Automating Physiological Data Collection from Human Reactions to Environmental Stimuli

Name of author

field and department

2021

# Abstract

Many face micro expressions mimic facial expressions to represent a broad range of emotions. People utilise the same facial expressions in public to indicate what they are feeling. At a broad level, we can hide our sentiments, yet it is possible to uncover what we are genuinely feeling and thinking on a smaller scale. One of the most serious worries regarding both people and technology is the expanding utilisation of micro expressions. As business buildings, hotels, restaurants, and educational institutions increase their use of facial expressions for identification, the trend is moving in that direction. The purpose of the project is to build a microscale feature identification infrastructure that works all the way from end to end. A distinct second purpose of the study is to analyse the different face microexpressions in detail. Our proposed Lossless Attention Resident (model) technology is scrutinised alongside numerous other face expressions, all of which have been taught to mimic state-of-the-art techniques. However, the purpose of this work is to provide a comprehensive study of micro expressions that may be observed on faces. To gain a closer, deeper look at a person's characteristics, CNN-based algorithms focus on using characteristics at a local level, such as facial pixels. Image coordinates are put on the locations where you need to do image-to-feature extraction, such as the nose, cheekbones, lips, and eyes. model is now slightly beating the existing state-of-the-art techniques due to its capability to detect microexpressions in real time. With more annotated data training, the model might be much better.

Facial expressions provide emotional information. Microexpressions are very brief facial expressions that appear seemingly involuntarily, and that last between 0.05 and 0.2 seconds. The ideas and genuine psychological conditions of the individual who employs a customary term can be accurately transmitted. A micro expression is difficult to detect, as it is fleeting and visible for only a brief period on the face. These researchers have been interested in automated micro expression identification over the previous decade. We also wish to highlight that a significant number of specialized micro expression datasets are accessible for public use at this time. The article's purpose is to comprehensively cover the many types of facial micro expression recognition algorithms in use today. The explanation contains all of the aforementioned features and micro expression learning approaches, as well as all the micro expression data sets that are available.

Contents

[Abstract 1](#_Toc74126869)

[Table of Figures 3](#_Toc74126870)

[Chapter 1 4](#_Toc74126871)

[Introduction 4](#_Toc74126872)

[1.1 Background 6](#_Toc74126873)

[1.2 Aims & Objectives 7](#_Toc74126874)

[1.3 Research Question 7](#_Toc74126875)

[Chapter 2 8](#_Toc74126876)

[Background and Related Work 8](#_Toc74126877)

[2.1 Introduction 8](#_Toc74126878)

[2.2 Traditional Approaches 8](#_Toc74126879)

[2.2 Deep Learning Approaches 8](#_Toc74126880)

[2.3 Closing Remarks 13](#_Toc74126881)

[Chapter 3 14](#_Toc74126882)

[Technical Approach 14](#_Toc74126883)

[3.1 Attention Residual Learning 15](#_Toc74126884)

[3.2 Mask Branch Block 15](#_Toc74126885)

[3.3 Affective Dimension Recognition 16](#_Toc74126886)

[3.4 System Overview 16](#_Toc74126887)

[3.5 Image Feature Extraction 17](#_Toc74126888)

[3.6 k – NN Algorithm for Regression Modelling 17](#_Toc74126889)

[3.7 Cross-Validation: Evaluating Estimator Performance 18](#_Toc74126890)

[Chapter 4: Implementation & Results 18](#_Toc74126891)

[4.1 Dataset 18](#_Toc74126892)

[4.2 Results & Analysis 19](#_Toc74126893)

[4.2.1 Implementation 19](#_Toc74126894)

[4.2.2 Performance Metrics 20](#_Toc74126895)

[4.2.3 Comparison & Analysis with the State-of-the-Art (SOTA) Models 21](#_Toc74126896)

[4.3 Software Requirements Specifications Document 21](#_Toc74126897)

[4.3.1 Introduction 21](#_Toc74126898)

[4.3.2 Purpose 21](#_Toc74126899)

[4.3.3 Scope 21](#_Toc74126900)

[4.3.4 Project Modules 21](#_Toc74126901)

[4.3.5 Background 22](#_Toc74126902)

[4.3.6 Development Technologies 22](#_Toc74126903)

[4.3.7 Overall Description 22](#_Toc74126904)

[4.3.8 Specific Requirements 22](#_Toc74126905)

[4.4 Interfaces 23](#_Toc74126906)

[4.5 Architecture Diagram 24](#_Toc74126907)

[4.6 Use Cases 24](#_Toc74126908)

[4.6.1 Use Case Diagram 24](#_Toc74126909)

[4.7 Use Case Descriptions 24](#_Toc74126910)

[4.7.1 UC-01: Initiate the Machine Learning Model 24](#_Toc74126911)

[4.8 Sequence Diagrams 25](#_Toc74126912)

[4.8.1 Sequence Diagram 1 – Take Image and Save to Local Storage 25](#_Toc74126913)

[4.8.2 Sequence Diagram 2 – Machine Learning Model Prediction 26](#_Toc74126914)

[Chapter 5 26](#_Toc74126915)

[Conclusions & Future Work 26](#_Toc74126916)

[5.1 Conclusion 26](#_Toc74126917)

[5.2 Limitations and Future Scope 27](#_Toc74126918)

[References 28](#_Toc74126919)

# Table of Figures

[Figure 1: The number of micro expression recognition publications from 2011 to 2020 (Data 5](#_Toc74126920)

[Figure 2:Process of video annotation using the software FeelTrace. The video was displayed on the left and emotion labels were produced with the right mouse moved. 12](#_Toc74126921)

[Figure 3: Real-time emotion detection system from facial expression on embedded devices. 17](#_Toc74126922)

[Figure 4: An image with an LBP histogram that contains the image. To identify the pixel orientation and corner patterns, use LBP. When the histogram was calculated, patterns could be created. 17](#_Toc74126923)

[Figure 5: the implementation of the CNN and the definition of each of its parameters using the functions written in the code. The number of epochs and the performance of the algorithm is also shown. 19](#_Toc74126924)

[Figure 6 - Comparison of the Performance of the Model on Training data and validation data. 20](#_Toc74126925)

[Figure 7 - Accuracy Scores in Percentages for the Model 21](#_Toc74126926)

[Figure 8 - Architecture Diagram 24](#_Toc74126927)

[Figure 9 – Use Case Diagram 24](#_Toc74126928)

[Figure 10 - Use Case 1 Description 25](#_Toc74126929)

[Figure 11 - Sequence Diagram 1 25](#_Toc74126930)

[Figure 12 - Sequence Diagram 2 26](#_Toc74126931)

