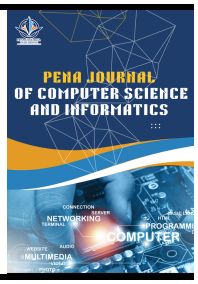
PENA CENDEKIA PUBLISHING
NURTURING IDEAS, ENRICHING KNOWLEDGE

Pena Journal of Computer Science and Informatics

Journal homepage:
https://penacendekia.com.my/index.php/pjcsi/index
ISSN: XXXX-XXXX



Deep Learning-based Blind Image Quality Assessment using Extreme Learning Machine

Nutveesa Verak[1,*], Phaklen Ehkan[1,2,*], Ruzelita Ngadiran[1,2], Suwimol Jungjit[3], Siraya Sitthisarn[3], Mohd Nazri Mohd Warip[1,2], Mohd Zaizu Ilyas[1,2], Fazrul Faiz Zakaria[1,2]

[1] Faculty of Electronic Engineering & Technology, Universiti Malaysia Perlis (UniMAP), Arau, Perlis, Malaysia
[2] Advanced Computing Centre of Excellence, Universiti Malaysia Perlis (UniMAP), Arau, Perlis, Malaysia
[3] Department of Computer and Information Technology, Faculty of Science, Thaksin University, Phatthalung, Thailand

ARTICLE INFO

Article history:
Received 29 January 2025
Received in revised form 25 February 2025
Accepted 1 March 2025
Available online 20 March 2025

Keywords:
Image quality assessment; Convolutional Neural Network (CNN); Extreme Learning Machine (ELM)

ABSTRACT

In many image processing applications, Blind Image Quality Assessment (BIQA) plays a crucial role when the reference image is unavailable. However, existing BIQA methods often lack consistency with human perception and are limited to specific types of distortion. This research introduces a novel approach for image quality assessment by combining Convolutional Neural Network (CNN) feature extraction and Extreme Learning Machine (ELM) regression. By leveraging a pre-trained ResNet CNN model, features are extracted from distorted images and used to train the ELM model. The proposed method achieves accelerated convergence in deep model training and significantly enhances accuracy in predicting image quality compared to traditional regression methods. Experimental results demonstrate the superiority of the proposed approach over state-of-the-art BIQA methods, making it a promising solution for automated image quality evaluation in various applications such as image processing and computer vision.

1. Introduction

Accurately predicting human quality judgments in picture quality models is crucial for enhancing consumer satisfaction and monitoring the quality of widely distributed visual content. These models also serve as benchmarks for picture processing algorithms such as compression engines, denoising algorithms, and super-resolution systems, which significantly impact the perceived quality of images. However, modelling these algorithms to align with the human visual system poses a significant challenge, as computers store data as bits and pixels without a holistic understanding of the larger

* *Corresponding author.*
E-mail address: nutveesa23@gmail.com

* *Corresponding author.*
E-mail address: phaklen@unimap.edu.my

https://doi.org/10.37934/pjcsi.1.1.112

picture. In the past decade, various methods have been proposed by the researchers to address this challenge, including subjective Image Quality Assessment (IQA) relying on human observers and objective IQA employing mathematical models [1]. Objective IQA can be categorized into three types: Full Reference (FR-IQA), Reduced Reference (RR-IQA), and No Reference (NR-IQA) or Blind IQA as shown in Figure 1. Blind image quality assessment (BIQA) remains a highly challenging problem due to the absence of a reference image. Existing BIQA methods typically follow a flowchart, as depicted in Figure 2.

