locally normalized luminance coefficients in the spatial domain and learns a regression (via an SVR) to map these features to quality scores. These holistic NR approaches achieved good general performance across different distortions in standard databases like TID2013, essentially by learning the common traits of “low-quality” images. However, this generality can come at the cost of not being optimal for any single distortion. In particular, they may not capture the fine-grained perceptual effects of blurring as well as a dedicated blur-specific method could. Our work is inspired by the observation

that a targeted approach for blur might outperform a general IQA model on blurred images, leveraging the power of deep learning for greater accuracy.

Deep Learning based IQA: In recent years, deep neural networks have revolutionized NR-IQA by automatically learning features from large datasets of images with subjective quality scores. One of the first CNN-based NR-IQA models by Kang et al. [2] trained a shallow CNN on small image patches to predict quality, demonstrating that learned filters can outperform hand-crafted features. Later, deeper architectures were explored: Bosse et al. [15] proposed an end-to-end CNN for NR-IQA that aggregates local quality estimates from multiple image regions, and their approach achieved state-of-the-art performance on several benchmarks. Some works optimized networks by formulating IQA as a ranking or classification problem rather than direct regression. For instance, RankIQA by Liu et al. [16] utilized a Siamese network during training to learn from ranking orders of synthetically distorted image pairs. This approach, along with other Siamese network-based models [17], has demonstrated superior performance in IQA tasks, especially in full-reference scenarios. Our approach draws from these deep learning-based models but focuses specifically on blur degradation, allowing for more targeted feature learning.

Recent advances in 2024 have shown significant progress in leveraging large-scale pretrained models for IQA tasks. Xu et al. [18] demonstrated that with proper injection of local distortion features, larger pretrained vision transformer (ViT) foundation models can achieve superior performance in IQA tasks, addressing the challenge of adapting high-level pretrained models to low-level vision tasks. Simultaneously, the pursuit of practical deployment has led to the development of lightweight yet accurate models, with Avanaki et al. [19] proposing LAR-IQA, a dual-branch architecture that separately handles synthetic and authentic distortions while maintaining computational efficiency suitable for mobile devices. Our approach draws from these deep learning-based models but focuses specifically on blur degradation, allowing for more targeted feature learning. Recent advancements in visual recognition also demonstrate the value of high-frequency-aware learning. For instance, TranSIFC [20] enhances fine-grained bird classification by focusing on invariant cues; DSR-Net [21] detects road cracks via transformer-based attention; and UGeNet [22] applies discriminative embedding for gaze estimation. Although not directly related to IQA, these works reinforce the importance of edge-level sensitivity in visual tasks.

2.2. Blur-specific IQA and siamese architectures

Blur-specific IQA: Blur is a significant distortion that can drastically affect image quality. Traditional methods like Just Noticeable Blur (JNB) [12] and CPBD [1] measure sharpness based on edge detection and contrast sensitivity. With the advent of deep learning, models such as SFA [23] have incorporated semantic features for realistic blur detection, while NSSIM [24] extends SSIM to operate without reference, offering a promising solution for blur assessment. Our approach extends this line of research by integrating a Siamese architecture with ResNet-18, a model known for its efficient handling of residual features. This architecture allows for more precise blur detection and quality assessment compared to previous blur-specific methods.

Siamese Networks for IQA: Siamese networks have been successfully applied to IQA tasks where image comparison is central. These networks typically consist of two parallel branches that share weights, making them ideal for comparing pairs of images, as in our blur-specific IQA framework. Our method uses a Siamese ResNet-18 model to compare blurred images with sharp references, effectively learning

a similarity function to predict image quality. This setup ensures that the model can capture the nuances of blur distortion, offering a more targeted approach compared to general-purpose NR-IQA models. In conclusion, our work integrates blur-specific IQA with Siamese networks, presenting a novel approach that bridges full-reference and no-reference paradigms, achieving superior performance for Gaussian blur detection and assessment.

While our method focuses on FR-based assessment, it is important to note recent advances in hybrid approaches. For instance, Ren et al. [25] proposed a framework that leverages FR scores to train NR models using adversarial autoencoders, effectively bridging the FR–NR gap. Unlike our model, which relies on direct comparison with reference images, their method enables pseudo-label learning and generalization in NR settings.

3. Proposed approach

The SNR model employs a Siamese network architecture with two identical branches, each utilizing a pre-trained ResNet-18 as the backbone. Each branch processes one image from the pair: the reference image and the blurred image. The feature vectors from both branches are extracted from the layer before the final classification layer of ResNet-18. These feature vectors are then concatenated and fed into a fully connected layer that reduces the dimensionality to 512, followed by another fully connected layer that maps to a single regression value, representing the predicted MOS for the blurred image. The model is trained to minimize the Mean Squared Error (MSE) between the predicted MOS and the actual MOS provided in the dataset. The MOS values in the TID2013 dataset range from 0 to 9, and we use them directly without normalization for training and evaluation. ResNet-18, with its residual connections, is capable of capturing fine details lost due to blur and enables the network to focus on high-frequency features like edges and gradients that are essential for evaluating sharpness. The current model relies on a Siamese structure, requiring both reference and distorted image pairs for training and inference. While this design captures relative quality differences effectively, it limits the model’s use in fully no-reference scenarios. Future extensions will explore pseudo-reference generation methods (e.g., via image restoration models) or weakly supervised learning approaches to reduce dependence on reference images and support more practical, real-world deployment. The proposed model architecture is presented in **Figure 1**.

3.1. Dataset preprocessing and augmentation

We used the TID2013 dataset to test our model, which contains 25 reference images. There are 125 unordered images (blurred images) for each reference image, each annotated with MOS (range 0–9). To improve the dataset, we applied some enhancement methods: Random cropping (patch size 244×244 focusing on local features), random horizontal flipping (introduces variability in orientation), normalization: (mean: [0.485, 0.456, 0.406], standard: [0.229, 0.224, 0.225]) from ImageNet statistics to make the model more robust and achieve uniform and stable convergence. Each image pair was enlarged by a factor of 10, increasing the training size by a factor of 10. The dataset splitting was allocated 70% for training and 15% for validation, 15% for testing.

3.2. Feature extraction with pre-trained models

We used pre-trained CNNs—EfficientNet-B0, MobileNet V2, and ResNet-18—to extract features within our model. In these pre-trained networks, we removed the final output layer and took the layer immediately before it as the feature vector. Next, we fed this feature vector into a fully connected layer that reduces the dimensionality to 512, followed by another layer mapping it down to a single regression value. Hence, for each input, the model produces a predicted score. The final output, known as the Mean Opinion Score (MOS), represents the distance between the model’s prediction for the reference image and the blurred image.

