


An Optimized Deep Learning Approach for Robust Image Quality Classification



Ahmed Elaraby^{1,2*}, Aymen Saad³, Hanen Karamti⁴, Madallah Alruwaili⁵

¹ College of Engineering and Information Technology, Buraydah Private Colleges, Buraydah 51418, Saudi Arabia

² Department of Computer Science, Faculty of Computers and Information, South Valley University, Qena 83523, Egypt

³ Department of Information Technology, Management Technical College, Al-Furat Al-Awsat Technical University, Kufa 54003, Iraq

⁴ Department of Computer Sciences, College of Computer and Information Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

⁵ College of Computer and Information Sciences, Jouf University, Sakaka, Aljouf 73211, Saudi Arabia

Corresponding Author Email: ahmed.elaraby@svu.edu.eg

<https://doi.org/10.18280/ts.400425>

ABSTRACT

Received: 23 January 2023

Revised: 28 March 2023

Accepted: 17 May 2023

Available online: 31 August 2023

Keywords:

image quality, deep learning, convolutional neural network, image classification

This study presents a novel methodology for robust classification of image quality, a critical task in the domain of computer vision. The ability to accurately and promptly classify an image as being of inferior quality, due to factors such as lighting, focus, encoding, and compression, is crucial for a wide range of applications, including autonomous vehicles, web search technologies, smartphones, and digital cameras. Moreover, this capability holds significant potential for numerous industrial applications, particularly in the realm of quality assurance in manufacturing processes or outgoing inspections. In response to this requirement, a novel automated system is proposed herein, employing an optimization algorithm to categorize images into six distinct classes: motion blur, white noise, Gaussian blur, poor illumination, JPEG 2000, and high-quality reference images. The proposed framework is evaluated against existing methodologies using a selection of publicly available datasets. Both subjective and objective assessment results will be presented to demonstrate the efficacy of the proposed framework. This work underscores the potential of leveraging optimized deep learning techniques for robust and automatic image quality classification, thereby paving the way for improved quality assurance across diverse industries.

1. INTRODUCTION

High-quality images are indispensable in both professional and personal spheres of life. Unfortunately, during the processes of image acquisition, transmission, and processing, images can become susceptible to distortion from environmental factors such as poor lighting conditions, device noise, and compression damage. These factors can lead to significant information loss and a consequent reduction in image quality. Images of suboptimal quality not only fail to satisfy human visual requirements but also pose challenges to computer processing and analysis. Consequently, a pressing need exists for the restoration of essential content and details from degraded images [1].

Throughout the stages of image acquisition, processing, compression, transmission, and reproduction, both images and videos can experience a myriad of distortions. Although deteriorations in visual quality are easily discernible to the human eye, objective evaluation of perceived image quality is challenging. For instance, consider a mobile application where customers upload "before" and "after" photos of an insurance event. If the photos are of poor quality, the system should be capable of highlighting these low-quality images and, if necessary, prompt the user to recapture the object [2].

Over the past decade, an increased interest has been observed in objective image quality assessment methodologies.

Such methodologies can optimize a wide range of image and video processing techniques, in addition to monitoring image quality deterioration and benchmarking image processing systems.

In recent years, various types of classification networks have emerged, achieving high classification accuracy. More sophisticated models, such as deep convolutional neural networks (DCNNs), have been employed to tackle this challenge. However, real-time image capture under varying conditions such as weather, noise, and motion can result in low-quality images, leading to a significant drop in network accuracy. This is primarily because image degradation obstructs the structural statistical properties of pixels in the neighborhood [3].

In the realm of computer vision, the categorization of low-quality images holds substantial importance. Automated identification of such images can enable numerous practical applications. For instance, search engines can discard low-quality images, digital and phone cameras can alert users of poor-quality shots, and autonomous driving technology can avoid using poorly captured frames, thereby reducing the likelihood of errors. The issues that may arise in an image can stem from a broad range of problems, such as inadequate lighting or blur due to improper photography techniques, or even encoding issues.

Approaches to address this problem vary. One solution

involves enhancing the visibility of the image before performing classification. However, such methods often demonstrate low accuracy and poor robustness. Another potential solution involves transforming the problem into a domain adaptation problem. However, most existing domain adaptation approaches require either complex architectures or continuous target domains.

In this study, our focus is on the design of a novel SimpleNet for robust image quality classification. It is demonstrated that a specific classification task can be accomplished with high accuracy simply by calibrating the model for the task using transfer learning.