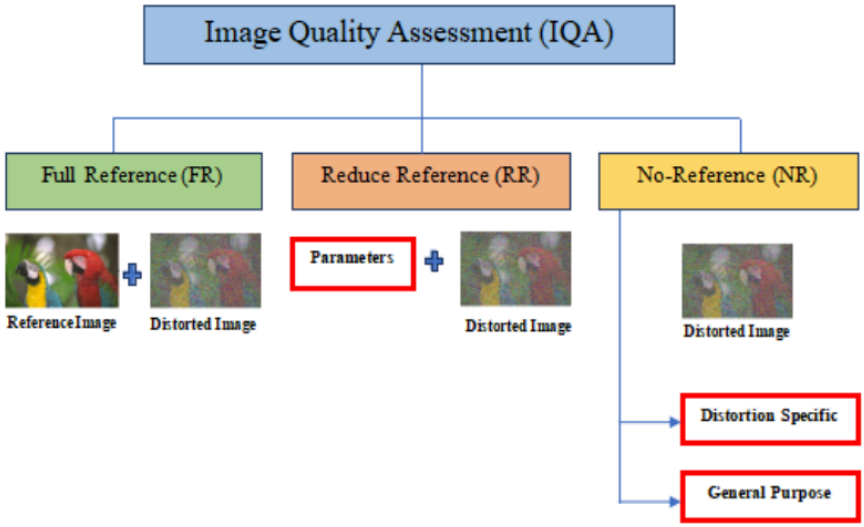


Fig. 1. Objective image quality assessment algorithms classification [2]

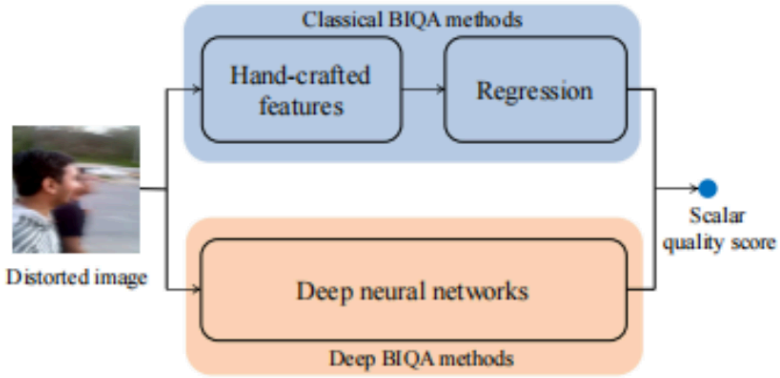


Fig. 2. Block diagram of existing BIQA [3]

Classical BIQA approaches [4-8] typically extract handcrafted features, derived from models based on natural scene statistics to represent the distorted image. These features are then used to train a regression model, such as support vector regression (SVR), to map the feature representations to subjective quality scores. In recent years, deep learning based BIQA methods have gained increasing attention due to the remarkable capability of deep neural networks in learning discriminative features. The Convolutional Neural Network (CNN)-based approaches [9], which automatically learn features from raw images, consistently outperform these handcrafted feature-

based algorithms. CNN-based algorithms could select features that aid in distortion detection and quality prediction.

This research focuses on a Blind Image Quality Assessment (BIQA) framework that combines the power of CNN Resnet and the Extreme Learning Machine (ELM) algorithm. By leveraging the capacity of CNN Resnet to extract discriminative features from distorted images and the advantages of the ELM algorithm as a regression model, the proposed framework aims to provide objective image quality measurements consistent with human perceptual measures. The CNN Resnet extracts complex representations from the distorted images, which are then input to the ELM algorithm for efficient feature mapping and regression. The ELM algorithm is chosen due to its benefits, such as avoiding random initialization of weights and biases, faster learning speed, smaller norm weights, and fewer neuron nodes compared to traditional methods like Support Vector Machine (SVM). The significance of this research lies in its contribution to the advancement of BIQA techniques through the integration of deep learning and efficient feature mapping. By addressing the limitations of using only scalar quality scores and considering the divergent subjective scores from different human raters, the proposed framework harnesses the power of deep learning and efficient feature mapping to enhance the accuracy and efficiency of image quality assessment. The ultimate goal is to provide reliable and automated image quality evaluation, overcoming the challenges of blind image quality assessment. Overall, this research aims to contribute to the development of effective and efficient BIQA methods by combining the CNN Resnet and ELM algorithm. The proposed framework holds promise in overcoming the limitations of existing methods and providing objective image quality measurements consistent with human perception. By leveraging the advancements in deep learning and feature mapping, this research strives to advance the field of BIQA and enable reliable automated image quality evaluation for various applications.