# Chapter 1

# Introduction

Facial expressions are one of the most important aspects of human communication, especially in commercial spaces. These expressions contribute to communicating and understanding not only the emotional take of a person but also the person’s actual ideas and thoughts, which he/she may not be willing to share. A crucial feature of facial expressions or facial emotions that makes this study so valuable that they are almost the same universally, irrespective of geography. Facial macro expressions are easily identified by humans and easily displayed. This fact results in questioning the genuineness of emotions, as those are easy to generate and hence can be used in deception. This is where micro expressions come into the picture. According to psychologists and researchers, microexpressions are facial expressions that show the true emotions of a person. Those facial expressions can also be termed as emotional leakage. Micro expressions are spontaneous and reveal a person’s true emotion in that context. This display of true emotions through micro expressions surfaces for a brief time only, for 1/15th to 1/25th of a second. This is so quick and spontaneous that it can hardly be noticed with the naked eye. The main challenge is that micro expressions are complex to identify with the naked eye, yet no one can hide them. Below Figure shows a sample of micro expressions of negative feelings. A micro expression classification would classify the sample as normal, but deep inside, the real emotion is that of a negative feeling. Hence, an in-depth analysis is needed for real feedback from customers maintaining a smooth commercial operation with a profit margin. As there is a tremendous increase in retail space, actual customer feedback on any product is a deciding factor for a product to be manufactured with the attached features. Although commercial complexes have deployed facial recognition and implemented facial expression recognitions, often, facial macro expressions do not indicate true emotions. Facial micro expressions are organic, i.e., spontaneous and will be upheld only for a very minute fraction of a second. However, they display the actual emotions and are crucial for feedback on products or situations. Hence, facial micro expressions are widely promoted for research and commercial usages. This model is even tested with a lie detector using a vision-based approach, as micro expressions properties can properly define actual emotion. Hence, a vision era on detecting real emotions extracted from micro-scale features are in growing demand in commercial, research, and defense fields(Adegun, 2016).

Many image classification architectures have been developed in the recent past and proven to provide a satisfactory result on macro expressions. However, they fail to work when accurately identifying facial micro expressions, as micro expressions are held just for a micro fraction of a second and need a depth micro-scale feature extraction for training. This work summarizes the depth in which residual attention networks perform on micro expressions and how they extract micro-scale features from a dataset(Meyer, 2015).

Insert here image, (caption: Sample micro expression from the selected dataset)

Reading people's faces is essential for better understanding their emotions and mental processes. The micro-movements of these quick, quick facial expressions are different from the type of facial expressions that can be regulated because the individual is unaware of the muscle contractions that follow the expression. This might mean that others could perceive someone's emotions even if the individual is concealing them. This statement is based on Haggard et al. (Haggard, 2015)in their study, 'Theory of a Microexpressions' which says that the concept of a micro expression was originally thought of in 1966. According to Ekman et al., (Ekman, 1972)the notion of their research is also depicted through a case study which is provided below in the same article. The correct detection of people's sentiments may be attained if micro expression patterns are developed and the relationship between such patterns and relevant emotional connections is understood. Because of this, the detection rate of lies increases dramatically. Another way to state this is to say that a patient showing a micro expression of joy during a mental diagnostic evaluation would be an excellent piece of evidence supporting this theory. A patient who has concluded the exam in a good manner is a patient who has "passed the test." Patients' concern about divulging information they want to keep private may be reflected in their facial expressions. The expression could reflect a lack of attention to critical questions or misunderstandings. While microexpressions assist in the identification of someone's genuine feelings as well as deceit, these little emotions often go unnoticed by most people. While the previously mentioned advantages of keeping a database of microexpressions are indisputable, there is also a worthwhile reason for doing so: It allows us to practise communication activities, such as negotiating a contract or reviewing a teacher.

There has been a significant growth in the number of studies on microexpressions in recent years, notably among academics. Start with the number of micro expression papers published in 2013, then draw a curve, moving ahead to demonstrate how the number of micro expression papers published in 2013 coincided with the emergence of two open-source micro expression databases in 2013. The following Micro-Expression Workshop has proved particularly efficient in circulating knowledge on face and gesture recognition among computer vision and machine learning professionals.

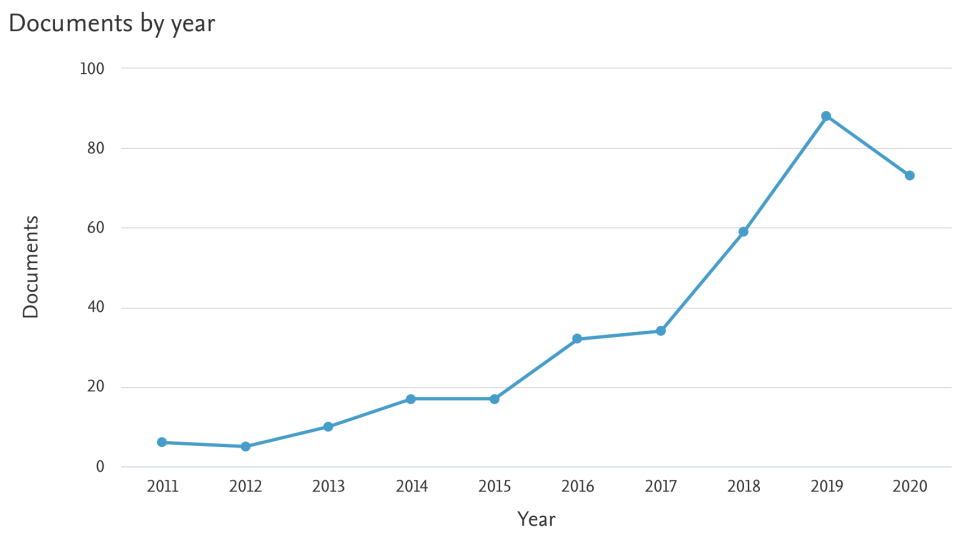


Figure 1: The number of micro expression recognition publications from 2011 to 2020 (Data

There have been a number of different but interrelated study avenues throughout the years as a result of micro expression study. The capacity to identify and differentiate different microexpressions is referred to as micro expression spotting, detection, or differentiation in all three of these phrases. As a result, it may be determined whether or not the person is experiencing an emotion based on the microexpressions. The latest impediment has attracted the greatest interest, the system known as Micro-Expression Recognition (MER). A challenge of the work is to be able to distinguish the numerous, subtle microexpressions contained inside a string of facial expressions. As a consequence of recent breakthroughs in computer vision, some academics have started using computer vision to generate automated MERs. For computer technology to be beneficial in the MRF, it offers several advantages in terms of increased efficiency, effectiveness, and material usage. It doesn't matter if your facial movements are barely detectable since computers can identify even the slightest alterations in facial expressions. To successfully run big data projects that use efficient and consistent machine learning methods, a significant amount of data is required. Computer analysis of data is cheaper than that required by humans in such circumstances, resulting in an economic advantage for computers.

This page includes all one needs to know about face meridians, explains the methods used and major data sets, examines some of the face meridian study's difficulties, and gives some possible study avenues. This study is unlike any other previous assessments of microexpressions since it studies a diverse assortment of individuals. We go in-depth, instead of just picking out microexpressions. This allows us to examine other, more "manually constructed" characteristics, as well as newer deep learning strategies. In addition, we provide academic libraries a complete database of relevant data sets, along with a step-by-step analysis of evaluation processes and data-gathering strategies. By employing innovative methodologies to tackle open research challenges, and approaches from related domains, we are able to give the best possible approach for resolving these problems.

