



FACIAL EMOTION RECOGNITION BY USING CONVOLUTION NEURAL NETWORKS

CH. SAI VENKATA GANESH¹, D. SNIGDHA¹, J. TRINADH¹, G. GANESH¹, MR. CH. SAMSONU²

¹UG students, ²Assistant Professor, ^{1,2}Department of Computer Science and Engineering

^{1,2} Kallam Haranadhareddy Institute of Technology, Chowdavaram, Guntur, Andhra Pradesh, India

ABSTRACT

Automatic emotion recognition based on facial expression is an interesting research field, which has presented and applied in several areas such as safety, health and in human machine interfaces, as a human, we can recognise emotions easily. There are several emotions that can be represented by us such as happiness, surprise, anger, fear, sadness, disgust, contempt. Emotion recognition with a machine is difficult job as it is unable to differentiate the human emotions. So, to overcome this we use machine learning and deep learning algorithms. Hence, we proposed a method to classify the human emotions, we underline on these contributions treated, the architecture and the databases used and we present the progress made by comparing the proposed methods and the results obtained.

Keywords: Facial Emotion Recognition, Deep learning, automatic recognition, Database, TensorFlow, Convolution Neural Networks.

1. INTRODUCTION

Affective computing deals with machines and emotions. Emotion recognition is very essential to develop effective Human Computer Interaction.

Human emotions are recognized by various non-verbal cues like facial expressions, gestures, body posture or speech. Among them facial expressions are easy to obtain.

Facial expressions can be used to obtain 7 categories of expressions like neutral, happiness, surprise, disgust, fear, anger, and surprise.

The general procedure of determining facial expressions has three important steps. An image is given as input.

The first step is the detection of face in the image in which important features are extracted and then face is identified.

The second step is to extract the expression features from the image. Then extracted features are given to the classifier to identify the expressions as output.

Facial Emotion Recognition (FER) is the technology that analyses facial expressions from both static images and videos in order to reveal information on one's emotional state.

Facial Emotion Recognition (FER) is a flourishing study topic in which many breakthroughs are being made in industries, such as automatic translation systems and machine-to-human contact.

Facial emotional expression is a part of face recognition, it has always been an easy task for humans, but achieving the same with a computer algorithm is challenging.



The aim of facial emotion recognition is to help identify the state of human emotion (e.g.; neutral, happy, sad, surprise, fear, anger, disgust, contempt) based on facial images.

2. LITERATURE REVIEW

For the development of a system that can recognize emotions through facial expressions, previous research on the way humans reveals emotions as well as the theory of automatic image categorization is reviewed. In the first part of this section, the role of interpreting facial expressions in emotion recognition will be discussed. Emotion Recognition System involves the process of acquiring the images, processing the images, detection of faces then extracting the expression features. The System consists of three main steps. First step is to identify the face region from the acquired image and then preprocessed to minimize the environmental and other variations in the image. The next step is to extract expression features which are then classified in the third step. The classifier provides the output of the expression which is recognized.

2.1 HUMAN EMOTIONS

A key feature in human interaction is the universality of facial expressions and body language. Already in the nineteenth century, Charles Darwin published upon globally shared facial expressions that play an important role in non-verbal communication. In 1971, Ekman & Friesen declared that facial behaviors are universally associated with emotions. Apparently, humans, but also animals, develop similar muscular movements belonging to a certain mental state, despite their place of birth, race, education, etcetera. Hence, if properly modelled, this universality can be a very convenient feature in human machine interaction: a well-trained system can understand emotions, independent of who the subject is.

One should keep in mind that facial expressions are not necessarily directly translatable into emotions, nor vice versa. Facial expression is additionally a function of e.g., mental state, while emotions are also expressed via body language and voice. More elaborate emotion recognition systems should therefore also include these latter two contributions. However, this is out of the scope of this research and will remain a recommendation for future work. Readers interested in research on emotion classification via speech recognition are referred to Nicholson et al. As a final point of attention, emotions should not be confused with mood, since mood is a long-term mental state. Accordingly, mood recognition often involves longstanding analysis of someone's behavior and expressions, and will therefore be omitted in this work.