2. Related Theory and Fundamentals

2.1 Convolutional Neural Networks (CNNs)

The framework of image feature extraction using the proposed CNN is shown in Figure 3.

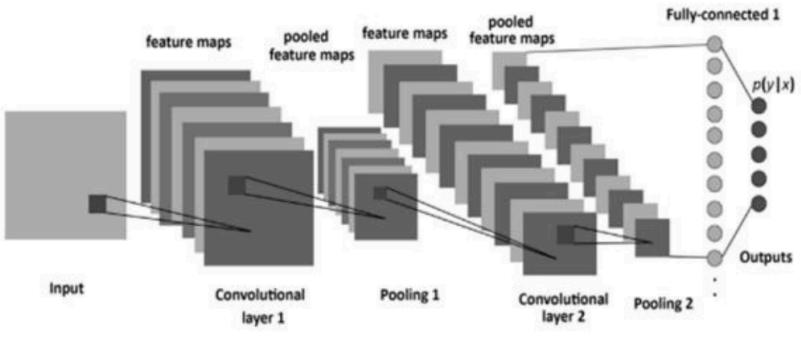


Fig. 3. Framework of feature extraction using CNN [9]

The CNNs are a class of deep neural networks commonly used for processing visual data, including images and videos. CNNs have revolutionized various fields, including computer vision and image processing, due to their ability to automatically learn hierarchical representations from raw input

data. The structure of a CNN is inspired by the organization of the animal visual cortex. It consists of multiple layers, including convolutional layers, pooling layers, and fully connected layers [9]. Convolutional layers: The core building blocks of CNNs. It is responsible for extracting local features from the input images. In the proposed metric, the ResNet model performs the convolutional operations as part of its architecture [9]. The filters (kernels) in the convolutional layer slide over the input image, performing element-wise multiplications and summations to produce feature maps. The size of the filters determines the receptive field, which defines the local region of the input that the filter considers. The convolutional layer applies non-linear activation functions to introduce nonlinearity into the output. Based on Eq. (1), the output $yi\ l$ of the ith filter in a convolutional layer is calculated as follows:

$$y_i^l = s\left(\sum_{j=1}^{C^{l-1}} f_{i,j}^l y_i^{l-i} + b^l\right) \quad (1)$$

where, $b\ l$ represents the bias vector for the convolutional layer, $fi,j\ l$ refers to the ith kernel of the convolutional layer (layer l) that is connected to the jth feature map of the previous layer (layer l-1), $yi\ l{-}i$ represents the output of the ith filter in the previous layer (layer l-1), and s denotes the activation function applied element-wise to the computed sum [2]. Pooling Layers: The pooling layers are inserted after convolutional layers to down sample the spatial dimensions of the feature maps while retaining important information. The two common pooling techniques: max pooling and average pooling are used in this proposed architecture. Max pooling outputs the maximum value within each pooling region, and average pooling computes the average value. Pooling helps reduce the spatial resolution of the feature maps, making the network more computationally efficient and invariant to small spatial shifts in the input. Fully Connected Layers: Unlike convolutional layers, fully connected layers do not have parameter sharing. The first fully connected layer is connected to all the activations from the previous layers. In a fully connected layer, weight (W) and bias (b) vectors are learned from the input of the previous layer. The computed values are then passed forward to the next layer, continuing until the output layer. The output of a fully connected layer is determined by the Eq. (2):

$$Output = W * Input + b \quad (2)$$

This Eq. (2) represents the linear transformation of the input data using the learned weights and biases in the fully connected layer. The output is then used for further processing or as the final prediction of the network.

