Facial Emotion Recognition and Impact Analysis of Iteration Based on Convolutional Neural Networks

Anna Li
The Faculty of Science, The University of Hong Kong, Hong Kong, China

Keywords: Facial Emotion Detection, CNN, Algorithms, Human Emotions.

Abstract: Many disciplines, including AI, psychology, and computer science, see considerable value in the ability to recognize human facial expressions. The primary goal of this research is to analyze the performance of the original Convolutional neural networks (CNN) model by examining how alternative CNN models perform in facial expression recognition under varying iteration cycles. This approach recognizes the intricate relationships between macro and micro expressions to present a more complete picture of the wide range of human emotions conveyed through facial expressions. A key takeaway from the experiments is that with repeated iterations, the model becomes increasingly accurate. This approach to facial emotion detection exemplifies the feasibility of combining various neural network architectures, allowing people to delve even deeper into the nuances of human emotion. Therefore, this research has made major contributions to the field of facial expression recognition by displaying the effectiveness of incorporating multi-scale feature extraction technologies to enhance the performance of the model. This study establishes the groundwork for future research avenues in the area and enables the development of more sophisticated emotion recognition algorithms for practical implementations.

1 INTRODUCTION

Since Darwin's early work, which spurred a frenzy of academic interest in the field, there has been a great deal of progress made in the study of human emotions (Ali t al 2020). Specifically reflected it can be achieved to identify seven separate emotions plainly by observing a person's face: neutral, anger, disgust, fear, happy, sad, and surprise. Then, people can comprehend others' intentions by observing their facial expressions, which are vital parts of human communication (Ko 2018). However, it is undeniable that identifying emotions has been a challenging issue for a long time since it is a subjective phenomenon that calls for the utilization of knowledge and scientific data behind labels to extract components (Dachapally 2017).

Multiple approaches have been put forth over the past several decades for recognizing emotions. In addition, conventional techniques often take into account a facial photograph that has been segmented from a source image, followed by detect facial components in the fragmented face regions (Ali t al 2020). As an illustration, the gradient histogram of the image will be computed to extract facial features by applying the combination of gradient graph (HoG) features and support vector machine (SVM) classifiers, afterwards the face areas and feature point locations can be effortlessly and precisely identified (Liu et al 2013). Since the advent of deep learning, the research topic of emotion recognition in computer vision has become mostly resolved. Meanwhile, in terms of accuracy and efficiency, it continues to be outpacing other machine-learning approaches (Giannopoulos et al 2018). With Frank Rosenblatt's 1957 idea of perceptron, a simple type of neural network, and the subsequent development of multi-layer perceptron (MLP), a more complex type of neural network that added techniques to differentiate non-linearly separable data based on single-layer primitives, facial characteristics can be extracted from images (Boughrara et al 2016). In the case of image processing, however, the MLP model may have limitations. To process complicated visual data, a significant number of parameters is typically required. More hidden layers and neurons are needed for high-resolution picture data, which extends MLP's training and inference times (Panchal et al 2016). Because every neuron in the MLP is attached to each neuron in the layer below, this fully interconnected technique may not be effective at processing picture data. On the

Li, A.
Facial Emotion Recognition and Impact Analysis of Iteration Based on Convolutional Neural Networks.
DOI: 10.5220/0012799000003885

In *Proceedings of the 1st International Conference on Data Analysis and Machine Learning (DAML 2023)*, pages 5-9
ISBN: 978-989-758-705-4

other hand, lacks a weight-sharing mechanism, which results in an excessive number of parameters for the model (Tang et al 2022). The Deep belief networks (DBN) aimed to optimize the matching of facial expression recognition datasets by resolving the training problem of deep neural networks (Terusaki, and Stigliani & Goodfellow et al 2013). It even needs dedicated hardware to accelerate the training process. As with other types of neural networks, such as the convolutional neural network (CNN) is currently frequently employed within particular settings, such as those concerning recognition of photographs and classifying. Owing to its capacity to automatically learn features in images, thus enables them to effectively capture and describe the image's details.

