A
Project Report
On
"Generate Avatar using Facial Expressions"

Prepared by

Kshitij Antani (18DCS005)

Under the guidance of

Krishna Patel

Assistant Professor cum Research Fellow

Vidhi Pandya

Assistant Professor

A Report Submitted to

Charotar University of Science and Technology

for Partial Fulfillment of the Requirements for the

6th Semester Software Group Project-IV (CS349)

Submitted at

CSE

DEPSTAR

At: Changa, Dist: Anand – 388421

APRIL 2021

A
Project Report
On
"Generate Avatar using Facial Expressions"

Prepared by

Rutika Mehta (18DCS045)

Under the guidance of

Krishna Patel

Assistant Professor cum Research Fellow

Vidhi Pandya

Assistant Professor

A Report Submitted to

Charotar University of Science and Technology

for Partial Fulfillment of the Requirements for the

6th Semester Software Group Project-IV (CS349)

Submitted at

CSE

DEPSTAR

At: Changa, Dist: Anand – 388421

APRIL 2021

CERTIFICATE

This is to certify that the report entitled "**Generate Avatar using Facial Expressions**" is a bonafied work carried out by **Ms. Kshitij Antani (18DCS005)** under the guidance and supervision of **Assistant Prof. cum Research Fellow Krishna Patel and Assistant Prof. Vidhi Pandya** for the subject CS349 -**Software Group Project-IV** (CSE) of 6th Semester of Bachelor of Technology in **DEPSTAR** at Faculty of Technology & Engineering – CHARUSAT, Gujarat.

To the best of my knowledge and belief, this work embodies the work of candidate himself, has duly been completed, and fulfills the requirement of the ordinance relating to the B.Tech. Degree of the University and is up to the standard in respect of content, presentation and language for being referred to the examiner.

Krishna Patel
Assistant Professor cum Research Fellow
CSE
DEPSTAR, Changa, Gujarat.

Vidhi Pandya
Assistant Professor
CSE
DEPSTAR, Changa, Gujarat.

Prof. Parth Goel
I/C Head of Department
CSE
DEPSTAR, Changa, Gujarat

Dr. Amit Ganatra
Principal, DEPSTAR
Dean, FTE
CHARUSAT, Changa, Gujarat

CERTIFICATE

This is to certify that the report entitled "**Generate Avatar using Facial Expressions**" is a bonafied work carried out by **Ms. Rutika Mehta (18DCS045)** under the guidance and supervision of **Assistant Prof. cum Research Fellow Krishna Patel and Assistant Prof. Vidhi Pandya** for the subject CS349 -**Software Group Project-IV** (CSE) of 6th Semester of Bachelor of Technology in **DEPSTAR** at Faculty of Technology & Engineering – CHARUSAT, Gujarat.

To the best of my knowledge and belief, this work embodies the work of candidate himself, has duly been completed, and fulfills the requirement of the ordinance relating to the B.Tech. Degree of the University and is up to the standard in respect of content, presentation and language for being referred to the examiner.

Krishna Patel
Assistant Professor cum Research Fellow
CSE
DEPSTAR, Changa, Gujarat.

Vidhi Pandya
Assistant Professor
CSE
DEPSTAR, Changa, Gujarat.

Prof. Parth Goel
I/C Head of Department
CSE
DEPSTAR, Changa, Gujarat

Dr. Amit Ganatra
Principal, DEPSTAR
Dean, FTE
CHARUSAT, Changa, Gujarat

ACKNOWLEDGEMENT

We are privileged to have this opportunity to express our gratitude and acknowledge everyone's never ending support and valuable contributions for our project.

We would like to express our sincere gratitude to our advisor **Assistant Prof. cum Research Fellow Krishna Patel** & **Assistant Prof. Vidhi Pandya** for the continuous support of our project and related research, motivation and immense knowledge.

Our sincere thanks also go to **Dean Dr. Amit Ganatra and HoD Prof. Parth Goel** who provided us an opportunity to work on a project and be able to present the same.

