International Journal of Mathematics and
Computer Science, **19**(2024), no. 4, 1389–1399

The Impact of the Number of Training Categories on the CNN Performance

Noaman Abduljabbar Ramadhan[1], Maha Ammar Mustafa[2]

[1]Ministry of Education
General directorate of educational planning
Baghdad, Iraq

[2]Branch of Mathematics and Computer Applications
Department of Applied Sciences
University of Technology-Iraq
Baghdad, Iraq

email: num_k@yahoo.com, mahaalkhazraji89@gmail.com,
100384@uotechnology.edu.iq

(Received April 22, 2024, Revised May 24, 2024,
Accepted May 29, 2024, Published June 1, 2024)

Abstract

The use of image classification have been hugely increased in the last decade and specifically with the large improvement in deep learning. Many recent fields use image classification applications which make them a motivating area for business, education, and research fields. In this work, convolution neural network (CNN) model is trained specifically on human face detection and face emotion recognition, the tests implemented several times on different datasets and different number of categories or classes to test the impact of the categories number on the convolution neural network performance. Four tests are implemented on the convolution neural network, in each test of them, the convolution neural network is trained on different number of categories or classes, and the best obtained accuracy is recorded. In

Key words and phrases: Convolution neural network, Face detection, Emotion recognition, Image classification.
AMS (MOS) Subject Classifications: 68R01.
ISSN 1814-0432, 2024, http://ijmcs.future-in-tech.net

addition, this work provide an elementary comprehending of the Convolutional Neural Network, their affecting factors, capabilities, and limitations. Such an effort can help later in improving it.

1 Introduction

Many image classification methods are introduced and all of them are mainly competed on improving the classification performance either by increasing the accuracy rate or reduce the time consumed while training and detection/ recognition rate [1]. CNNs are considered as a type of neural networks that are widely used for image classification and detection problems.

2 Convolution Neural Networks

As the most successful recent computer vision algorithm, Convolution Neural Networks (CNN) have been widely adopted in image classification field as the core algorithm. Convolutional Neural Network is more suitable for computer vision problems, such as image classification and object detection problems when compared with other classic neural networks because its neurons in the convolutional layer is only connected to a part of the input data, unlike in other neural networks, neurons are fully-connected to the input data. In fact, the CNN unique architecture makes CNN requires less pre-processing. Therefore, a Convolutional Neural Networks is the most effective learning algorithm for understanding picture material [2].
Convolutional neural networks have a wide range of applications in many fields, such as medical applications [3], Self-driving or autonomous cars, Auto translation, and much more. However, they mostly used in solving problems related to face recognition, face or object detection, image analysis, and image segmentation. Convolutional Neural Networks is considered as a supervised deep learning based algorithm [4]. Since 2016, together with the deep learning evolution, CNN became widely used in computer vision problems. It is one of the feed-forward neural networks that utilize the technique of back-propagation for training. It is capable of learning spatial hierarchies of features in an automatic and adaptive way through back-propagation using several building blocks, such as convolution, pooling, and fully-connected layers [5]. Convolutional Neural Network models are excelled in fields, such as object detection and classification. Convolutional Neural Networks have the ability of performing both feature extraction and classification handcraft-

ing features [6]-[7]. In addition, the number of parameters remains constant whatever the input image grows, and this enable them to train more difficult and important networks [8]. Therefore, Convolutional Neural Networks is considered as a powerful automatic feature extractor [9].

3 Convolution Neural Network Components

In general, convolutional neural networks are consist of two main components, these are, feature extraction part and classification part [10].

3.1 The feature extraction part

Features are detected and extracted in this part, through a sequence of convolution as well as pooling operations. Low level features, like simple shapes, are detected in the first layers, such as edges. While higher level features, which is more complicated shapes, are detected in the next layers, such as human face.

3.2 The classification part

Classification is implemented after the first part, i.e. after the sequence of convolution and pooling layers. Classification is implemented through a few dense layers. These dense layers are only use one dimension data. Consequently, a flatten function is used to convert the two dimensional to one dimensional data. Indeed, these dense layers consist neurons that are fully connected to all activations within the previous layer.

4 Some convolution neural networks challenges

- CNN requires large and diverse dataset.
- Training CNN requires high computation power such as GPU to reduce the training time.
- Selecting the hyper parameters can highly affect the convolutional neural network performance. A slightly change in these values could influence the overall CNN performance.

5 The motivation of using convolutional neural network for human facial recognition

The CNNs were chosen for this study, because they are especially effective at image classification problems due to their concept of using dimensionality reduction that suits the huge number of parameters in images, and provide high accuracy. CNNs have advantages over other methods in image classification field because they are considered as a multi-classification classifier, which means it can classify instances into one of three or more classes. In addition, CNNs learn distinctive features for each class automatically.

