International Journal of Engineering & Technology, 11 (1) (2022) 59-62

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Implementation of a convolutional neural network method to predict university students' emotional extent in online learning

Baharuddin [1], S Gultom [2], M Restuati [3], A Mansyur [2], H Fibriasari [4]*, I I Pane [5], S Pujiastuti [6], H D Hutahaean[7], B D Waluyo [1], V S G Galamgam [8], R P Hacla [8], M A Rais [7], R Effendi [7], M R Pane[7]

[1] *Departement of Electrical Engineering Education, Faculty of Engineering, Universitas Negeri medan, Medan, 20221, Indonesia*
[2] *Departement of Mathematics, Faculty of Mathematical and Natural Science, Universitas Negeri Medan, Medan, 20211, Indonesia*
[3] *Departement of Biology, Faculty of Mathematical and Natural Science, Universitas Negeri Medan, Medan, 20221, Indonesia*
[4] *Departement of French Language Education, Faculty of Languages and Literature, Universitas Negeri Medan, Medan, 20221, Indonesia*
[5] *Departement of English Language Education, Faculty of Languages and Literature, Universitas Negeri Medan, Medan, 20221, Indonesia*
[6] *Departement of Germany Language Education, Faculty of Languages and Literature, Universitas Negeri Medan, Medan, 20221, Indo-nesia*
[7] *Computer Science and Informatics Education Study Program, Faculty of Engineering, Universitas Negeri Medan, Medan, 20221, Indonesia*
[8] *Computer Engineering Department, College of Engineering and Architecture, Pangasinan State University Urdaneta City, Pangasinan, Philippines*
**Corresponding author E-mail: hesti@unimed.ac.id*

Abstract

A learning system that was directly face-to-face in class, had to be substituted with a learning system that is virtually integrated with the inter-net (online learning). The system uses online learning that connects teaching staff with the learning participants. Tiredness and lazi-ness start-ed to be felt among the participants. This is worried to affect nonoptimal learning outcomes. During lessons, teaching staff often focuses on learning materials without knowing the emotions of their learning participants. Indirect online learning becomes a new challenge in adaptation; teaching staff and learning participants are required to adapt to existing conditions. With the problems above, the writer has an idea in devel-oping a prototype system that is developed to be able to help teachers in evaluating manners and emotions of students in online learning. Facial detection is a technology for detecting facial features such as the nose, eyes, and mouth. Development of facial detection that has un-dergone upgrades until now can differentiate if something is a facial feature or not and detect more than one face. This research aims to help teaching staff in evaluating learning participants' affective behavior online by applying a Convolutional Neural Network method that has been trained to be able to predict emotional extents through facial expression.

Keywords: *Emotional Extent; Convolutional Neural Network; Learning; Face Detection.*

1. Introduction

The effects of the spread of COVID-19 on education obliges teachers and learners to be able to adapt with existing changes. A learning system that was directly face-to-face in class, had to be substituted with a learning system that is virtually integrated with the internet (online learning) which is learning that is carried out by online applications that connect teaching staff with learning participants; video conference. Video conference applications such as Zoom, Google Meet, Webex and others, can directly connect teaching staff with learning participants without direct contact. This learning method has been massively done since the Ministry of Education and Culture of Republic of Indonesia issued Circular Letter Number 4 Year 2020 (Surat Edaran Nomor 4 Tahun 2020) about the Execution of Education Policies in COVID Spreading Emergency Period (Pelaksanaan Kebijakan Pendidikan dalam Masa Darurat Penyebaran COVID) on March 24, 2020, where it said that learning processes are to be carried out at home using long-distance learning to give meaningful learning experiences for students[8,11].

