Effect of Image Degradation on Performance of Convolutional Neural Networks

Inad A. Aljarrah

Department of Computer Engineering, Jordan University of Science and Technology, Irbid 22110, Jordan

***Abstract*:** The use of deep learning approaches in image classification and recognition tasks is growing rapidly and gaining huge importance in research due to the great enhancement they achieve. Particularly, Convolutional Neural Networks (CNN) have shown a great significance in the field of computer vision and image recognition recently. They made an enormous improvement in classification and recognition systems' accuracy. In this work, an investigation of how image related parameters such as contrast, noise, and occlusion affect the work of CNNs is to be carried out. Also, whether all types of variations cause the same drop to performance and how they rank in that regard is considered. After the experiments were carried out, the results revealed that the extent of effect of each degradation type to be different from others. It was clear that blurring and occlusion affects accuracy more than noise when considering the root mean square error as a common objective measure of the amount of alteration that each degradation caused.

***Keywords*:** Convolutional Neural Networks, deep learning, image degradation, image recognition**.**

1. Introduction

Digital imaging is becoming the most important and valuable source for extracting insights and information. Many of today's applications use digital images such as: medical media, social images, surveillance systems, and consumer image. However, the demands of these applications are growing in order to make accurate and fast actions. Image classification and recognition are considered of the most significant research topics in the multimedia community; which answer the question of how to label an image into one of a set of predefined classes [1]. Deep learning is considered as a promising tool in the image classification and recognition field, particularly Convolutional Neural Networks (CNN) which are a special form of neural networks that extract high-level representation by using deep architectures comprising several layers [2]. By training CNN networks using huge volume of data, the network learns common features that are able to recognize even the untrained samples [3].

The idea of neural networks in general was inspired by how our brains work, and that was discovered in 1959 by Hubel and Wiesel [4] by testing how cats brains react to small visual regions. But, until 2012 there was no big enhancement in this field mainly due to the lack of big datasets and the lack in computational power at that period. In 2012, the huge break through took place by Krizhevsky et al. [5] when they worked on the huge ImageNet [6] dataset provided publicly by Stanford University. And of course, by that time, computational power has increased significantly due to the employment of Graphical Processing Units (GPUs).

The construction of Convolutional Neural Networks mainly consists of three levels of layers which include: convolutional, pooling, and fully connected layers. A detailed description of each layer and how it works is presented in the next subsection. It worth mentioning that the main addition that a CNN provide over a classical Neural Network is the ability of the first two set of layers (convolutional and pooling) of extracting features that will be ready for the fully connected layers to use as an input. This way the fully connected layer deals with much less size of data that decreases its complexity and makes training more attainable.

The three main set of layers of Convolutional Neural Networks are: convolutional, pooling, and fully connected layers. The main purpose of the convolutional layers is to highlight areas of input images that have the potential of serving as good features. Usually edge detection filters are used as the convolutional filters for this purpose as edges of objects serve as good features. Pooling follows convolutional layers, where a common function that is usually used in this step is the max function. It chooses the maximum value among each window of values, while eliminating all the reminding values. This process reduces the size of the input data which in turn reduces the number of connections in the neural network. Finally, the selected features are combined and provided to the fully connected layer to perform the classification.

To train a Convolutional Neural Network, a random initialization of the convolutional filters weights is performed in addition to the weights of the fully connected layer. A ReLU operation is usually used to trim any negative values and inject nonlinearity in the system, the operation has proven to improve the overall performance of a Convolutional Neural Network. All steps from taking an input image to reaching the output are iterated and repeated in the same manner until all input images have passed through the network. Training Convolutional Neural Networks usually uses datasets that don't include a lot of degraded images, hence when testing those networks with degraded images their performance is likely to drop. In this work we will consider, blurring, noise (Gaussian), and occlusion to study the effect of degradation on performance. A set of degraded images with different levels of degradation is generated and included in the experiments. Blurring is achieved by smoothing the image with a Gaussian low pass filter. For the noise part, an additive zero mean Gaussian noise is used.