2. RELATED WORK

Due to its utility in a wide range of applications, including assessing picture capture pipelines, storage methods, and sharing media, automatically learned quality evaluation for photos has lately become a hot issue.

Kang et al. [1] demonstrate that actual blind quality assessment performance may be achieved by extracting high-level characteristics using CNNs. The fundamental benefit of utilizing CNNs for pixel-level quality assessment tasks seems to be that end-to-end feature learning systems replace hand-crafted features [2]. Bosse et al. [2] use a deep CNN with 12 layers to improve image quality forecasting. Both techniques need score aggregation throughout the entire image because of the small input size. Bianco et al. [3] propose a deep-quality predictor based on Krizhevsky et al. [4]. Multiple CNN features are retrieved from pictures and then regressed to the human scores. The Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [5] is used by "BRISQUE" to compute the no-reference image quality score for a picture. A support vector regression (SVR) model trained on an image database and matching differential mean opinion score (DMOS) values are used to compute the BRISQUE score. Photos with known distortions such as compression artifacts, blurring, and noise are included in the database, as are clean copies of the afflicted photographs. The image to be assessed must contain at least one of the distortions for which the model was trained. Talebi and Milanfar [6] proposed a network that may be utilized to help with the adaption and optimization of picture editing/enhancement algorithms in a photographic pipeline in addition to scoring photos consistently and with good correlation to human perception. All of this may be done without the requirement for a "golden" reference picture, making it possible to judge quality using a single image that is both semantically and perceptually aware.

For low-quality picture classification, Wang et al. [7] provide a unique deep degradation prior that may be utilized to lessen feature mismatches between clear photos and variously degraded ones. The developed feature de-drifting module is resilient and efficient even when trained on a tiny dataset, such as 10 photos from CUB-C. Feature de-drifting modules may easily be "plugged in" to already-existing classification networks.

Past research has primarily concentrated on conducting a binary classification for one image quality issue at a time. Nevertheless, the approach presented in this study can identify motion blur, Gaussian blur, inadequate lighting, white-noise, and JPEG-2000 compression errors in an image. As far as our knowledge goes, there are no current studies that undertake a comprehensive classification of images based on the category

of image quality problems.

Recently, great advances [8-14] have been made to improve the quality of images based on different techniques. Non-learning-based methods for detecting blurry images have been in use for numerous years. Although these methods yield satisfactory outcomes in identifying blurriness, they lack the ability to classify an image into various categories of poor image quality or identify other quality issues such as white noise or compression artifacts. Nonetheless, we have incorporated the previous work that employs these non-learning-based methods for the sake of comprehensiveness.

3. METHODOLOGY

Based on the SimpleNet and InceptionResNetV2 models, we describe the suggested framework for undesirable picture categorization (bad lighting, fastfading, gblur, jp2k, jpeg, refimgs, and wn). The suggested CNN-based classifiers framework was constructed using "Google-Colab".

3.1 Dataset (Unwanted image)

In this study, a primary dataset containing 1,165 poor quality images as shown in Table 1, we have focused on seven categories (bad_lighting, fastfading, gblur, jp2k, jpeg, refimgs and wn) as shown in Figure 1 (a, b, c, d, e, f and g) one example for each class.

Table 1. Details of photo quality dataset

No. of Class	Name of Class	No. of Image
1	bad_lighting	22
2	fastfading	174
3	gblur	207
4	jp2k	227
5	jpeg	233
6	refimgs	128
7	wn	174

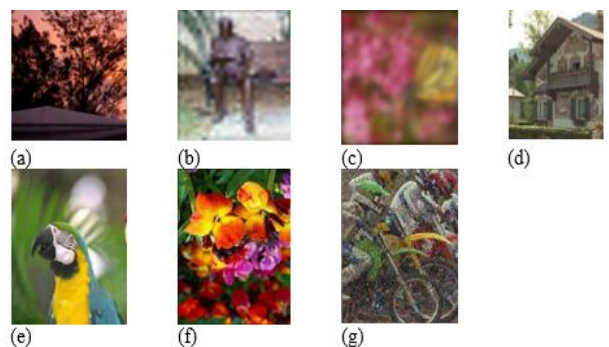


Figure 1. Example for each class of photo quality

There are several reasons why an image may be regarded to be of low quality [15-19]. In the following, we explain all types of poor images:

- Bad Lighting (bad lighting) Images that were captured without enough light for the camera's timing and aperture. This might make photos appear drab or dark.
- Fast fading (fastfading) is the effect of constructive and destructive interference patterns which are caused due to multipath in channels of the image.

