**Facial Emotion Recognition based on Deep Learning Convolutional Neural Network**

Tran Anh Dung

Department of Computer Science, UIT

**Abstract**

The purpose of this research is to introduce a method based on emotion recognition which can be detected using facial expression, body language, and so on. This study field proposes a learning emotion identification model, which consists of three stages: Feature extraction, subset feature, and emotion classifier. A Haarcascades method is used to detect input face images which disseminated face, eyes, mouth, and other characteristics. Through Neural Network classifier training classify seven types of emotion to obtain the results based on seven different kinds of emotional categories which are trained properly by Convolutional Neural Network. Experimental facial emotion recognition using Facial Emotion Recognition available in Kaggle. This report demonstrates the effectiveness of facial emotion recognition is feasible.

In this research, utilizing deep learning convolutional network involves feature learning which seems promising where defining features is not trivial. The Deep Learning CNN architecture providing the most optimal accuracy for testing and curbing the loss between actual and predicted value. Testing processes comprise both images and videos where collect randomly in the testing dataset or get from Internet, the videos can be testing in the actual time or available video. Although the neural network model is not state-of-art but accomplishes auspicious results when tested on a different dataset. Therefore, the deep learning convolutional neural network has the potential to augment the ability interaction between human and machine.

**1. Introduction.**

In the past decades, much research has focused on developing technology to enforce the capability of technology especially robots, AI fields which play a crucial role in every aspect of life. In the process of face-to-face human interaction, the emotion revealing through face, body language to express feeling as well as the deeper in words. In the virtual learning environments, the computers learned experiences by emotional images, and videos as well.

The emotional state of humans can obtain from verbal and non-verbal information captured by the various sensors, especially facial changes, tone of voice, and physiological signals. Ongoing conversation, the face changes constantly expressing by the gesture of mouth, eyes, or even ears to transmit the emotional state

Extracting features from one face to another is a tough and sensitive task in order to have a better classification. In this study, a haarcascades method used as a feature extraction model which can detect correctly face in several images, or videos, which allow the automatic extraction of features with high accuracy, this architecture prompted to start using this technique to recognize human emotions.

The model using in the article is deep CNN which is one of the models to recognize and classify images appropriately solving facial emotion recognition effectively. The model contains two main features: Feature learning and classification.

The purpose of this research is to find the best solution for emotion recognition based on facial recognition when training the deep learning convolutional neural network model and apply this to optimize as could as be possible for both accuracy and efficiency of the systems.

**2. Literature Review and Methodology**

The goal of facial emotion recognition is to gather data and analyze the feelings of subjects make appropriate responses related to input images putting on. On the basis, emotion divided into seven kinds of categories: angry, disgust, fear, happy, sad, surprise, and neutral. In each category has a bunch of images used to train the CNN model, which enforces the computer can recognize emotional feelings based on the face.

The facial recognition process consists of three main stages: data augmentation, feature extraction, and emotion classification. This research will come up alongside the three stages.

**2.1. Data Augmentation**Data localization using for training and testing deep learning model in facial emotion recognition need plenty of data to identify. The more data the more accurate the model gets. The number of images required for training depends on the type of data a neural network will evaluate, which correctly assesses categories on each image putting on. Deploying this method generally has 6 types typical as Flip, Rotation, Scale, Crop, Translation, and Gaussian Noise.

Pre-processing data has a crucial role to define before training the network in order to have a better emotion classification. Data augmentation down sampling with 48x48 pixels and intensity normalization are the steps that were applied before training CNN, which consists of three convolution-pooling layers ending with three fully connected with 256, 512, and 7 neurons. The best weight results gained at the training stages are used at the test stage.  
In each image, converting RGB image into grey image convenience to process  
**2.2. Feature Extraction**Feature extraction methods depend on the classification method and the application environment. In this process, extracting face component features such as eyes, nose, mouth, and so on from the face input image. It plays a vital initialization of processing techniques for facial expression recognition. Haarcascades model is the best technique for solving the classification task including this task.  
The haar features used as convolutional kernel, each feature is a single value obtained by subtracting sum of pixels under white rectangle from sum of pixels under black rectangle.  
A picture containing square

Description automatically generatedA picture containing icon

Description automatically generated  
The characteristic value detection using Haar Cascades method to identify whether a face exists in the images, which locate face if it is exists, the eyes and mouth regions need to be cropped. Otherwise, if the face is not existing then return to the start and input the image frames. We train the feature extraction using the neural network method.  
**2.3. Emotion classification**The deep learning CNNs model using consists of three convolutional layers and end up with three fully connected layers. In the end, the SoftMax activation used to find appropriate argument numbers and image classification. The other activation is ‘ReLU’ contributed as the most important part, which is converted to non-linear in ConvNet, the value we able to use is non-linear positive values.  
A picture containing text, shoji, crossword puzzle

Description automatically generated  
First, all the images are all defined at 48x48 pixels and make a variation in size and number of filters also the type of optimizer chosen (Adam) on a simple CNN, which contain two successive convolution layers, the second layer play the role the max pooling, then the SoftMax function for classification. The model train constantly until reaches the best weight that used to recognize emotion facial-based. Classifying the basic emotions such as: angry, disgust, fear, happy, sad, surprise, and neutral.

**3. Result and Discussion**This section tests the implementation of the proposed emotion recognition system based on facial recognition. This proposed environment on Google Colab platform using python 3 with state-of-art OpenCV 2.0 library. It is mainly used to detect the object via both images and video.After training continuously, the deep learning CNNs model reached 83% that is quite optimal for correct classification. Testing stage by several images toke by hand, fit the resolution as possible to ensure neither corrupts nor unexpected errors.  
For testing images, this study intakes images captured by mobile phone, and testing in the actual time.

|  |  |  |
| --- | --- | --- |
| Results | | |
| Sample Images | Proposed Solution (HaarCascades) Facial | |
| Detected Emotion | Category |
|  |  | Neutral |
|  |  | Happy |
|  |  | Happy |
|  |  | Sad |
|  |  | Fear |
|  |  | Happy |

The table above shows that the results from the proposed solution. This test used seven expressions (angry, disgust, fear, happy, sad, surprise, and neutral). There are a total 6 images put to test and seemingly the result is quite optimistic.

**4.Conclusion**

In conclusion, this report provided a specific visualization by using a Deep Convolutional Neural Network, which contributes a step to enhance the capability of recognition of machine proposed by different researchers, and presented some different image collected randomly from the real world, in order to have and accomplish an accurate detection of human emotions and takes a step that is what highlight that machines today will be more capable of interpreting emotions, which implies that the interaction human-machine becomes and more natural.

Using HaarCascade to detect face and identify all kinds of emotion through the neural network method.