In this work, our contribution will be to introduce a unified quantitative comparison between different types of degradation based on the root mean square error metric. The rest of the paper is organized as follows: Section 2 presents the related work. Section 3 introduces the methodology. Section 4 provides the results and discussion. Finally, the conclusion is presented in Section 5.

2. Related work

Degraded image classification has been studied in several recent works. In [7], the effect of degradations on face identification was studied. Kai et al. [8] worked on images with white Gaussian noise and how to implement a CNN de-noisers to elevate the effect of the noise on the accuracy of the classification. Kalalembang et al. [9] offered a technique of spotting undesirable motion blurring, while Ramakrishnan et al. [10] presented a way of de-blurring such image. Pei et al. [11] studied the effect of image degradations on the performance of CNN-based image classification and whether degradation removal helps to improve accuracy. But since each degradation type is controlled by different parameters, a unified quantitative comparison was lacking.

Although convolutional neural network is extensively used in numerous image classification tasks, the effect of image degradation is not thoroughly studied. Therefore, in presence of different image degradations, the performance of different deep learning designs in classification tasks containing different challenging images is considered in this work. Several kinds of image degradations, like noise due to sensor malfunctions, down sampling, and motion blurring can significantly affect the quality of an image, which results in hindering the performance of any system that uses input images that are infected with those degradations.

In [12] a method for the classification of degraded images with various levels of degradation is tackled. The work proposes a convolutional neural network based system to classify degraded images by using a restoration network and ensemble learning. They claim that their results demonstrate that the proposed network can classify degraded images over various levels of degradation. In [13] the performance of convolutional neural network of images degraded by compression artifacts, blur or noise. They analyze some of the common degradations in images like: Gaussian noise, blur and compression. Authors in [14] propose a system for searching surveillance video contents that uses CNN for object recognition and classifications. Such a system deals with images that are extracted from video frames, those videos are usually stored after being compressed to reduce their enormous size, but such reduction in size comes at a cost of reducing the quality.

In [15] an image enhancement convolutional neural network (CNN) based on ResNet is proposed to learn the implicit mapping model between degraded and non-degraded images in the spatial domain. So, multiple degradations in images are considered instead of a single degradation. The author in [16] proposes a deep learning approach for handwritten offline Arabic digit recognition using convolutional neural networks. Such system performance is affected by any degradation in the input images, so studying this topic will enhance the modeling and training of such systems. The authors in [17] employ fully convolution network (FCN) for the detection of text in an image and identification of its language, they use FCNs for both model enhancement and classification.

Machine learning classification techniques have been implemented and proposed for many applications like age or gender estimation. Some of these techniques tackle different challenges related to image conditions and suggest different approaches to handle them [18]. In [19] the authors propose a visibility detection method using deep convolutional neural networks that handles issues caused by the absence of adequate visibility labeled datasets. In the work proposed, each input image was first divided into several areas, which were encoded to extract visual features using a pre-trained unidentified image quality assessment neural network. Lastly, the neural network was fine-tuned to fit the problem of visibility detection using the current recognition results equally.

Dodge and Karam [20] provide an evaluation of the effect of image distortions on convolutional neural networks image classification capabilities. They consider the following degradations: blur, noise, contrast, and compression (JPEG). They show that blur and noise are more affecting than other types. Low quality images should improve testing results on low quality images, but perhaps this may also result in decreased performance on high quality images. Furthermore, training with increased number of samples leads to longer training times. An investigation of the benefits of training with low quality samples is left for future work.

