Evaluating Convolutional Neural Networks for No -Reference Image Quality Assessment

Kyriakos D. Apostolidis, Theodore Polyzos, Ioannis Grigoriadis, George A. Papakostas*
MLV Research Group, Department of Computer Science, International Hellenic University, 65404 Kavala, Greece

kyriapos1@cs.ihu.gr, theroly@teiemt.gr, j.grhgoriadhs@gmail.com, gpapak@cs.ihu.gr
**corresponding author*

2021 4th International Conference on Signal Processing and Information Security (ICSPIS) | 978-1-6654-3796-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICSPIS53734.2021.9652176

***Abstract*— In the past years, deep learning evolution has helped the development of computer vision systems. However, the quality of images plays a significant role in the effectiveness of these systems and it would be useful to know the quality of the images that are imported into our systems. No-reference image quality assessment is a challenging procedure, which tries to predict the quality of an image without using any reference image. In this paper, we evaluate the performance of widely used deep learning models in no-reference image quality assessment. To that end, we used transfer learning on 8 pre-trained models which we fit into 3 datasets related to image quality assessment. The performance of these models was studied in terms of mean absolute percentage error (MAPE). Although most of the models performed reasonably well in a MAPE range of 15% to 40% depending on the dataset used, the best performing one in a single dataset was DenseNet201 with a MAPE of 9.8%, while the overall best performing model was ResNet50.**

Keywords*— Image quality assessment, convolutional neural networks, transfer learning

I. Introduction

Computer Vision has become an integral part of intelligent systems such as biometrics [1], autonomous cars [2], medical imaging [3], feature extraction [4], etc. However, the good quality of images is crucial for the robustness of these systems and especially in medical image analysis [5], which deals with human lives. Image quality can be referred to "as the weighted combination of all of the visually significant attributes of an image" [6]. The importance and number of these attributes may vary depending on the methodology used but they usually consist of sharpness, brightness, contrast, color, noise, and artifacts.

Assessing the quality of an image can be done in two ways, using subjective or objective methods. Objective methods consist of full-reference, where the image quality is assessed in comparison to a reference image that is considered to have perfect quality, reduced-reference, where the image quality is assessed by extracting and comparing features from both images, and no-reference, where the image quality is assessed without an original image to be compared to. Subjective methods on the other hand consist of studying how a group of people perceives the quality of an image and mapping those perceptions onto numerical values.

Unlike image detection, segmentation and classification, when it comes to building datasets for image quality assessment, complex psychometric tests are needed, which makes it difficult to create datasets because it is time consuming and costly. Regardless of the methodology that is followed, experts are required to oversee the process and make sure that everything is correctly executed. That is why it is difficult to find diverse, quality data to work with. In this study, we used 3 datasets for experiments, KoniQ-10k [7], TID2013 [8] and LIVE In the Wild [9], [10].

Through this paper, we contribute in two ways: (1) by presenting a novel performance evaluation of some of the best publicly available convolutional neural networks (CNNs) in image quality assessment and (2) by studying how well they generalize on problems they weren't designed for while comparing their performance between datasets.

In section 2, we briefly summarize noteworthy work done in image quality assessment and deep learning. Section 3 presents an overview of the models used and a brief rundown of their technical descriptions, while Section 4 goes into more detail about the findings of the experiments that we conducted, examining each model performance per dataset. Lastly, Section 5 concludes this paper.

II. Related Works

Image quality assessment has been highly discussed and researched over the past few decades, with an extensive amount of research being done on the topic. Advances in technology along with the emergence of Digital Imaging have created higher definition visual content, which has been further pushed by the development of image processing. Some of the first attempts at image quality assessment came from Ian R. L. Davies [11], who developed an automatic "image quality meter" mimicking the human visual system, based on neurophysiological and psychophysical studies. Its purpose was to assess the degree of impairment in broadcast TV images, captured directly from the TV using a CCD camera and a digital sampling hardware, using spatial and temporal filters and a neural network with three layers. W. Osberger [12] presented a system based on the human visual system, consisting of an early vision model and a visual attention model, which pinpoints regions of interest in a scenery using Importance Maps. M. Carnec [13] introduced a new method of evaluating the quality of distorted images, using reduced references containing perceptual structural information. The method is based on an implementation of a model of the Human Visual System and has been established based on neurophysiological descriptions. The full process can be divided into two stages. The first stage includes building the perceptual representation of the original and the distorted images, while the second stage compares the representations to compute a quality score.