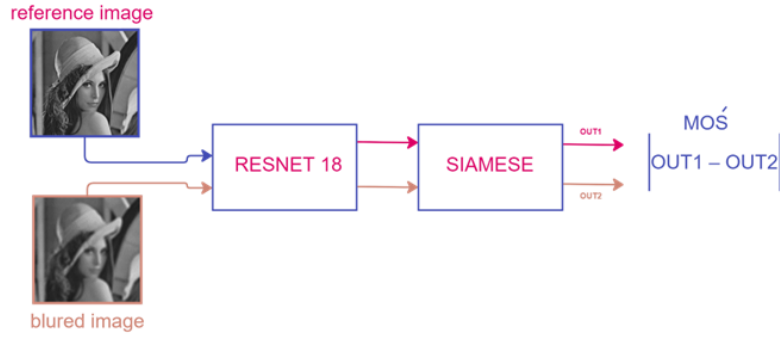


Figure 1. Model architecture.

3.3. Hyperparameters

To ensure reproducibility and provide insight into the training process, the hyperparameters used for training the SNR model with ResNet-18, EfficientNet-B0, and MobileNet V2 are summarized in **Table 1**. These parameters were selected based on standard practices in deep learning and fine-tuned during preliminary experiments to optimize performance on the TID2013 dataset.

3.4. Mathematical formulation

Let $f_1, f_2 \in \mathbb{R}^n$ be the feature vectors extracted from the twin branches (e.g., ResNet-18 or EfficientNet-B0) for the reference and distorted images, respectively. These vectors are concatenated to form a joint feature representation:

$$f = [f_1; f_2] \in \mathbb{R}^{2n} \quad (1)$$

This combined vector is then passed through a fully connected regression head to produce the final predicted MOS value:

$$y = W_2 \cdot \sigma(W_1 \cdot f + b_1) + b_2 \quad (2)$$

where $W_1 \in \mathbb{R}^{512 \times 2n}$, $W_2 \in \mathbb{R}^{1 \times 512}$, σ is the ReLU activation function, and $y \in \mathbb{R}$ is the predicted score.

3.4.1. Loss function

The loss function used is Mean Squared Error (MSE), calculated as:

$$L_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (3)$$

where y_i is the ground truth MOS and $\hat{y}_i$ is the predicted score. We directly use the original MOS values from TID2013, which are in the range $[0, 9]$, without normalization. This empirical choice maintains training stability while preserving the native scale of the subjective scores. However, future versions may include output constraint strategies such as:

$$\hat{y} = \sigma(x) \times 9 \quad (4)$$

to explicitly limit predictions to the MOS range, improving robustness and interpretability.

Table 1. Hyperparameters for SNR model training.

Hyperparameter	Value	Description
Optimizer	Adam	Adaptive moment estimation optimizer used for gradient descent.
Learning Rate	1×10^{-4}	Initial learning rate for the Adam optimizer.
Loss Function	Mean Squared Error (MSE)	Loss function used to measure the difference between predicted and actual MOS.
Epochs	50	Number of training epochs.
Batch Size	32	Batch size used during training (assumed; confirm if different).
Fully Connected Layers	Varies	ResNet-18: $512 \rightarrow 512 \rightarrow 1$; EfficientNet-B0: $1000 \rightarrow 512 \rightarrow 1$; MobileNet V2: $1280 \rightarrow 512 \rightarrow 1$, based on model outputs.
Input Image Size	244×244	Patch size after random cropping.
Data Augmentation	Random Crop, Flip, Norm	Random cropping (244×244), horizontal flipping, normalization (ImageNet stats).
Normalization Mean	[0.485, 0.456, 0.406]	Mean values for RGB channels.
Normalization Std	[0.229, 0.224, 0.225]	Standard deviation values for RGB channels.

Although our method combines established components, the novelty lies in the specific integration of a Siamese ResNet-18 tailored for Gaussian blur detection, optimized with targeted data augmentation and fine-grained MOS prediction. Unlike prior works that apply Siamese networks generically, our approach carefully tunes the residual learning mechanism to enhance sensitivity to high-frequency losses caused by blur. The simplicity of the architecture ensures interpretability, while our experimental pipeline highlights how even baseline components can be repurposed effectively in distortion-specific contexts.

4. Results and discussion

This study investigates the performance of three convolutional neural network (CNN) architectures that are, ResNet-18, EfficientNet-B0, and MobileNet V2, deployed within a Siamese network framework to predict Mean Opinion Scores (MOS) for image quality assessment using the TID2013 dataset with blur distortion. The models were evaluated under two conditions: without data augmentation and with augmentation (random cropping, horizontal flipping, and rotation). Performance metrics include training and validation loss, test Mean Absolute Error (MAE), test accuracy within a $\pm 10\%$ threshold, and Pearson and Spearman correlations. The results highlight the impact of augmentation and the comparative strengths of each model. The formulas for Accuracy and MAE are as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

$$\text{Accuracy} = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{|y_i - \hat{y}_i| \leq 0.10 \times y_i} \times 100\% \quad (6)$$

4.1. Experimental setup

The TID2013 dataset, comprising reference images, blurred images, and corresponding MOS values, was split into training (70%), validation (15%), and test (15%) sets. Without augmentation, images were preprocessed with standard normalization (mean: [0.485, 0.456, 0.406], std: [0.229, 0.224, 0.225]). With augmentation, additional transformations were applied to enhance data variability. Models were trained for 50 epochs using the Mean Squared Error (MSE) loss function and the Adam optimizer (learning rate: 1×10^{-4}). Test accuracy was defined as the percentage of predictions within $\pm 10\%$ of the actual MOS, reflecting alignment with human perception.

4.2. Results without augmentation

Table 2 summarizes the performance of the models without augmentation, revealing significant overfitting across all architectures. **Figure 2** illustrates the performance comparison of the three models trained with data without augmentation.

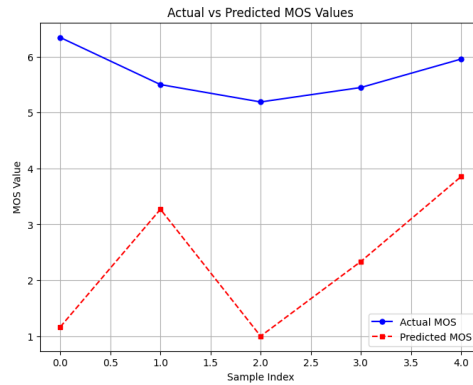
- **ResNet-18:** Achieved a near-perfect training loss (0.0076) and accuracy (100.00%), but a high validation loss (7.1223) and modest validation accuracy (33.33%) indicate overfitting. The test MAE (3.4546) was the highest, with a test accuracy of 0.00%, reflecting poor generalization. Weak positive correlations (Pearson: 0.1416, Spearman: 0.4000) suggest limited alignment with actual MOS values.
- **EfficientNet-B0:** Recorded a training loss of 0.2285 and accuracy of 76.47%, less overfitting than ResNet-18 and MobileNet V2 on the training set. However, a validation loss of 8.8761 and accuracy of 0.00% confirm poor generalization. The test MAE (2.5341) was moderate, and a test accuracy of 20.00% was the highest among non-augmented models, though negligible negative correlations (Pearson: -0.0429 , Spearman: -0.0513) indicate random predictions.
- **MobileNet V2:** Demonstrated exceptional training performance (loss: 0.0213, accuracy: 100.00%) but the highest validation loss (12.6013) and a validation accuracy of 0.00%, signaling severe overfitting. The test MAE (2.3092) was the lowest, yet a test accuracy of 0.00% and negative correlations (Pearson: -0.4324 ,

Spearman: -0.1000) reveal an inverse relationship between predicted and actual MOS values.