3. Methodology

In this work, GoogLeNet [21] is used to represent deep neural networks. The testing will be limited to only this neural network so that the work will concentrate on studying how a deep neural network is affected by different types of degradation rather than emphasizing on the differences between different networks. GoogLeNet was trained using the ImageNet database. ImageNet consists of about one million images that are categorized into 1000 classes, for each class hundreds if not thousands of images are used to train the network. For a given input image the network generates (in its final layer) probability values for that image chance of belonging to each of the 1000 classes it trained to recognize. The size of the input image has to be 224 x 224 x 3, otherwise it is resized to match those values of the GoogLeNet input layer. GoogLeNet consists of 22 layers thus it is considered deep enough to learn enough features to classify images with high accuracy.

For testing data a subset of the validation data of the ImageNet dataset [22,23] will be used to speed up the process. 5,000 of the available 50,000 images will be chosen randomly under the constraint that each of the 1,000 classes is represented by 5 images. From each of those chosen images, a set of degraded images with varying levels of degradation is generated to run the experiments. The parameters chosen to alter the quality of the image are: blurring, noise (Gaussian), and occlusion. Blurring is achieved by smoothing the image with a Gaussian low pass filter. The value of σ (sigma) of the filter will be changed to apply different levels of blurring. For the noise part, an additive zero mean Gaussian noise with different variance values is applied. It should be noted that the value of square root of variance applied should be equal to the root mean square error that we will use and define later. The top-5 accuracy is considered for the accuracy calculation. It considers a result as right if the right class is in the top 5 predicted classes. Top-5 accuracy is usually preferred over top-1 accuracy due to the fact that for many images in the database there is more than one object in the image. The label of that image is one of the objects in the image (probably the largest), so top-5 accuracy enables the system to guess some of the other objects in the image as a correct result.

The quality of an image will be altered by the addition of the noise, so in order to quantify that distortion a quality measurement is required, in this paper the root mean square error (RMSE) will be used for that purpose. The RMSE is calculated using equation 1.

$$RMSE = \sqrt{\frac{1}{MN}\sum\sum(f(x,y) - f^{\backslash}(x,y))^2} \quad (1)$$

Where: M is the length of image, N is the width of image

f(x,y) is the pixel value of original image at x and y coordinates and f'(x,y) is the corresponding pixel value of noisy image. Figure 1 shows two degraded images with the left having a Gaussian noise added o it while the one to the right blurred by a low pass filter.

Figure 1. Sample degraded images infected with noise and blurring respectively.

4. Results and Discussion

The experiments were carried out and the following results were achieved; Figure 2 shows the accuracy of the CNN versus the added noise, the RMSE is used to measure the degradation of the image. As expected, the recognition rate decreases as the variance of the noise increases.

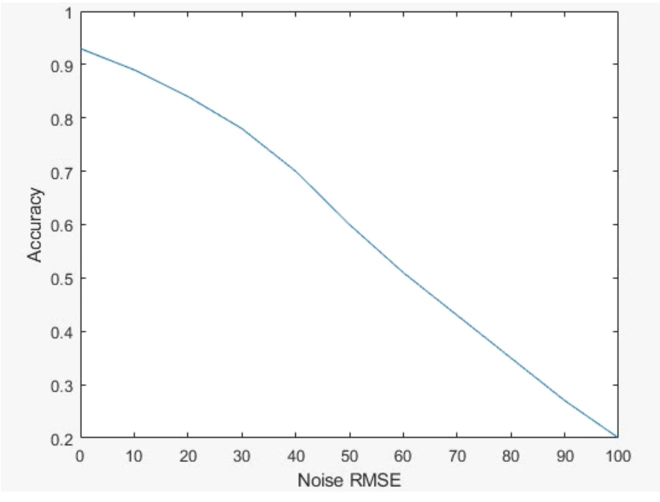

Figure 2. The accuracy of the CNN versus the added noise.

The second experiment deals with altering the contrast of the image while measuring the accuracy of the CNN. The variance of the Gaussian filter is used to measure the amount of blurring induced. Figure 3 shows the accuracy of the CNN versus the added blurring.