- An out-of-focus camera causes Gaussian Blur (gblur). This form of blur might be caused by a camera's malfunctioning auto-focus system, a poorly designed lens, or a picture focused on the wrong topic.
- JPEG-2K In the year 2000, the JPEG-2K picture format was released. While this picture compression format has numerous advantages, one downside is that it is significantly less content adaptable than the earlier JPEG format, which means that image quality might vary greatly given the same bit rate for different materials.
- Reference Image (refimgs) is used to refer to any visual resource such as a diagram, graph, illustration, design, photograph, or video. They may be found in books, journals, reports, web pages, online videos, DVDs and other kinds of media.
- White Noise (wn) is the appearance of random white grains in a picture. This noise is frequently created by film grain, different sensors and circuits such as CCDs in digital cameras and detectors in scanners, or by the communications channel or signal quantization.

3.2 Deep convolutional neural networks

Deep artificial neural networks (DANNs) or deep convolutional neural networks (CNNs) are used to classify images, group similar images together, and identify objects in scenes [20]. CNN is made up of convolutional and subsampling layers, which are then followed by one or more fully linked layers. CNN's architecture is intended to capitalize on the 2D structures of input pictures. Moreover, CNNs have fewer parameters and are simpler to train than fully connected networks. In order to train and evaluate the CNN model, each input image is routed through a series of convolution layers

and pooling for feature learning. Additionally, an activation function like Softmax, Sigmoid, or ReLU is employed to classify an item. In the area of computer vision, CNNs and other cutting-edge machine learning algorithms have become crucial. We create and train a CNN with three convolutional layers to categorize pictures into six predetermined categories in our method. We demonstrate that our technique can achieve reasonably high accuracy when compared to a very sophisticated model (ResNet) with only a small number of parameters.

3.2.1 Our Model (SimpelNet)

The network consists of seven blocks. Each block is made up of a fully connected layer with a Sigmoid activation function, followed by a convolutional layer with a Relu activation function. There are also six node output layers with sigmoid activation functions that output the six required classes. The first convolutional layer takes a 256 256 3 image array and applies 32 filters to create 256 256 3 feature maps. We begin with the bigger filter size since it delivers the best results for picture classification tasks in the higher layers [20]. The feature maps are then normalized, the Relu activation function is used, and a 2 2 Max Pooling is conducted to create separate image domains for each block, beginning with the second Conv2D and ending with the last Conv2D.

In addition, we started from 20% Dropout from the second Conv2D to the third Conv2D and a 30% Dropout from the fourth Conv2D to the fifth Conv2D. Then a 40% Dropout from the sixth Conv2D to the seventh Conv2D.

The goal of this significant dropout is to keep the network from overtraining on the training set and aid in greater generalization when used with other types of images. The output layer, which has six nodes and is connected to every node from the layer before it, has sigmoid activation functions. A total of 1,425,606 parameters makes up the networks. The architecture of our CNN is depicted in Figure 2, and Table 2 lists the layers and parameters for each layer.

Table 2. Model SimpelNet

# Of Layer	Layer (Type)	Output Shape	# Of Param
Input Layer	img_input (Input Layer)	[(None, 256, 256, 3)]	0
Block -1	11 (Conv2D)	(None, 256, 256, 32)	896
	12 (Conv2D)	(None, 256, 256, 64)	18496
Block -2	13 (MaxPooling2D)	(None, 128, 128, 64)	0
	dropout_43 (Dropout)	(None, 128, 128, 64)	0
	14 (Conv2D)	(None, 128, 128, 64)	36928
Block -3	15 (MaxPooling2D)	(None, 64, 64, 64)	0
	dropout_44 (Dropout)	(None, 64, 64, 64)	0
	16 (Conv2D)	(None, 64, 64, 128)	73856
Block -4	17 (MaxPooling2D)	(None, 32, 32, 128)	0
	dropout_45 (Dropout)	(None, 32, 32, 128)	0
	18 (Conv2D)	(None, 32, 32, 128)	147584
Block -5	19 (MaxPooling2D)	(None, 16, 16, 128)	0
	dropout_46 (Dropout)	(None, 16, 16, 256)	295168
	110 (Conv2D)	(None, 16, 16, 128)	0
Block -6	111 (MaxPooling2D)	(None, 8, 8, 256)	0
	dropout_47 (Dropout)	(None, 8, 8, 256)	0
	112 (Conv2D)	(None, 8, 8, 256)	590080
Block -7	113 (MaxPooling2D)	(None, 4, 4, 256)	0
	dropout_48 (Dropout)	(None, 4, 4, 256)	0
	fc1 (Flatten)	(None, 4096)	0
	114 (Dense)	(None, 64)	262208
	dropout_49 (Dropout)	(None, 64)	0
	predictions (Dense)	(None, 6)	390
Total params: 1,425,606, Trainable params: 1,425,606			