To further enhance both the effectiveness and accuracy of face emotion detection characteristics. Consequently, the core objective of the current investigation is to figure out the ways that how the various numbers of iterations impact CNN's ability to detect a person's expressions. In this research, the Facial Expression Recognition 2013 Dataset (FER-2013) utilized to achieve both training and assessment purposes, which is available at Kaggle (Giannopoulos et al 2018). CNNs are primarily utilized, which are divided into three stages: image processing, feature extraction, and performance analysis. Specifically, first, before computing features, images must undergo pre-processing, which mostly entails data augmentation and normalization to remove noise and other distracting elements from the image that are not connected to the face. Second, is feature extraction, which is the process of extracting face image-related feature data by generating CNN models (convolutional kernels), thereby offering useful data features for subsequent facial expression identification. The third is performance analysis, observing the impact generated by varying iteration cycles on the recognition of facial expressions in Convolutional Neural Networks. The experiments conducted reveal that the accuracy of CNN's face expression recognition technology improves with increased training iterations, and the loss subsequently diminishes. Additionally, it is essential to understand that some factors, like the network structure, the dataset, and others may all impact the effects associated with various iteration cycles on CNN facial expression identification. Therefore, if the goal is to figure out the optimal amount of iteration cycles for any given project must be determined through experimentation and evaluation.

2 METHODOLOGY

2.1 Dataset Description and Preprocessing

FER-2013, the dataset utilized for this analysis, has been obtained from Kaggle (Giannopoulos et al 2018). Google Image Search API is used to crawl images that matched emotive keywords, providing the dataset's foundation. The dataset contains 35887 photos of face expressions, split into three groups of 28709 for training, while the public test and private test each contain 3589 images. The photos have been digitally manipulated to guarantee that is approximately aligned and the visage of the subject covers about the identical proportion of the area for every pictures.

Images' pixel values and an emotion label are the two fundamental components of the collection. People face in the grayscale images, all of which are 48 pixels on a side, and are composed of seven varieties of emoticons, which correspond to digital labels 0 through 6. Each of these emoticons has a corresponding name in both Chinese and English: Angry (0), Disgust (1), Fear (2), Happy (3), Sad (4), Surprise (5), and Neutral (6). An example of a data point in this dataset would be a 48x48 pixel grayscale image of a face, associated with an emotion label. Fig. 1 showcases some instances from the dataset.

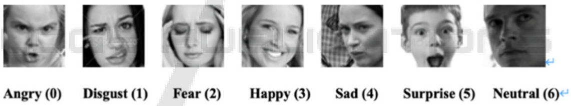

Figure 1: Images from the FER-2013 dataset (Original).

2.2 Proposed Approach

The focus of this proposed method for facial emotion detection is the CNN model. However, there is an extended procedure required before the model can be used. It is essential to first preprocess and reshape the data while importing the required dataset. After that, the generator is fitted to the data, and the images used for facial emotion detection are enhanced before being transferred to the model. In the following phase, CNN is applied to generate a model for detecting facial emotions, which is then trained. This is exactly the key objective of this research: explore the consequences of CNN on facial emotion detection by modifying the iteration period constantly. The pipeline is shown in Fig. 2.

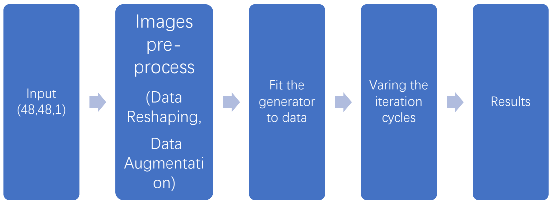
Input (48,48,1)
Images pre-process (Data Reshaping, Data Augmentation)
Fit the generator to data
Varing the iteration cycles
Results

Figure 2: The pipeline (Original).