Last but not the least; we would like to thank our friends and family for supporting us spiritually throughout this project and for always being a constant source of inspiration.

We also place on record, our sense of gratitude to one and all, who directly or indirectly, have lent their hand in this venture.

ABSTRACT

Nowadays, each and every person is willing to express emotions graphically. With such a good environment that the social interactive sessions become more interactive and effective in day-to-day life. Thus, this approach helps to improvise a person's bonding skills in terms of innovativeness, transformation and refactoring. Here, we use communication over a computer network, with the help of abstract faces such as computer graphics avatars. These avatars convey user's emotions, expressions and enrich their communications. Seeing some avatars showing expressions the same as users, we feel that they are more vivid and lively than photorealistic faces. Some recent advancements in facial tracking systems provide accurate facial movements tracking so that to get accurate data. However, the existing systems are lacking in context of converting or expressing the data acquired from face tracking to other sources. In this paper we are trying an approach of development in an immensely innovative way to turn facial data into graphical representation via showing the avatar nearly matching the expression of a person using Convolutional Neural Network (CNN). CNN helped to achieve the accuracy of 92% which is higher than the traditional approaches. Thus, results in development of a system lead to development of tools to express emotions digitally.

Table of Contents

List of Figures

CHAPTER 1: PROJECT DEFINITION

Communication is an important part of everyday life whether it is verbal or non-verbal, it allows one to engage in conversations with the world. With the help of communication technologies that is the internet and other communication devices have made it possible to engage in fast, dynamic, and effective communication. Emojis are being used for the visual depictions of human emotions through their expressions. The identification of facial expressions plays an important key role in identifying patterns and image processing. Also helps identifying facial expressions and generates the emoji that almost matches the person's emotions. Some of the application, related to face and its expressions include person identification and access control, video call and teleconferencing, forensic applications, human-computer interaction, automated surveillance, cosmetology, and so on. In this system, it is provided for dynamically generating and displaying emojis. During operation, the system receives a sequence of images capturing a user's head and face. For each image, the system determines the user's head position in the image; applies a machine-learning technique, which comprises a convolutional neutral network (CNN), on the image to determine a facial-expression class associated with the user's face; and generates an emoji image based on the user's head position and the determined facial-expression class. The system then applies a morph-target animation technique to a sequence of emoji images generated for the sequence of images to generate an animated emoji and displays the animated emoji.

CHAPTER 2: PROJECT DESCRIPTION

2.1 Project Summary

To build our Facial Expression Recognition module, we used Convolutional Neural Networks (CNNs). Deep learning-based approaches, mainly which comprises CNNs have a high success rate for tasks which are related to images as they are proficient in extracting adequate representations from the data. Convolutional Neural Networks is a special type of neural network that is very effectively used for image recognition and classification. It is highly proficient in areas like identification of objects, faces, traffic signs. They are also used in self driving cars and robots. Convolutional is a simple application of a filter to an input that results in an activation. So, there is some sort of input and there are some certain sorts of thresholds, and when the input meets those thresholds then there is an activation. Now, repeated application of the same filter to an input results in a map of activations which is called a feature map. If there is a certain type of input that goes in and that input repeats itself in a certain way then we can see there is a feature map forming that indicates their locations and strengths of a detected feature in input. We can then train the model with help of those features. Facial expressions are arrangements of facial muscles which are used for conveying emotional state to the one who is observing. Emotions can be divided into seven broad categories—Anger, Disgust, Fear, Happiness, Neutral, Sad and Surprised. We will build a deep learning model to classify facial expressions from the images. Then we will map the classified emotion to a predefined avatar.

The neural network needs sets of training data and teaching signals for training. In the proposed method, we selected 7 facial expressions (neutral, happy, angry, surprised, sad) as the recognition target. We have adopted a 5-layer neural network. The learning rate was set to 0.0001, epoch was set to 75 and decay was set 1^{e}-6. We have used two activation functions one Relu and Softmax. Relu is better activation function then other.