6 The used convolutional neural network model structure

In this work, the CNN model consist of eleven layers divided into two main parts. The first part is the feature extraction part, which consist of eight layers (eight convolution layers and four max pooling layers). For more none linearity, the ReLU activation function is used on the top of the convolution. ReLU is efficiently higher than other activation functions because it only allows some (not all) neurons to be activated at a time [11]. After feature extraction part comes the second part, which is the classification part that contains only three dense layers (flatten layer, dense layer, and the (last) output layer). The output layer use softmax activation function, which is used for multi-class classification problems. The CNN model was implemented using Python and the Keras library, leveraging GPU for enhanced performance.. FER2013 dataset as well as images from the internet is used for the training.
The CNN architecture composed of eleven layers. The layers are organized in seven blocks to make it easier. The structure is the same for all the initial four blocks. Each block of them composed of two convolution layers and one maxpooling layer, as well as ReLU activation function, batch normalization, and dropout. The convolution layers in the first block used 32 kernels of size (3x3), while 64 kernels are used for the convolution layers in the second block, Then, 128 kernels are used for the third block, and 256 kernels are used for the last convolution layers. Additionally, the activation function (ReLU) is used at the end of each convolution layer. ReLU activation function is very important to eliminate the exponential growth occurs in the computations needed for operating the neural network, because it will activate only

a certain number of neurons at a time. Add to that, the (2x2) max pooling layer is used in each block after implementing the second ReLU activation function. The pooling operation is important to extract sharp and smooth shapes. The max pooling used to extract low level features, such as edges, and point.

The fifth block comes after the previous layers, the fifth block is performed by applying flatten and fully-connected layers as well as the ReLU activation function, batch normalization, and dropout. Flattening is implemented by converting the two dimensional arrays resultant from the pooled feature maps into a one dimension linear vector. The resultant image matrix from the forth block is converted into a flattened vector in the fifth block, this conversion is obtained by converting into a flattened vector, which contain (64) neurons, in the fifth block. The fifth block output, which is a dense layer of 64 neurons, is fed as an input to the sixth block, which is also a dense layer that composed of 64 neurons. Lastly, the output of the latter block is fed into the last block. At last, the last (seventh) block is also composed of fully connected layer that composed of only seven neurons (each neuron of the output refer to one class). The output uses Softmax activation function because the model is a multi-class. The CNN model is implemented after specifying the classes/ categories number (7), image dimensions (48x48), and batch size (32). The paths of the train and validation datasets are also specified. After that, the ImageDataGenerator in Keras is used to increase the dataset by creating new images from the existing images. It takes a set of images from the training dataset, and implement a number of random manipulations on each image of the set. The used image transformations are: (Rotation range =30, Shear range = 0.3, Zoom range = 0.3, Width shift range = 0.4, Height shift range = 0.4, Horizontal flip = True, Fill mode = nearest.)

Moreover, the optimizer is used in order to decrease the losses by changing the model attributes, such as the learning rate and the weights. The used optimizer in this work is Adam. Next to that, actions at different training stages are performed by specifying Keras callbacks. For instance, at the beginning or at the end of an epoch. The utilized callbacks are: a (Model Check Point) is utilized for saving the model or the weights at some period of time in a check point file, so the model or weights can be loaded later to continue the training from the state saved. Moreover, (Early Stopping) is utilized for monitoring the trigger by specifying the performance measurement, when the trigger occur, the train process will end. Add to that, if the metric improvement has stopped, the (Reduce LROn Plateau) is implemented for decreasing the learning rate. After that, the loss function, optimizer, and metrics are

used by calling the compiler. In this case, loss function = categorical cross entropy, optimizer = Adam, and metrics= accuracy.
Another things in the model are specified, such as the amount of validation and training samples, and the epochs number. The CNN model also used (Model fitting) to show how well a machine learning model generalizes to data similar to that on which it was trained. In fact, a model must produce more accurate outcomes to be considered as a well-fitted. When the model just produce data that is similar to the training data, the model is called over-fitted.

7 The CNN implemented tests

In this work, the Convolutional Neural Network model is trained on detecting human face as well as on recognizing face emotions. The model is performed using Python software with the help of a GPU and Keras library. FER2013 dataset is used to train the model. The Convolutional Neural Networks model is trained 3 times. Each time Convolutional Neural Networks model is trained on different number of categories.
1. **First test** was implemented by training the Convolutional Neural Network model on seven categories/classes of emotions (from FER2013 dataset), emotions appear on the human face, such as happiness, sadness, surprise, neutral, anger, fear, and disgust. The total training set was equal to (26,144) samples and the total validation set was equal to (7,066) samples. Epochs and learning rates values were changing many times to accomplish the best accuracy, which in this case equal to %63. Fig.(1). Shows the highest obtained value for train accuracy per epoch which was (%55), and the highest validation accuracy per epoch which was (%63).