2.2.1 CNN

CNN is a widely employed feature extractor utilized in neural networks for applications such as image recognition and classification. The input size is set to 48 by 48 grayscale images. These pictures need to be processed in a certain way to improve the data before they are supplied into the network. Improving the model's performance and generalization ability via picture data improvement is possible. In this investigation, the CNN model consists of two convolutional layers, followed by two pooling layers, also one flattening layer, and at last two fully connected layers. Fig. 3 shows the process.

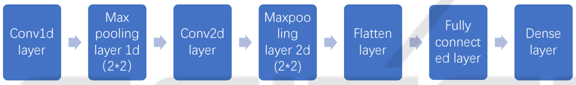

Figure 3: The structure of CNN (Original).

Given an explanation first, the Convolutional layers function as the data source used by the pooling layer, utilizing the backpropagation approach to fine-tune the parameters of each of the multiple convolutional units that collectively make up the neural network. The pooling layer, which comes after the convolutional one, likewise uses several feature planes. There is a one-to-one correspondence between feature planes throughout layers; the number of feature planes does not increase or decrease. The dense layer of neurons is an initial layer of neurons in which all the neurons in the preceding layer send input to each of the neurons in the current layer. Plus, the whole convolutional neural network relies on the "classifier" function provided by the fully connected layer. Therefore, the Conv2D layer generates a convolutional kernel that convolves with the layer input, making it simple to produce tensors. The pooling layer plays a role in secondary feature extraction, with each neuron executing pooling operations on the local receptive domain.

In the pooling layer, each neuron in the network applies its local receptive region to perform a pooling operation, making it is beneficial to secondary feature extraction. Here, the most frequently used pooling method, max pooling, can be used to take the point with the highest median in the local acceptance domain. So, the MaxPooling2D layer down samples the input along the spatial dimension. The Rectified Linear Unit (ReLu) function was also used as a neural network activation in the aforementioned procedures, as it would speed up calculations and prevent the dreaded gradient vanishing problem. Convolutional layers, pooling layers, and activation function layers assist with mapping the raw data to the hidden layer feature space; the fully connected layer then contributes to translating the learned "distributed characteristic expression" to the specimen labelling area.

2.2.2 Output Layer

The dense layer of neurons is an initial layer of neurons in which all the neurons in the preceding layer send input to each of the neurons within the existing layer. Here, the dense layer flattens the input, and it is then input to a completely connected layer. In basic terms, it is a linear transfer from one feature space to another. After non-linear changes in the dense, the dense layer's objective is to back-map the given input domain to its characteristics that have been extracted from it in the preceding layers.

The SoftMax function is applied to multi-categorization issues with a single right solution as output, which is the same as mutually exclusive output. It is developed as a probabilistic representation of multiclassification outcomes, and it is an extension of the Sigmoid function for binary classification. In typical neural network structures, it operates as the entire ultimate layer, the SoftMax layer transforms numerical values into probability distributions. This implies that the SoftMax function can be applied to generate and ascertain the probability distribution of seven distinct emotional categories.

2.2.3 Loss Function

The final model outputs probability distributions on 7 distinct emotion categories through SoftMax, along with the optimizer, loss function, and evaluation metrics specified during training. The loss function is the classification cross-entropy, and precision serves as the benchmark. In order to motivate the model to allocate a higher probability to the appropriate category while dealing with multi-class classification problems, the method of classification cross-entropy frequently acts as the loss function for estimating the discrepancy between the predicted label and the actual label. For each image, the model computes its loss and then compares the predicted likelihood of each emotion category to a hot-coded real label, as,

$$Loss = -\frac{1}{n}\sum_i x_i lny_i \qquad (1)$$

$$A = -\frac{1}{n}\sum_j [alnb + (1-a)\, ln(1-b)] \qquad (2)$$

In the formula, j stands for the number of samples, a for the actual label, b for the expected result, and n for the total number of samples.