For almost a year, the learning participants undergo long-distance learning or online learning that was caused by the COVID-19 pandemic. Tiredness and laziness started to be felt among the participants. This is worried to affect nonoptimal learning outcomes. Uninterest in learning is one of the problems that is often felt by learning participants that causes learning motivation to go down, the emergence of

laziness, and the drop of participants' achievements. This means the emergence of laziness happens since learning participants experience tiredness in learning.
Because of tiredness, student motivation in learning goes down. This happens continuously in online learning. During lessons, teaching staff often focuses on learning materials without knowing the emotions of their learning participants. This becomes a problem in learning where the teacher doesn't pay attention to the development of his or her students. The neglect of manner evaluation and the lack of teaching staffs' attention towards learning participants gives a bad effect on the development of learning participants. Indirect online learning becomes a new challenge in adaptation; teaching staff and learning participants are required to adapt to existing conditions. With the problems above, the writer has an idea in developing a prototype system that is developed to be able to help teachers in evaluating manners and emotions of students in online learning. [1], [2].
Happiness, anger, sadness, confusion, are a few means of facial expression. These expressions can usually be seen in the looks of someone's face. When someone is happy, he/she will tend to smile, and on the other hand, when some are sad he/she will show a frowny face. [3], [4], [6], [8]. On and on, these expressions will produce different looks on the face. These facial expressions can be read through a few changes of the face such as the lips and forehead. Facial detection is a technology for detecting facial features such as the nose, eyes, and mouth. Development of facial detection that has undergone upgrades until now can differentiate if something is a facial feature or not and detect more than one face. [7].
Facial emotion detection is a process of classifying faces according to the emotions shown on them. This classification is started with recognition of the face on the picture, then classification is done to show emotion prediction on said picture. With facial detection, classification will sharpen the process. This classification of course can be based on real-time video since videos are a collection of frames or pictures.
A trained Convolutional Neural Network Method can predict emotions through facial expression. In the component process model structure, emotion is defined as episodes of change that are mutually related and synchronized to the state of all or most of five subsystems of organisms in responding to external or internal stimulus relevant to the main matters of organisms. [5] Different stimuli will show different emotions. This research aims to help teaching staff in evaluating learning participants' affective behavior online.

2. Research method

2.1. Facial recognition

Facial recognition is one of the applications of image detection [1], [2]. Object recognition done by machines is more complex than object recognition done by humans, even though fundamentally both have the same goal, which is detecting an object that was seen. To be able to detect facial features (objects), the computer needs a few steps to go through for the object to be detected accurately. The steps can be seen in Image 1. In the process of detecting emotions, a dataset is needed as a systematic guide for classifying expressions [3], [4].

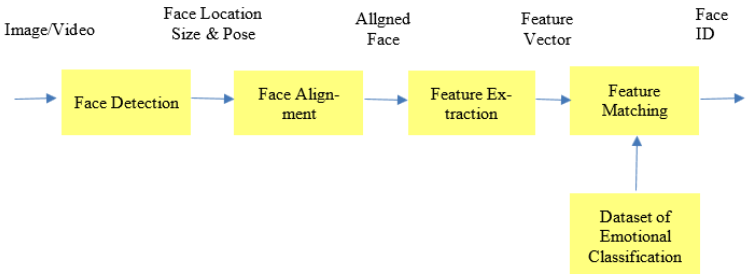

Fig. 1: Facial and Emotional Identification.

2.2. Convolutional neural network (CNN)

At this step of the research, the Convolutional Neural Network was used without the addition of other features such as extraction or other classification methods. The CNN that was used was a pre-trained model which has been trained before to show the working process in accuracy and time used for processing. The researchers used AlexNet that has been trained by the database ImageNet. AlexNet is one of the Convolutional Neural Networks that has an 8 layer depth.[5].
The use of this architecture is very crucial because AlexNet is one of the Convolutional Neural Networks that has a simple pattern in its use. AlexNet has been trained to use the ImageNet dataset that has more than 1 million images and can classify pictures into more than 1000 object categories. Because of that, this network has learned many feature representations for various pictures. The use of this pre-trained network will help in the training process where a small dataset is used.[6].
By using input that consists of labeled pictures, with the probability of each object category, with picture input network size being 227x227, the researchers used an open database which is dataset FER-2013 that is provided by Kaggle, with each picture sized 48x48 pixels and in grayscale. The following Image 2 shows the concept of pre-trained CNN usage in this research.

