Machine Learning-based Automated Facial Emotion Recognition

Terms of Reference

Alexandru-Adonis Neagu

18017575

Computer Science

Supervisor: Dr Ossama Alshabrawy

Second marker: Dr Paul Vickers

**Background to Project**

Every human being has felt that in one way or another, they express these emotions as a way of communication. People can show our feelings in different ways as body language, voice tonality and facial expression. ( Happy, S. and Routray, A., 2015), (Majumder, A., Behera, L. and Subramanian, V., 2018).

           Recognising facial expression has been a subject that arouses the curiosity of many researchers and philosophers. In 1872 Charles Darwin was the first one who suggested that there are six universal emotions that humans can read and understand all over the world. (Matsumoto, D., Hwang, H.S., 2011) This suggestion has been rejected by many researchers such as Paul Ekman, Wallace Friesen and Phoebe Ellsworth (Matsumoto, D., Hwang, H.S., 2011). However, later Silvan Tomkins recruited Paul Ekman and Carroll Izard to continue Charles Darwin hypothesis, and they concluded that regardless of location and culture people can express and understand the following emotions: surprise, fear, disgust, anger, happiness and sadness. (Happy, S. and Routray, A., 2015), (Fasel, B. and Luettin, J., 2019)

           Ekman found that by reading facial emotion, we would be able to figure out patterns that can help us predict behaviours or thoughts. (Ekman, P. and Friesen, 1984) Nowadays, the advancing of technology opens new pathways in how we can use facial emotion recognition, and it has many applications in artificial intelligence domain.

           In the field of robotics, facial emotions play a massive role in human-computer interaction. In order to develop a robot that can simulate human behaviour, it has to learn by itself how to associate the inner feeling and the facial movements. ( Zhang, L. et al., 2013) A great example would be Sophia, a very advance robot that can learn and display facial emotions using artificial intelligence. ( https://www.hansonrobotics.com/sophia/)

           Teaching has changed lately, and e-learning is the most common way to teach due to the pandemic. The technology of facial emotion recognition can assist teachers by measuring the learning levels of the students, but it can also predict satisfaction and understanding when they participate in a lecture session. (Lansley, J., 2020), (Wang, H. and Gu, J., 2018)

           In the medical field, facial emotion recognition has begun to be used by doctors in the treatment of mental and psychiatric diseases. ( Samadiani, N. et al., 2019) The system could use a mix of facial emotion recognition, body recognition and voice tone recognition to expose subtitle information that will help doctors to make judgements on the emotional and mental state of people with autism, depression, schizophrenia and so on. (Lansley, J., 2020), (Wang, H. and Gu, J., 2018)

           Large companies are taking advantage of facial emotion recognition to increase their selling. They are collecting the facial expression of people when revelling a new product such as a phone or a car. Using this method, they can obtain a hones reaction when presenting the design and specification and predict whether the product will be profitable or not. Also, they are looking forward to developing a new way of recruiting trustworthy employees. Such a process implies a recorded questionnaire taken by the potential employees, which is then analysed using face emotion detection. (Wang, H. and Gu, J., 2018)

           Airports, military, security firms are also showing interest in researching the use of cameras and software to pair everything that is known about facial emotion recognition and deception in order detect any suspicious movements that can indicate drug traffic, terrorist attacks or any other potential crimes that can occur in a large and busy space. (Lansley, J., 2020),( Wang, H. and Gu, J., 2018)

           One of the reasons why I chose this subject is the growing interest that the companies are showing in this technology. I find this subject attractive because I find fascinating the idea that we can identify and predict the behaviour of human being. Moreover, I want to bring a small contribution to all the efforts made by other researchers and push myself to learn new stuff and develop my skills as a computer scientist.