In our proposed metric, the focus is on using pre-trained ResNet features for image feature extraction. The proposed method utilizes an Extreme Learning Machine (ELM) model, which consists of a single fully connected layer. While traditional CNN architectures often include fully connected layers after the convolutional and pooling layers to learn high-level features and make predictions, our approach takes a different path. Instead of explicitly incorporating fully connected layers as part of the CNN, we leverage the extracted features from the ResNet model directly as inputs to the ELM model. In this framework, the ResNet features serve as the learned representations of the input images, which are then fed into the ELM model acting as a regression model for predicting image quality scores. By bypassing the conventional fully connected layers, we can harness the power of the pre-trained ResNet features to achieve accurate image quality prediction.

2.2 Extreme Learning Machine (ELM)

ELM is a highly efficient learning algorithm for single-hidden layer feedforward neural networks (SLFNs) [10]. It outperforms other techniques like NNs and SVR in terms of both learning accuracy and speed. The purpose of the Extreme Learning Machine (ELM) in the study is to minimize the deviation between the subjective quality scores, denoted as yi , and the predicted scores, represented by the function $f(Ci$). Given that the function $f(Ci$)is expressed as the summation of weighted activation functions, where β is the output weighing vector and gj (Ci)represents the activation function. Each training sample consists of a feature vector, Ci , corresponding to an original or distorted image, and its subjective quality score, yi :

$$f(C_i) = \sum_{j=1}^{L} \beta_j g_j(C_i), i = 1, \dots, N \quad (3)$$

The ELM aims to approximate the N training samples with zero error, such that the difference between $f(Ci$)and yi is minimized ∑ ‖$f(Ci) - yi$‖ N i=1 = 0.

$$g_j(C_i) = g(w_j . C_i + b_j) \quad (4)$$

The activation function gj (Ci) in Eq. (4) can approximate the N training samples with zero error by applying an inner product between the input weight vector $wj$and Ci , and adding the threshold bj (input node). The input weight vectors $wj$and bias terms $bj$are randomly generated based on a continuous probability distribution, and the hidden nodes are also randomly generated and independent of each other. Based on theory [9], it has been proven that the wj (input weighing vector) and bias term bj can be randomly generated based on the continuous probability distribution. Besides, all hidden nodes are also randomly generated as well and independent each other. Hence, the β in Eq. (3) is only the parameter to be measured. This is one reason why ELM has fast learning speed than other feedforward neural networks algorithms [11]. For N training sample in Eq. (3) can be written compactly as:

$$Y_H \beta = Y \quad (5)$$

$$Y_H = \begin{pmatrix} g(w_1 . C_1 + b_1) & \cdots & g(w_L . C_1 + b_L) \\ \vdots & \vdots & \vdots \\ g(w_1 . C_N + b_1) & \cdots & g(w_L . C_N + b_L) \end{pmatrix}_{NxL} \quad (6)$$

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_L \end{pmatrix}_{Lx1}, Y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}_{Nx1} \quad (7)$$

In order to solve Eq. (3) efficiently, the ELM employs the Moore-Penrose generalized inverse concept. The hidden layer output matrix, YH, is computed as shown in Eq. (6), where YHconsists of the activation function values of the hidden nodes for each input feature vector. The output weights vector, β, is estimated analytically using the Moore-Penrose generalized inverse of YH in Eq. (8) as β = YY † H as follow. This approach aims to minimize the norm of the output weights, resulting in

improved generalization performance and higher learning accuracy compared to other neural network algorithms.

$$\beta = YY^{\dagger}{}_{H} \tag{8}$$

In summary, the ELM in the provided code utilizes random generation of input weight vectors, bias terms, and hidden nodes to approximate the subjective quality scores using activation functions. The Moore-Penrose generalized inverse is employed to estimate the output weights, leading to a minimized norm and improved learning accuracy.