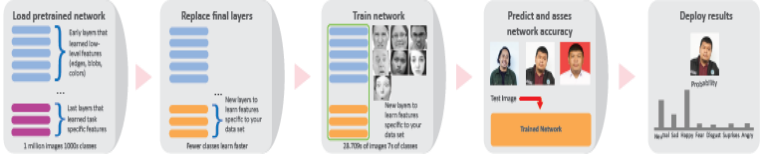

Fig. 2: Usage of Pre-Trained CNN to Support New Training.

Image 2 is a summary of the AlexNet architecture used in this research. AlexNet is used as a foundation of an architecture that consists of 5 convolutional layers, 3 fully connected layers, and other supporting layers with a total of 25 layers, although it's structured with 8 inner layers. The convolutional layer in Image 1 has performance that is supported by 4 separate poll layers. The kernel size of convolutional layer 1 to 5 stays at a consistent 3x3 size and is followed by 3 fully connected layers (fc6, fc7, fc8) Further information can be seen in

Table 1 that consists of a summary of the AlexNet that was used on this research. This architecture has been trained by using MATLAB that is supported by simple devices like computers.

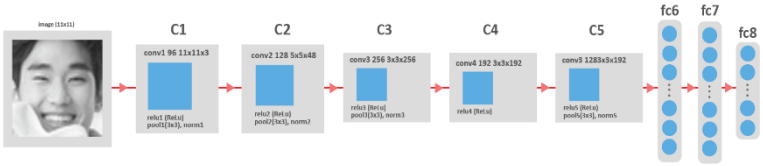

Fig. 3: Alexnet Layers with 8 Inner Layers.

Table 1: Summary of CNN Architecture Using Alexnet

No	Deep Layer	Type Layer	Layer	Info
1		data	Image Input	227x227x3 images with 'zerocenter' normalization
2		conv1	Convolutional	96 11x11x3 convolutions with stride [4 4] and padding [0 0 0 0]
3	C1	relu1	ReLu	ReLU
4		norm1	Cross Channel Normalization	cross channel normalization with 5 channels per element
5		pool1	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0 0 0]
6		conv2	Convolutional	2 groups of 128 5x5x48 convolutions with stride [1 1] and padding [2 2 2 2]
7	C2	relu2	ReLu	ReLU
8		norm2	Cross Channel Normalization	cross channel normalization with 5 channels per element
9		pool2	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0 0 0]
10	C3	conv3	Convolutional	384 3x3x256 convolutions with stride [1 1] and padding [1 1 1 1]
11		relu3	ReLu	ReLU
12	C4	conv4	Convolutional	2 groups of 192 3x3x192 convolutions with stride [1 1] and padding [1 1 1 1]
13		relu4	ReLu	ReLu
14		conv5	Convolutional	2 groups of 128 3x3x192 convolutions with stride [1 1] and padding [1 1 1 1]
15	C5	relu5	ReLu	ReLu
16		pool5	Max Polling	3x3 max pooling with stride [2 2] and padding [0 0 0 0]
17		fc6	Fully Connected	4096 fully connected layer
18	F6	relu6	ReLu	ReLu
19		drop6	Dropout	50% dropout
20		fc7	Fully Connected	4096 fully connected layer
21	F7	relu7	ReLu	ReLu
22		drop7	Dropout	50% dropout
23	F8	fc8	Fully Connected	1000 fully connected layer
24		prob	Softmax	Softmax
25		output	Classification Output	crossentropyex with 'tench' and 999 other classes

3. Results and discussion

In this part, the result of the method that the researchers had applied, first will show the explanation of the dataset that was used, and the result of training until testing with the pre-trained model. This research used the dataset FER-2013 that was provided by Kaggle [https://www.kaggle.com/msambare/fer2013]. The pictures that were used were arranged photos of facial expressions from various groups of people. The pictures in FER-2013 had been segmented with the same size and made grayscale. This research uses the dataset below in helping to classify categories of objects that are going to be read, which consists of categories of facial expression such as anger, disgust, fear, happiness, neutral, sadness, and surprise with a total of 3.589 examples. In the image below a few examples of the dataset that are going to be used are shown [4].