Recently, a popular approach to image quality assessment is using deep learning along with convolutional neural networks, given the rapid growth of the field and its application to computer vision. A no-reference as well a full-reference approach was presented by Bosse [14]. The network does not rely on hand-crafted features and uses a purely data-driven approach and allowing joint learning of local quality and local weights. Y. Li [15] presented a general-purpose framework based on deep CNNs, using as an input a raw image and providing the quality score of the image. To avoid hand-crafted features, the framework integrates feature learning and regression into one optimization process. J. Li [16] proposed combining convolutional neural networks with the Prewitt magnitude of segmented images to take into consideration both the human visual system and the mean of all of the image patch scores. Bare [17] introduced an accurate deep CNN model that used image patches as the input. The model achieves an end-to-end method that doesn't require handcrafted features or pre-processing procedures. Zhang [18] proposed a deep bilinear convolutional neural network model that works for synthetically and authentically distorted images. The model specializes separately for each type of distortion.

III. Model Description

The performance of a set of 8 pre-trained models was evaluated at assessing the quality of an image with no-reference. Table 1 gives a brief summary of the models used in this study.

TABLE I. PARAMETERS AND DEPTH OF MODELS USED

Models	Parameters	Depth
Xception	22,910,480	126
VGG16	138,357,544	23
ResNet50	25,636,712	168
InceptionV3	23,851,784	159
MobileNet	4,253,864	88
DenseNet201	20,242,984	201
EfficientNetB0	5,330,571	-
NasNetLarge	88,949,818	-

Xception [19] is inspired by Inception and attempts a step in-between regular convolution and depthwise separable convolution operation. This architecture is highly efficient and performs better on large image classification datasets. VGG [20] uses very small (3x3) convolution filters and implements higher depth weight layers. ResNet50 [21] reformulates the layers as learning residual functions with reference to the layer inputs. It is easier to optimize and gains accuracy with increased depth. InceptionV3 [22], instead of focusing on increased model size and computational cost, explores ways to scale up efficiency by suitably factorized convolutions and aggressive regularization. MobileNet [23] uses depth-wise separable convolutions to build lightweight neural networks that efficiently trade between accuracy and resources. DenseNet201 [24] connects each layer to every other layer in a feed-forward way, featuring more direct connections between layers. EfficientNetB0 [25] uniformly scales all dimensions of depth/width/resolution using a highly effective coefficient and it achieves high efficiency while maintaining accuracy. NASNet Large [26] uses a resource expensive approach that learns the model architectures directly on the dataset of interest. All of the above models were trained on the ImageNet [27] dataset.

IV. Experimental Study

The experimental study was performed on pre-trained versions of the 8 models in Table 1, as are implemented in the Keras [28] library. These experiments were carried out on a desktop computer with 16GB of DDR4 RAM, an Intel i7 9700k and a GTX 1070 NVIDIA GPU. All models were trained for 20 epochs to prevent over-fitting and the only layers left trainable were the ones related to the output. The models were compiled using the Adam optimizer with a learning rate of 0.0001, while the loss function chosen was mean squared error (MSE). Each model was separately trained for each dataset and the metric used was absolute percentage error (MAPE). Below, Table 2 presents the results for each model in further detail.

TABLE II. MEAN ABSOLUTE PERCENTAGE ERROR (MAPE) FOR EACH MODEL AND DATASET

Models	KonIQ-10k	LIVE In the wild	TID2013
Xception	15.7%	41.3%	24.0%
VGG16	18.4%	38.3%	15.7%
ResNet50	12.4%	25.5%	16.5%
InceptionV3	12.8%	37.1%	20.4%
MobileNet	13.0%	35.2%	24.8%
DenseNet201	9.8%	32.0%	18.5%
EfficientNetB0	20.1%	44.1%	23.0%
NasNetLarge	13.5%	39.8%	20.2%

From these results, we can interpret that every model has a hard time assessing the quality of an image when the datasets are lacking in size. We can see that regardless of how resource expensive a model is, each one had bad results for the Live In the wild dataset (1162 images), mediocre results for the TID2013 dataset (3,000 images) and decently good results for the KonIQ-10k dataset (10,073 images). It should also be noted that: (1) each of these models has its unique pre-processing function, which is a significant factor in the quality assessment process, and: (2) the best overall performing models were not the ones with more parameters. ResNet50 performed, overall, the best out of all of the convolutional neural networks, closely followed by DenseNet201. Below, Figures 1, 2 and 3 present examples of the difference of an image's mean opinion score (DMOS) compared to the predicted value.