2.2 IMAGE CLASSIFICATION TECHNIQUES

The growth of available computational power on consumer computers in the beginning of the twenty-first century gave a boost to the development of algorithms used for interpreting pictures. In the field of image classification, two starting points can be distinguished. On the one hand pre-programmed feature extractors can be used to analytically break down several elements in the picture in order to categorize the object shown. Directly opposed to this approach, self-learning neural networks provide a form of 'Blackbox' identification technique. In the latter concept, the system itself develops rules for object classification by training upon labelled sample data. An extensive overview of analytical feature extractors and neural network approaches for facial expression recognition is given by Fazel and Luetin. It can be concluded that by the time of writing, at the beginning of the twenty-first century, both approaches work approximately equally well. However, given the current availability of training data and computational power it is the expectation that the performance of neural network-based models can be significantly improved by now. Some recent achievements will be listed below.



(i) A breakthrough publication on automatic image classification in general is given by Kievsky and Hinton. This work shows a deep neural network that resembles the functionality of the human visual cortex. Using a self-developed labelled collection of 60000 images over 10 classes, called the CIFAR-10 dataset, a model to categorize objects from pictures is obtained. Another important outcome of the research is the visualization of the filters in the network, such that it can be assessed how the model breaks down the pictures.

(ii) In another work which adopts the CIFAR-10 dataset, a very wide and deep network architecture is developed, combined with GPU support to decrease training time. On popular datasets, such as the MNIST handwritten digits, Chinese characters, and the CIFAR-10 images, near-human performance is achieved. The extremely low error rates beat prior state-of-the-art results significantly. However, it must be mentioned that the network used for the CIFAR10 dataset consists of 4 convolutional layers with 300 maps each, 3 max pooling layers, and 3 fully connected output layers. As a result, although a GPU was used, the training time was several days.

(iii) In 2010, the introduction of the yearly ImageNet challenge boosted the research on image classification and the belonging gigantic set of labelled data is often used in publications ever since. In a later work of Kievsky et al, a network with 5 convolutional, 3 max pooling, and 3 fully connected layers is trained with 1.2 million high resolution images from the ImageNet LSVRC-2010 contest. After implementing techniques to reduce overfitting, the results are promising compared to previous state-of-the-art models. Furthermore, experiments are done with lowering the network size, stating that the number of layers can be significantly reduced while the performance drops only a little.

(iv) With respect to facial expression recognition in particular, Lev et al. present a deep belief network specifically for use with the Japanese Female Facial Expression (JAFPE) and extended Cohn-Kanade (CK+) databases. The most notable feature of the network is the hierarchical face parsing concept, i.e., the image is passed through the network several times to first detect the face, thereafter the eyes, nose, and mouth, and finally the belonging emotion. The results are comparable with the accuracy obtained by other methods on the same database, such as Support Vector Machine (SVM) and Learning Vector Quantization (LVQ)(v) Another work on the Cohn-Kanade database makes use of Gabor filtering for image processing and Support Vector Machine (SVM) for classification. A Gabor filter is particularly suitable for pattern recognition in images and is claimed to mimic the function of the human visual system. The emotion recognition accuracies are high, varying from 88% on anger to 100% on surprised. A big disadvantage of the approach however is that very precise pre-processing of the data is required, such that every image complies to a strict format before feeding it into the classifier.

(vi) One of the most recent studies on emotion recognition describes a neural network able to recognize race, age, gender, and emotion from pictures of face. The dataset used for the latter category is originating from the Facial Expression Recognition Challenge (FERC-2013). A clearly organized deep network consisting of 3 convolutional layers, 1 fully connected layer, and some small layers in between obtained an average accuracy of 67% on emotion classification, which is equal to previous state-of-the-art publications on the same dataset. Furthermore, this thesis lays down a valuable analysis of the effect of adjusting the network size, pooling, and dropout.

