PAPER • OPEN ACCESS

Emotion Recognition Using Convolutional Neural Network (CNN)

To cite this article: Nur Alia Syahirah Badrulhisham and Nur Nabilah Abu Mangshor 2021 *J. Phys.: Conf. Ser.* **1962** 012040

Emotion Recognition Using Convolutional Neural Network (CNN)

Nur Alia Syahirah Badrulhisham[1] and Nur Nabilah Abu Mangshor[2]

[1,2]Fakulti Sains Komputer dan Matematik, Universiti Teknologi MARA Cawangan Melaka Kampus Jasin, 77300 Merlimau, Melaka, MALAYSIA

nurnabilah@uitm.edu.my

Abstract. Emotion is an expression that human use in expressing their feelings. It can be express through facial expression, body language and voice tone. Humans' facial expression is a major way in conveying emotion since it is the most powerful, natural and universal signal to express humans' emotion condition. However, humans' facial expression has similar patterns, and it is very confusing in recognizing the expression using naked eye. For instance, afraid and surprised is very similar to one another. Thus, this will lead to confusion in determining the facial expression. Hence, this study aims to develop a mobile based application for emotion recognition that can recognize emotion based on facial expression in real-time. The Deep Learning based technique, Convolutional Neural Network (CNN) is implemented in this study. The MobileNet algorithm is deployed to train the model for recognition. There are four types of facial expressions to be recognized which are happy, sad, surprise, and disgusting. As the result, this study obtained 85% recognition accuracy. In the future, the developed application could be improved by adding more face expression categories.

1. Introduction

Facial expression is a valuable expression that portrays human emotion [1][2][3]. Emotion expression a natural ability in each human being and people use emotion to communicate their feeling directly. It is also known as a process of non-verbal communication [1][4] Based on the emotion portrays on a person face, indirectly other people could have a cue on how to communicate with the person. In addition, human emotion recognition is a key technique in human-computer interaction [2].

Emotion is highly dependent on the persons' body condition and mental state. Hence, it is addressed as the inter-personal communication [1][2][5]. Although human can directly express their emotion via facial expression, there exists similar patterns between different facial expressions that contributes to the difficulty in recognizing the emotion correctly [6][7]. For instance, surprise and fear expression are like one another. Other than that, afraid and surprised expression also exhibits similar expression. Hence, these similarities could lead to the false recognition using the naked eyes [4][5]. Due to the challenges in recognizing the emotion traditionally, an automatic emotion recognition application highly needed.

In the literature, K-Nearest Neighbors is widely used in automatic emotion recognition [4][7][8] applied KNN and achieved accuracy more than 85%. However, the implementation of KNN requires high memory and it is slow in performance [10]. On the other hand, the Deep Learning based techinique, Convolutional Neural Network (CNN) offeres high accuracy performance and fast recognition [3][11]. CNN has been widely used in automatic emotion recognition such in [1][2][12][13][14] and obtained promising results with the accuracy obtained more than 90%.

Hence, with the great potential of the CNN, this study proposed a mobile-based emotion recognition using Convolutional Neural Network (CNN). The outline of this paper is as follows. section 2 describes on Methodology and followed by section 3 on Results and Analysis. Subsequently, section 4 and section 5 present Conclusion and Acknowledgement, respectively.

2. Methodology

This section describes the methodology proposed in this study. This study starts with inputting the real-time images, followed by the implementation of Convolutional Neural Network (CNN) for recognizing the emotion. Succeeding, the recognized emotion will be displayed. Figure 1 shows the proposed flowchart in this study. Further explanation will be elaborate in the sub section accordingly.