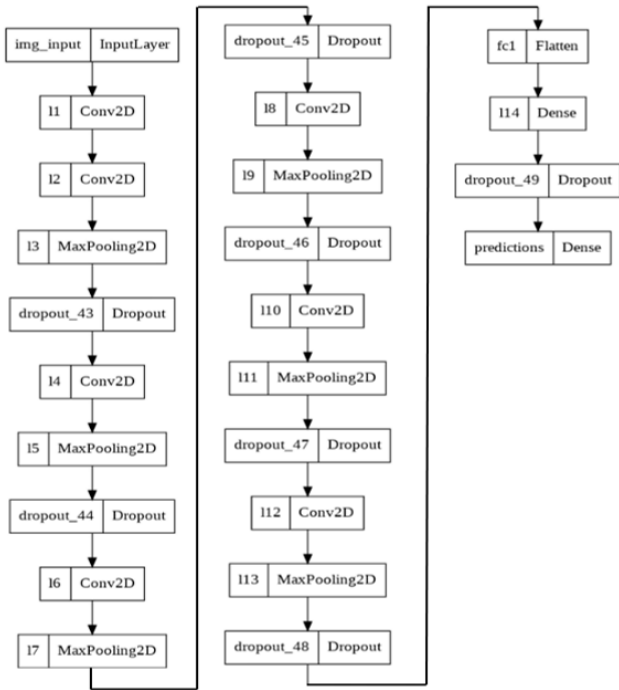


Figure 2. The architecture of our CNN SimpleNet

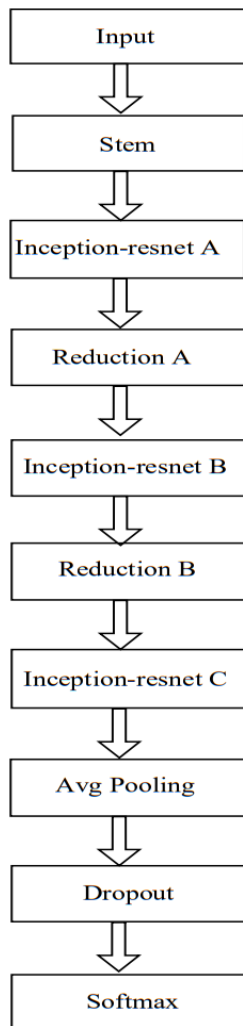


Figure 3. Basic block diagram of InceptionResNetV2 model

3.2.2 InceptionResNetV2

Inception ResNetV2 is a hybrid of the Inception and ResNet families, with 164 layers for picture object identification and feature extraction [21-23]. Only the last layer is applied to examine the results. To train the InceptionResNetV2 CNN model, we explicitly replicate resolution inconsistency in affine face wrappings during video editing. Using a trained model reduces the size and difficulty of training. InceptionResNetV2 is one of famous pre-trained networks.

When there are a limited number of training samples available, tuning a pre-trained network is the best option. In this example, the top layers of a pre-trained network's parameters are fine-tuned, but the initial layers, which represent generic characteristics, are frozen. When weights for a layer or group of layers are not changed during the training stage, this is referred to as freezing. Importantly, this method uses the parameters learnt from a network that has already been trained on a certain dataset and then modifies the parameters for the dataset of interest. As a result, fine-tuning modifies the parameters of the repeated model, making it more relevant to the dataset under consideration. The top layers of a pre-trained network can be fine-tuned or all layers; nevertheless, the latter technique is favoured [24].

This is due to the fact that the earliest levels of a structure encode general, reusable properties, and the later layers encode highly specialized traits. As a result, fine-tuning those individual traits is more efficient. Furthermore, because of the enormous number of parameters that must be established during this procedure, fine-tuning all layers produces overfitting. As a result, only the top three levels of pre-existing convnets were fine-tuned in this study. These might be the completely linked layers alone, as in InceptionResNetV2. Figure 3 depicts the architecture of the InceptionResNetV2 model's fundamental block diagram.