Underlined by some other literature, the most promising concept for facial expression analysis is the use of deep convolutional neural networks. However, the network from (ii) is too heavy for our limited amount of available processing resources. The original network from (iii) is large as well, but smaller versions are claimed to be equally suitable. Furthermore, due to their somewhat analytical and unconventional approaches, we will not evaluate (iv) and (v). Hence, in the next section, three deep



architectures in total will be subjected to an emotion classification problem. These architectures are derived from, but not necessarily equal to, the networks described at items i, iii, and vi.

3. PROPOSED SYSTEM

In our proposed method we are detecting human emotions from the given dataset.

We are proposing a method which is detecting the emotions from the dataset given.

Here in this method, we are using cascade network of the deep learning which is a Convolutional Neural Networks (CNN) and a machine learning method that is TensorFlow library.

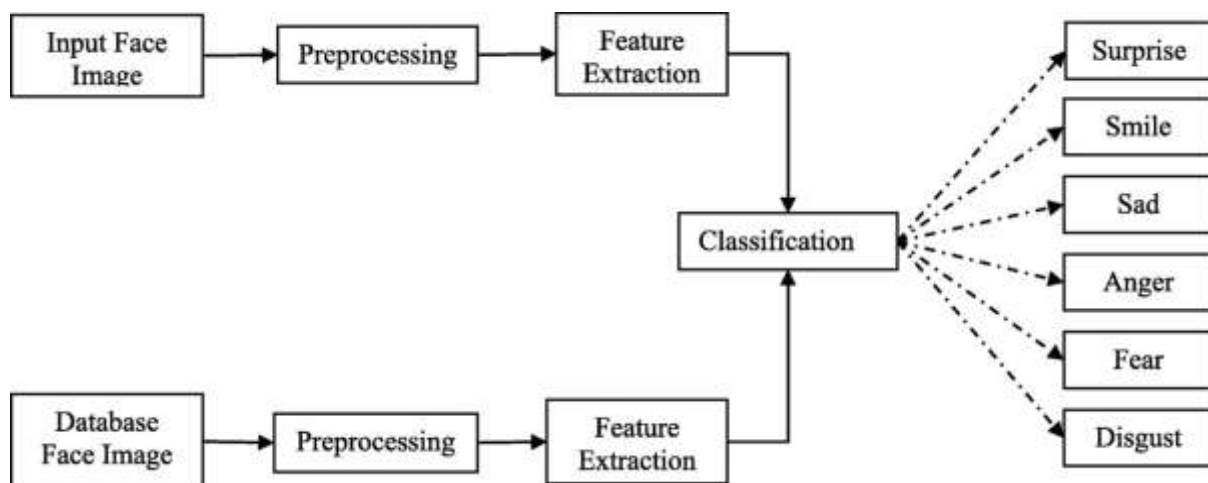
These algorithms are been used to train the face image dataset, upon where the classification will be performed along with the face recognition.

ADVANTAGES

Less complexity due to Transfer learning.

High performance.

Accurate classification.



3.1 DATASET

The data consists of 48x48 pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less centered and occupies about the same amount of space in each image. Each image corresponds to a facial expression in one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). The dataset contains approximately 36K images.

3.2 DATA VISUALIZING

Can you guess which images are related to which expressions?

This task is quite easy for a human, but it may be a bit challenging for a predictive algorithm because: the images have a low resolution



the faces are not in the same position

some images have text written on them

some people hide part of their faces with their hands

However, all this diversity of images will contribute to make a more generalizable model.

3.3 MODEL

We chose to use a Convolutional Neural Network in order to tackle this face recognition problem. Indeed, this type of Neural Network (NN) is good for extracting the features of images and is widely used for image analysis subjects like image classification.

Convolutional Neural Networks also have Convolutional layers that apply sliding functions to group of pixels that are next to each other. Therefore, those structures have a better understanding of patterns that we can observe in images. We will explain this in more details after.

We define our CNN with the following global architecture:

4 convolutional layers

2 fully connected layers








The convolutional layers will extract relevant features from the images and the fully connected layers will focus on using these features to classify well our images.

Let's focus on how our convolution layers work. Each of them contains the following operations:

A convolution operator: extracts features from the input image using sliding matrices to preserve the spatial relations between the pixels.