2. Second test was implemented using just five categories from FER2013 (happiness, sadness, surprise, neutral, and anger). The overall training and validation samples equal to (24,282) and (5,937) respectively. This test was also implemented many times until obtain the best accuracy, which is equal to %71. Fig.(2) shows the case of the highest achieved accuracy for both train and validation phases. The train accuracy was (%63) while the validation accuracy was (%71).
3. Third test was implemented by training the CNN model on only two categories/classes to achieve face detection, i.e. the first class set consist of

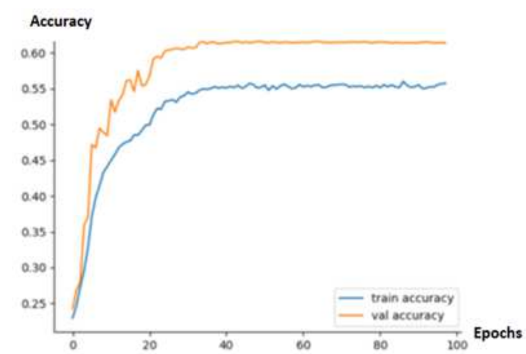

Figure 1: The plot of the highest obtained accuracy per epoch for both training and validation phases in the case of (7) classes, The orange curve shows the validation accuracy while the blue curve shows the train accuracy.

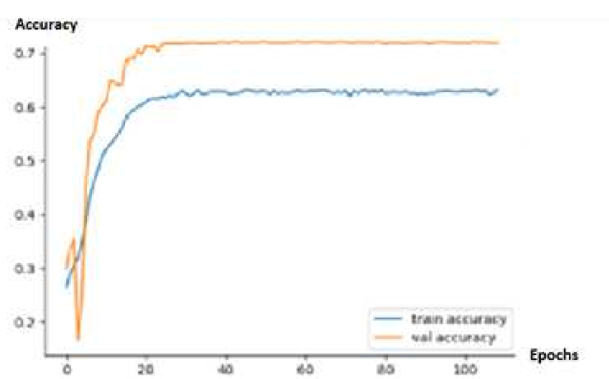

Figure 2: The plot of the highest obtained accuracy per epoch for both training and validation phases in the case of (5) classes, The orange curve shows the validation accuracy while the blue curve shows the train accuracy.

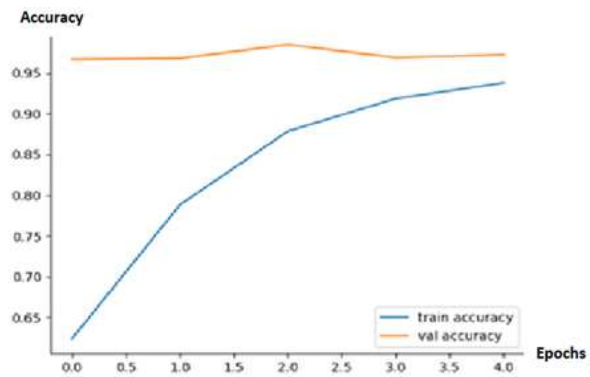

Figure 3: The plot of the highest obtained accuracy per epoch for both training and validation phases in the case of (2) classes, The orange curve shows the validation accuracy while the blue curve shows the train accuracy.

images of faces, and the second class set consist of images of different subjects (such as nature and cars). However, images does not contain any human face. (7,629) samples were used for training and (5,156) samples were used for validation. In this case, the best accuracy achieved was %97. The graph plot in Fig.(3). shows the highest train and validation accuracies achieved in the case of (2) classes. the train accuracy was (%94) and the validation accuracy was (%97).

4. The last implemented test was similar to the previous test, i.e. CNN model was also trained on face detection (classifying images into faces and non faces). But this time, the model trained on another dataset, which consisted of (15300) training samples and (14000) validation samples. Although, this test differs from the previous one in both training dataset and the number of training and validation samples, but the best achieved accuracy was close to the previous one, which in this case is equal to %98. In addition, this test implemented twice. Once the model was trained with just %50 of the dataset. The other one was trained with %100 of the dataset. The detection rate in the second test was much better.