3. Experimental

3.1 Experimental Setup

In the proposed metric, the experiment involves evaluating the performance of blind image quality assessment (BIQA) models using four representative IQA databases: LIVE [13], LIVE-C[12], CSIQ[14], and TID2013[15]. These databases contain image data that are crucial for assessing the effectiveness of the models. Three different CNN models were considered: pre-trained AlexNet [16], pre-trained ResNet50 [17], and VGG-16[18]. During the fine-tuning process, these models were used. To enhance their performance, image crops were randomly extracted from the training images. For most databases, 50 crops of size 227 × 227 for AlexNet and VGG16, while 224 × 224 in ResNet50 were extracted per image. However, for TID2013, which has more distorted images, only 25 crops per image were extracted. The learning rate for the fine-tuning process was set as a logarithmically spaced vector within the range of [1e-3, 1e-4] for both models. In the testing stage, overlapped image patches were extracted from each test image. The stride, which determines the step size between consecutive patches, was set to 64 for the fine-tuned deep models in AlexNet, ResNet50 and VGG16. Two metrics, Spearman's rank correlation coefficient (SRCC) and Pearson's linear correlation coefficient (PLCC), were used to evaluate the performance of the learned BIQA models. To ensure reliable evaluation, the samples in each database were randomly divided into three sets: 80% for training, 10% for testing, and 10% for validation. The division was performed in a way that avoided overlap in image content between the sets. This allows for training the models on a large portion of the data while still having separate datasets for testing and validation. The entire experiment was repeated 10 times, and the median SRCC and PLCC values were reported as the final results, providing a robust evaluation of the models' performance. By considering multiple databases, random data division, and repeated experiments, the approach ensures the reliability and accuracy of the assessment.

3.2 Comparison among Different CNN Models

The purpose of comparing different CNN models in blind image quality assessment is to evaluate their effectiveness and suitability for this task. The goal is to determine which model performs best in predicting image quality and to understand their capabilities in extracting relevant features for quality assessment.

Table 1
Comparing the best performance of proposed metric using different CNN

Models	PLCC	SRCC
Alexnet	0.5863	0.6333
Resnet	0.8078	0.803
Vgg16	0.6846	0.7707

Based on the PLCC results in Table 1, ResNet-50 achieved the highest value of 0.8078, indicating a stronger linear correlation with the ground truth scores compared to AlexNet (0.5863) and VGG-16 (0.6846). This suggests that ResNet-50 has a better ability to capture the linear relationship between predicted and ground truth scores, making it more effective for quality prediction in the context of the LIVE database. Moving on to the SRCC results, ResNet-50 again demonstrated the highest value of 0.803, indicating a stronger monotonic correlation with the ground truth scores. VGG-16 followed closely with an SRCC of 0.7707, while AlexNet had a lower SRCC value of 0.6333. The higher SRCC values for ResNet-50 and VGG-16 suggest that they capture the underlying monotonic relationship between predicted and ground truth scores more effectively than AlexNet. In summary, the evaluation results suggest that ResNet-50 (Figure 4) outperformed AlexNet (Figure 5) and VGG-16 (Figure 6) in terms of both PLCC and SRCC metrics for the LIVE database. It exhibited a stronger linear and monotonic correlation with the ground truth scores, indicating its superior performance in blind image quality assessment.