The following image summarizes how it works:

The green matrix corresponds to the raw image values. The orange sliding matrix is called a 'filter' or 'kernel'. This filter slides over the image by one pixel at each step (stride). During each step, we multiply the filter with the corresponding elements of the base matrix. There are different types of filters and each one will be able to retrieve different image features:

Operation	Filter	Convolved Image
Identity	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	
Edge detection	$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$	
	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	
	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$	
Sharpen	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	
Box blur (normalized)	$\frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$	
Gaussian blur (approximation)	$\frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$	



We apply the ReLU function to introduce non-linearity in our CNN. Other functions like tanh or sigmoid could also be used, but ReLU has been found to perform better in most situations.

Pooling is used to reduce the dimensionality of each features while retaining the most important information. Like for the convolutional step, we apply a sliding function on our data. Different functions can be applied: max, sum, mean... The max function usually performs better.

We also use some common techniques for each layer:

Batch normalization: improves the performance and stability of NNs by providing inputs with zero mean and unit variance.

Dropout: reduces overfitting by randomly not updating the weights of some nodes. This helps prevent the NN from relying on one node in the layer too much.

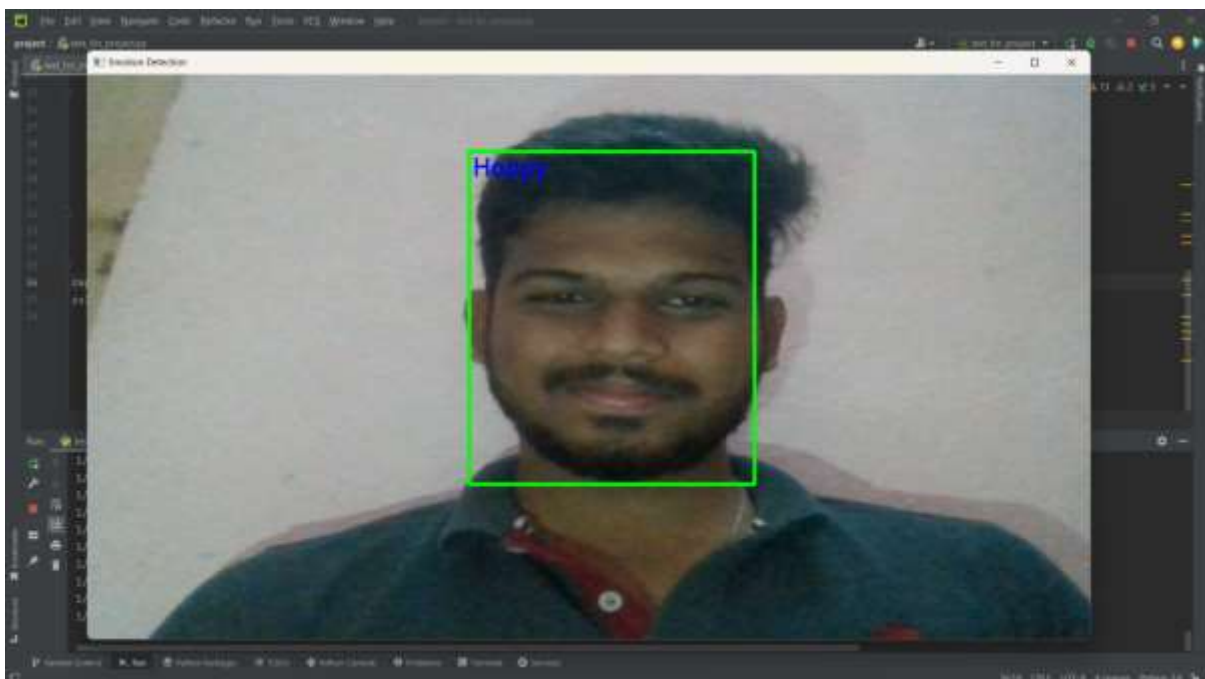
We chose softmax as our last activation function as it is commonly used for multi-label classification.