## 1.1 Background

Micro-expressions, the short-lived and compulsory outward appearance, regularly happening in high-stake circumstances when individuals attempt to cover or veil their actual emotions, turned out to be notable since the 1960s. Micro-expressions are excessively short (1/25 to 1/2s) and unpretentious for natural eyes to see. Study shows that for miniature articulation acknowledgement undertakings, normal individuals without preparing just perform somewhat better compared to risk overall. So, PC vision and AI techniques for programmed miniature articulation examination become engaging (Zhao and Li, 2019)

## 1.2 Aims & Objectives

This research project aims to create a web-based application that combines a Machine Learning (ML) algorithm/ensemble trained on the OASIS dataset to categorize micro-expressions and a camera that can provide real-time micro-expressions as input for the ML. This web-based app will record real-time micro-expressions and classify them based on training from the OASIS dataset and create a dataset that can be used for behavioural mechanic’s study. To achieve this, the following steps are traced:

* Literature Review is done to understand the previous approaches to this problem and to understand in-depth micro-expressions.
* Designing the Front-end of the project. This involves designing mockups that will make up the GUI of the final app.
* Designing the Machine Learning Algorithm/Ensemble and training it on the OASIS dataset. The machine-learning algorithm has to be very efficient in training since there isn’t much data to train on.
* Testing the algorithm performance using Performance metrics like confusion matrices. The evaluation of the dataset created will also be an evaluation point for this project.

## 1.3 Research Question

The following research questions will guide the direction this project will take:

1. Can machine learning effectively automate the process of collecting physiological data?
2. Will the dataset created by machine learning be good enough to provide good insights for behavioural sciences?

# Chapter 2

# Background and Related Work

## 2.1 Introduction

Due to its genuineness and diversified use, research on micro expression have gained momentum in the recent years. The field of computer vision and pattern recognition has attracted many researchers to work on this topic due to its sparse usage in the commercial and psychological spaces.

The pattern recognition of the micro expressions has been mainly analyzed based on major six emotions. Micro Expression testing was first done on the database presented by Polikovsky [[4](#_bookmark22)], York Deception Test [[5](#_bookmark23)], and USF-HD [[6](#_bookmark24)]. But these datasets being insufficient were soon overtaken by SMIC [[3](#_bookmark21)], CASME II [[7](#_bookmark25)], CASME [[7](#_bookmark25)], and CAS(ME)2 [[8](#_bookmark26)]. The main reason the former did not gain popularity because the datasets were created by asking the participants to mimic or create emotions which as explained before does not generate micro expression. These were mainly artificial type of emotions and not the real ones. Hence, no fruitful results can be concluded using the former datasets. The York DDT contained very few expressions which were clearly insufficient for the research. The dataset SAMM, which stands for spontaneous actions and micro movements, consisted of 32 participants from nearly 13 different cultures. These datasets, rather than focusing on emotion recognition, focused on micro movement identification.

## 2.2 Traditional Approaches

The feature extraction technique evolved over the years due to easily available dataset and ever going research in face forensics field. Among these techniques is the LBP (Local Bi- nary Pattern) introduced by Ojala et al. (Ojala, 1996) LBP produced a remarkable result on monotonic illumination variation but limited to spatial data. So, to gain results in low intensity value Local Binary Patterns on Three orthogonal Planes (LBP-TOP) was introduced. LBP-TOP is basically the upgraded form of the first introduced LBP which now works on both temporal and spatial feature extraction simultaneously. Li et al (Li, 2014), Yan et al (Yan, 2014), Pfister et al (Pfister,2017), Guo et al, House and Meyer, and Adegun and Vadapalli implemented LBP- TOP features extraction with different facial detection and classification method for micro expression detection. The main drawback of TOP model was the computational complexity and hence efforts to improve the performance led to development of LBP-SIP or Linear Binary Pattern with Six Intersection Points and LBP-MOP (with Mean Orthogonal Planes). The drawbacks of these methods are the accuracy in extracting micro-scale level features on facial micro expressions due to limited scale features being extracted and trained.

## 2.2 Deep Learning Approaches

Deep learning-based approaches have gained attention in face forensics recently, particularly in the detection fields. A high-level representation of micro expressions can be extracted from Convolution Neural Network (CNN)-based algorithms. Patel et al. (Patel, 2017) were the first to introduce a CNN model in facial micro expressions detection. Due to fewer usable datasets, the researchers used pre-trained ImageNet weights with the Visual Geometry Group (VGG) architecture model. Mayya et al. (Mayya, 2016) introduced another method in their proposed model by combining temporal interpolation with a deep CNN (DCNN) for recognition. Later, it was fed to support vector machine (SVM) for classification and for faster performance using a Caffe library, which was used for feature extraction along with a Graphics Processing Unit (GPU) unit. The advantages of image classification using transfer learning containing feedforward convolution networks are using very deep structures and decoder functionality in auto encoder which is later taken from the feedforward mechanism. Further, several methods have been proposed for improving the discriminative ability of deep convolutions, such as VGG, Inception, and residual learning. To avoid overfitting and to exploit regularization for convergence, functions, such as stochastic depth, batch normalization, and dropout, have been initialized. However, all of the above models could not capture critical micro-scale movements in micro expressions datasets(Meyer, 2015).

Hence, deep learning-based approaches have gained potential in the face forensics in the recent past. The first framework in the field of face recognition was introduced by Jones – Viola. Their framework detected faces in an image using machine learning approach in real time. After that a large number of CNN-based face detection methods have been developed including Normalized Pixel Difference (NDP) face. Among them was one proposed by Ranjan et al. (Ranjan, 2017) which used a selective search algorithm for face detection. It was although not able to localize well with the actual face region. The deep learning mechanisms have gained lot of attraction in various detection fields. Facial recognition and micro expression field is not less in this. The high-level representation of micro expressions is extracted from convolution neural networks-based algorithms. Patel et al. (Patel, 2017) were the first to introduce CNN model in facial micro expressions detection. Due to less usable datasets, the researchers used pre-trained ImageNet weights with VGG architecture model. Mayya et al. (Mayya, 2016), in their proposed model, introduced another method by combining temporal interpolation with deep convolutional neural network (DCNN) for recognition. Later, it was fed to SVM for classification for a faster performance using Caffe library which was used for feature extraction along with GPU unit. Recent advantages on image classification using transfer learning containing feedforward convolutions networks are using very deep structures and the decoder functionality in auto encoder which is later taken from the feedforward mechanism. Several methods have further been proposed to improve the discriminative ability of deep convolutions, such as VGG, Inception, and residual learning. To avoid overfitting, functions, like stochastic depth, batch normalization, and dropout, have been initialized and to exploit regularization for convergence. However, all the above models could not capture the critical micro-scale movements of micro expression datasets.

In recent times, region proposal networks has been successfully adopted in object detection applications. In image classification, an additional region proposal stage is added before feedforward mechanism. The proposed regions contain useful information and are hence used for feature learning in the further stages. Unlike object detection, in which its region proposals rely the ground truth bounding boxes or detailed segmentation masks, unsupervised learning is usually used to generate region proposals for image classification. But, due to the heavy complexity of bringing-in segmentation masks and boundary boxes, especially for image classification tasks, this model is completely unnecessary.

Peng et al. proposed a model called dual temporal scale CNN for recognizing spontaneous micro expressions. This network works in two streams. These streams are used to process multiple frame rates of a micro expressions dataset. Each stream contains an independent shallow network to estimate overfitting. Inputs can be optical flow sequences, so that features can be produced by a shallow network. After learning, a linear SVM feature classifier is used to classify the output. The model has been proven to show decent performance compared with the conventional naive SVM and LBP methods, but it experiences the same problem with lagging in the extraction of critical micro-scale features in the model because of which its accuracies is not high enough to proceed.