In the third experiment, the effect of occlusion on the accuracy is studied. The testing images where occluded by a varying percentage and the accuracy where recorded. Figure 4 shows the accuracy of the CNN versus the percentage of the occlusion. It is obvious that the recognition rate decreases rapidly as the percentage of the occlusion increases

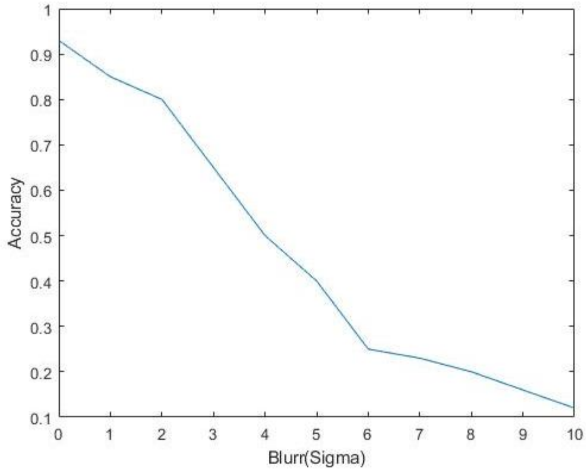

Figure 3. The accuracy of the CNN versus the added blur.

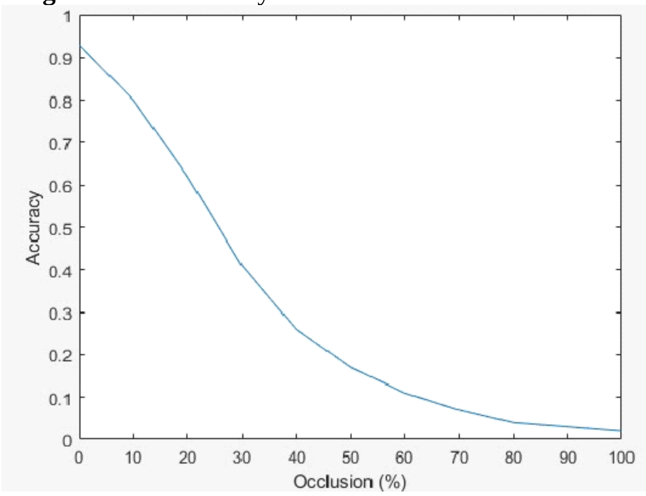

Figure 4. The accuracy of the CNN versus the percentage of the occlusion.

To allow a fair quantitative comparison between the three types of degradations, all types will be measured by the RMSE equation mentioned previously. For the occlusion degradation, the occluded part will be replaced by the average of the image pixel values, so that an occluded featureless area will not contribute much to the RMSE. All the curves were redrawn to match the equivalent RMSE for each corresponding type of degradation. The results are shown in Figure 5.

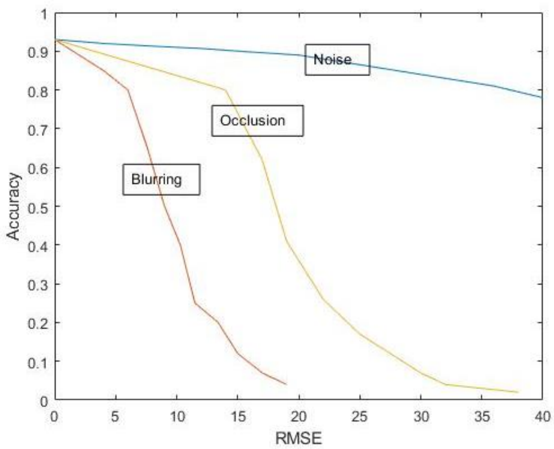

Figure 5. The accuracy of the CNN of different types of degradations versus the RMSE.

The results show that the severity of effect of each degradation type is different from others. It is obvious that blurring and occlusion affects accuracy more than noise.