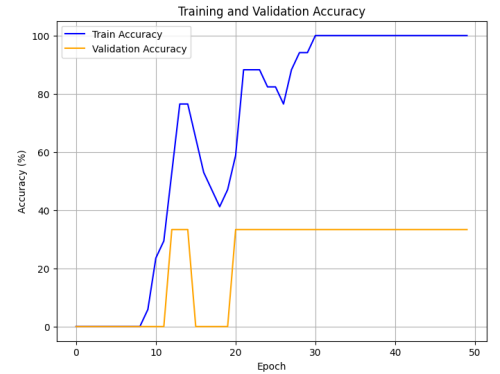
Table 2. Performance metrics without augmentation.

Model	Train Loss	Val Loss	Train Acc (%)	Val Acc (%)	Test MAE	Test Acc (%)	Pearson Corr	Spearman Corr
ResNet-18	0.0076	7.1223	100.00	33.33	3.4546	0.00	0.1416	0.4000
EfficientNet-B0	0.2285	8.8761	76.47	0.00	2.5341	20.00	-0.0429	-0.0513
MobileNet V2	0.0213	12.6013	100.00	0.00	2.3092	0.00	-0.4324	-0.1000

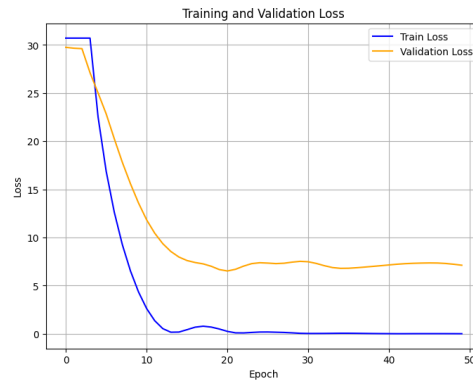
Discussion: Without augmentation, all models overfit, with training performance far exceeding validation outcomes. MobileNet V2’s low MAE (2.3092) suggests closer absolute predictions, but its negative correlations and zero test accuracy indicate systematic errors. EfficientNet-B0’s non-zero test accuracy (20.00%) offers a slight advantage, though its predictions lack meaningful correlation. ResNet-18’s balanced architecture yields weak positive correlations but the highest MAE, reflecting poor predictive accuracy. The absence of data variability likely causes these models to memorize the training set rather than learn generalizable features.



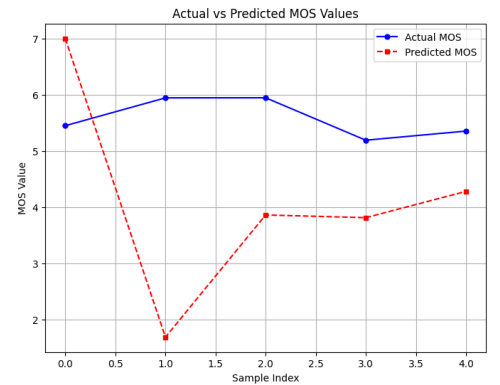
(a) ResNet-18: MOS



(b) ResNet-18: Acc



(c) ResNet-18: Loss



(d) MobileNet: MOS

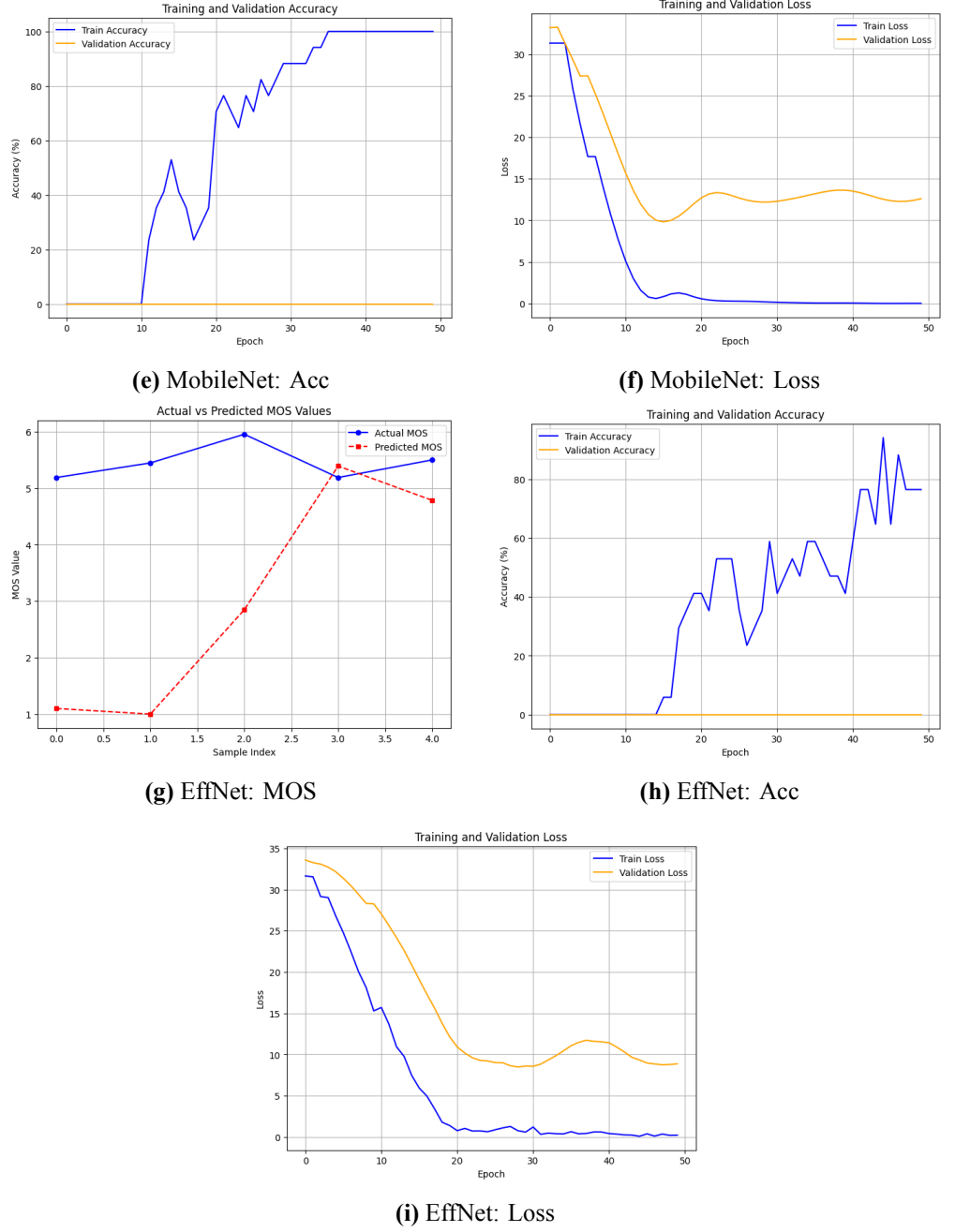


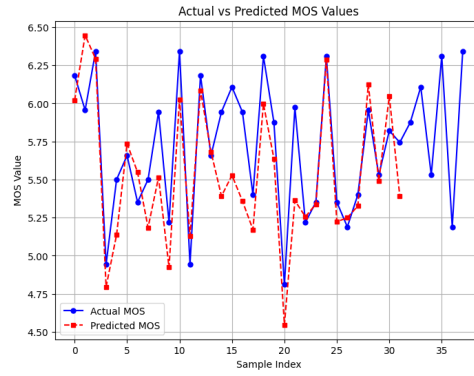
Figure 2. Performance comparison of ResNet-18, MobileNet V2, and EfficientNet-B0 without augmentation