**Data**

Automated affective computing within the wild setting is a challenging issue in computer vision. Existing annotated databases of facial expressions within the wild are little and mostly cover discrete emotions. There are exceptionally constrained commented on facial databases for personal computing within the continuous dimensional show. For this project will be used data from the AffectNet database. Since the database consists only in facial images, there will not be any other kind of data that will need to be collected. This database is by far the largest database of facial expressions, valence and in arousal in the wild enabling research in automated facial expression recognition in two different emotion models. It was created by collecting and annotating facial images. The word Affect is a mental term used to portray the outward articulation of feeling and emotions.

The AffectNet database contains about 1M facial pictures gathered from the internet by questioning three significant web search engines utilising 1250 emotions related watchwords in six different languages. There are approximative 550000 images that were automatically annotated using ResNext Neural Network. These samples were obtained by using the other samples as a training set. The other 420000 samples are manually annotated for the presence of seven discrete facial expressions and the intensity of valence and arousal. The database provides eleven annotated emotions: Neutral, Happiness, Sadness, Surprise, Fear, Disgust, Anger, Contempt, None, Uncertain and No-Face.

|  |  |
| --- | --- |
| Neutral | 75,374 |
| Happy | 134,915 |
| Sad | 25,959 |
| Surprise | 14,590 |
| Fear | 6,878 |
| Disgust | 4,303 |
| Anger | 25,382 |
| Contempt | 4,250 |
| None | 33,588 |
| Uncertain | 12,145 |
| Non-Face | 82,915 |
| Total | 420,299 |

Features provided by the AffectNet database:

* Images of the faces
* Location of the faces in the images
* Location of the 68 facial landmarks
* Eleven emotion and non-emotion categorical labels (Neutral, Happiness, Sadness, Surprise, Fear, Disgust, Anger, Contempt, None, Uncertain, No-Face)
* Valence and arousal values of the facial expressions in a continuous domain

**Proposed Work**

In this project, the main idea is to create a deep learning model to classify the faces images in terms of motions. In the past years, many researchers tried to develop their version of how emotions can be detected on the face. Besides the actual code that I am going to create, I will conduct some tests and see how my model performs in comparison with some other similar programs. The comparison will be related to the build of the program, meaning that I will describe the feature extraction process, classification, training and testing processes. A Deep Convolutional Neural Network (CNN) method will be used, which has an architecture that consists of filter layers and a classification layer. This approach has been chosen for its ability to deal with raw data and to overcome the disadvantage of feature-based methods that require a significant exertion should be put on to plan and utilise different include extraction techniques which are human-made features. CNN input and output are array vectors called a feature map. Since this project will deal with images, the array dimension will be 2D.

The code that is going to be build will use images with people showing the six universal emotions, which will be the dataset. First, the images from the dataset will be pre-processed. In order to process the pictures, a face detection code will be used. This will help the program to recognise the human face from the picture,

The second stage will be facial expression extraction. Basically, the program will have to recognise patterns that will indicate a specific emotion. The program will need to store just the optimal key points like eyes, eyebrows and mouth. After the extraction, the features are stored into a feature vector.

The next step would be the classification. The problem with the classifiers is that they assume that there is no correlation between features. The program will need to learn how to recognise the difference between the images. To do so, labels will be added, matching the six emotions.

Once all steps have been carefully completed, the training of the systems can begin. The dataset that will be used is AffectNet dataset which contains over one million facial images collected from the internet. This is by far the largest database of facial expression.

For the training bit, only half of the pictures will be used. The training process accuracy will depend on how correctly the system can classify the images so that the training process will proceed optimally. Moreover, to accelerate and reduce the cost of training, a pre-existing network model will be used. The idea behind transfer learning for image classification is that an effective generic model of the visual world can be obtained by training a model on a large and general enough dataset. You can then take advantage of these learned feature maps without having to start from scratch by training a large model on a large dataset. The only modification that needs to be made is to unfreeze a few of the top layers of a frozen model base and jointly train both the newly-added classifier layers and the last layers of the base model. This will allow to "fine-tune" the higher-order feature representations in order to make them more relevant to this task.