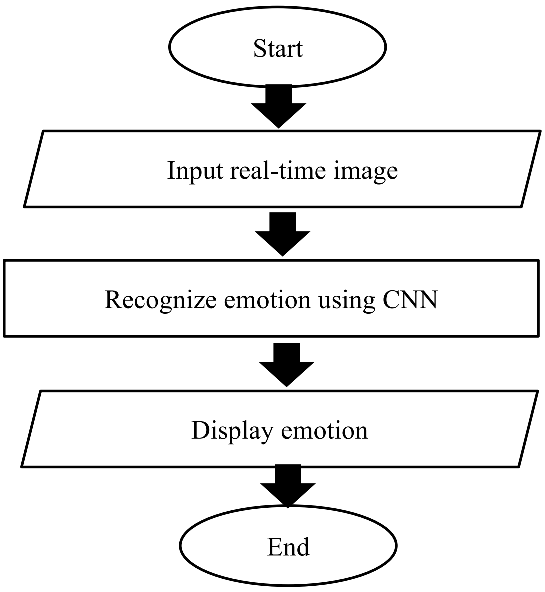

Figure 1. Proposed flow chart for this study.

2.1 Facial Expression Real-Time Images

This study covers four types of facial expression which are happy, sad, surprise and disgusting. Figure 2 shows example of these four types of facial expression.

Figure 2. Facial expression for happy, surprise, sad and disgusting.

2.2 Recognize Emotion Using Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is a Deep Learning based technology that has the capabilities to achieve high precision in recognition (Liam Schonevel 2021). CNN has multiple layers where each layer performs a specific transformation function. Convolutional is the first layer to extract features from the input image. The convolutional will then preserves the the relationship between pixels by learning image feature using small squares of input data. Convolution of an image with different filters can perform operations such as edge detection, blur and sharpen by applying filters. ReLu purpose is to introduce non-linearity in ConvNet. The real data would want our ConvNet to learn would be non-negative linear values.

Next, the pooling layer functions' to reduce the number of parameters when the image is too large. Spatial pooling also called subsampling or down sampling which reduces the dimensionally of each map but retains important information. Spatial pooling can be of different types which is max pooling, average pooling or sum pooling. Full connected layer is flattened the matrix into vector and feed it into a fully connected layer like a neural network. Figure 3 shows CNN architecture.

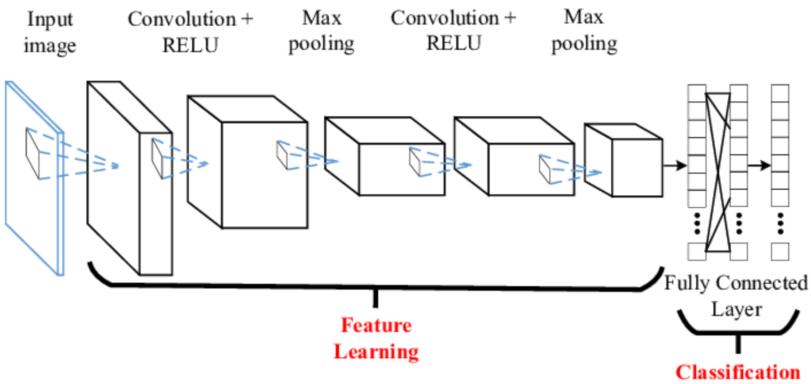

Figure 3. Convolutional Neural Network (CNN) achitecture.

2.2.1 Training the CNN model

Before starting the training process, all images will be transformed on the Roboflow platform using the argumentation process. The augmentation process included flipped horizontal or vertical, rotation, saturation, and blur process for the purpose to enhance the image.

After that, the training process executes. The best accuracy is 80% for training and it is run using 100 epochs. The value of epochs gives the effect of value accuracy. If the epochs values are below 100, the accuracy did not reach 80% and it may result in poor recognition. Table 1 shows the analysis of the accuracy of training, respectively.

Table 1. Analysis of the accuracy of training.

	Splitting images (%)	Accuracy (%)	Loss (%)	Error Rate	Error Rate (%)
Train	**90**	**94**	**53**		
Test	**5**	**80**	**94**	**9/46**	**19.56**
Validation	**5**	**84**	**66**		
Train	80	94	54		
Test	10	78	79	19/91	20.87
Validation	10	89	66		
Train	70	93	53		
Test	15	78	79	28/100	28
Validation	15	79	79		

Table 1 shows the analysis conducted during training process. This analysis is conducted to find the most ideal and optimal ratio for the splitting image into train, test, and validation images separately.