2.3 Implementation Details

When putting into practice the suggested model, it is crucial to remain in consideration that all the photographs in the information set are grayscale facial images with the subject's face roughly centered in the middle. Therefore, this necessitates no special processing, as stated. With respect to hyperparameters, limiting suggested initial learning rate to 0.001 whilst applying Adam as an optimizer. As the model is being trained, the optimizer can adjust its settings. The gradient of the loss function can be utilized by the Adam optimizer to adjust the learning rate.

3 RESULTS AND DISCUSSION

The fundamental intention of this investigation is to investigate the influences generated by varying iteration cycles on the recognition of facial expressions in CNN. Therefore, selecting the epoch=20,40,60,80,100 to run the model and analyze the impact generated by increasing iteration cycles and the relationship between loss and accuracy.

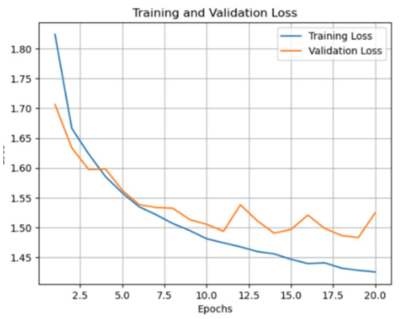

Figure 4: The loss curve in 20 epochs (Original).

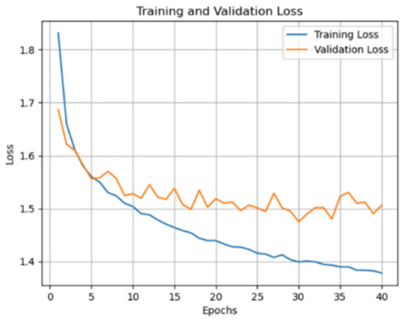

Figure 5: The loss curve in 40 epochs (Original).

After many hours of training, the CNN model was able to pass validation, as shown in the Fig. 4-8. The FER-2013 dataset is used for the experiment, and its results are analyzed to figure out the effectiveness of the network model. With regular iteration, the CNN model improved its accuracy and its loss rapidly diminished. In this case, the CNN model's accuracy exemplifies rising fluctuations that eventually normalize to a reduced loss. In furtherance of indicating that the network model is being trained normally and optimally, this also shows that its performance is steadily rising. Growing the overall amount of iteration cycles has the potential to optimize the CNN's parameters and build up convergence, a term describing the point at which the CNN's performance becomes stable and further training does not improve recognition accuracy a great deal. To accomplish optimal efficiency while eliminating computation consumption of resources, discovering a proper balance between iteration cycles is indispensable. Overall, the results of facial emotion detection are greatly enhanced when the CNN model is subjected to several iteration cycles. This performance advancement, as well as an improvement in accuracy and a reduction in losses, are all expressed in the chart. Such training tactics can be incorporated into subsequent research to further improve the accuracy of emotion recognition.

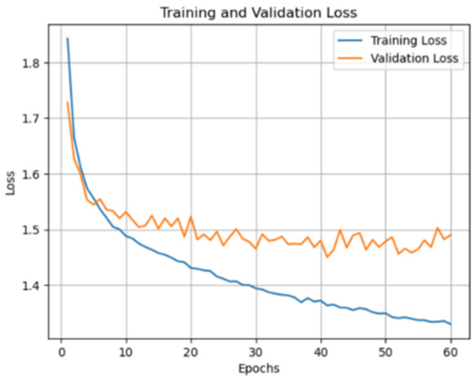

Figure 6: The loss curve in 60 epochs (Original).

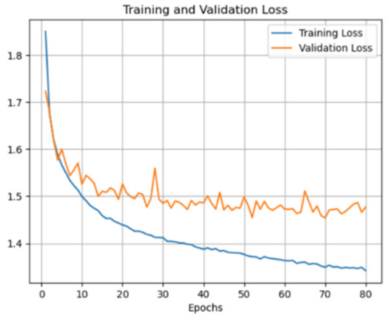

Figure 7: The loss curve in 80 epochs (Original).