Finally, the last bit consists of testing. Once the training has been done, we can see how well the system can recognise the emotions from new pictures that were not used previously in the training process. The accuracy for each of the six emotions will be measured to see if there could be any modifications that can be done to improve the overall score so that the error percentage to be as small as possible.

Besides all the work done on the actual code, research will be carried out to identify the most suitable ways to develop a better understanding of what machine learning/deep learning is and how it works. Also, the language used for the code will be python which is a high-level program language. To master this type of language, extra time will be spent on reading and practising some exercises

**Aims**

* Develop a program using machine learning/deep learning that is capable to automatically detect and interpret the emotions expressed by a human face.
* Optimise the program in order to deliver accurate predictions.

**Objectives**

1. Carry out the literature review
   1. Research face detection technology
   2. Research facial emotion technology
2. Investigate suitable tools, techniques and methods to carry out the project
3. Finding the best suitable methods for building the program
   1. Find a face detection algorithm
   2. Find a feature extractor
   3. Find a classifier
4. Training the program to detect and recognise facial emotion from the pictures
5. Testing how accurate the trained program can detect facial emotions from new pictures
6. Conduct an analysis of how the system performed
7. Evaluate how the overall system works
   1. Discuss the results and highlight the good practice
   2. Discuss the flows of the system
8. Give Recommendations
   1. Conclude and recommend what I could have done differently and future work and research that can be done within the field.

**Skills**

|  |  |  |
| --- | --- | --- |
| Skills | Current skill description | Skills improvement via |
| Python programming | Skill level: low | Online lectures and practice sessions |
| Machine learning | Skill level: medium  Acquired via Machine Learning module during 3rd year | Continuing to follow the module and studying deeper my free time |
| Deep learning | Skill level: medium  Acquired via Machine Learning Module during 3rd year | Continuing to follow the module and studying deeper my free time |
| Data processing | Skill level: low | Utilisation of various Python libraries to facilitate the processing and analysis of datasets. |
| Time management | Skill level: medium  Acquired during other projects development |  |

**Resources**

Hardware:

* A suitable, modern and powerful CPU to facilitate the software requirements.
* A suitable, modern and powerful GPU to accelerate the machine learning training.
* Sufficient RAM memory to facilitate the loading of the dataset into program memory.

Software:

* Windows 10 as an operating system
* Machine learning/Deep Learning packages in python (Scikit-learn, tensorflow, keras,Numpy CUDA)
* Images dataset (AffectNet dataset)
* Anaconda3
* Jupyter notebook
* Pyton 3.8
* Github for cloud storage

**Report structure**

Abstract

1. Introduction
   1. Background presentation
   2. Aims
   3. Objectives
   4. Structure of the dissertation
2. Literature review
   1. General face detection
   2. General facial emotion detection
3. Machine Learning Techniques
   1. Datasets
   2. Feature selections
   3. Classifiers comparison
   4. Deep Learning exploration
4. Analysis
   1. Face detection
   2. Feature extraction
   3. Classification
   4. Training results
5. Evaluation, Results and Discussion
6. Conclusion and future work

**Marking scheme**

1. Project Type

Investigative Project

|  |  |
| --- | --- |
| Report | 90% |
| Viva | 10% |
| Total | 100% |

Report

|  |  |
| --- | --- |
| Abstract & Introduction | 5% |
| Analysis | 20% |
| Synthesis | 50% |
| Evaluation & Conclusion | 20% |
| Presentation | 5% |
| Total | 100% |

Viva

|  |  |
| --- | --- |
| Presentation | 50% |
| Discussion | 50% |
| Total | 100% |