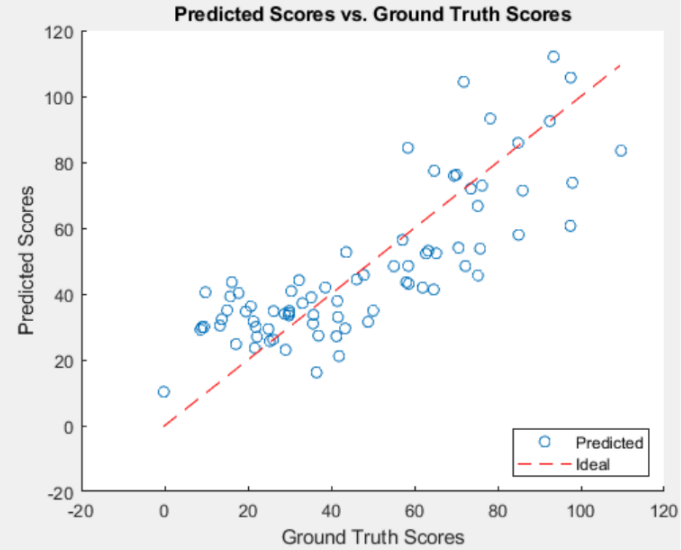


Fig. 4. Resnet

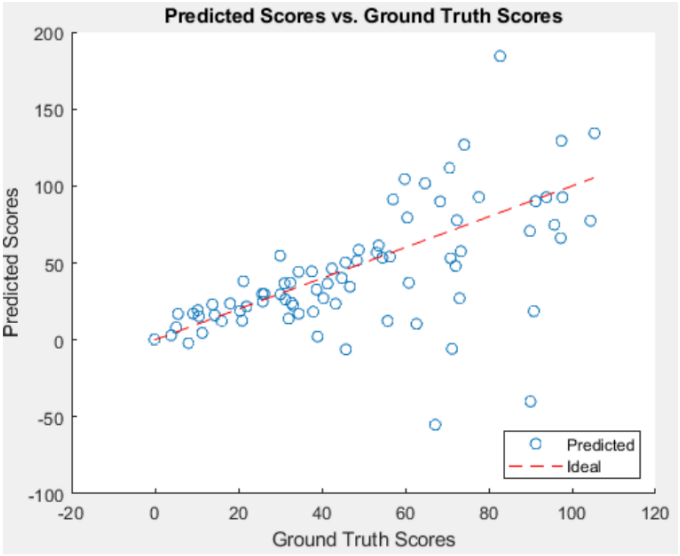


Fig. 5. Alexnet

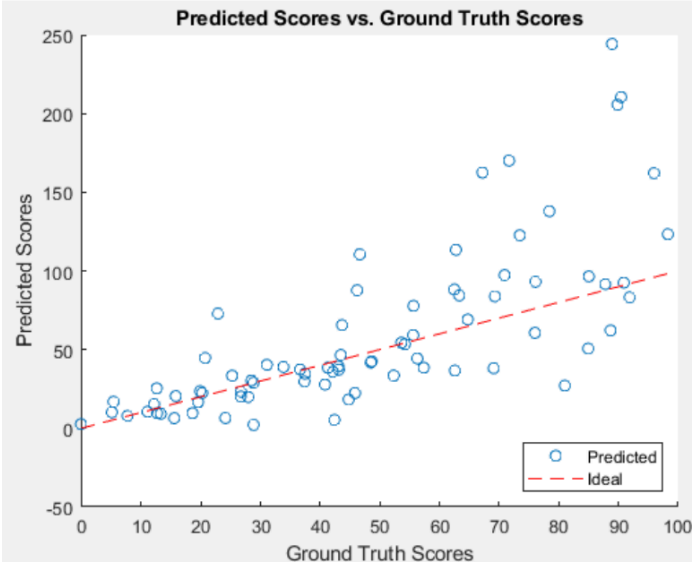


Fig. 6. VGG16

3.3 Comparison with the State-of-the-Art on LIVE Database

In this research, the focus was on evaluating the performance of CNN-ELM based models for blind image quality assessment. To ensure a fair comparison, the proposed models were compared against existing BIQA methods, including DIIVINE [19], CORNIA [20], BRISQUE [21], NIQE[22], HOSA[23], FRIQUEE-ALL[24], and Bosse et al [25] as shown in Table 2. The source codes of these methods were re-run using the same training and testing splits as the CNN-ELM models to avoid any potential bias.

Table 2
Comparison with existing BIQA models in LIVE database

Methods	SRCC	PLCC
DIIVINE [19]	0.58	0.60
CORNIA [20}	0.63	0.66
BRISQUE [21]	0.61	0.65
NIQE [22]	0.43	0.48
HOSA [23]	0.66	0.68
FRIQUEE-ALL [24]	0.69	0.71
Bosse *et al*., [25]	0.67	0.68
Proposed method	0.78	0.84

The performance evaluation of the models employed two key metrics, namely Spearman'sRank Correlation Coefficient (SRCC) and Pearson's Linear Correlation Coefficient (PLCC). These metrics quantified the correlation between the predicted quality scores and the ground truth scores of the images in the LIVE database. Higher SRCC and PLCC values indicate a stronger correlation and better performance in predicting image quality.