Based on the understanding of how CNN work, the early convolutional layers purpose is to extract the important features that are supposed to distinguish an object from other objects. An occlusion that will result in masking a good portion of those features will definitely alter the extracted features resulting in an incorrect match. Also, blurring usually eliminates edges that are mainly used by CNN in the early stages of convolution to extract features.

To further emphasize this idea, the Sobel filter which is one of the most commonly used operators for edge detection will be considered. It is a gradient based edge detection filter that consists of two kernels Gx and Gy as shown in Figure 6. The purpose of having two kernels is to capture the change of pixel value in both the horizontal and vertical directions. The two kernels weights represent an approximation of each component of the gradient, the approximation is favored over the actual value (which represent the magnitude of the gradient) because it doesn't include the calculation of the square root that the magnitude of the gradient include.

$$
\begin{array}{ccc}
1 & 0 & -1 \\
2 & 0 & -2 \\
1 & 0 & -1
\end{array}
\qquad
\begin{array}{ccc}
1 & 2 & 1 \\
0 & 0 & 0 \\
-1 & -2 & -1
\end{array}
$$

Gx Gy

Figure 6. The kernels of the Sobel filter

Motion blur is a usual type of blur usually caused by camera shaking or fast-moving of the pictured objects. Such blurring affects the selected feature points from the subject image by the edge detection filter resulting in lowering the similarity measure produced by the fully connected neural network.

The similarity measure reduction is logically proportional to the percentage of the lost feature points due to the degradation. In the occlusion case, the percentage can be simply assumed to be equal to the percentage of occlusion, while in the additive noise case it will be approximately equal to the percentage of noisy pixels. But in the case of blurring it is not obvious how many feature points will be lost as the blurring do affect all the image pixels and it may or may not resulting in that point being left out.

5. Conclusions

In this work, an analysis of how image related parameters such as contrast, noise, and occlusion affect the work of convolutional neural networks was carried out. To aid that, a series of experiments were implemented. Those experiments concentrated on calculating the recognition rate drop with each type of image degradation. The results revealed that the extent of effect of each degradation type is different from others. It was obvious that blurring and occlusion affects accuracy more than noise. As a future work, including the degraded images in the training set should be considered, even though it is not guaranteed that such an approach should improve the performance or increase the recognition accuracy, but it should give an insight on how deep neural networks react to it. A further milestone could be the study of CNN adaptation to overcome those degradations and if the human brain uses the same mechanism.

6. Acknowledgement

Inad A. Aljarrah thanks North Carolina State University for hosting him as a visiting scholar during the completion of this work. Special thanks go to the department of Electrical and Computer Engineering (ECE) and prof. Hamid Krim as the hosting professor. Also, I want to mention that this work was supported by Jordan University of Science and Technology.

References

[1] P. Kamavisdar, S. Saluja, and S.Agrawal, " A survey on image classification approaches and techniques," International Journal of Advanced Research in Computer and Communication Engineering, 2(1), pp.1005-1009. 2013, 22-A Survey on Image Classification.pdf (ijarcce.com)

[2] Y. Bengio, A. C. Courville, and P. Vincent. "Unsupervised Feature Learning and Deep Learning: A Review and New Perspectives," CoRR, abs/1206.5538, 2012. http://arxiv.org/pdf/1206.5538v1.pdf

[3] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. "CNN features off-the-shelf: an astounding baseline for recognition," In Computer Vision and Pattern Recognition Workshops (CVPRW), 2014. 1, 2.

[4] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurons in the cat's striate cortex," The Journal of physiology, vol. 148, no. 3, pp. 574–591, 1959.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, pp. 1097–1105, 2012.

[6] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., "Imagenet large scale visual recognition challenge," International Journal of ComputerVision, vol. 115, no. 3, pp. 211–252, 2015.