1. Project report

|  |  |  |
| --- | --- | --- |
| Report | Section |  |
| Abstract & Introduction | 1.Abstract  2.Introduction | 5% |
| Analysis | 3.Literature review  4. Machine learning techniques | 20% |
| Synthesis: discussion of methods & results | 5.Description and justification of the approach to be taken  6.Description of the tools to be used  7.Testing and results discussion | 20% |
| Synthesis: quality of practical work | 8.Quality of deliverables  9.Demonstration of effective use of practical skill | 30% |
| Evaluation & Conclusion | 10.Conclusion and future work | 20% |
| Presentation | \* | 5% |
| Total | \* | 100% |

**Project plan**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Semester 1** | | | | | | | | | | | | | | | | | | | | | | | |
| Aim | Objectives | Tasks | Estimate Hour Required | | W1 | W2 | W3 | | W4 | | W5 | W6 | | W7 | | W8 | W9 | W10 | | | W11 | | W12 |
| Project Initiation | Project Initiation Document | PID Creation | 10 | |  | | | |  | |  |  | |  | |  |  |  | | |  | |  |
| PID Submission |  | |  |  | Submission | |  | |  |  | |  | |  |  |  | | |  | |  |
| Project Research Plan | Terms of Reference | Project Research | 15 | |  |  |  | | | | | | | | |  |  |  | | |  | |  |
| Project Planning | 15 | |  |  |  | |  | |  | | | | |  |  |  | | |  | |  |
| TOR Review | 5 | |  |  |  | |  | |  |  | |  | |  |  |  | | |  | |  |
| TOR Submission | \* | |  |  |  | |  | |  |  | |  | | Submission |  |  | | |  | |  |
| Investigating facial emotion detection algorithms and develop future knowledge | Research the principles of face detection and facial emotion detection | Chapter Planning | 2 | |  |  |  | |  | |  |  | |  | |  | |  | | |  | |  |
| Research face detection | 10 | |  |  |  | |  | |  |  | |  | |  | |  | | |  | |  |
| Research face emotion detection | 10 | |  |  |  | |  | |  |  | |  | |  |  | |  | |  | |  |
| Chapter write up | 6 | |  |  |  | |  | |  |  | |  | |  |  |  |  | |  | |  |
| Research various code components | Chapter planning | 2 | |  |  |  | |  | |  |  | |  | |  |  |  |  | |  | |  |
| Dataset reading and research | 10 | |  |  |  | |  | |  |  | |  | |  |  |  |  | | |  |  |
| Feature extraction reading and research | 10 | |  |  |  | |  | |  |  | |  | |  |  |  |  | | |  |  |
| Classification reading and research | 10 | |  |  |  | |  | |  |  | |  | |  |  |  | | |  |  | |
| Deep learning reading and research | 20 | |  |  |  | |  | |  |  | |  | |  |  |  | | |  |  | |
| Chapter write up | 6 | |  |  |  | |  | |  |  | |  | |  |  |  | | |  |  | |
| Analysis chapter submission | Analysis chapter submission | \* | |  |  |  | |  | |  |  | |  | |  |  |  | | |  |  | Submission |
| **Semester 2** | | | | | | | | | | | | | | | | | | | | | | | |
| Aim | Objectives | Tasks | Estimate Hour Required | W1 | | W2 | W3 | W4 | | W5 | W6 | | W7 | | W8 | | W9 | W10 | | W11 | | | W12 |
| Implementing all the data acquired from research by developing the detection engine and discuss the results | The creations of relevant algorithms for incoming data pre-processing | Chapter planning | 3 |  |  |  |  |  | |  |  | |  | |  | |  |  | |  | | |  |
| Pseudo-code for algorithms | 8 |  |  |  |  |  | |  |  | |  | |  | |  |  | |  | | |  |
| Code for algorithms | 10 |  | |  |  |  | |  |  | |  | |  | |  |  | |  | | |  |
| Chapter write up | 10 |  | |  |  |  | |  |  | |  | |  | |  |  | |  | | |  |
| Perform an analysis of a test of the detection engine on an unseen dataset | Chapter planning | 3 |  | |  |  |  |  |  |  | |  | |  | |  |  | |  | | |  |
| Testing methodology | 5 |  | |  |  |  |  |  |  | |  | |  | |  |  | |  | | |  |
| Product testing | 10 |  | |  |  |  | |  |  | |  | |  | |  |  | |  | | |  |
| Collecting testing result | 5 |  | |  |  |  | |  |  |  |  | |  | |  |  | |  | | |  |
| Chapter write up | 10 |  | |  |  |  | |  |  |  | | |  | |  |  | |  | | |  |
| Carry out an evaluation of the final detection engine and identify areas of improvement and future work | Chapter planning | 3 |  | |  |  |  | |  |  | |  | |  |  |  |  | |  | | |  |
| Results analysis | 8 |  | |  |  |  | |  |  | |  | |  |  |  |  | |  | | |  |
| Evaluation of results | 8 |  | |  |  |  | |  |  | |  | |  | |  |  | |  | | |  |
| Identification of areas of improvement | 8 |  | |  |  |  | |  |  | |  | |  | |  |  | |  | | |  |
| Chapter write up | 15 |  | |  |  |  | |  |  | |  | |  | |  |  | |  | | |  |
| Report and product submission |  |  |  | |  |  |  | |  |  | |  | |  | |  |  | | Submission | | |  |