4.3. Results with augmentation

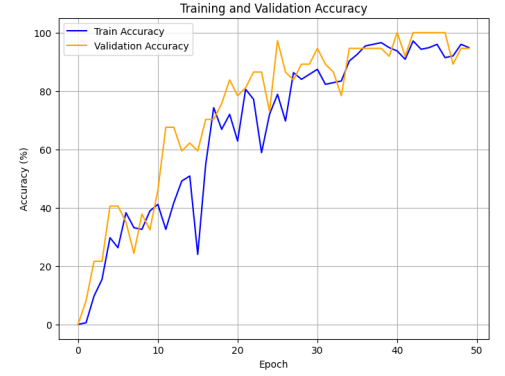
Table 3 presents the performance metrics with augmentation, showing marked improvements in generalization and predictive accuracy of ResNet-18 where augmentation has been added to the model. **Figure 3** illustrates the performance comparison of the three models trained with data augmentation; ResNet-18 had superior MOS that was the closest to the actual value compared to the other 2 models of MobileNet V2 and EfficientNet-B0; also, the training and validation accuracy showed higher values while having the lowest loss among all. The values for actual versus predicted MOS in **Table 4** prove the fact that our proposed model could gain the highest possible similarity of actual versus predicted MOS compared to other models. In conclusion, ResNet-18 showed superior performance, showing training and validation loss convergence, improved accuracy over epochs, and strong alignment

between actual and predicted Mean Opinion Score (MOS) values, highlighting the model's excellent generalization capabilities.

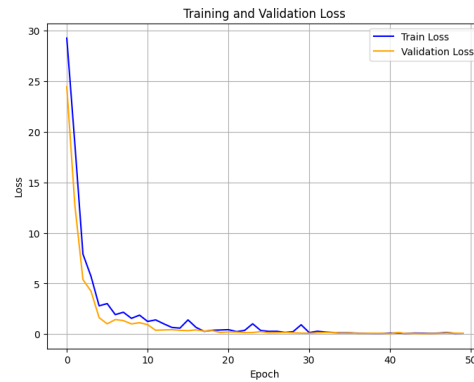
- **ResNet-18:** Exhibited near-identical training (0.0848) and validation (0.0844) losses, with high accuracies (94.86% and 94.59%), indicating excellent generalization. The test MAE (0.2408) and test accuracy (97.37%) were the best overall, supported by strong correlations (Pearson: 0.8364, Spearman: 0.8354), reflecting robust MOS prediction.
- **EfficientNet-B0:** Improved with augmentation, but a training loss of 0.9418 and validation loss of 2.0890 suggest residual overfitting. Low accuracies (45.71% train, 32.43% validation) and a test MAE of 0.9635 with 39.47% test accuracy indicate limited performance. Moderate correlations (Pearson: 0.4651, Spearman: 0.3303) show some predictive capability, though inferior to the other models.
- **MobileNet V2:** Achieved closely aligned training (0.1133) and validation (0.1334) losses, with high accuracies (92.00% train, 89.19% validation). The test MAE (0.2214) was the lowest, and a test accuracy of 92.11% was strong, with robust correlations (Pearson: 0.8285, Spearman: 0.8166), closely rivaling ResNet-18.