From the findings, the most ideal and optimal ratio for splitting images is 90% for train images, and 5% each for test and validation images in which achieved the lowest error rate of 19.56%.

3. Result and Analysis

The testing process executes to ensure the objective of this application is achieved. The accuracy is determined on emotion based on facial expression image that has been entered into an application through the mobile phone camera. The images used for testing the accuracy of application are 80 images.

A total number of 80 images have been used during the accuracy testing conducted. From Table 2, there exist a FALSE result to indicates the application is wrongly recognize the emotion. The reason of obtain false result is because the application failed to recognize the correct emotion based on the image of facial expression. For example, the expected result should be sad, but the application recognizes it as happy. Hence, the similarities of the facial expression cause the application a false result. To conclude overall accuracy performance, the average accuracy is calculated. Equation 1 shows the formula for accuracy calculation.

$$Accuracy = \frac{Number\ of\ correct\ prediction}{Total\ number\ of\ all\ cases} * 100\% \quad (1)$$

From the equation, prediction that predict by the application, while the total number of all cases of prediction is the number of images has been tested. The quotient is multiple with 100% to get the percentage of accuracy. The overall accuracy result for this application is 85%.

3.1 Confusion Matrix

The accuracy testing of the application is conducted using 80 testing images which composed of 20 images from each emotion category. Table 2 tabulates the confusion matrix result from the testing conducted.

Table 2. Confusion matrix result.

		Actual				
		Happy	Sad	Surprise	Disgusting	Total
Predicted	Happy	**18**	0	1	1	20
	Sad	1	**16**	0	3	20
	Surprise	1	0	**17**	2	20
	Disgusting	2	0	1	*17*	20
	Total	22	16	19	23	**80**

From Table 2, only 18 testing images of happy facial expression were correctly recognized as happy emotion and the remaining 1 image each is misrecognized as surprise and disgusting emotion, respectively. Next, for sad emotion only 16 images are correctly recognized as sad emotion while 3 images were misrecognized as disgusting, and 1 image was misrecognized as happy emotion. However, for the surprise emotion, 17 images were correctly recognized as surprise emotion while 1 image was misrecognized as happy, and 2 images were misrecognized disgusting emotion. Lastly, for the disgusting emotion, 17 images were correctly recognized as disgusting emotion while 2 images were misrecognized as happy, and 1 image was misrecognized as surprise emotion Table 3 tabulates summarization result of confusion matrix.

Table 3. Summarization of confusion matrix result.

	Happy	Sad	Surprise	Disgusting
True Positive (TP)	18	16	17	17
True Negative (TN)	56	60	58	54
False Positive (FP)	4	0	2	6
False Negative (FN)	2	4	3	3

Therefore, the sensitivity and specificity calculation can be performed based on the figure obtained in Table 3 earlier. In general, the sensitivity measures the correctly positive result by testing against all the images in the application and specificity is to measure all correctly negative result by testing against all the images in the application [15]. Equation 2 and Equation 3 shows formula to calculate the sensitivity and specificity accordingly. Next, Table 4 shows the results for accuracy, sensitivity and specificity calculation for each class of emotion.

$$Sensitivity = \frac{TP}{TP\ \ FN} \tag{2}$$

$$Specificity\ = \frac{TP}{TP\ \ FN} \tag{3}$$

Table 4. Accuracy, sensitivity, and specificity result.

Type of Emotion	Accuracy(%)	Sensitivity(%)	Specificity(%)
Happy	92.50	90	93.33
Sad	95.00	80	100.00
Surprise	93.75	85	96.66
Disgusting	88.75	85	90.00
Average	**92.50**	**85.00**	**95.00**

From Table 4, the average accuracy, average sensitivity, and average specificity of the application based on the confusion matrix are calculated. The average accuracy obtained high value at 92.5% in recognizing the emotion. However, the value of average sensitivity is a bit low at 85.00% implying that the application generates some false-negative value while the average specificity comes out good at 95.00% implying that the application produces a low false-positive value.