**References**

Matsumoto, D., Hwang, H.S., 2011, Reading facial expressions of emotion, Psychological Science Agenda

Ekman, P. and Friesen, W., 1984. Unmasking the Face. Palo Alto, Ca.: Consulting psychologists Press.

Lansley, J., 2020. Facial Expressions · EIA. [online] The Emotional Intelligence Academy. Available at: <https://www.eiagroup.com/knowledge/facial-expressions/> [Accessed 14 November 2020].

Wang, H. and Gu, J., 2018. The Applications of Facial Expression Recognition in Human-computer Interaction. 2018 IEEE International Conference on Advanced Manufacturing (ICAM).

Mollahosseini, A., Hasani, B. and Mahoor, M., 2017 Affectnet: A Database for Facial Expression, Valence, And Arousal Computing In The Wild. [online] Mohammadmahoor.com. Available at: <http://mohammadmahoor.com/wp-content/uploads/2017/08/AffectNet\_oneColumn-2.pdf> [Accessed 14 November 2020].

Samadiani, N., Huang, G., Cai, B., Luo, W., Chi, C., Xiang, Y. and He, J., 2019. A Review on Automatic Facial Expression Recognition Systems Assisted by Multimodal Sensor Data. Sensors, 19(8), p.1863.

Happy, S. and Routray, A., 2015. Automatic Facial Expression Recognition Using Features of Salient Facial Patches. [online] Ieeexplore.ieee.org. Available at: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6998925> [Accessed 14 November 2020].

Fasel, B. and Luettin, J., 2019. Automatic Facial Expression Analysis: A Survey. [online] Available at: <https://doi.org/10.1016/S0031-3203(02)00052-3> [Accessed 14 November 2020].

Majumder, A., Behera, L. and Subramanian, V., 2018. Automatic Facial Expression Recognition System Using Deep Network-Based Data Fusion - IEEE Journals & Magazine. [online] Ieeexplore.ieee.org. Available at: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7747479> [Accessed 14 November 2020].

Burkert, P., Trier, F., Afzal, M., Dengel, A. and Liwick, M., 2016. Depression: Deep Convolutional Neural Network For Expression Recognition. [online] Arxiv.org. Available at: <https://arxiv.org/pdf/1509.05371.pdf> [Accessed 14 November 2020].

Zhang, L., Jiang, M., Farid, D. and Hossain, M., 2013. Intelligent facial emotion recognition and semantic-based topic detection for a humanoid robot. Expert Systems with Applications, 40(13), pp.5160-5168.

Mollahosseini, A., Hasani, B. and Mahoor, M., 2017. *Affectnet: A Database For Facial Expression, Valence, And Arousal Computing In The Wild*.