Kim et al. (Kim, 2017) proposed a model consisting of CNN and long short-term memory (LSTM) to manage spatial and temporal information. Instead of using full movement inten- sity, each expression stage is learned by the network in the spatial domain. The variation in expression classes, state, and state continuity results in making features resistant to variation in illumination. LSTM helps in learning the CNN spatial information and its temporal characteristics. The LSTM approach can extract temporal information through distinct frame rate video datasets. The developed model obtained better accuracy than the old LBP techniques and subsequent variant models. Although, the imbalance in the dataset samples affected the confusion matrix results. Control gates have been used extensively in LSTM networks. In the process of feedforward training, updates are made in control gates for neurons using the helpful information. Further, the control gates have a direct influence in this process. Choi et al. (Choi, 2020) proposed LFM-based CNN-LSTM hybrid method to recognize facial micro expressions from video frames. Landmark feature maps (LFM) extracts landmarks from all parts of the face and is then fed to the CNN-LSTM hybrid architecture to compute and classify the facial micro-expressions. Although the architecture is computationally strong enough to dig deeper into the frames, the major drawback is it equally focus on all parts instead of the parts which change with respect to emotions frame-wise.

Recently, Yu et al. (Yu, 2018) introduced a deep cascaded peak pilot network to learn and determine weak expressions. Apex, i.e., peak expressions were used to supervise onset/offset non-peak expressions. The addition of backpropagation and a cascaded fine-tuned algorithm improved the overfitting problem and performance simultaneously. However, the authors tested macro expressions, which resulted in a best performance of approximately 90%.

Soft attention networks developed in recent times and soft attention modules are employing residual attention networks to develop a feedforward neural network. This approach has been adopted by the authors for this work. Recently proposed spatial transformer modules by Jaderberg et al. achieved contemporary results on almost all visual recognition tasks. An affine transformation is produced by a residual network that captures useful information available in the encoder section. Then, the input image patch is processed with the affine transformation to determine the attended region. Further, it is fed to the residual network for feature extraction.

This process is performed in an end-to-end residual attention framework that performs spatial transformations. This work has been inspired by Wang et al. regarding the design of soft attention networks with encoders and decoders as the pipeline for extracting top feature maps from both global and local information. Long et al. performed skip connections, which were used within the top and bottom features and reached state-of- the-art image segmentation results. Although this approach works satisfactorily, image classification does not require high weight structures that consume high computation power. Hence, much into local information as image segmentation, this work focuses on global and local information as far as micro-scale features from the face are included. The dataset consists of several videos, and each video is only a few seconds long, i.e., when a specific expression is seen, a video is recorded. This temporal information is considered for model training. Hence, the dataset is well refined, as micro expressions cannot be easily identified by cropping a video to the particular segment which contains the expression.

Regardless of their effectiveness in facial recognition-specific applications, deep learning and deep neural networks are different from MER in the sense that they have yet to be fully used in this field. Microexpressions have been noted to appear for the first time in the physical world thanks to the incorporation of deep learning in 2016, as seen in the figure below. Year after year, and exponentially. For the Kim et al. study, the researchers used deep learning for feature extraction, which is done using feature representations obtained from expression states and CNNs (start, start to apex, apex, apex to end, and end). The ultimate purpose of spatial learning is to help users express themselves better by maximizing the separability of their expression classes. Qualities are connected with temporal scales using an LSTM network. Sadly, their answer was worse than the custom-crafted feature-based alternatives they've had for the last decade. For these findings to be evaluated with caution, it is important to keep in mind the qualifications stated below. The limitations of the study, including a small sample size and lack of data variety, make it extremely unreliable (a single data set, CASME II, is discussed shortly). While this information points to the substantial room for growth in the field of deep learning micro expression innovation, there is some uncertainty around these findings.

In this way, Peng et al. leverage optical flow data as input, using known image processing techniques that have already been employed in other imaging applications. MER and just four layers for convolutional and pooling. In the CASME and CASME II data sets, there were four micro expression classifications: Negative, Positive, Surprise, and Other. DTSCNN beat alternative techniques including STCLQP, MDMO, and FDM in that data set.

They designed the Enhanced Long-term Recursive Coevolutionary Network (ELRCN), which was developed by Donahue et al. and used by Khor etc. for micro pressure identification. A sophisticated spatial extractor and cutting-edge temporal extractor are included in the ELRCN model. Input channels with numerous input ports raise the spatial dimension and a better understanding of both types of networks improves the time dimension with the profundity characteristics of various depth channels. Experimental study has shown that two different modules support spatial and temporal intelligence, each depending on each other for its proper functioning. Everything else was done with data already tested, but all the training sets and test sets were from standard data sets. More accurately, CASME II training was carried out, whereas SAMM tests were conducted.

A brand-new kind of cloud service is used to discern distinct emotional states from facial expressions in the Microsoft Oxford API, which was made available in 2015. You can locate the expression in images using this API, as well as the facial features that correspond to the expression. The diverse emotions, like anger, contempt, disgust, fear, happiness, neutrality, sadness, and surprise, that were detected have been assigned these emotion tags. International and cross-cultural facial expressions are thought to be expressed through facial expressions, according to research. It is both experimental and inaccurate. In combination with a stationary wavelet entropy technique, the system uses feed-forward neural networks (FFNNs) and fuzzy support vector machines (FSVMs) to attain image resolution accuracy in the neighbourhood of 96%.

A variety of methods were developed to reliably record an individual's views of four different emotional dimensions(Dai, 2015). FeelTrace is a two-dimensional tool that makes it possible to "monitor the emotional content of a stimulus, as you perceive it over time." It is based on the grid for activation and assessment. In a rectangle structure, this CAD software gives two basic emotional aspects: the duality of feeling. FeelTrace works by having the user watch stimuli, for example a television show or listen to emotional music and then move the mouse pointer to the emotional position. An indirect approach to the time measure is to draw a circle and how long the mouse takes to cross a certain distance. Labels that could be used for tests reported in this research have been created by categorizing FeelTrace. A FeelTrace exhibition with a Figure session can be found here. This color-coding technique is used to determine the emotional condition of a person. The user's overall sentiments in the (extremely active) activation dimensions are seen in the pure red color, while the user's overall sentiments in the (extremely passive) activation dimension are most positively classified as pure green.

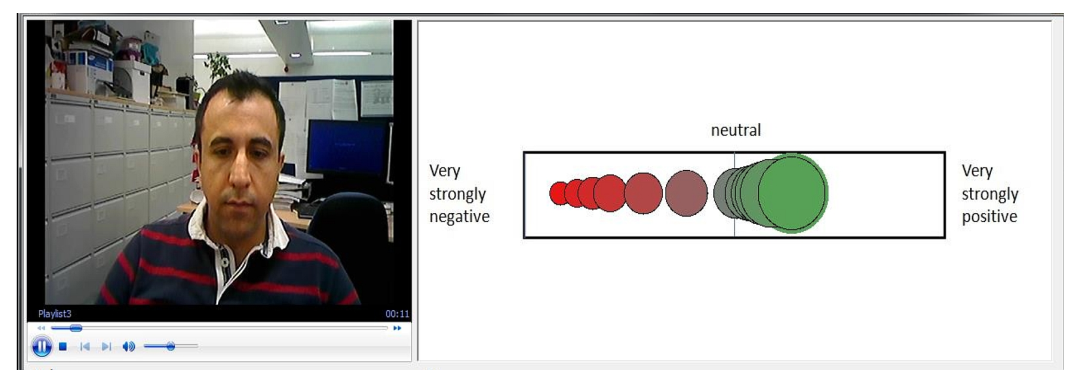


Figure 2:Process of video annotation using the software FeelTrace. The video was displayed on the left and emotion labels were produced with the right mouse moved.