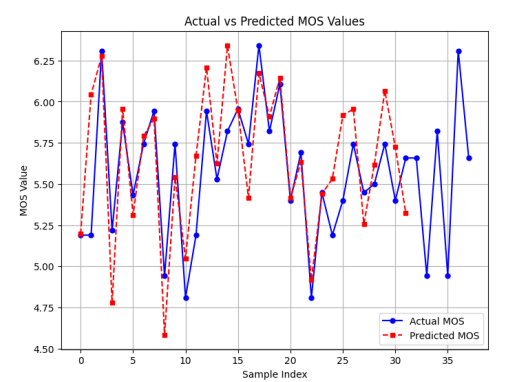
(a) ResNet-18: MOS



(b) ResNet-18: Acc



(c) ResNet-18: Loss



(d) MobileNet: MOS

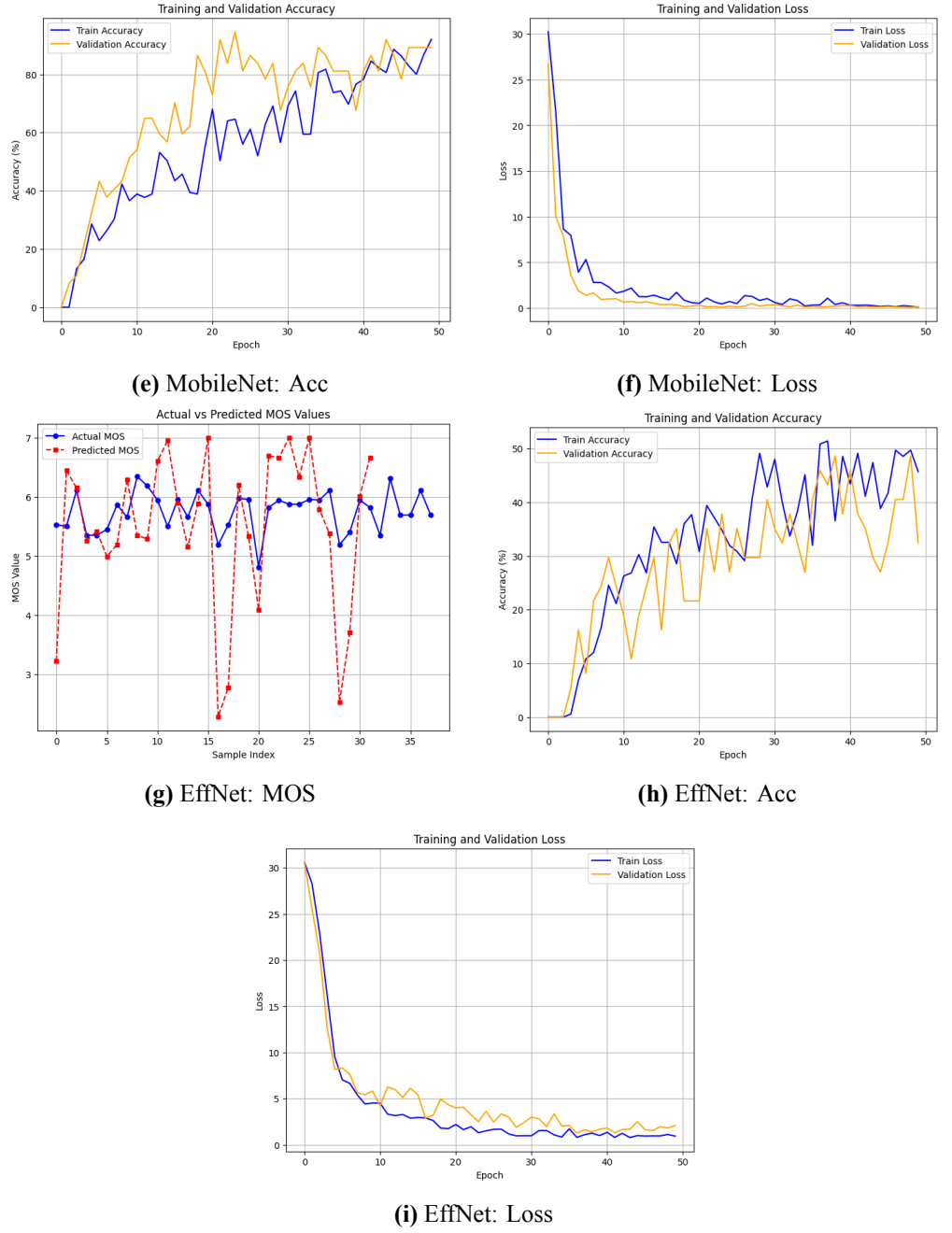


Figure 3. Performance comparison of ResNet-18, MobileNet V2, and EfficientNet-B0 with augmentation (continued).

Table 3. Performance metrics with augmentation

Model	Train Loss	Val Loss	Train Acc (%)	Val Acc (%)	Test MAE	Test Acc (%)	Pearson Corr	Spearman Corr
ResNet-18	0.0848	0.0844	94.86	94.59	0.2408	97.37	0.8364	0.8354
EfficientNet-B0	0.9418	2.0890	45.71	32.43	0.9635	39.47	0.4651	0.3303
MobileNet V2	0.1133	0.1334	92.00	89.19	0.2214	92.11	0.8285	0.8166

Table 4. Actual vs predicted MOS for ResNet-18, EfficientNet-B0, and MobileNet-V2.

Index	Actual MOS (ResNet-18)	Predicted MOS (ResNet-18)	Actual MOS (EfficientNet-B0)	Predicted MOS (EfficientNet-B0)	Actual MOS (MobileNet-V2)	Predicted MOS (MobileNet-V2)
0	6.1842	6.0425	5.5294	2.1207	5.1892	5.8625
1	5.9556	6.1534	5.5000	6.1655	5.1892	5.7402
2	6.3421	5.9351	6.1081	4.9898	6.3095	6.2027
3	4.9429	5.1443	5.3514	5.2576	5.2188	4.7955
4	5.5000	5.1450	5.3514	5.8140	5.8750	5.8701
5	5.6579	5.9283	5.4474	5.5743	5.4324	4.9960
6	5.3514	5.4924	5.8636	4.8459	5.7436	5.8146
7	5.5000	5.2407	5.6579	6.1790	5.9444	5.8032
8	5.9444	5.5125	6.3421	4.9240	4.9429	4.8057
9	5.2188	4.9055	6.1842	4.3309	5.7436	5.8251
10	6.3421	6.1974	5.9429	6.6586	4.8095	4.5717
11	4.9429	5.1440	5.5000	5.9791	5.1892	4.9739
12	6.1842	5.8779	5.9556	5.9445	5.9429	5.8684
13	5.6579	5.5333	5.6579	5.9366	5.5294	5.8548
14	5.9444	5.7376	6.1081	5.1396	5.8205	6.0183
15	6.1081	5.5540	5.8750	5.5415	5.9556	5.9419
16	5.9429	5.3263	5.1892	2.2698	5.7436	5.6769
17	5.4000	5.1983	5.5294	2.5789	6.3421	6.0322
18	6.3095	5.9978	5.9737	6.6496	5.8205	6.3114
19	5.8750	5.7634	5.9556	6.5464	6.1081	5.9189
20	4.8095	4.5185	4.8095	3.9665	5.4000	5.4168
21	5.9737	5.4357	5.8205	7.3701	5.6905	5.8520
22	5.2188	4.9005	5.9429	6.6461	4.8095	4.5113
23	5.3514	5.5541	5.8750	7.1598	5.4474	5.1918
24	6.3095	5.7613	5.8750	6.5885	5.1892	5.8471
25	5.3514	5.2667	5.9556	5.9426	5.4000	5.5687
26	5.1892	5.2101	5.9444	5.4595	5.7436	6.1152
27	5.4000	5.3021	6.1081	5.0008	5.4474	5.1761
28	5.9556	5.9390	5.1892	2.4646	5.5000	5.4826
29	5.5294	5.6041	5.4000	3.9795	5.7436	5.7536
30	5.8205	5.9217	5.9444	6.0653	5.4000	5.3728
31	5.7436	5.5792	5.8205	6.6049	5.6579	5.7726
32	5.8750	5.5171	5.3514	4.9411	5.6579	5.4657
33	6.1081	5.6796	6.3095	6.6076	4.9429	4.7879
34	5.5294	5.5396	5.6905	5.3899	5.8205	6.0629
35	6.3095	6.3463	5.6905	4.7841	4.9429	5.0802
36	5.1892	5.3153	6.1081	5.4395	6.3095	6.6357
37	6.3421	6.1061	5.6905	4.0222	5.6579	5.9006

Discussion: Augmentation significantly enhances performance by reducing overfitting and improving generalization. ResNet-18 emerges as the top performer, with the highest test accuracy (97.37%) and strong correlations, leveraging its balanced depth effectively with augmented data. MobileNet V2 follows closely, with the lowest MAE (0.2214) and high accuracy (92.11%), benefiting from its lightweight design. EfficientNet-B0, despite improvement, underperforms, likely due to its complexity requiring more data or optimization.

Although we report 97.37% accuracy within a $\pm 10\%$ MOS threshold, we acknowledge that this range (± 0.9 on a 0–9 scale) may mask fine-grained prediction errors. Therefore, we emphasize that MAE (0.2408) and correlation scores (Pearson: 0.8364) offer a more precise view of prediction performance. For context, prior work (e.g., Bosse et al. [26]) reports comparable MAE on TID2013, suggesting our results are competitive but not a dramatic leap beyond the state of the art.

4.4. Comparative analysis

Here, we will discuss the impact of augmentation on model performance, highlight the strengths of each architecture, and analyze the improvements in correlation trends, particularly how augmentation aligns predictions with human perception. The Comparison of all the results of all the various methods for IQA is shown in the **Table 5**.