2.2 Project Related Works

2.2.1 Maker- Free Personal Animated Emojis:

Phone supports and enables users to use a new type of emojis called Animated Emoji. It also supports characters like fox, pig, dog, cat and different animals. This app allows users to send and receive animated emojis through smartphones. One can express himself with 3D animated stickers and emojis featuring yourself as a cartoon avatar maker.

2.2.2 Bitmoji – Your Personal Emoji:

People can create an expressive cartoon avatar a lot like themselves. One can choose from a huge and various library of stickers all featuring your looks. One can use this application wherever he wants like third party apps.

2.2.3 GIFs, Facemoji Keyboard – Emoji Keyboard, and Stickers:

The First ever Emoji-Centric Android keyboard. It holds over 3600+ Emoji, emoticons, GIFs, stickers on this Emoji keyboard. This application helps us to spice up and make chat more interesting on Facebook, Twitter, Instagram, Messenger, WhatsApp and for many more apps.

2.2.4 Elite Emoji:

This application has sensational emojis and attractive stickers to express emotions, moods, messages and ideas beautifully in your chat conversations. It contains almost 2000 exclusive high definition emojis and stickers to choose from. Also, with one more feature as it supports emojis with a single touch in Social networking sites. By using this application, we get access to a huge selection of gifs as well and even create your own animated images to make communication describe more perfectly

CHAPTER 3: SOFTWARE & HARDWARE REQUIREMENTS

3.1 Software Requirements

3.1.1 Python3:

Python is a general purpose and highlevel programming language. You can use Python for developing desktop GUI applications, websites and web applications. ... The simple syntax rules of the programming language further make it easier for you to keep the code base readable and application maintainable. Python offers concise and readable code. While complex algorithms and versatile workflows stand behind machine learning and AI, Python's simplicity allows developers to write reliable systems.

3.1.2 PythonIDE:

The python IDE we used for the project is PyCharm. PyCharm is an integrated development environment (IDE) used in computer programming, specifically for the Python language. It is developed by the Czech company JetBrains. It is cross-platform, with Windows, macOS and Linux versions.