The proposed local binary pattern (LBP) algorithm uses image changes to build a micropattern. The show is described through pictures of the faces of people. In the local area (the neighbourhood block) the usual LBP method performs computations of the pixels in grey scale images (which are first given in MATLAB as numerical rather than RGB intensity values and then translated into numeric grey scale intensity values). As a result of the technique the numerical values presented in a histogram alternate. The components of the alternation numbers of the histogram form a vector line. The line vectors of each block are horizontally linked to create a single functional vector to describe the whole image. There are different variations in LBP algorithms in the current research. The approach discussed in this study benefits from the same LBP properties. Uniform LBP is similar to basic LBP, but the size of feature vectors is reduced to save time and memory.

The procedure for determining the results of a numerical survey by distributing the results to a second dataset is called cross-validation. Cross validation is typically used when the aim is to predict and to determine how the predictive system functions in practise and in real time. Typically, a system has two datasets: a training data set and a test dataset used to validate the predictive models of the system. The aim of this strategy is to represent a data set to identify potential system faults. Also used in this study is the cross-validation approach, because it is an effective system performance estimator. The failure of the test data set reflects the system's performance rating correctly. This is because there are sufficient data and/or because the data is sufficiently distributed and divided into training sets. Cross-validation is thus a viable way of predicting the system's prediction accurately.

## 2.3 Closing Remarks

Computer vision methodology relies on localized (in time or space) appearance-based features to investigate microexpressions. With the conclusion of this part, we will be able to sum up and conclude that LBP-basic TOP, spatiotemporal 3DHOG, and HOOF may be compared. On the other hand, it's found in low-resolution pictures. These ideas were hinted at previously. A lesser resolution would decrease LBP performance because of the reliance on local spatial information. It is important to note that HOOF and 3DHOG are similarly reliant on temporal variation, but they are not reliant in the same way. Intermittent changes in picture resolution do not cause a significant impact on interframe information, although this is not to say that they have no effect. Due to this, the two photos appear to be in two separate locations even when viewed at different resolutions.

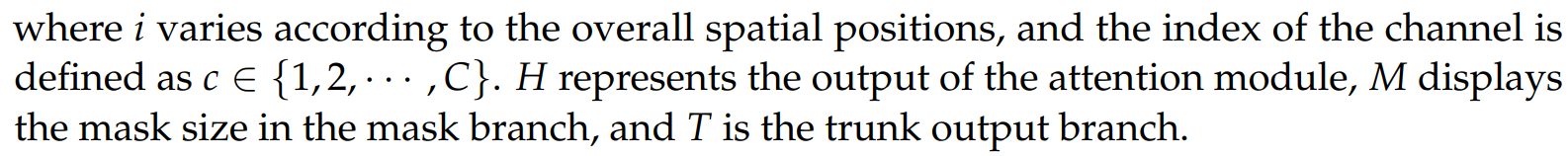
# Chapter 3

# Technical Approach

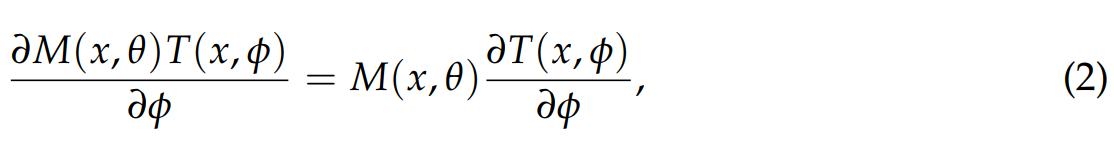
Facial micro expressions detection using Lossless Attention Residual Network (LAR- Net) is an end-to-end deep learning framework for classifying underlying facial microemo- tions. These expressions might not be captured by a human owing to their instantaneous change. Hence, the proposed model is fed with consecutive frames of the video, whereby each frame shows a very minute fraction of change. This change is the key to extracting information from the frames. model extracts this crucial information, which is available for a fraction of a second, and trains it accordingly under a specific class label. This is even applicable to detecting unconscious emotions.

model is constructed as a stack of multiple attention modules similar to the residual attention networks mentioned in Reference [[41](#_bookmark57)], whereby each branch is classified into two sub-branches, named as mask and trunk branches. Feature extraction processing is performed in the trunk branch and this block is adapted by comparison with other state of the art feature extraction processing methods. In this work, the authors have implemented two residual blocks, ResNet-56 and ResNet-92, concatenated with a custom-designed residual block built on ResNet, known as EmoResNet. The two residual networks were used as already built, and the authors froze their last layers and concatenated them with the upcoming layers, in this case, the next blocks. The outputs of each residual block are fed as inputs to the other, and the latter is fed to EmoResNet. Input *x* is given to the trunk branch which produces an output *T*(*x*). The mask branch computes a generation of masks on each image using a bottom-up and top-down approach, which mimics the feedforward and feedback attention process. Control gates of neurons in the trunk branch are the result of the outputs of the mask branch, i.e., mask outputs are bridges with control gates similar to a highway network [[41](#_bookmark57)]. Attention network outputs are represented as:





The backbone of the mask branch serves as the feature selector during the feedforward mechanism and as a gradient update filter during the backpropagation process. The gradient for input feature selection in the mask branch is defined as

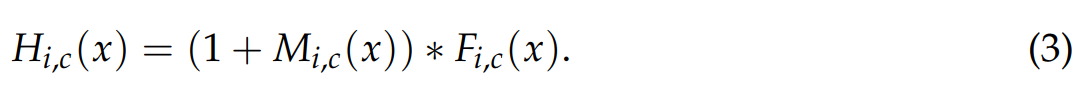


where *θ* and *φ* are the mask and trunk branch parameters, respectively. The trunk branch parameter mainly consists of a convolution filter. The advantages of having a mask branch in the attention network are that the wrong gradients are prevented from the dataset and trunk parameters are updated if noisy labels are present in the dataset [[41](#_bookmark57)]. The mask branch uses up-sampling and down-sampling computation to prevent any wrong gradients. The top-bottom approach can then identify wrong gradients and update the trunk branch accordingly. The authors have created a soft weight mask by implementing a three-network residual branch that is identical to the layer of spatial transformer. As the network clusters the features from the face, the drawbacks faced with existing state-of-the- art models, such as cluster background, complex scenes, zoomed appearance, etc., would require considerable attention, thereby making the network more complex. The main drawback of existing attention models that they can modify features only once using the backpropagation channel. The network does not have a scope for further modification if it fails either in some part or entire image. This results in false features and inaccurate results. Hence, the authors have introduced three residual blocks that alleviate the single check feature extraction. Each trunk branch in the attention model uses its mask branch for feature learning.