4. EXPERIMENTS AND PERFORMANCE ANALYSIS

This section discusses the training results of our DCNN models. This study used datasets [15]. There are 1,165 low-quality photos retrieved from the training dataset in the proposed classifier, with various numbers of images derived for each class. These photos were analyzed during the CNN model training stage. The CNN model employs 50-300 epochs. The performance details of the results for InceptionResNetV2 and our model SimpleNet models respectively are shown in Table 3.

Table 3. Comparison between the results for our CNN models

Models	Ac-Train	Ac-Test	Time (min)	# Of Epoch
SimpleNet	82	75	12.5	50-300
InceptionResNetV2	80	71	13.28	50-300

The accuracy and loss are observed during the training process for the InceptionResNetV2 and SimpleNet models, as illustrated in Figure 4 and Figure 5, respectively. As the period lengthens, the accuracy grows until it reaches a saturated level, when an error is minimal and varies at a specific level [25]. It is deteriorating, although the loss diminishes with decreasing epochs until it reaches a saturation level. In other words, accuracy indicates how excellent a model is, whereas loss

determines how terrible a model is. A wonderful model will often have high accuracy and little loss.

The proposed classifier contains 233 photos for two classes in the validation or testing dataset, where each class has distinct images. A confusion matrix demonstrates how well the model predicts the class used in supervised learning. The suggested model predicts the corresponding classes of the two classes in the validation test. The number of predictions for

each class is shown in each column of the matrix, whereas the number of cases in the real class is represented in each row. The outputs of each process are recorded in the confusion matrix as shown in Figure 6 (a and b) for 50-epoch and Figure 7 (a and b) for 300-epoch SimpleNet and InceptionResNetV2 models respectively.

More specific details of each class are presented for this study as shown in Table 4.