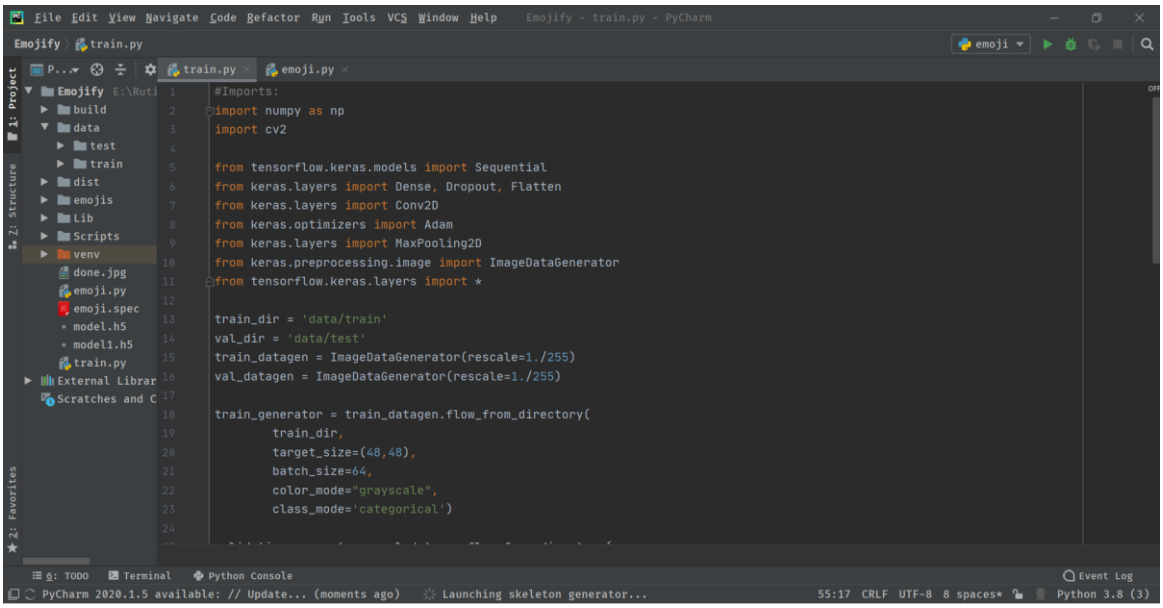

Fig 3.1.2 PyCharm IDE Screen

3.1.3 Avatar Maker:

Online Avatar maker's sites are most useful to create numerous types of avatars as per one's facial type. We can modify eyes, nose, hairs, mouth, pupil's size, eye colors, ear, haircut, eyebrows, ears, and gender. These many features help creating

accurate expression same as user. Link: https://www.cartoonify.de/anime-character-creator/

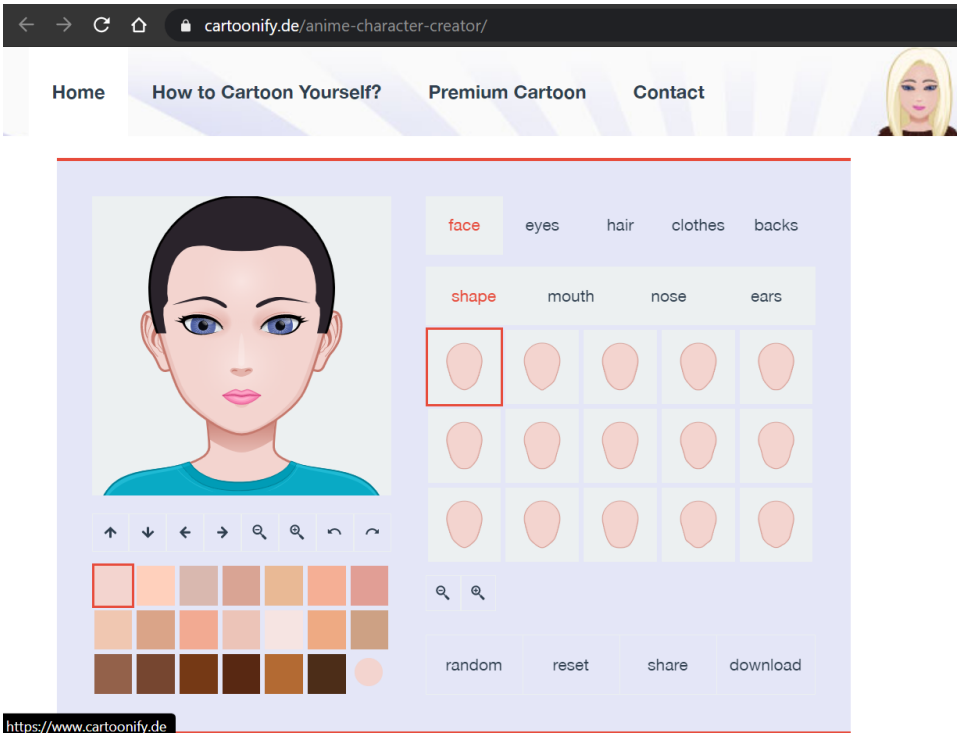

Fig 3.1.3 Online Avatar Maker Site

3.2 Hardware Requirements

3.2.1 Laptop/PC:

A laptop or PC that supports the webcam to detect user's face.

CHAPTER 4: MAJOR FUNCTIONALITY

4.1 Dataset

The data consists of 48x48 pixel grayscale images of faces. The faces in the dataset have been automatically recorded so that the face occupies the same amount of space in each image. The dataset is categorized into each face based on the emotion shown in the facial expression into one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).