## 3.1 Attention Residual Learning

This section describes the feature learning methodology of attention modules. There is an apparent performance drop in naive attention networks. This apparent drop takes place owing due to the degradation of matrix values of features in the hidden layers caused by a repeated dot product of the mask range from zero to one. There is a conception that masks branch breaks identical mapping of a residual unit of the trunk branch in naive attention modules.

These problems can be eradicated if the output from the attention network can be modified as follows.



Here, *M*(*x*) varies from [0, 1], such that *M*(*x*) and *H*(*x*) approximating 0 and features *F*(*x*), respectively. This is a representation of residual learning.

The original concept of residual learning proposed through ResNet was formulated as   
*Hi*,*c*(*x*) = *x* + *Fi*,*c*(*x*), designating *Fi*,*c*(*x*) as the residual function. This proposal slightly tweaks the function in mapping the features generated by the ConvNets, inspired by Wang et al. [[41](#_bookmark57)]. This implicates the mask branch being identical in terms of feature mapping and selectors to increase good features and removes the noise from extracted features with the trunk branch. Unlike a single run feature modification, stacking attention network backs up in tweaking its weights in an incremental manner. This network extracts good properties from extracted valid features, bypasses the soft mask branch, and then weakens the mask branch’s feature extractor. This gives the network the ability to go deeper into the features, thereby consistently increasing accuracy. A similar type of work implemented by Wang et al. (Wang, 2020)surpassed the performance of other residual networks by 452 times.

## 3.2 Mask Branch Block

Moving forward with the idea of an attention mechanism as proposed by Larochelle et al. (Larochelle, 2010), the authors have instigated fast feedforward and top-down feedback steps for extracting good features and valid weights to attain a near-zero error rate. As mentioned in the previous section, the mask branch plays an important role in the feature extraction process. The feedforward block accumulates global information from the image and the top-down feedback block combines this global information with the feature maps. The max-pooling function is used in the input block in all small residual modules to increase the receptive field swiftly. When the images reach the lowest resolution while feature extraction is similar to an encoder network, the global information is drastically expanded symmetrically by the top-down feedback block to direct the input features at each pixel block level. The sigmoid activation function is then attached at the branch end, and it normalizes the output to the range of [0, 1] coming from two consecutive 1 × 1 Conv layers. Skip connections are added between fast feedforward or top-down and bottom-up layers to capture information through features from different scales. Top-down and bottom-up networks in the residual attention module gear the entire network to learn features better for micro-scale level feature learning through branch blocks.

## 3.3 Affective Dimension Recognition

It is possible to experience certain emotions (such as pleased, sad, disgusted, furious, and afraid) through a variety of different representations, as described above. Each dimension of arousal, dominance, and valence has a unique magnitude assigned to it in the 3D model. To provide an accurate assessment of the emotional state of video content, an automatic system that measures each change in video presentation and translates those measurements into values that reflect arousal, dominance, and valence is required. The projection of these values, however, should be understood to be actual frame numbers in the image sequence, rather than a regression problem.

Volunteer input is used to help with the merger system. It resulted in videos that were six times longer than those who watched the videos, and six videos were simultaneously filmed. Ten videos are shot at a frame rate of 25 frames per second, which translates to one second per frame (52,500 frames in total). The MATLAB system was used to implement the two-fold cross-validation approach, which used both 5-fold and 2-fold cross-validation. The primary goal of GTrace was to procure and construct training datasets. The characteristics of LBP can be created by extracting footage from the videos. LBP is used frequently in a variety of emotional state identification systems, and we think that is because it is widely used in various feature extraction applications. We have chosen to use the k-Nearest Neighbors algorithm because it is a little bit more difficult to build an algorithm on FPGA. A subset of the training videos' imagery was used to extract the features. A number of calculations were performed on the photos, specifically feature extraction calculations to create feature values that could be used in regression calculations. The values were then given to the k-NN for use in activation and valence computation. To create a robust training set, we needed to ensure that the entire system consistently produced consistent results every day.

## 3.4 System Overview

The diagram below illustrates the emotional state detection system's framework. The computer accepted the frames as input that were obtained from the camera. We were able to find training videos and datasets by using FeelTrace. The videos used to create LBP assets were recorded. The K-Nearest Neighbors algorithm was applied. Training images from the LBP video sequences were extracted, and these images were used for training. Two aspects of the LBP implementation were employed for the k-NN regression with dimensional emotion labels: the LBP feature set and the FPGA-attached camera video recordings.

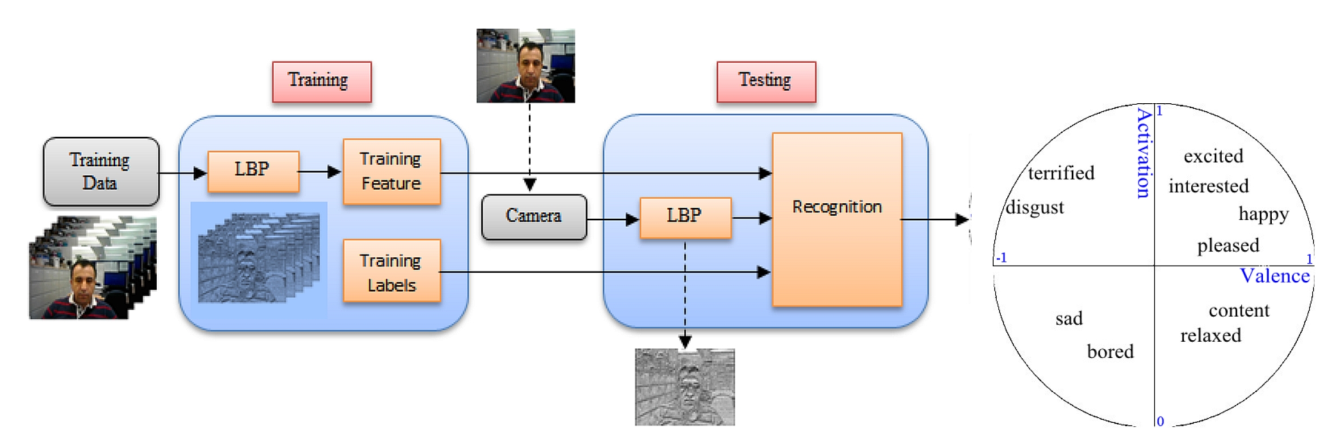


Figure 3: Real-time emotion detection system from facial expression on embedded devices.

## 3.5 Image Feature Extraction

Local block in the grayscale image is converted into an alternation counts histogram by the LBP technique. A feature vector was created for each block, and all of the block's feature vectors were concatenated into a single value to represent the complete image. The field of LBP algorithms has seen substantial progress over the last decade. This paper demonstrates the algorithm, which utilises LBP characteristics, in action. Dimensionally reduced feature vectors are used to speed up processing and reduce memory requirements. Feature extraction was implemented using the conventional LBP algorithm, also known as the histogram of an LBP image. This histogram shows that a small number of pixels used the majority of the pixel values, while the bulk of the pixel values were in a smaller area.