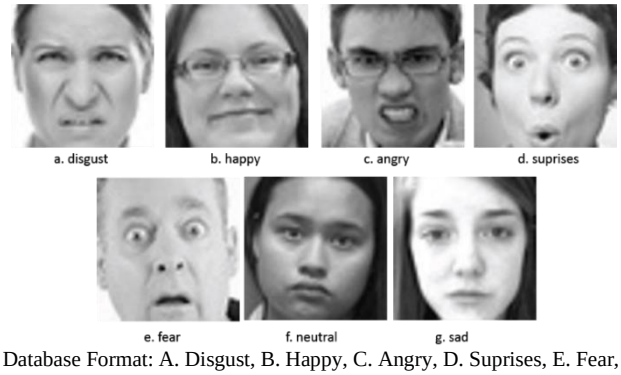

Fig. 4: The Image Database Format: A. Disgust, B. Happy, C. Angry, D. Suprises, E. Fear, F. Neutral, G. Sad.

The training process is done in MATLAB with the details that were shown in Image 3, with 20 epochs. This process took approximately 174 minutes with a single GTX 1050 2GB GPU. The validation accuracy was 62.96% of the training done. With 40180 iterations that consist of 2009 every epoch, each epoch forward and backward passes correspond to its total iterations. After the training process is done, the researchers specify and arrange folders according to the expected category. Every folder consists of a certain category, such as the "happy" folder will consist of all pictures with a happy expression. After every category that consists of each folder is done being trained, a new database from the processing results that has been done will be made, and then this database can be called in a testing process.

To test the performance and function of this device, we used a picture of someone. This picture is photographed using a digital camera and will be put into the CNN. In this experiment, the photo is taken spontaneously in normal conditions. At this stage, the test is done and a result akin to image 3 was got. Looking at the tested examples, the test was successfully done to detect a facial expression in the picture.

In Image 5, a square label on the face, the label “happy” and a “happy” presentation description can be seen. To see dataset testing look at the following Image 5:

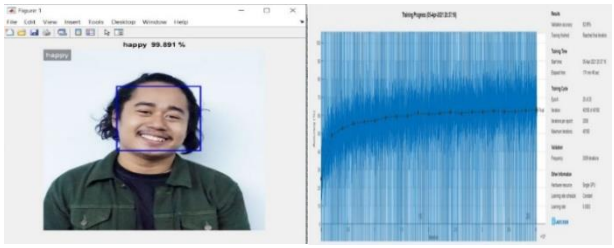

Fig. 5: Test Results by Using Alexnet Model.

After the method testing is successfully done, the next step is online system development as seen on Image 6, the emotional detection development scheme. The summary of that image is a few processes and subprocesses were developed to increase accuracy from the detection that was applied on the output after CNN. Development became an application platform that can be applied to online learning particularly through video conference, using a live recording as an input on directly analyzing learning participants’ emotions without losing focus on teaching.

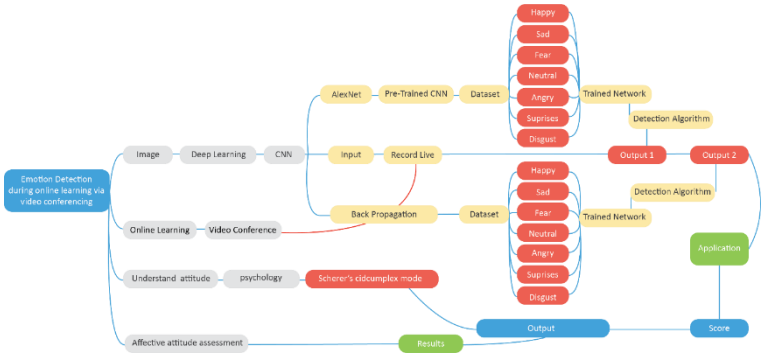

Fig. 6: Emotional Detection Development Scheme on Online Learning.