4.2 Data Pre-processing:

Data preprocessing refers to a process of developing the raw data and making it befitting for a machine learning model. It is the first and foremost step while building the machine learning model. When building machine learning projects, using a clean and well formatted dataset is a predominant factor. The data gathered from real-world generally consists of missing values, noises or can be in format which cannot be used directly. Hence, data preprocessing plays a crucial role by cleaning the data and making it appropriate for machine learning models. A well formatted dataset also increases the efficiency and accuracy of the model. In our project, we have used the dataset FER- 2013 in which the training set consists of 28,709 examples and the public test set consists of 3,589 examples.

4.3 Model Architecture

We will generate datasets for training data and validation data from image files in our directory from scale of 1 to 255, since we define our images in pixels which is an array consisting of values from 1 to 255. Then we'll create our generator from the imported data. The generator will be specified with target size, batch size, color mode and class mode. The target size refers to the pixel size of the data, batch size is a term used in machine learning and refers to the number of training examples utilized in one iteration, the color mode specifies the mode of the image dataset, and here the class mode will be categorical since we have categories of angry, disgust, fearful, happiness, sadness, surprised and neutral.

We will use Keras library to create a model which uses Sequential Convolutional Network. The Sequential Convolutional Network allows to create layers of a model in a linear stack of layers i.e., layer by layer, which is the preferred way for most problems.[5] The convolutional network is divided into following components:

4.3.1 Convolutional Layer:

It is of two dimensions. It creates a convolutional kernel that is convolved with the layer input to produce a tensor of outputs. Here we will provide input shape, height and width of convolutional network, activation function and input shape of the image.

4.3.2 Activation functions:

In the neural network, activation function decides whether a neuron should be activated or not by calculating the weighted sum and adding it further with bias. Bias is an additional parameter used to adjust the weighted sum of inputs along with outputs to the neuron. In our project, we will use two activation functions, Relu and SoftMax. Relu stands for rectifier which is an activation function used in a neural network. It is one of the most commonly used activation functions. Here, Softmax function is used for normalizing the outputs by converting them from weighted sum values into probabilities that aggregates to one.

4.3.3 Pooling Layer:

The pooling layers down sample the data. Down sampling means training on a disproportionately low subset of the classic samples. We are down sampling to include the lower subset of the data. In other words, pooling is the process of merging data. In this project, max pooling is used where only maximum value is selected from the input.

4.3.4 Dropout Layers:

The dropout layers randomly set input units to zero with the frequency rate at each during the training time which helps to prevent overfitting. Overfitting occurs when a model has learned the noise instead of signal that is considered to be considered to be the required signal for learning. Thus, overfitting means training the model more on the noise than the actual data we need.

4.3.5 Flatten Layer:

Here, the data is converted into a 1-dimensional array so that it can be used as input for the next layer. The output of the convolutional layer is flattened to create a single long feature vector. We need to flatten the values before we turn all those neurons into the next layer.

4.3.6 Dense Layer:

Dense layer is the hidden layer. In this project, the first dense layer is going to contain 1024 neurons. And the second dense layer contains seven, so this layer is going to contain the exact number of possibilities that we can have. Since in our dataset, we have expressions divided into seven categories. So, we have to convert 1024 neurons from the first hidden layer into those seven categories.