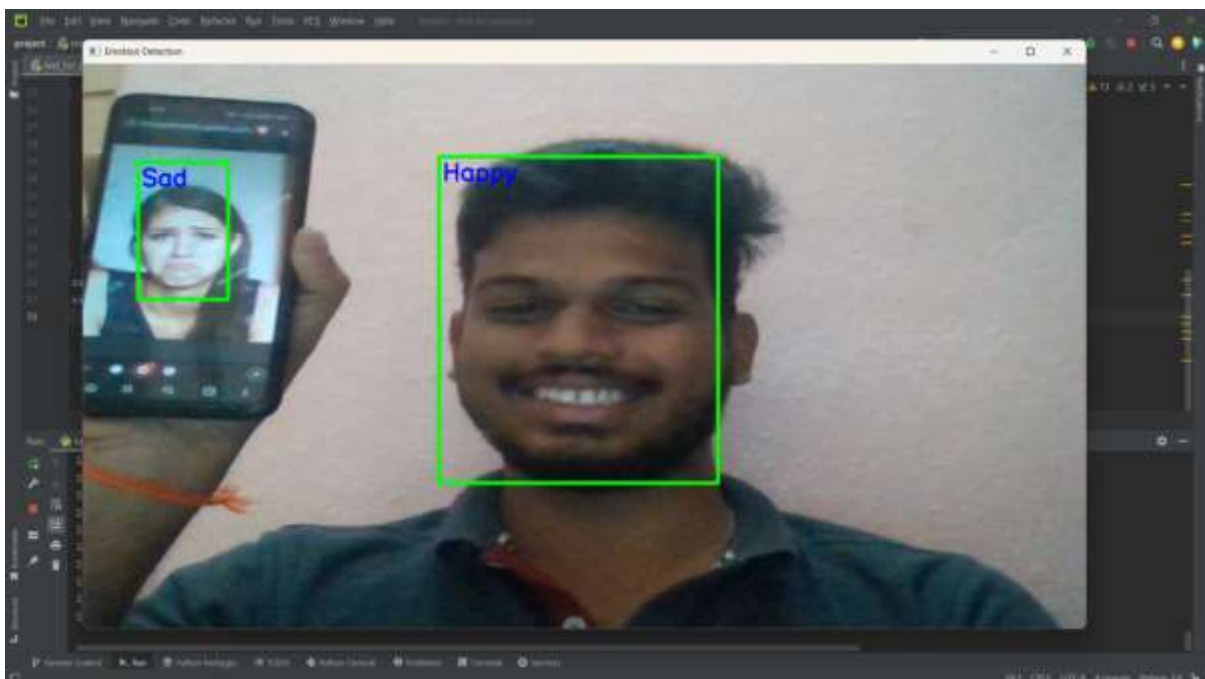
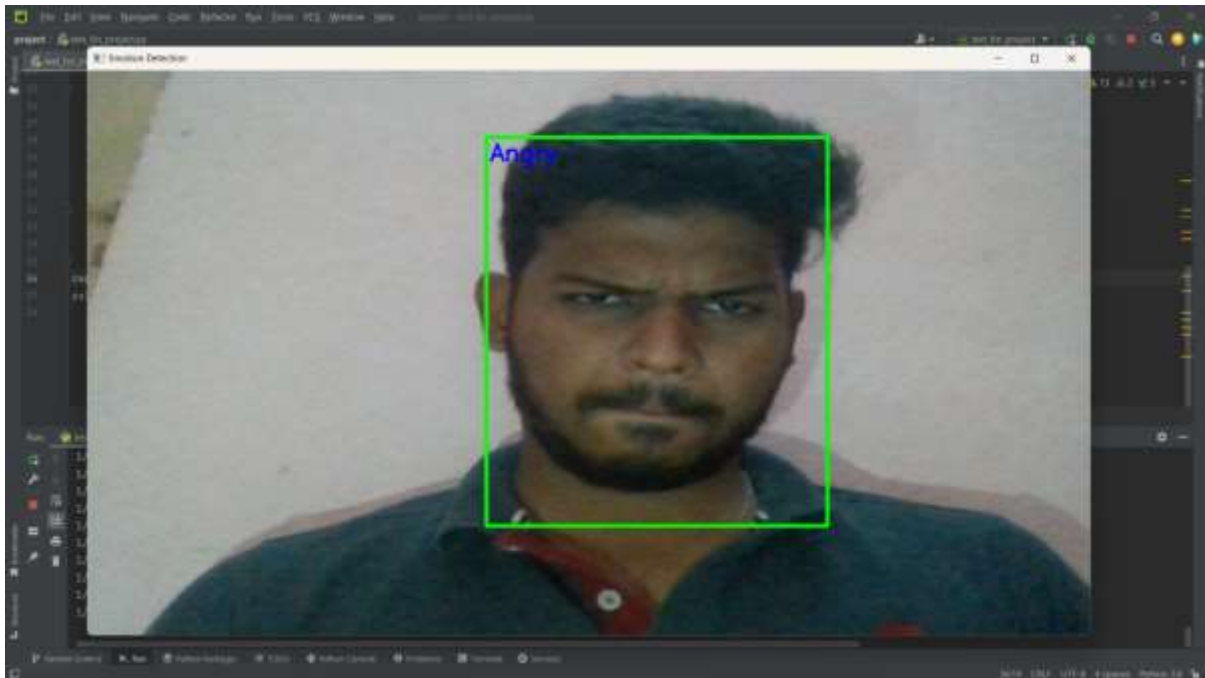
Now that our CNN is defined, we can compile it with a few more parameters. We chose the Adam optimizer as it is one of the most computationally effective. We chose the categorical cross-entropy as our loss function as it is quite relevant for classification tasks. Our metric will be the accuracy, which is also quite informative for classification tasks on balanced datasets.