4.4.1. Impact of augmentation

Without augmentation, models overfit severely (e.g., MobileNet V2: training loss 0.0213 vs. validation loss 12.6013). Augmentation aligns training and validation performance (e.g., ResNet-18: 0.0848 vs. 0.0844) and boosts test accuracy (e.g., from 0.00% to 97.37% for ResNet-18), demonstrating its role in simulating real-world variability. Notably, the performance leap from 0% to 97.37% accuracy for ResNet-18 is driven primarily by data augmentation. Similar improvements across all models highlight the importance of sample diversity. While our architecture benefits from this, its raw performance without augmentation is poor, indicating potential fragility. We thus view augmentation not as a novel contribution but as a necessary preprocessing step. In contrast, Li et al.’s self-supervised aggregation offers a more principled approach to addressing data scarcity [23]. **Figure 4** shows the significant improvement in model accuracy achieved through data augmentation, highlighting its critical role across all evaluated architectures. This Figure demonstrates the significant role that augmentation plays in achieving high performance across all tested architectures.

4.4.2. Model strengths

ResNet-18 excels with augmentation due to its balanced architecture, while MobileNet V2 offers efficiency and competitive accuracy, making it suitable for resource-constrained applications. EfficientNet-B0’s deeper structure yields the weakest results, suggesting a mismatch with the dataset size or augmentation strategy.

4.4.3. Correlation trends

Augmentation improves correlations dramatically (e.g., MobileNet V2: Pearson from -0.4324 to 0.8285), aligning predictions with human perception, whereas without augmentation, negative or weak correlations dominate.

4.5. Qualitative analysis

To support our claim that the network captures high-frequency features, we visualize activation maps using Grad-CAM. As shown in **Figure 5 (5a–5c)**, the network focuses on edges and texture-rich regions, which are most impacted by Gaussian blur. This visual evidence highlights the model’s sensitivity to detail loss, which validates the role of residual connections in enhancing perceptual sharpness detection. The network (ResNet-18) consistently focuses on edge and texture areas, confirming its sensitivity to high-frequency degradation caused by Gaussian blur.

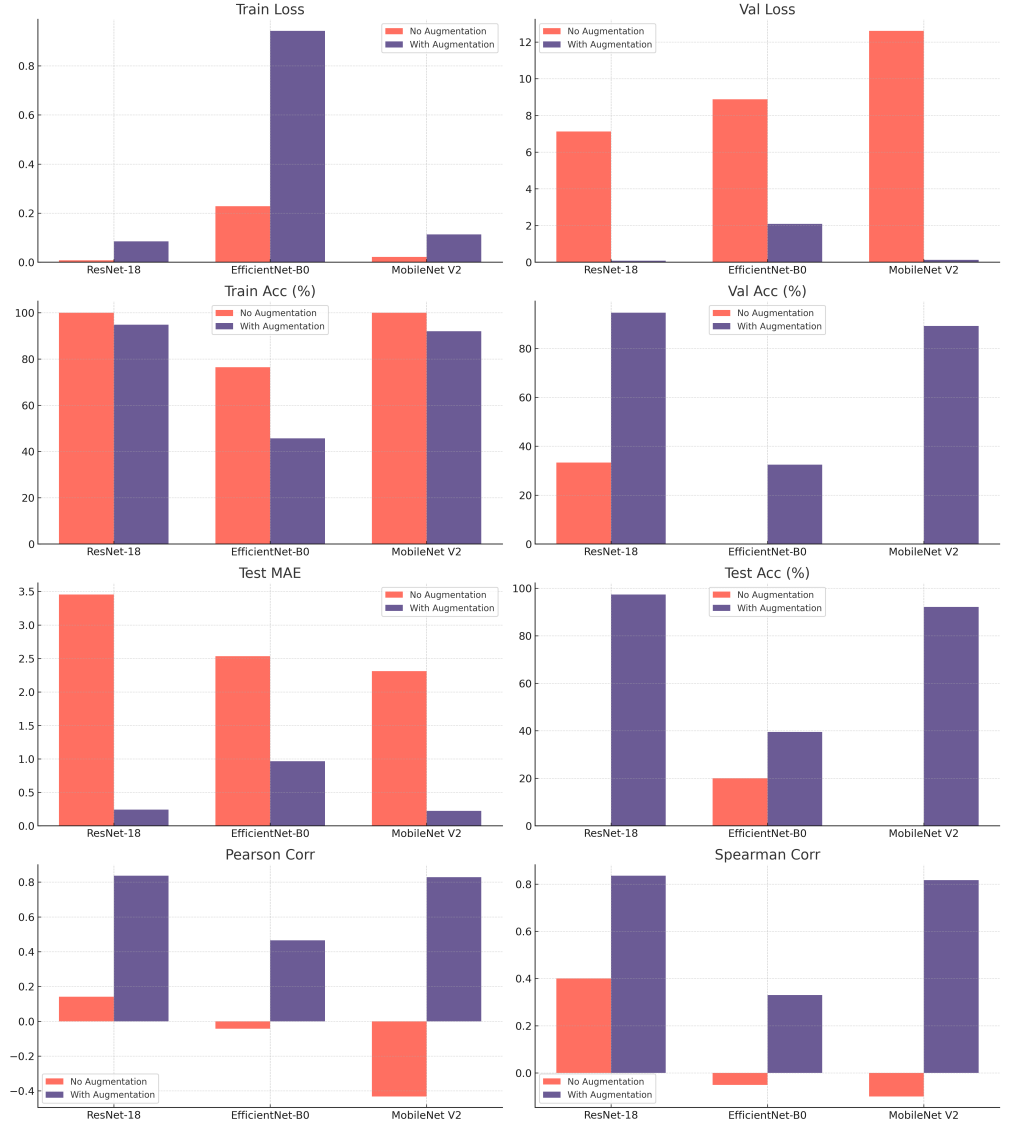


Figure 4. Accuracy comparison before and after augmentation.

While we compared SNR against widely used full-reference (FR) methods (PSNR, SSIM), integrating comparisons with blur-specific no-reference (NR) metrics such as CPBD(cumulative probability of blur detection) [27] and advanced deep methods like SFA (Semantic Feature Aggregation) would strengthen our evaluation. Moreover, the hallucinated-IQA approach proposed by Lin and Wang) [28] demonstrates how adversarial learning can generate pseudo-references to achieve high no-reference performance by leveraging implicit FR information. Including such hybrid baselines in future benchmarks would offer a more comprehensive perspective on performance and reference-dependency.