CHATER 5: FLOW CHART

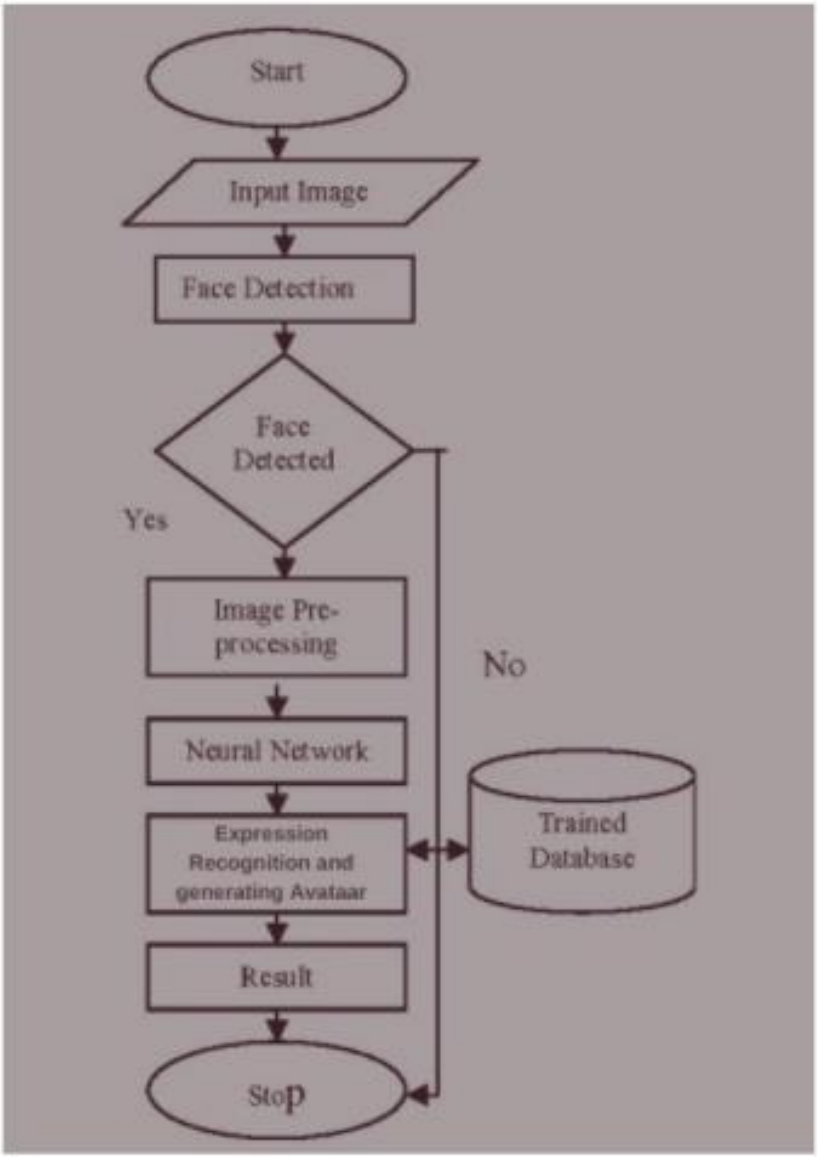

Fig 5.1 Flowchart

CHAPTER 6: SCREENSHOTS OF PROJECT OUTPUT

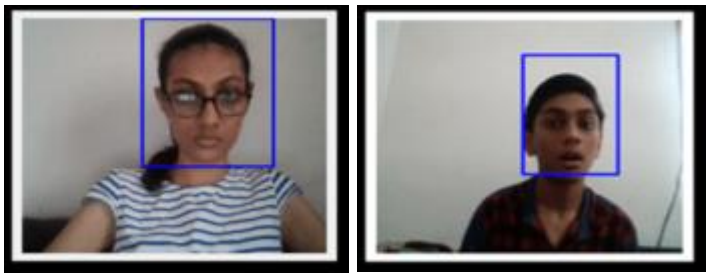

Fig 6.1 Face Detection

Fig 6.2 Generated Avatar from Facial Expression

CHAPTER 7: LIMITATIONS OF PROJECT

Although our study contributes new findings, it has necessary limitations.

- The facial expressions recognized by the software are limited to angry, disgust, fear, happy, sad, surprised and neutral. Many other expressions are to be added.
- The efficiency and performance of the software can be improved.
- The project is available as desktop application only, web application and mobile application can be built.
- The accuracy of the application can also be improved.

CHAPTER 8: PROJECT OUTCOME

In the below figure, i.e., the sample output window the user expressions are captured on the left-hand side of the window through the webcam.

On the right-hand side of the output window, the corresponding avatar is displayed on the basis of the user's facial expression.