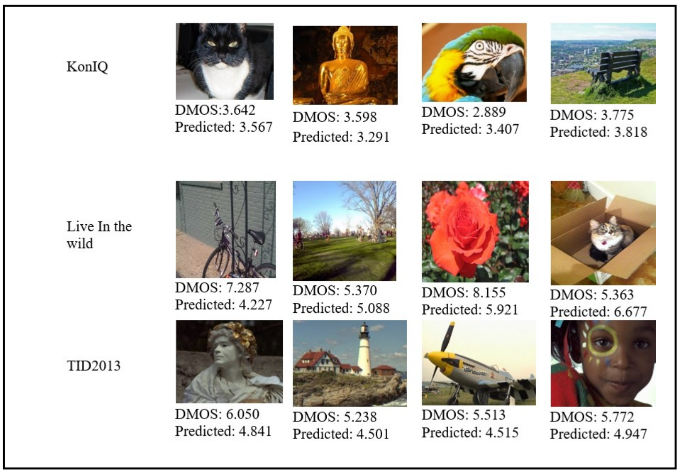

Fig. 1. Examples of an image's DMOS compared to its predicted value for the DenseNet201model.

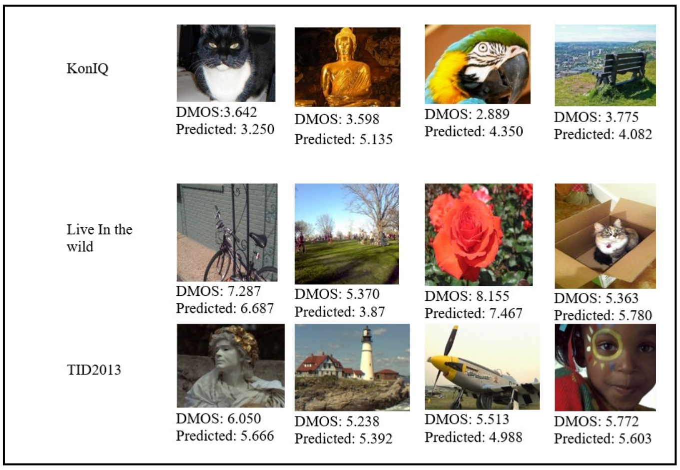

Fig. 2 Examples of an image's DMOS compared to its predicted value for the VGG16 model

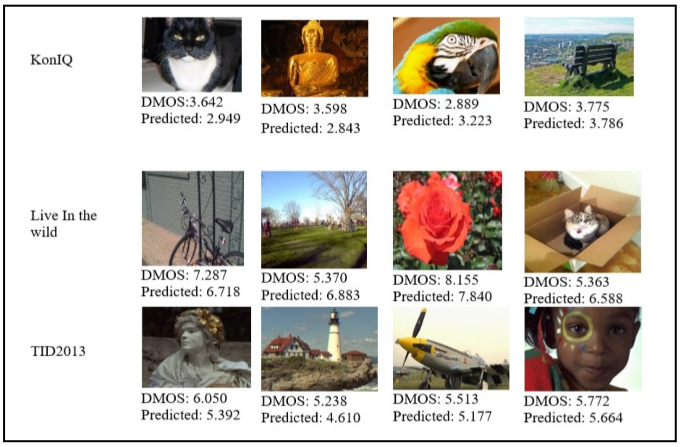

Fig. 3 Examples of an image's DMOS compared to its predicted value for the ResNet50 model.

V. CONCLUSION

In this paper, we studied the performance of 8 pre-trained convolutional neural networks and their ability to assess the quality of an image with no-reference point used. Our experiments showed that models can provide varied results depending on what kind of dataset is used, but can reach a very low Mean Absolute Error Percentage of up to 9.8% (DenseNet201 on KonIQ-10k). We can conclude that the image pre-processing function plays an important role in a model's ability to generalize, and computationally expensive models do not necessarily mean better results. Despite all this, more data is needed to arrive at conclusions that are more definitive, so more experiments should take place in the future.

ACKNOWLEDGMENT

This work was supported by the MPhil program "Advanced Technologies in Informatics and Computers", hosted by the Department of Computer Science, International Hellenic University, Kavala, Greece.

REFERENCES

[1] K. Apostolidis, P. Amanatidis, and G. Papakostas, "Performance Evaluation of Convolutional Neural Networks for Gait Recognition," in 24th Pan-Hellenic Conference on Informatics, Athens Greece, Nov. 2020, pp. 61–63. doi: 10.1145/3437120.3437276.

[2] S. Grigorescu, B. Trasnea, T. Cocias, and G. Macesanu, "A Survey of Deep Learning Techniques for Autonomous Driving," J. Field Robotics, vol. 37, no. 3, pp. 362–386, Apr. 2020, doi: 10.1002/rob.21918.

[3] G. Litjens et al., "A Survey on Deep Learning in Medical Image Analysis," Medical Image Analysis, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

[4] G. K. Sidiropoulos, P. Kiratsa, P. Chatzipetrou, and G. A. Papakostas, "Feature Extraction for Finger-Vein-Based Identity Recognition," J. Imaging, vol. 7, no. 5, p. 89, May 2021, doi: 10.3390/jimaging7050089.