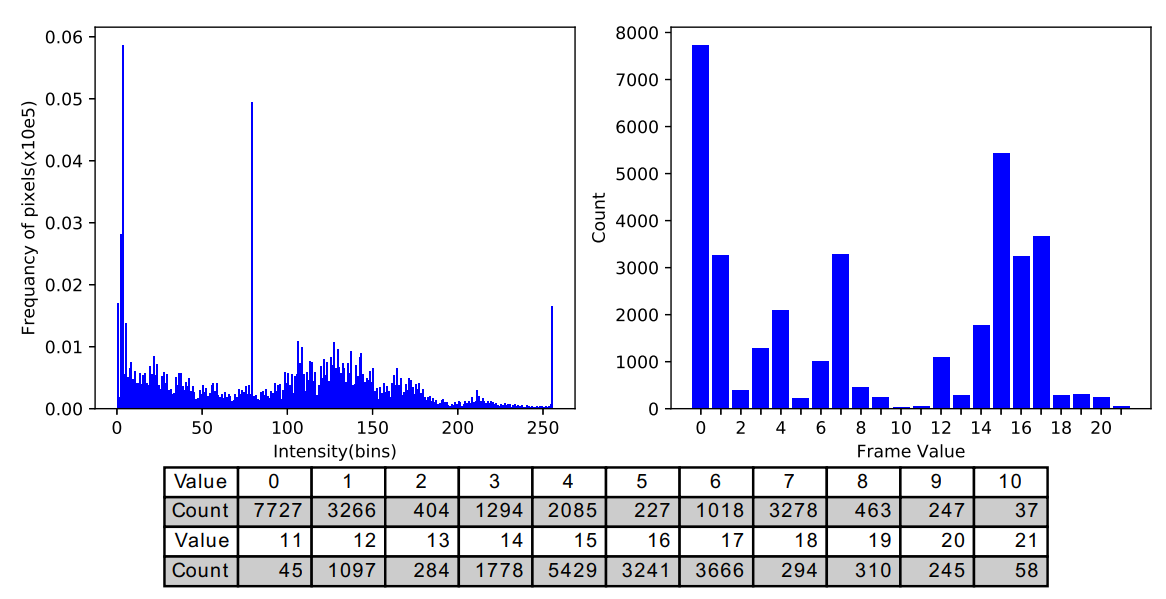


Figure 4: An image with an LBP histogram that contains the image. To identify the pixel orientation and corner patterns, use LBP. When the histogram was calculated, patterns could be created.

## 3.6 k – NN Algorithm for Regression Modelling

The k-NN technique was used to perform regression analysis. The four images which are adjacent to each other have LBP feature inputs. The k-NN classification system classifies instances according to whether or not they belong to a specific class. Usually, k is quite small. When k = 1, the test results are assigned to the nearest-neighbour class. Instead of concentrating on how much money each neighbour contributes, use a statistic that measures the cumulative effect of nearby contributions(Huang, 2016). One of the most common uses of k-Nearest Neighbour (k-NN) regression is determining the distance between a query point (an example test sample) and the list of training samples (training dataset). When it comes to regression problems, the winner is almost always assigned the role of the query, and then it is utilised as the forecast for new cases.

## 3.7 Cross-Validation: Evaluating Estimator Performance

Cross-validation is a commonly used tool for assessing the accuracy of systems on smaller datasets. Instead of analyzing the whole dataset, the organization chose to focus on breaking it down into smaller portions and removing data that would be erased in testing(Srivastava, 2014). The algorithm's last step computes an average result from all the trained models' outputs.

# Chapter 4: Implementation & Results

## 4.1 Dataset

The data are derived from the OASIS study. In response to a planned stimulus that is not deprived of the participants, the OASIS project will study a broad range of micro-expressive reactions. The training of the Machine Learning Algorithm is based on this dataset.

The dataset is a collection of images of different facial expressions that depict the following emotions:

* Anger
* Disgusted
* Fearful
* Happy
* Neutral
* Sad
* Surprised

The dataset is taken from Kaggle. The dataset already divided into a train and test set during the time of download. The number of images for each emotion in the train and test dataset are inconsistent.

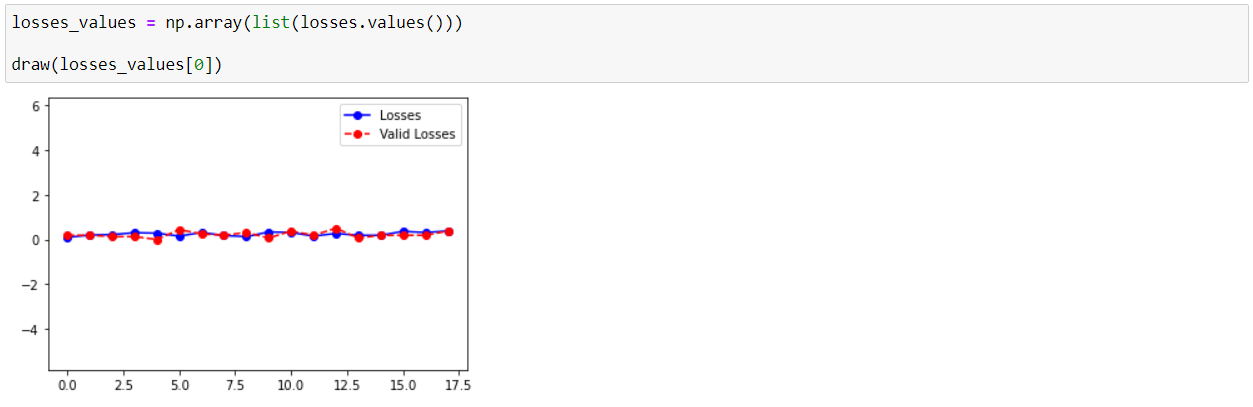
## 4.2 Results & Analysis

### 4.2.1 Implementation

The objective of this project is to create a web app that can capture the facial expression of a person based on a stimulus provided by a slideshow on the web app. The images captured will then be forwarded to the algorithm where it will classify the image and save it in the local directory.

The project will have two modules; the frontend module that will be made using flask and jinja, and the machine learning module made using a Convolutional Neural Network from PyTorch. This document will only cover the machine learning aspect of the project as I will handle the software aspect myself.

The code is divided into various functions, The **EmotionsDataset** function is the function that handles the importing and the preprocessing of the dataset for the code. All the other functions define each parameter of the Conv2D convolutional neural network from PyTorch. There aren’t many graphs in the code. The only graph in the code is below which graphically explains the performance of the Conv2D.