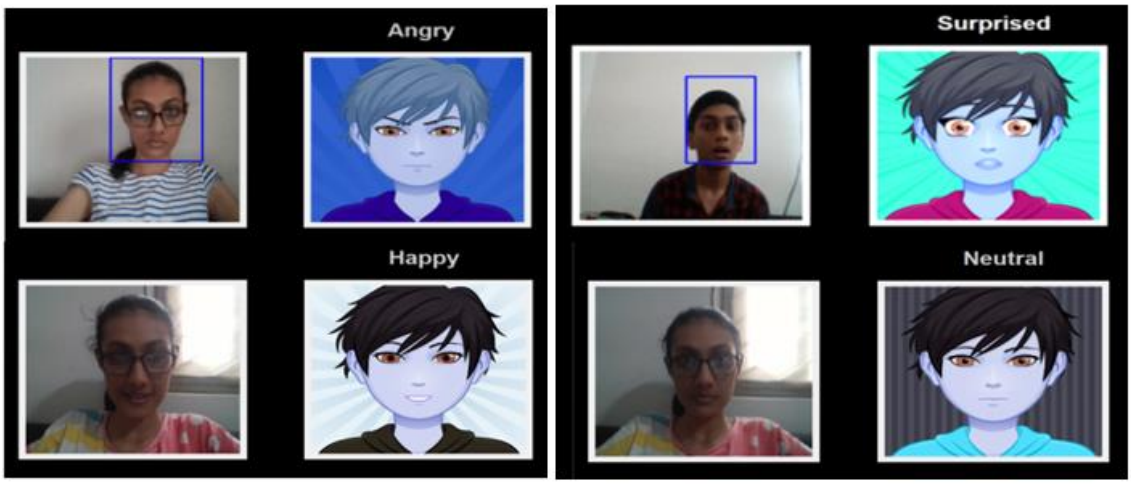

Fig 8.1 Results of project

On top of the avatar, the respective emotion text is displayed. The avatar changes when the expression of the user changes and that is displayed on the output screen. This real time application can be used in the fields of computer science, neuroscience, psychology and various other fields.

CHAPTER 9: FUTURE ENHANCEMENT

- In a further variation, the method further comprises aggregating facial-expression-detection outputs from the multiple specialized convolutional networks by applying weights to the facial-expression-detection outputs from the multiple specialized convolutional neural networks
- The determined facial-expression class can be selected from a group consisting of: neutral, happy, sad, fear, angry, surprise, disgust, tongue out, kiss, wink, eyebrow raise, and nose wrinkle.
- Determining the user's head position comprises determining facial landmark points in the image using a shape modelling algorithm.
- Receiving the sequence of images comprises receiving a live video feed from a camera associated with computer and sampling frames from the received live video feed.

REFERENCES

- Guifang Lin, WeiShen. Research on convolutional neural networks based on improved Relu piecewise activation function. Procedia Computer Science, volume = 131, doi: https://doi.org/10.1016/j.procs.2018.04.239
- K. Suzuki et al., "Recognition and mapping of facial expressions to avatar by embedded photo reflective sensors in head mounted display," 2017 IEEE Virtual Reality (VR), Los Angeles, CA, USA, 2017, pp. 177-185, doi: 10.1109/VR.2017.7892245
- Pradnya Bhor, Ashwini Raut, Sachin Kate, Anurag Salve, Chandan Prasad. “Research paper on Dynamic Emotion Recognition and Emoji Generation”, 2019 IRJET.
- N. Swapna Goud, K. Revanth Reddy, G. Alekhya, G. S. Sucheta. “Research Paper on Facial Emoji Recognition”, volume = 3, doi: https://doi.org/10.31142/ijtsrd23166
- N. Swapna Goud, K. Revanth Reddy, G. Alekhya, G. S. Sucheta, “Facial Emoji Recognition”, Volume= 3, Issue= 3, doi: 10.31142/ijtsrd23166
- Abdallah El Ali, Torben Wallbaum, Merlin Wasmann, Wilko Heuten, “Face2Emoji: Using Facial Emotional Expressions to Filter Emojis”, doi: http://dx.doi.org/10.1145/3027063.3053086
- https://www.kaggle.com/msambare/fer2013/metadata