From the results, it can be observed that the proposed method achieved the highest performance with an SRCC of 0.78 and a PLCC of 0.84. These values indicate a strong correlation between the predicted quality scores and the ground truth scores, demonstrating the effectiveness of the proposed CNN-ELM based model for blind image quality assessment on the LIVE database. Comparing the proposed method with the existing methods, it outperformed all the competitors. FRIQUEE-ALL had the closest performance with an SRCC of 0.69 and a PLCC of 0.71. Other methods, such as DIIVINE, CORNIA, BRISQUE, NIQE, HOSA, and Bosse et al., achieved lower correlation coefficients, indicating less accurate predictions of image quality compared to the proposed method. These results highlight the superiority of the proposed method in terms of its ability to capture and predict image quality accurately. The higher correlation coefficients signify its effectiveness in assessing the perceptual quality of authentically distorted images. 3.4 Computation complexity between ELM and SVR Based on the experimental results in Table 3, ELM demonstrates a lower computational complexity compared to SVR. This is supported by the training time and overall process running time results. The training time for ELM is 0.5509 seconds, while for SVR it is 1.1649 seconds. This indicates that ELM requires less time to train the model. Similarly, the process running time for ELM is 291.7603 seconds, whereas for SVR it is 292.3127 seconds. Although the difference in process running time is minimal, it still suggests that ELM has a slightly lower computational burden.

These results suggest that ELM has a simpler and more efficient learning algorithm compared to SVR. ELM's one-pass learning approach, where the weights are randomly generated and fixed during training, contributes to its faster training time and overall lower computational complexity. On the other hand, SVR involves solving a quadratic programming problem, which can be more computationally intensive, leading to longer training and process running times. In summary, the experimental results indicate that ELM (Figure 7) exhibits a lower computation complexity than SVR (Figure 8). ELM's faster training time and slightly shorter process running time suggest that it is a computationally efficient approach for the given experimental setup.

Table 3
Comparison of ELM with SVR

Evaluation	ELM	SVR
PLCC	0.7716	0.3456
SRCC	0.8436	0.3177
Training Time, s	0.5509	1.1649
Overall Process Running Time, s	291.7603	292.3127

Fig. 7. ELM

Fig. 8. SVR

4. Conclusions

In conclusion, the results of our experiments demonstrate the effectiveness of the proposed blind image quality assessment metric. The metric surpasses several existing methods by achieving higher correlation coefficients (SRCC and PLCC) with ground truth quality scores. When combined with the

Extreme Learning Machine (ELM) regression model, the proposed metric provides more accurate predictions of perceptual image quality compared to Support Vector Regression (SVR). Moreover, the ELM model exhibits shorter training time, resulting in improved computational efficiency. These findings highlight the potential of the proposed metric in practical image quality assessment applications. By incorporating this metric into image Journal of Advanced Research in Applied Sciences and Engineering Technology Volume XX, Issue X (2022) XX-XX 28 quality assessment systems, their performance can be enhanced, leading to more reliable evaluation of image quality. For future work, to further improve the performance, the pooling algorithms in ELM architecture can be enhanced. Specifically, exploring adaptive pooling, spatial pooling, and attention-based pooling techniques can lead to more effective and better model performance. In addition, this algorithm can be expanded to video quality assessment. These approaches enable the network to dynamically adjust pooling operations, capture spatial relationships, and focus on salient features, respectively. Additionally, extending the algorithm to video quality assessment can broaden its applicability and enable more comprehensive evaluation of visual content.