[7] S. Karahan, M. Kilinc Yildirum, K. Kirtac, F. S. Rende, G. Butun and H. K. Ekenel, "How Image Degradations Affect Deep CNN-Based Face Recognition?," 2016 International Conference of the Biometrics Special Interest Group (BIOSIG),pp. 1-5, 2016 doi: 10.1109/BIOSIG.2016.7736924.

[8] Z. Kai, W. Zuo, S. Gu, and Z. Lei, "Learning deep cnn denoiser prior for image restoration," in IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[9] E. Kalalembang, K. Usman, and I. P. Gunawan, "Dct-based local motion blur detection," in International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering. IEEE, pp. 1–6. 2009.

[10] S. Ramakrishnan, S. P. A. Gangopadhyay, and S. Raman, "Deep generative filter for motion deblurring," in IEEE International Conference on Computer Vision Workshop, 2017.

[11] Y. Pei, Y. Huang, Q. Zou, X. Zhang and S. Wang, "Effects of Image Degradation and Degradation Removal to CNN-Based Image Classification," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 4, pp. 1239-1253, 1 April 2021, doi: 10.1109/TPAMI.2019.2950923.

[12] K. Endo, M. Tanaka and M. Okutomi, "Classifying Degraded Images Over Various Levels Of Degradation," 2020 IEEE International Conference on

Image Processing (ICIP), pp. 1691-1695, 2020,doi:10.1109/ICIP40778.2020.9191087.

[13] S. Ghosh, R. Shet, P. Amon, A. Hutter and A. Kaup, "Robustness of Deep Convolutional Neural Networks for Image Degradations," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2916-2920, 2018 doi: 10.1109/ICASSP.2018.8461907.

[14] D. Mohammad, I. Aljarrah, and M. Jarrah "Searching surveillance video contents using convolutional neural network," International Journal of Electrical and Computer Engineering (IJECE),Vol. 11, No. 2, April 2021, pp. 1656~1665,ISSN: 2088-8708, DOI: 10.11591/ijece.v11i2.pp1656-1665.

[15] K. Wang,, L. Zhuo, J. Li, et al. "Learning an Enhancement Convolutional Neural Network for Multi-degraded Images," Sens Imaging 21, 25 (2020). https://doi.org/10.1007/s11220-020-00289-0

[16] J. Alkhateeb, "Handwritten Arabic Digit Recognition Using Convolutional Neural Network," International Journal of Communication Networks and Information Security (IJCNIS), 12(3),2020. Retrieved from https://www.ijcnis.org/index.php/ijcnis/article/view/4807.

[17] A. Khalil, M. Jarrah, M. Al-Ayyoub, and Y. Jararweh, "Text detection and script identification in natural scene images using deep learning," Computers & Electrical Engineering, Volume 91, 2021, 107043, ISSN 0045-7906,https://doi.org/10.1016/j.compeleceng.2021.10704

[18]] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," IEEE Transactions on Information Forensics and Security, vol. 9, no. 12, pp. 2170–2179, 2014.

[19] Q. Li, S. Tang, X. Peng, and Q. Ma, ."A Method of Visibility Detection Based on the Transfer Learning," Journal of Atmospheric and Oceanic Technology, 36(10), pp. 1945-1956. https://journals.ametsoc.org/view/journals/atot/36/10/jtech-d-19-0025.1.xml

[20] S. Dodge and L. Karam, "Understanding how image quality affects deep neural networks," 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX), pp. 1-6, 2016, doi: 10.1109/QoMEX.2016.7498955.

[21] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. "Going deeper with convolutions," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 1–9, 2015.

[22] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," International Journal of Computer Vision (IJCV) , vol. 115, no. 3, pp. 211–252, 2015.

[23] J. Deng, W. Dong, R. Socher, L.-j. Li, K. Li, and L. Fei-Fei. "Imagenet: A large-scale hierarchical image database," In 2009 IEEE conference on computer vision and pattern recognition ,pp. 248–255, 2009. https://image-net.org/papers/imagenet_cvpr09.pdf