8 Tests results

Some of the concluded points regarding the previous implemented tests, Convolutional Neural Networks can be utilized for both object detection and object classification. CNNs are flexible in their structure, which consist of two parts, feature extraction and classification, and this is make them suitable for various models. The implemented tests show that the accuracy declined when the classes/categories number increased and they may overlap. For instance, as shown in the previous tests, the CNN model was implemented in four cases with the following different number of classes (7, 5, 2 and 2), and the achieved accuracies were %63, %73, and %97, and %98 respectively. Using Convolutional Neural Networks require a very large and diverse training data sets as well as a very high computation power is needed. Consequently, it is important to use high computation platforms with GPUs of computation capability equal to (3.5) or higher.

Convolutional Neural Network performance in emotion recognition is affected by many factors, such as the diversity and size of dataset. In addition, images contain faces with high expression intensity can be recognized much better than images with low expression intensity. Moreover, some images contain faces expression with mixed or compound emotions, such as faces show two basic emotions like happily surprised or sadly surprised, in this case the recognition accuracy might be incorrect. Add to that, low resolution images and/ or illumination with high variations might cause incorrect recognition too. Some recent studies show that expression intensity in females differs from males, females usually express emotions more than males, and this might influence emotion recognition accuracy. Other studies show that some emotion nature is better recognized than others. i.e. (happiness) is among the easiest emotions to recognize while (disgust) is among the difficult emotions to recognize. Some important advantages and disadvantaged of using Convolutional Neural Networks in image classification field are listed in Table (1):

Table 1: Advantages and Disadvantages of using Convolutional Neural Networks in image classification.

9 The conclusions

This paper introduced some important factors that affecting the Convolutional Neural Network performance and specifically concentrate on the impact of the number of categories (classes) used in the training dataset. In

No.	Advantages	Disadvantages
1	CNN is capable of extracting features automatically.	To obtain high recognition accuracy, CNN requires huge and diverse training dataset.
2	CNN accuracy is very high in the case of image classification problems.	CNN is unable to realize the location and orientation of the objects within the image [12].
3	CNN consume less time for training compared to other face detection algorithms such as Viola-Jones algorithm.	CNN models improving requires effort and experience for specific cases.
4	Weight sharing used in CNN decrease the number of parameters in the trainable network, i.e. enhance the network generalization as well as avoid overfitting [4].	

this work, the Convolutional Neural Network model implemented in various face detection and emotion recognition tests. The implemented tests used different number of dataset classes or categories in the CNN training stage, as well as used different datasets. It is shown from these implemented tests that, the accomplished accuracy might be influenced. The accuracy could be declined when the number of trained classes or categories is increased.

References

[1] B. I. Farhan, A. D. Jasim, Survey of intrusion detection using deep learning in the internet of things, Iraqi Journal for Computer Science and Mathematics, **3**, no. 1, (2022), 83–93.

[2] A. Khan, A. Sohail, U. Zahoora, A. S. Qureshi, A survey of the recent architectures of deep convolutional neural networks, Artificial Intelligence Review, **53**, no. 8, (2020), 5455–5516.

[3] R. S. Saeed, B. K. O. Chabor Alwawi, A binary classification model of COVID-19 based on convolution neural network, Bulletin of Electrical Engineering and Informatics, **12**, no. 3, (2023), 1413–1417.

[4] L. Alzubaidi et al., Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, Journal of Big Data, **8**, no. 1, (2021).

[5] R. Yamashita, M. Nishio, Convolutional neural networks, an overview and application in radiology, Insights into Imaging, (2018).

[6] V. H. Phung, E. J. Rhee, A High-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets, Applied Sciences, **9**, no. 21, (2019).

[7] K. Liu, M. Zhang, Z. Pan, Facial Expression Recognition with CNN Ensemble, Proc. Int. Conf. Cyberworlds, CW (2016), 163–166.

[8] S. O. Zaben, A. E. Ali, Detection Covid-19 Based on Chest X-ray Images Using Convolution Neural Networks, Iraqi J. Comput. Commun. Control Syst. Eng., **22**, no. 1, (2022), 34–42.

[9] K. Komala, D. Jayadevappa, G. Shivaprakash, K. Komala, Human Emotion Detection and Classification Using Convolution Neural Network, Eur. J. Mol. Clin. Med., **7**, no. 6, (2020), 237–245.

[10] A. S. Dawood, Z. M. Faris, Architecture of Deep Learning and Its Applications, **23**, no. 1, (2023), 35–56.

[11] S. Sharma, S. Sharma, A. Athaiya, Activation Functions in Neural Networks, Int. J. Eng. Appl. Sci. Technol., **4**, no. 12, (2020), 310–316.

[12] A. Asmaa, K. Maisa, A. Ahmed, Object Detection Using Deep Learning Methods: A Review, Iraqi J. Comput. Commun. Control Syst. Eng., **23**, (2023).