4. Conclusion

As a conclusion, an emotion recognition application using the Convolutional Neural Network (CNN) was successfully developed. The application able to recognize four types of emotions which are happy, sad, surprising, and disgusting. The Convolutional Neural Network (CNN) used the MobileNet algorithm with a custom dataset and evaluated using a confusion expression is a valuable expression that portrays human. The developed application achieved an average accuracy of 92.50%. in term of the sensitivity and specificity, it able to achieve 85.00% and 95.00% respectively. Hence, the implementation of CNN in recognizing emotion successfully achieved promising results and could be able to contributes to the succession work in CNN. In future, the integration of CNN with any other Artificial Intelligence (AI) method is believed to improve the performance of the application.

References

All authors would like to express high gratitude to Universiti Teknologi MARA Cawangan Melaka for funding this research.

References

[1] Liam Schoneveld, Alice Othmani, Hazem Abdelkawy. (2021). Laveraging recent advances in deep learning for audio-Visual emotion recognition. *Pattern Recognition Letters 146*, 1-7.

[2] Yishu Liu, Guifang Fu. (2021). Emotion recognition by deeply learned multi-channel textual and EEG features. *Futures Generation Computer Systems 119,* 1-6.

[3] Neha Jain, Shishir Kumar, Amit Kumar, Pourya Shamsolmoali, Masoumeh Zareapoor. (2018). Hybrid deep neural networks for face emotion recognition. *Pattern Recognition Letters 115*, 101-106.

[4] Paweł Tarnowski, Marcin Kołodziej, Andrzej Majkowski, Remigiusz J. Rak. Emotion recognition using facial expressions. *International Conference on Computational Science* 2017; (): . (accessed).

[5] Shervin Minaee, Amirali Abdolrashidi. Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network. 2019; (): . (accessed).

[6] Nithya Roopa. S. Emotion Recognition from Facial Expression using Deep Learning. *International Journal of Engineering and Advanced Technology (IJEAT)* 2019; 8(6S): .

[7] Arkaprabha Bhandari, Nikhil R Pal. (2021). Can edges help convolutional neural networks in emotion recognition?. *Neurocomputing* 433, 162-168.

[8] Poorna S S, Anjana S, Parvathy Varma, Anjana Sajeev, Arya K Ca , Renjith Sb , G J Nair a. Facial Emotion Recognition using DWT based Similarity and Difference features. *Proceedings of the Second International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)* 2018; (): .

[9] Firoz Mahmud, Bayezid Islam, Arfat Hossain, Pushpen Bikash Goala. Facial Region Segmentation Based Emotion Recognition Using K-Nearest Neighbors . *Proceeding of the International Conference on Innovation in Engineering and Technology (ICIET)* 2018; ():

[10] Marina Chatterjee. *A Quick Introduction to KNN Algorithm. https://www.mygreatlearning.com/blog/knn-algorithm-introduction/ (accessed 10 May 2020).*

[11] Wafa Mellouk, Wahida Handouzi. Facial emotion recognition using deep learning: review and insights. *Procedia Computer Science* 2020; 175(): .

[12] Ratnalata Gupta, Vishwamitra LK. (2020). Facial expression recognition from videos using CNN and feature aggregation. *Materials Today: Proceedings.*

[13] Jagli, P. M., & Shetty, M. P. (2019). Human Emotion Recognition using Machine Learning. *International Journal of Trend in Scientific Research and Development*, 101-103.

[14] Rassadin, & Savchenko. (2017). Compressing deep convolutional neural networks in visual emotion recognition. *National Research University Higher School of Economics, Laboratory of Algorithms and Technologies for Network Analysis*, 207-213.

[15] Stephanie Glen. *Sensitivity vs Specificity and Predictive Value. https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/sensitivity-vs-specificity-statistics/ (accessed 15 June 2020).*