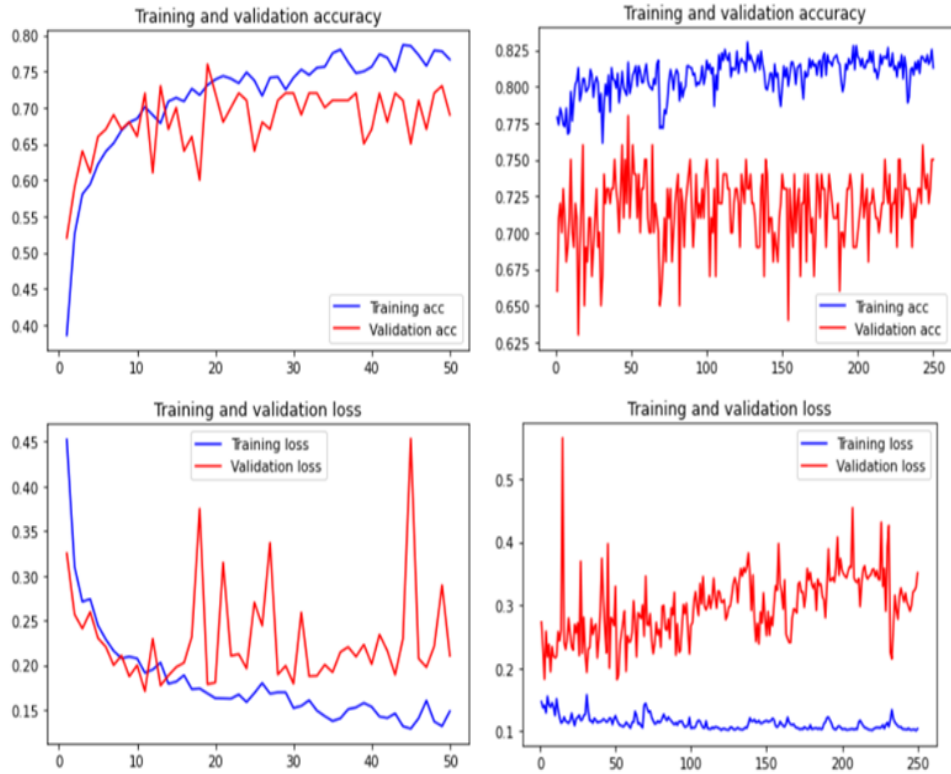


Figure 4. Accuracy and loss of training for SimpleNet model

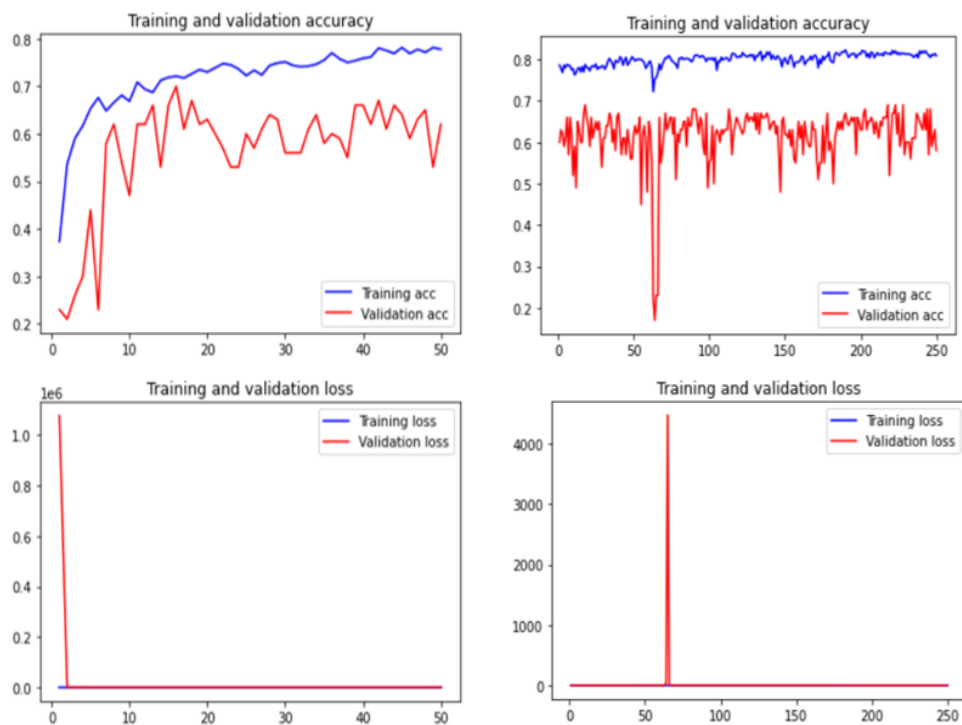


Figure 5. Accuracy and loss of training for InceptionResNetV2

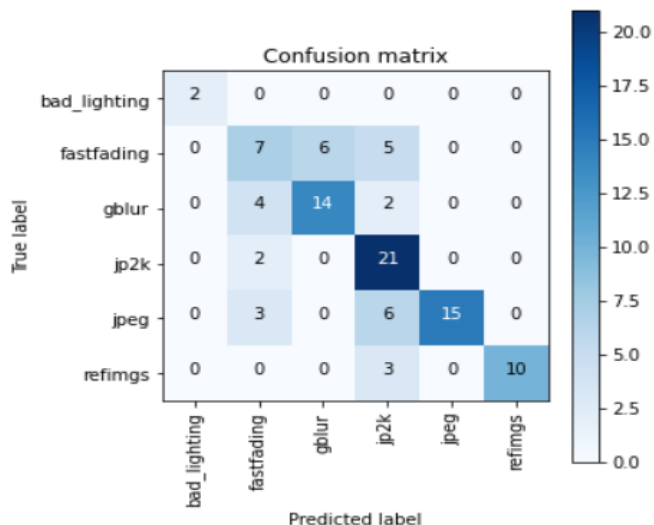


Figure 6. (a) SimpelNet model 50-epoch

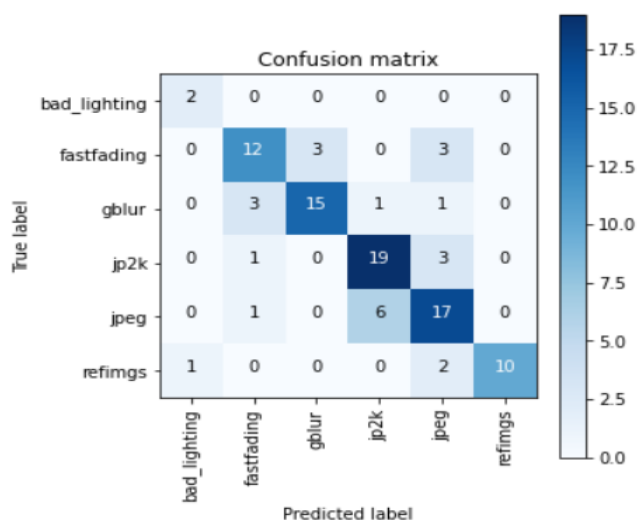


Figure 6. (b) SimpelNet model 300-epoch

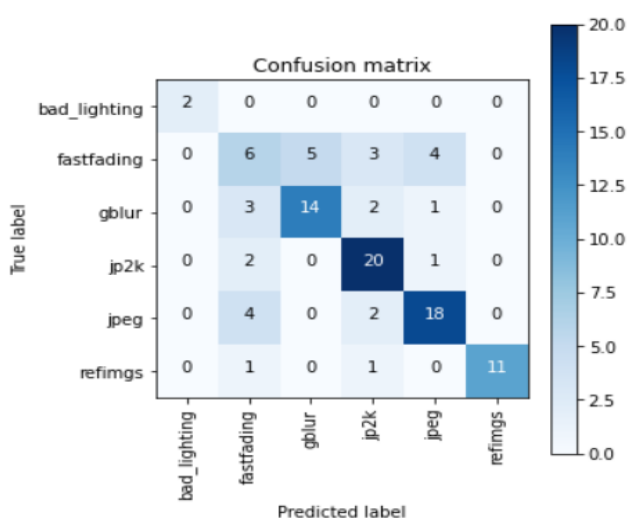


Figure 7. (a) InceptionResNetV2 model 50-epoch

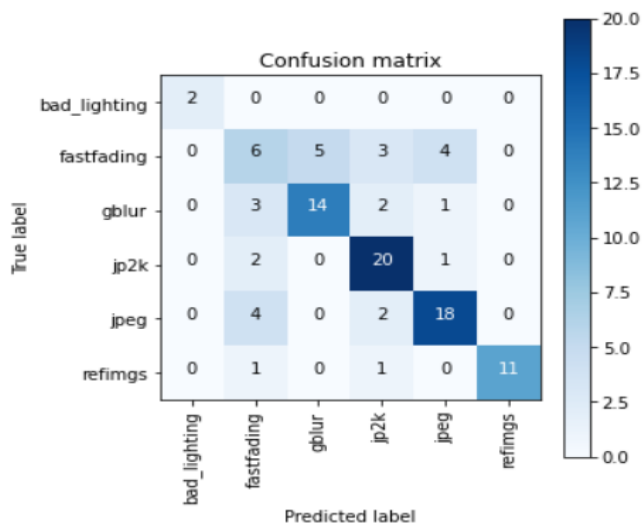


Figure 7. (b) InceptionResNetV2 300-epoch

Table 4. Comparison between results for our CNN models

Class	SimpleNet		InceptionResNetV2	
	Sensitivity	Specificity	Sensitivity	Specificity
bad_lighting	1	1	1	1
fastfading	0.38	0.90	0.33	0.90
gblur	0.70	0.93	0.7	0.93
jp2k	0.91	0.79	0.86	0.9
jpeg	0.62	1	0.75	0.91
refimgs	0.76	1	0.84	1

5. CONCLUSION

Machine learning algorithms must be able to recognize when the outcomes of activity are undesirable given the rapid improvement of machine learning and the expanding use of autonomous technologies. The findings of this study demonstrate that our suggested CNN can accurately identify low-quality photos. Finally, the generated model's application was evaluated to find photos that would not be appropriate for the sign categorization task needed by autonomous cars. We demonstrated that it is possible to attain extremely high accuracy for the given classification job by merely calibrating

the model for the task using transfer learning. This study was focused on designing a novel SimpleNet and comparing it with the InceptionResNetV2 model which is considered a highly complex model to detect poor-quality images. Moreover, we perform sensitivity and specificity analysis to determine how differences in camera hardware affect the accuracy of the model.