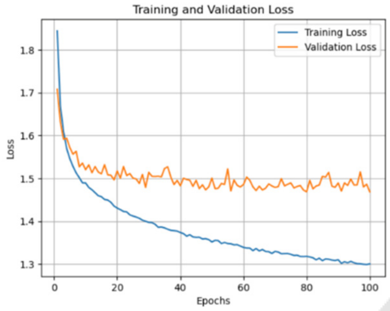

Figure 8: The loss curve in 100 epochs (Original).

4 CONCLUSION

The CNN model is an important boost within the wider context of this research. The performance of CNN facial emotion detection has been investigated through experiments carried out on the FER-2013 dataset, with a focus on studying its performance under varying iteration cycles. The efficiency in detecting emotional expressions in images can be greatly enhanced by using several iterations on object directions of various sizes and complexities, thereby increasing accuracy and impact. The training set is expected to achieve an accuracy of 0.5117. In comparison to the previous iteration with a lower number of iterations, the current result indicates a notable enhancement of 8.11%. Further, the degree of loss in the training set is consistently diminishing, suggesting that greater iterations contribute to better model performance. Variable iteration cycles may have an impact on CNN's generalization capacity, which might improve the accuracy of facial expression recognition on unobserved data. Consequently, if the CNN is trained with insufficient iteration cycles, it may not generalize effectively, which will result in inadequate recognition performance for novel face expressions. To a certain degree, fostering the iteration cycles could enhance generalization. In the meantime, it also implies that normal procedures are implemented in constructing the network model and reflect the optimal outcome. It's critically important to know that the dataset, network configuration, and other potential factors are capable of influencing exactly what consequences of various iteration cycles on CNN facial expression recognition. The subsequent research is expected to further increase even more thanks to tackling any potential constraints. Since the development potential of emotion recognition technology in the future is enormous. Furthermore, it will investigate the model's performance across various facial features and emotional states, with the objective of gaining a more thorough understanding of facial expression detection.

REFERENCES

M. Ali, M. Khatun, N. Turzo, “Facial emotion detection using neural network,” the international journal of scientific and engineering research, 2020

B. Ko, “A Brief Review of Facial Emotion Recognition Based on Visual Information,” Sensors, vol. 18, 2018, p. 401

P. Dachapally, “Facial emotion detection using convolutional neural networks and representational autoencoder units,” arXiv, 2017, unpublished

H. Liu, T. Xu, X. Wang, Y. Qian, “Related HOG features for human detection using cascaded adaboost and SVM classifiers,” In Advances in Multimedia Modeling: 19th International Conference, MMM 2013, 2013, pp. 345-355

P. Giannopoulos, I. Perikos, I. Hatzilygeroudis, “Deep learning approaches for facial emotion recognition: A case study on FER-2013,” Advances in Hybridization of Intelligent Methods: Models, Systems and Applications, 2018, pp. 1-16

H. Boughrara, M. Chtourou, C. Ben, L. Chen, “Facial expression recognition based on a mlp neural network using constructive training algorithm,” Multimedia Tools and Applications, vol. 75, 2016, pp. 709-731

G. Panchal, A. Ganatra, Y. Kosta, “Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers,” International Journal of Computer Theory and Engineering, vol. 3, 2016, pp. 332-337

C. Tang, Y. Zhao, G. Wang, “Sparse MLP for image recognition: Is self-attention really necessary?,” In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, 2022, pp. 2344-2351

K. Terusaki, V. Stigliani, “Emotion detection using deep belief networks,” unpublished

I. Goodfellow, D. Erhan, P. Carrier, “Challenges in representation learning: A report on three machine learning contests Neural Information,” Processing: 20th International Conference (ICONIP) Springer berlin Heidelberg, 2013, pp. 117-124