Our best model managed to obtain a validation accuracy of approximately 65%, which is quite good given the fact that our target class has 7 possible values!

At each epoch, Keras checks if our model performed better than during the previous epochs. If it is the case, the new best model weights are saved into a file. This will allow us to load directly the weights of our model without having to re-train it if we want to use it.

4. RESULTS AND DISCUSSION





5. CONCLUSION

The proposed system was successfully implemented to get expected output.

From the above output images, we can conclude that it predicts human emotions very well except for disgust category due to insufficient training data in that category.



REFERENCES

- [1] T. Ahsan, T. Jabid, and U.-P. Chong. Facial expression recognition using local transitional pattern n gabor filtered facial images. *IETE Technical Review*, 30(1):47–52, 2013.
- [2] D. Ciresan, U. Meier, and J. Schmidhuber. Multi-column deep neural networks for image classification. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3642–3649. IEEE, 2012.
- [3] C. R. Darwin. *The expression of the emotions in man and animals*. John Murray, London, 1872.
- [4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009.
- [5] P. Ekman and W. V. Friesen. Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2):124, 1971.
- [6] B. Fasel and J. Luetttin. Automatic facial expression analysis: a survey. *Pattern recognition*, 36(1):259–275, 2003.
- [7] A. Gudi. Recognizing semantic features in faces using deep learning. arXiv preprint arXiv:1512.00743, 2015.
- [8] Kaggle. Challenges in representation learning: Facial expression recognition challenge, 2013. [9] A. Krizhevsky and G. Hinton. Learning multiple layers of features from tiny images, 2009. [10] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [11] O. Langner, R. Dotsch, G. Bijlstra, D. H. Wigboldus, S. T. Hawk, and A. van Knippenberg. Presentation and validation of the radboud faces database. *Cognition and emotion*, 24(8):1377– 1388, 2010.
- [12] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 94–101. IEEE, 2010.
- [13] Y. Lv, Z. Feng, and C. Xu. Facial expression recognition via deep learning. In *Smart Computing (SMARTCOMP), 2014 International Conference on*, pages 303–308. IEEE, 2014. [14] J. Nicholson, K. Takahashi, and R. Nakatsu. Emotion recognition in speech using neural networks. *Neural computing & applications*, 9(4): 290–296, 2000.
- [15] OpenSourceComputerVision. Face detection using haar cascades. URL http://docs.opencv.org/master/d7/d8b/tutorial_py_face_detection.html.
- [16] TFlern. Tflern: Deep learning library featuring a higher-level api for tensorflow. URL <http://tflern.org/>