4. Conclusion

The use of deep learning that can help humans in facilitating matters has been widely used, one example is object identification. Development can vary on different needs. The focus of this research was the use of deep learning in analyzing learning participants’ emotions while online learning, when the teacher is focused on giving out learning material, oftentimes manner evaluation that is supposed to be done in class isn’t carried out. Development of facial detection that has undergone upgrades until now can differentiate if something is a facial feature or not and detect more than one face. Applying a Convotional Neural Network Method that has been trained with processed pictures or videos can predict emotional extents through facial expression which result will be able to help teaching staff in evaluating learning participants online.

References

[1] Goldman A I and Sripada C S 2005 Simulationist models of face-based emotion recognition *Cognition* 94 193–213 https://doi.org/10.1016/j.cognition.2004.01.005.
[2] Mano L Y et al. 2019 Using emotion recognition to assess simulation-based learning *Nurse Educ. Pract.* 36 13–19 https://doi.org/10.1016/j.nepr.2019.02.017.
[3] Wang W, Xu K, Niu H and Miao X 2020 Emotion Recognition of Students Based on Facial Expressions in Online Education Based on the Perspective of Computer Simulation *Complexity* 2020 https://doi.org/10.1155/2020/4065207.
[4] Engel D, Woolley A W, Jing L X, Chabris C F and Malone T W 2014 Reading the mind in the eyes or reading between the lines? Theory of mind predicts collective intelligence equally well online and face-to-face *PLoS One* 9 https://doi.org/10.1371/journal.pone.0115212.
[5] Scherer K R 2005 What are emotions? and how can they be measured? *Soc. Sci. Inf.* 44 695–729 https://doi.org/10.1177/0539018405058216.
[6] N. Mehendale, “Facial emotion recognition using convolutional neural networks (FERC),” *SN Appl. Sci.*, vol. 2, no. 3, pp. 1–8, 2020, https://doi: 10.1007/s42452-020-2234-1.
[7] D. Y. Liliana, “Emotion recognition from facial expression using deep convolutional neural network,” *J. Phys. Conf. Ser.*, vol. 1193, no. 1, 2019, https://doi: 10.1088/1742-6596/1193/1/012004.
[8] H. D. Hutahaean, S. Muhammad Aulia Rahman, and M. D. Mendoza, “Development of interactive learning media in computer network using augmented reality technology,” *J. Phys. Conf. Ser.*, vol. 2193, no. 1, p. 012072, 2022, https://doi: 10.1088/1742-6596/2193/1/012072.
[9] S. Razia, P. Swathi Prathyusha, N. Vamsi Krishna, and N. Sathya Sumana, “A Comparative study of machine learning algorithms on thyroid disease prediction,” *Int. J. Eng. Technol.*, vol. 7, no. 2.8, pp. 315–319, 2018, https://doi: 10.14419/ijet.v7i2.8.10432.
[10] S. Q. Ong, H. Ahmad, G. Nair, P. Isawasan, and A. H. A. Majid, “Implementation of a deep learning model for automated classification of Aedes aegypti (Linnaeus) and Aedes albopictus (Skuse) in real time,” *Sci. Rep.*, vol. 11, no. 1, pp. 1–12, 2021, https://doi: 10.1038/s41598-021-89365-3.
[11] H. Fibriasari, N. Harianja, D. Ampera, Z. Ramadhan, and Baharuddin, “E-Learning Learning Model to Improve the Quality of Student’s French Master,” *Rev. Int. Geogr. Educ. Online*, vol. 11, no. 5, pp. 3053–3063, 2021, https://doi: 10.48047/rigeo.11.05.197.