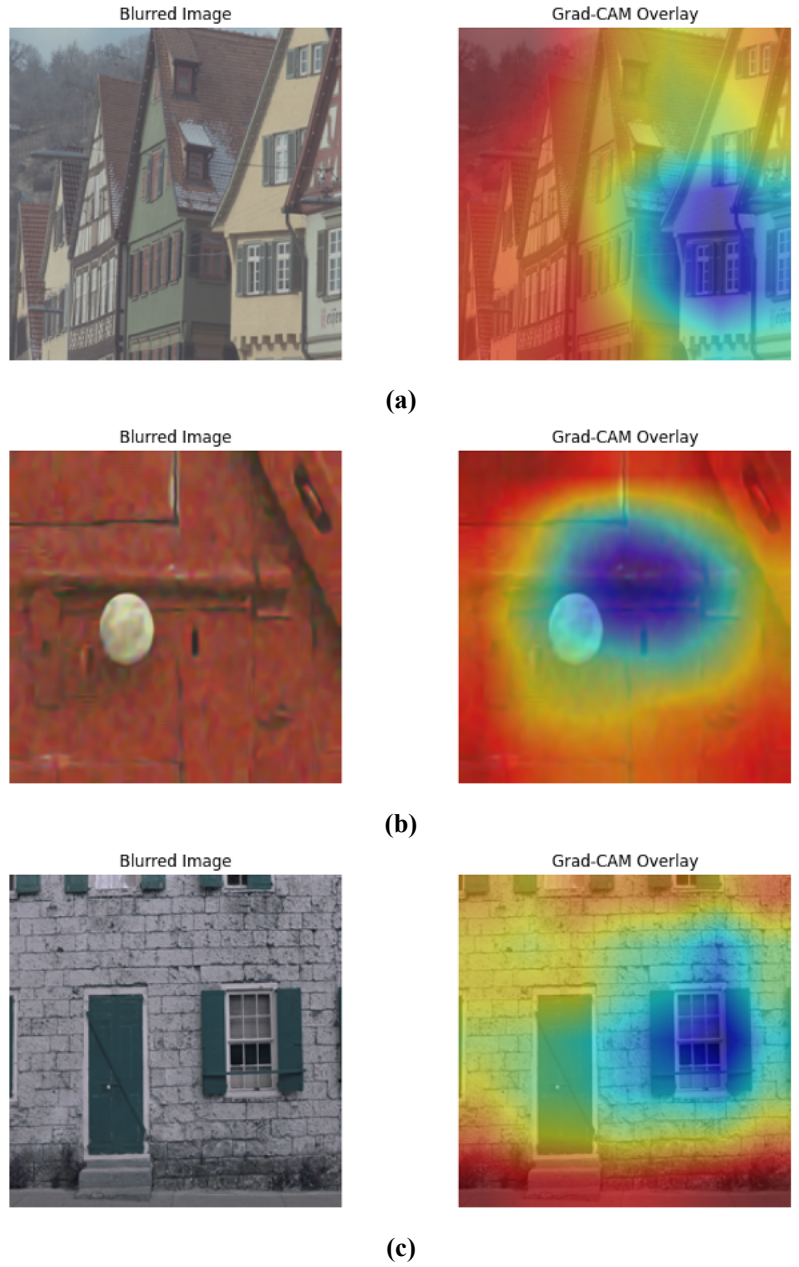


Figure 5. Grad-CAM visualizations over blurred image patches.

Table 5. Comparison of results across different methods for IQA.

Method	Type	MAE ↓	Pearson ↑	Notes
MobileNet V2	CNN	0.2214	0.8201	Outperforms ResNet-18 in MAE
ResNet-18 (Ours)	CNN	0.2408	0.8364	Main model used
CPBD	Blur-specific	—	0.70	Classical sharpness metric
SFA	Blur-specific	—	0.72	Spatial frequency-based blur detection

4.6. Limitations

Despite the promising results achieved with the proposed SNR model, several limitations remain. These limitations primarily stem from the model's specialization and the challenges inherent in its architecture and dataset dependencies, which may restrict its broader applicability.

- **Specialization:** SNR is restricted to Gaussian blur, limiting applicability to other distortions (e.g., noise, motion blur).
- **Reference Dependency:** The Siamese architecture relies on paired images, complicating deployment in true NR scenarios.
- **Computational Cost:** Dual ResNet-18 branches incur overhead, though mitigated by its lightweight design.
- **Dataset Constraints:** TID2013’s modest size may limit generalizability, particularly for complex models like EfficientNet-B0.

Despite ResNet-18 being relatively lightweight, the twin architecture effectively doubles the computational requirements due to the parallel processing of both input images. This results in increased inference time and memory footprint, which may be problematic for real-time applications in edge or mobile devices. Future efforts will focus on integrating lightweight architectures such as MobileNetV3 or applying model pruning, quantization, and knowledge distillation to reduce computational complexity.

Another limitation lies in the dataset size. TID2013 contains a limited number of reference-distorted image pairs, which may not be sufficient to train more complex models or ensure generalizability across real-world scenarios. While data augmentation and patch-based training were applied to mitigate this issue, expanding the training with additional datasets or cross-dataset evaluation would better reflect performance in practical applications.

Our model is currently constrained to Gaussian blur, and its dual-branch ResNet-18 design imposes computational overhead ($\sim 2\times$ compared to single-branch networks). While MobileNet V2 provides a lightweight alternative with competitive MAE (0.2214), its performance remains slightly lower in correlation metrics. Future work will include benchmarked efficiency testing, cross-distortion training (e.g., motion blur), and adaptation for true NR scenarios using pseudo-references or knowledge distillation.

4.7. Future directions

While the proposed Siamese Network Regression (SNR) model achieves state-of-the-art performance in Gaussian blur detection on the TID2013 dataset, as demonstrated in our experiments, it currently exhibits limitations in generalization to other distortion types such as noise, motion blur, and compression artifacts. To enhance its practical utility and robustness for broader real-world applications, future work will prioritize expanding the training data to encompass multiple distortion types and developing a distortion-aware or multi-branch architecture capable of handling diverse degradations within a unified framework. Additionally, addressing the model’s reliance on paired reference-distorted images is crucial for enabling true no-reference (NR) scenarios. We plan to explore pseudo-reference generation techniques, such as using pretrained image restoration networks or deblurred images, alongside semi-supervised learning and knowledge distillation to reduce reference dependency and improve scalability. To optimize computational efficiency, particularly given the twin-branch ResNet-18 design that doubles inference cost, we aim to integrate lightweight architectures like MobileNetV3 and apply model compression strategies including pruning and quantization. Further enhancements will involve testing advanced data augmentation methods (e.g., CutMix, RandAugment, and color jittering) to improve generalization, alongside hyperparameter tuning to refine performance. Finally, cross-dataset validation on larger datasets such as KADID-10k and evaluation in real-world blur scenarios will be pursued to ensure robustness across varied

conditions. These directions collectively aim to overcome the current constraints of specialization, reference dependency, computational overhead, and dataset limitations, paving the way for a more versatile and deployable IQA framework.