Acknowledgement
This author gratefully acknowledges financial support from the Malaysian Ministry of Higher Education (MOHE) under the Fundamental Research Grant Scheme (FRGS/1/2020/ICT02/UNIMAP/ 02/5) and Universiti Malaysia Perlis (UniMAP)

References

[1] Mohammadi, Pedram, Abbas Ebrahimi-Moghadam, and Shahram Shirani. "Subjective and objective quality assessment of image: A survey." *arXiv preprint arXiv:1406.7799* (2014).
[2] Ravela, Ravi S., "NO REFERENCE IMAGE QUALITY ASSESSMENT" (2019). Electrical Engineering Theses. Paper 43.
[3] Zeng, Hui, Lei Zhang, and Alan C. Bovik. "Blind image quality assessment with a probabilistic quality representation." In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 609-613. IEEE, 2018.
[4] Moorthy, Anush Krishna, and Alan Conrad Bovik. "Blind image quality assessment: From natural scene statistics to perceptual quality." *IEEE transactions on Image Processing* 20, no. 12 (2011): 3350-3364.
[5] Ye, Peng, Jayant Kumar, Le Kang, and David Doermann. "Unsupervised feature learning framework for no-reference image quality assessment." In *2012 IEEE conference on computer vision and pattern recognition*, pp. 1098-1105. IEEE, 2012.
[6] Mittal, Anish, Anush Krishna Moorthy, and Alan Conrad Bovik. "No-reference image quality assessment in the spatial domain." *IEEE Transactions on image processing* 21, no. 12 (2012): 4695-4708.
[7] Xue, Wufeng, Xuanqin Mou, Lei Zhang, Alan C. Bovik, and Xiangchu Feng. "Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features." *IEEE Transactions on Image Processing* 23, no. 11 (2014): 4850-4862.
[8] Ghadiyaram, Deepti, and Alan C. Bovik. "Perceptual quality prediction on authentically distorted images using a bag of features approach." *Journal of vision* 17, no. 1 (2017): 32-32.
[9] LeCun, Yann. "LeNet-5, convolutional neural networks". Retrieved 16 November 2013.
[10] Huang, Guang-Bin, Hongming Zhou, Xiaojian Ding, and Rui Zhang. "Extreme learning machine for regression and multiclass classification." *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 42, no. 2 (2011): 513-529.
[11] Le Moan, Steven, and Donald Bailey. "Comparison of machine learning-based feature pooling strategies for colour image fidelity assessment." In *2017 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, pp. 1-5. IEEE, 2017.
[12] Ghadiyaram, Deepti, and Alan C. Bovik. "Massive online crowdsourced study of subjective and objective picture quality." *IEEE Transactions on Image Processing* 25, no. 1 (2015): 372-387.
[13] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "Live image quality assessment database release 2," 2005.
[14] Larson, Eric C., and Damon M. Chandler. "Most apparent distortion: full-reference image quality assessment and the role of strategy." *Journal of electronic imaging* 19, no. 1 (2010): 011006-011006.
[15] Ponomarenko, Nikolay, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Lina Jin, Jaakko Astola, Benoit Vozel et al. "Color image database TID2013: Peculiarities and preliminary results." In *European workshop on visual information processing (EUVIP)*, pp. 106-111. IEEE, 2013.

[16] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25 (2012).
[17] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778. 2016.
[18] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).
[19] T. Vu, "DIIVINE: A Deep Invariant Indexing for Visual Quality Assessment," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
[20] A. D. Bagdanov, A. Del Bimbo, and G. M. Santini, "CORNIA: A No-Reference Metric for Visual Attention," Journal of Vision, 2017.
[21] Mittal, Anish, Anush Krishna Moorthy, and Alan Conrad Bovik. "No-reference image quality assessment in the spatial domain." *IEEE Transactions on image processing* 21, no. 12 (2012): 4695-4708.
[22] Mittal, Anish, Rajiv Soundararajan, and Alan C. Bovik. "Making a “completely blind” image quality analyzer." *IEEE Signal processing letters* 20, no. 3 (2012): 209-212.
[23] Z. Wang, Q. Li, and G. W. Winkler, "Image quality assessment: From error visibility to structural similarity," IEEE Transactions on Image Processing, 2004.
[24] S. Bosse, D. Maniry, and C. A. Poynton, "FRIQUEE: A New Biologically Inspired Image Quality Evaluation Algorithm," Journal of Imaging Science and Technology, 2017.
[25] S. Bosse, D. Maniry, and C. A. Poynton, "Image Quality Assessment: Unifying Local and Global Approaches.