ACKNOWLEDGMENT

The authors would like to thank Princess Nourah bint

REFERENCES

- [1] Kang, L., Ye, P., Li, Y., Doermann, D. (2014). Convolutional neural networks for no-reference image quality assessment. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1733-1740. <https://doi.org/10.1109/CVPR.2014.224>
- [2] Bosse, S., Maniry, D., Wiegand, T., Samek, W. (2016). A deep neural network for image quality assessment. In 2016 IEEE International Conference on Image Processing (ICIP), pp. 3773-3777. <https://doi.org/10.1109/ICIP.2016.7533065>
- [3] Bianco, S., Celona, L., Napoletano, P., Schettini, R. (2018). On the use of deep learning for blind image quality assessment. *Signal, Image and Video Processing*, 12: 355-362. <https://doi.org/10.1007/s11760-017-1166-8>
- [4] Krizhevsky, A., Sutskever, I., Hinton, G.E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6): 84-90. <https://doi.org/10.1145/3065386>
- [5] Mittal, A., Moorthy, A.K., Bovik, A.C. (2012). No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12): 4695-4708. <https://doi.org/10.1109/TIP.2012.2214050>
- [6] Talebi, H., Milanfar, P. (2018). NIMA: Neural image assessment. *IEEE Transactions on Image Processing*, 27(8): 3998-4011. <https://doi.org/10.1109/TIP.2018.2831899>
- [7] Wang, Y., Cao, Y., Zha, Z.J., Zhang, J., Xiong, Z. (2020). Deep degradation prior for low-quality image classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11049-11058. <https://doi.org/10.1109/CVPR42600.2020.01106>
- [8] Alireza Golestaneh, S., Karam, L.J. (2017). Spatially-varying blur detection based on multiscale fused and sorted transform coefficients of gradient magnitudes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5800-5809. <https://doi.org/10.1109/CVPR.2017.71>
- [9] Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y. (2020). Residual dense network for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(7): 2480-2495. <https://doi.org/10.1109/TPAMI.2020.2968521>
- [10] Zhang, H., Sun, L., Wu, L., Gu, K. (2021). DuGAN: An effective framework for underwater image enhancement. *IET Image Processing*, 15(9): 2010-2019. <https://doi.org/10.1049/ipr2.12172>
- [11] Li, F., Zheng, J., Zhang, Y.F. (2021). Generative adversarial network for low - light image enhancement. *IET Image Processing*, 15(7): 1542-1552. <https://doi.org/10.1049/ipr2.12124>
- [12] Li, S., Qin, B., Xiao, J., Liu, Q., Wang, Y., Liang, D. (2019). Multi-channel and multi-model-based autoencoding prior for grayscale image restoration. *IEEE Transactions on Image Processing*, 29: 142-156. <https://doi.org/10.1109/TIP.2019.2931240>
- [13] Jin, Z., Iqbal, M.Z., Bobkov, D., Zou, W., Li, X., Steinbach, E. (2019). A flexible deep CNN framework for image restoration. *IEEE Transactions on Multimedia*, 22(4): 1055-1068. <https://doi.org/10.1109/TMM.2019.2938340>
- [14] Huang, H., Schiopu, I., Munteanu, A. (2020). Macro-pixel-wise CNN-based filtering for quality enhancement of light field images. *Electronics Letters*, 56(25): 1413-1416. <https://doi.org/10.1049/el.2020.2344>
- [15] Tang, X., Luo, W., Wang, X. (2013). Content-based photo quality assessment. *IEEE Transactions on Multimedia*, 15(8): 1930-1943. <https://doi.org/10.1109/TMM.2013.2269899>
- [16] Choi, Y.S., Voltz, P.J., Cassara, F.A. (2001). On channel estimation and detection for multicarrier signals in fast and selective Rayleigh fading channels. *IEEE Transactions on Communications*, 49(8): 1375-1387. <https://doi.org/10.1109/26.939860>
- [17] Hsu, P., Chen, B.Y. (2008). Blurred image detection and classification. In *Advances in Multimedia Modeling: 14th International Multimedia Modeling Conference, MMM 2008, Kyoto, Japan, January 9-11, 2008. Proceedings*, Springer, Berlin, Heidelberg, pp. 277-286. https://doi.org/10.1007/978-3-540-77409-9_26
- [18] Wang, Z., Simoncelli, E.P. (2005). Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In *Human Vision and Electronic Imaging X*, 5666: 149-159. <https://doi.org/10.1117/12.597306>
- [19] Liu, W., Lin, W. (2012). Additive white Gaussian noise level estimation in SVD domain for images. *IEEE Transactions on Image Processing*, 22(3): 872-883. <https://doi.org/10.1109/TIP.2012.2219544>
- [20] Aghdam, H.H., Heravi, E.J. (2017). *Guide to convolutional neural networks*. New York, NY: Springer, 10(978-973): 51. <https://doi.org/10.1007/978-3-319-57550-6>
- [21] Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1): 4278-4284. <https://doi.org/10.1609/aaai.v31i1.11231>
- [22] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [23] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, pp. 1-9. <https://doi.org/10.1109/CVPR.2015.7298594>
- [24] Chollet, F. (2021). *Deep Learning with Python*. Simon and Schuster.
- [25] Powers, D.M. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*. <https://doi.org/10.48550/arXiv.2010.16061>