5. Conclusion

This study introduced SNR (Siamese Network with ResNet-18), a novel framework for Gaussian blur-specific Full Reference Image Quality Assessment (FR-IQA), addressing the gap in full-reference methods for blur degradation. By leveraging a Siamese architecture with ResNet-18 to process paired reference and blurred images, SNR achieves state-of-the-art performance on the TID2013 dataset. Data augmentation proved indispensable, transforming ResNet-18's test accuracy from 0.00% to 97.37% (MAE: 0.2408), while MobileNet V2 offered lightweight efficiency (92.11% accuracy, 0.2214 MAE). EfficientNet-B0 underperformed (39.47%), highlighting challenges with complex architectures on limited datasets. The Siamese design bridges full-reference (FR) and no-reference (NR) paradigms, using residual connections to capture blur features like edge sharpness. Although SNR excels in Gaussian blur assessment, its reliance on paired images and limited generalization to other distortions constrain broader applicability. Although the proposed Siamese Network Regression (SNR) architecture achieves state-of-the-art results in Gaussian blur detection, it exhibits several limitations. Chief among these is its reliance on paired reference-distorted images, which restricts its applicability in no-reference scenarios. Additionally, the use of a twin-branch ResNet-18 backbone effectively doubles the inference computational cost, posing challenges for deployment in real-time or low-resource environments. The relatively small size of the TID2013 dataset further constrains the training of deeper or more complex models. To address these issues and enable broader practical deployment, future research will focus on incorporating more efficient backbone architectures, data-efficient learning strategies, and synthetic pseudo-reference generation.

Author contributions: Writing—original draft preparation, OZ, EA, MB and AAB; methodology, OZ and JHJ; writing review and editing, ZM, BKK and JHJ; resources, BKK; conceptualization, OZ; visualization, JHJ; supervision, AAS. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

1. Narvekar, N.D., Karam, L.J. A no-reference perceptual image sharpness metric based on a cumulative probability of blur detection. In: Proceedings of the International Workshop on Quality of Multimedia Experience (QoMEX); 18 September 2009; San Diego, CA, USA. pp. 87–91. doi: 10.1109/QOMEX.2009.5246982
2. Kang L, Ye P, Li Y, et al. Convolutional neural networks for no-reference image quality assessment. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 23–28 June 2014; Columbus, OH, USA. pp. 1733–1740.
3. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2016; 38(9): 1777–1787.
4. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*. 2004; 13(4): 600–612. doi: 10.1109/TIP.2003.819861
5. Wang Z, Simoncelli EP, Bovik AC. Multiscale structural similarity for image quality assessment. *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*. 2003; 2: 1398–1402. doi: 10.1109/ACSSC.2003.1292216
6. Zhang L, Zhang L, Mou X, Zhang D. Fsim: A feature similarity index for image quality assessment. *IEEE Transactions*

- on Image Processing. 2011; 20(8): 2378–2386. doi: 10.1109/TIP.2011.2109730
7. Mittal A, Moorthy AK, Bovik AC. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*. 2012; 21(12): 4695–4708. doi: 10.1109/TIP.2012.2214050
8. Moorthy AK, Bovik AC. A two-step framework for blind image quality assessment. *IEEE Transactions on Image Processing*. 2010; 20(9): 2463–2474. doi: 10.1109/TIP.2011.2123162
9. Saad MA, Bovik AC, Charrier C. Blind image quality assessment: A natural scene statistics approach in the dct domain. *IEEE Transactions on Image Processing*. 2012; 21(8): 3339–3352. doi: 10.1109/TIP.2012.2191563
10. Allaert B, Mennesson J, Bilasco IM, et al. Impact of the face registration techniques on facial expressions recognition. *Signal Processing: Image Communication*. 2018; 61: 44–53.
11. Nawaz T, Poiesi F, Cavallaro A. Assessing tracking assessment measures. In: *Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP)*; 27–30 October 2014; Paris, France. pp. 441–445.
12. Ferzli R, Karam LJ. A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB). *IEEE transactions on image processing*. 2009; 18(4): 717–728.
13. Craciun P, Zerubia J. Unsupervised marked point process model for boat extraction in harbors from high resolution optical remotely sensed images. In: *Proceedings of the 2013 IEEE International Conference on Image Processing*; 15–18 September 2013; Melbourne, Australia. pp. 4122–4125.
14. Joshi P, Prakash S. No-reference image quality assessment using contourlet transform. *Multimedia Tools and Applications*. 2018; 77(13): 16977–16994.
15. Bosse S, Maniry D, Müller KR, et al. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Transactions on image processing*. 2017; 27(1): 206–219.
16. Liu X, Van De Weijer J, Bagdanov AD. Rankiq: Learning from rankings for no-reference image quality assessment. *arXiv preprint*. 2017. arXiv:1707.08347
17. Cong H, Fu L, Zhang R, et al. Image quality assessment with gradient siamese network. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 18–24 June 2022; New Orleans, LA, USA. pp. 1201–1210.
18. Xu K, Liao L, Xiao J, et al. Boosting image quality assessment through efficient transformer adaptation with local feature enhancement. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 16–22 June 2024; Seattle, WA, USA. pp. 2662–2672.
19. Avanaki NJ, Ghildyal A, Barman N, et al. Lar-iqa: A lightweight, accurate, and robust no-reference image quality assessment model. *arXiv preprint*. 2024; 17057.
20. Liu, H., Li, Z., Liao, S., et al. TransIFC: Invariant cues-aware feature concentration learning for efficient fine-grained bird image classification. *IEEE Transactions on Multimedia*, 2023.
21. Yi Deng, Jun Ma, Ziyi Wu, Wanhao Wang, Hai Liu. DSR-Net: Distinct selective rollback queries for road cracks detection with detection transformer. *Digital Signal Processing*. 2025; 164: 105266. doi: 10.1016/j.dsp.2025.105266
22. Zhang, Y., Li, G., Lu, J., et al. UGeNet: Uncertainty-aware gaze estimation with discriminative embedding. In: *Proceedings of the European Conference on Computer Vision (ECCV)*; 2022. pp. 253–270.
23. Li L, Wu D, Wu J, et al. Blind image quality assessment using semantics and perception. *IEEE Access*. 2017; 5: 24303–24312. doi: 10.1109/ACCESS.2017.2767858
24. Zhang W, Li L, Wu J, Lin W. Blind image quality assessment using a no-reference structural similarity index. *IEEE Signal Processing Letters*. 2018; 25(11): 1658–1662. doi: 10.1109/LSP.2018.2870778
25. Ren H, Chen D, Wang Y. RAN4IQA: restorative adversarial nets for no-reference image quality assessment. Available online: <http://arxiv.org/abs/1712.05444> (accessed on 2 June 2025).
26. Bosse S, Maniry D, Wiegand T, et al. A deep neural network for image quality assessment. In: *Proceedings of the IEEE international conference on image processing (ICIP)*; 25–28 September 2016; Phoenix, AZ, USA. pp. 3773–3777.
27. Li D, Jiang T, Jiang M. Exploiting high-level semantics for no-reference image quality assessment of realistic blur images. In: *Proceedings of the 25th ACM international conference on Multimedia*; 23–27 October 2017; New York, NY, USA. pp. 378–386.
28. Lin KY, Wang G. Hallucinated-IQA: No-reference image quality assessment via adversarial learning. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 18–23 June 2018; Salt Lake City, UT, USA. pp. 732–741.