[5] K. D. Apostolidis and G. A. Papakostas, "A Survey on Adversarial Deep Learning Robustness in Medical Image Analysis," Electronics, vol. 10, no. 17, p. 2132, Sep. 2021, doi: 10.3390/electronics10172132.

[6] "Image Quality Metrics - Burningham - - Major Reference Works - Wiley Online Library." https://onlinelibrary.wiley.com/doi/10.1002/0471443395.img038 (accessed Sep. 14, 2021).

[7] V. Hosu, H. Lin, T. Sziranyi, and D. Saupe, "KonIQ-10k: An ecologically valid database for deep learning of blind image quality assessment," IEEE Trans. on Image Process., vol. 29, pp. 4041–4056, 2020, doi: 10.1109/TIP.2020.2967829.

[8] N. Ponomarenko, "Image database TID2013_ Peculiarities, results and perspectives," Signal Processing, p. 21, 2015.

[9] D. Ghadiyaram and A. C. Bovik, "Massive Online Crowdsourced Study of Subjective and Objective Picture Quality," IEEE Trans. on Image Process., vol. 25, no. 1, pp. 372–387, Jan. 2016, doi: 10.1109/TIP.2015.2500021.

[10] "Laboratory for Image and Video Engineering - The University of Texas at Austin." https://live.ece.utexas.edu/research/ChallengeDB/index.html (accessed Sep. 14, 2021).

[11] Ian R. L. Davies, Dave Rose, and R. Smith, "Automated image quality assessment," Sep. 1993, vol. 1913. [Online]. Available: https://doi.org/10.1117/12.152711

[12] W. Osberger, N. Bergmann, and A. Maeder, "An automatic image quality assessment technique incorporating higher level perceptual factors," in Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No.98CB36269), Chicago, IL, USA, 1998, vol. 3, pp. 414–418. doi: 10.1109/ICIP.1998.727227.

[13] M. Carnec and P. L. Callet, “An Image Quality Assessment Method Based on Perception of Structural Information,” International Conference on Image Processing (Cat. No.03CH37429), pp. III-185, 2003. doi: 10.1109/ICIP.2003.1247212.

[14] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, “Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment,” IEEE Transactions on Image Processing, vol. 27, no. 1, p. 14, 2018. doi: 10.1109/TIP.2017.2760518.

[15] Y. Li, L.-M. Po, L. Feng, and F. Yuan, “No-reference image quality assessment with deep convolutional neural networks,” in 2016 IEEE International Conference on Digital Signal Processing (DSP), Beijing, China, Oct. 2016, pp. 685–689. doi: 10.1109/ICDSP.2016.7868646.

[16] J. Li, L. Zou, J. Yan, D. Deng, T. Qu, and G. Xie, “No-reference image quality assessment using Prewitt magnitude based on convolutional neural networks,” Signal, Image and Video Processing, vol. 10, no. 4, pp. 609–616, Apr. 2016, doi: 10.1007/s11760-015-0784-2.

[17] B. Bare, K. Li, and B. Yan, “An accurate deep convolutional neural networks model for no-reference image quality assessment,” in 2017 IEEE International Conference on Multimedia and Expo (ICME), Hong Kong, Hong Kong, Jul. 2017, pp. 1356–1361. doi: 10.1109/ICME.2017.8019508.

[18] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, “Blind Image Quality Assessment Using A Deep Bilinear Convolutional Neural Network,” IEEE Trans. Circuits Syst. Video Technol., vol. 30, no. 1, pp. 36–47, Jan. 2020, doi: 10.1109/TCSVT.2018.2886771.

[19] F. Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, Jul. 2017, pp. 1800–1807. doi: 10.1109/CVPR.2017.195.

[20] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” arXiv:1409.1556 [cs], Apr. 2015, Accessed: Jun. 04, 2021. [Online]. Available: http://arxiv.org/abs/1409.1556

[21] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[22] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 2818–2826. doi: 10.1109/CVPR.2016.308.

[23] A. G. Howard et al., “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” arXiv:1704.04861 [cs], Apr. 2017, Accessed: Jun. 04, 2021. [Online]. Available: http://arxiv.org/abs/1704.04861

[24] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” arXiv:1608.06993 [cs], Jan. 2018, Accessed: Sep. 14, 2021. [Online]. Available: http://arxiv.org/abs/1608.06993

[25] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” arXiv:1905.11946 [cs, stat], Sep. 2020, Accessed: Sep. 14, 2021. [Online]. Available: http://arxiv.org/abs/1905.11946

[26] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, “Learning Transferable Architectures for Scalable Image Recognition,” arXiv:1707.07012 [cs, stat], Apr. 2018, Accessed: Sep. 14, 2021. [Online]. Available: http://arxiv.org/abs/1707.07012

[27] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database,” p. 8.

[28] “Keras: the Python deep learning API.” https://keras.io/ (accessed Sep. 14, 2021).