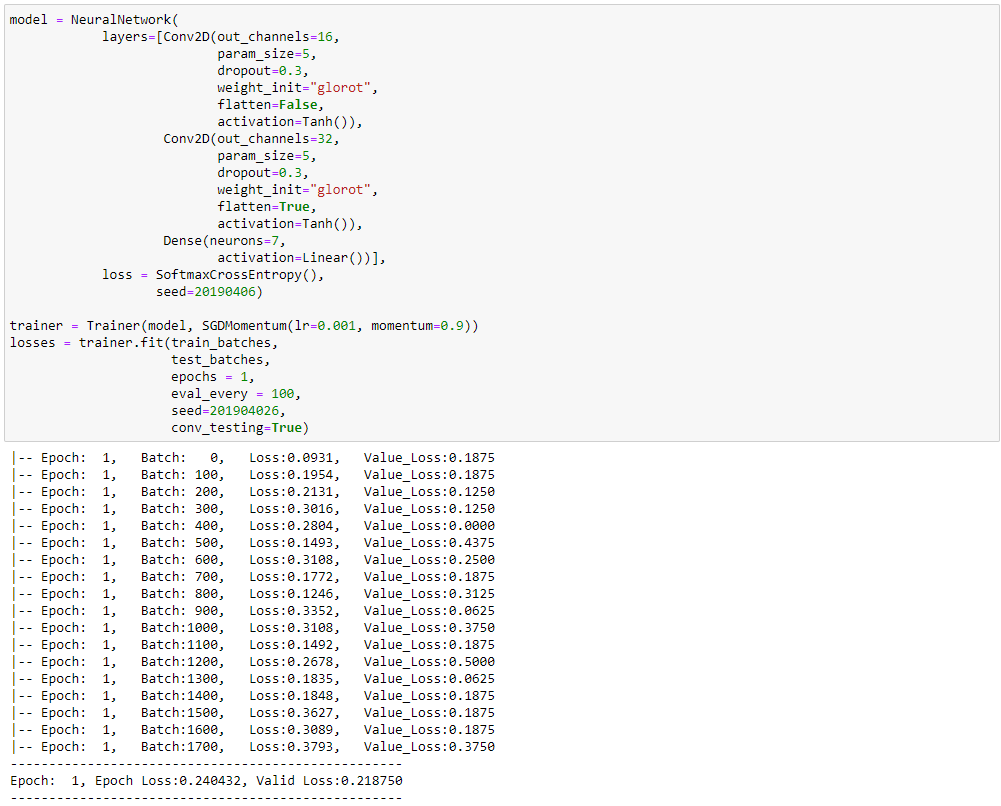


Figure 5: the implementation of the CNN and the definition of each of its parameters using the functions written in the code. The number of epochs and the performance of the algorithm is also shown.

### 4.2.2 Performance Metrics

The epochs of the code were run in such a way so as to provide four values per epoch. These four values are as follows:

* Training Accuracy
* Validation Accuracy
* Training Loss
* Validation Loss

These values are used to provide an insight into the performance of the model through graphical representation of the comparison between two values, namely Training Accuracy vs Validation Accuracy and Training loss vs Validation Loss. The graphs for both these comparisons are as follows:



Figure 6 - Comparison of the Performance of the Model on Training data and validation data.

These graphs provide a comparison of the performance of the model on both the training dataset and the validation dataset. As compared to the size of the training dataset, i.e., 22968 images across 7 output classes, the size of the validation dataset is small i.e., 1432 images across 7 output classes. While, this is the correct method of performing cross-validation for model performance testing, the difference in the size of the dataset has a significant effect on the shape of the graph. With lesser data in the validation set, the fluctuations in the graphs for both the validation accuracy and the validation loss are much more than their training dataset counterparts. The reason is that the patterns detected in the validation dataset are much more magnified in the graph due to lesser data in the validation set. This does not however affect the performance of the model in anyway.

Another performance metric was used to measure how well the model performs by providing an empirical value instead of a graphical one to get a much deeper insight and easily comprehend the model performance and accuracy The following code snippet provides the training accuracy and validation accuracy of the algorithm in percentages using an in-built function of Keras.

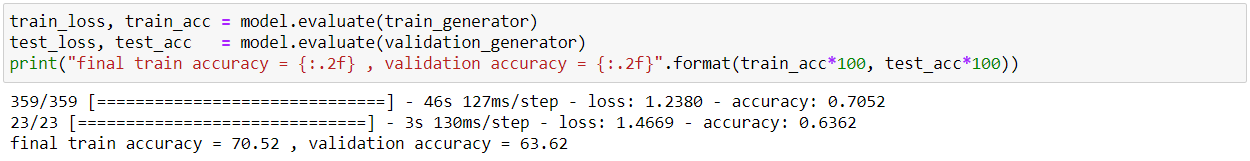


Figure 7 - Accuracy Scores in Percentages for the Model

### 4.2.3 Comparison & Analysis with the State-of-the-Art (SOTA) Models

## 4.3 Software Requirements Specifications Document

### 4.3.1 Introduction

This section sheds light on the purpose of this document to provide a basic understanding of the system for a head start. For vocabulary that might be too technical, an appendix section will be provided at the end of the document.

### 4.3.2 Purpose

With technological development at its all-time high, the environment now adapts to humans rather than the humans adapting to the environment. Experience defines everything and humans are constantly striving to create a better experience of every single activity performed. This requires an in-depth analysis of the way humans react and interact with the environment. One of the methods to study this is human involuntary reactions. Biomechanical researchers strive to collect facial data for involuntary expressions from observing the way humans react to several stimuli from the environment. This project makes it easier for the researchers to record facial data and classify it and store it in the database by automating the complete process using Machine Learning.

The purpose this document carries is to document the functionalities of the proposed system along with the design process required to build the system. The system is created from Flask and Jinja, and consists of a front-end module and the business logic integrated into it.

The document also provides the non-functional requirements for the system along with an overview of the system constraints and the major stakeholders in the system.

### 4.3.3 Scope

* The proposed system consists of a single interface with the local system storage treated as the database.
* The system is server-based and the server is located on the local machine the code is executed on.
* The primary language would be English.
* The system would be tested in a controlled environment as well as in a real environment.

### 4.3.4 Project Modules

#### Front End

Made with Jinja, which is the template engine of choice for this project paired with Flask. It uses basic HTML and inline CSS to provide a highly basic and functional UI which caters to the requirements.

#### Flask Server

Made with Flask, the server is based in the local machine and is created with Python. The back-end server includes the routes for the application and the business logic which includes data preprocessing, model initialization, prediction, webcam control and file I/O as functions that can be called once a specific route is accessed.

### 4.3.5 Background

Biomechanical researchers require a lot of facial data from involuntary expressions made by humans upon sudden interaction from environmental stimulus. Currently there are high-end systems available like iMotion that perform this on a very high level but none that are easily accessible and require less work input and are easier to use.

This project provides the ease of access to a good quality facial recognition algorithm that will perform with excellent accuracy and will provide the data required for biomechanical research.

### 4.3.6 Development Technologies

* Flask
* Jinja
* Bootstrap
* HTML
* Python

### 4.3.7 Overall Description

This app is built to provide a UI to the Machine Learning algorithm responsible for the classification of involuntary human facial expression into various emotions. The domain for this project would be Machine Learning and Facial Recognition. Although this is the current extent of the system, the system would be capable of scaling and expanding according to the requirements if necessary.

* Stakeholders: Biomechanical Researchers, Students, Data Scientists, Researchers
* Major Product Functions:
  + Data collection
  + Prediction using Machine Learning

### 4.3.8 Specific Requirements

The following are the functionalities of the system under consideration:

#### 4.3.8.1 Functional Requirements

FR-01: The system shall provide a way to receive live camera feed to record user facial expressions.

FR-02: The system shall provide a way to initialize the algorithm prediction procedure.

FR-03: The system shall provide a slideshow of images that will act as a surprise external stimulus for the user.

FR-04: The system shall take images of the user and store it in the local storage of the device that it is running on.

FR-05: The system shall save the prediction made by the model in a text file for future reference.

#### 4.3.8.2 Non-functional Requirements

Availability

NFR-01. The system shall be available to all potential stakeholders 99.03% of the time.

Usability

NFR-02. The system shall implement the principle of locality for the complete website.

NFR-03. The system shall be easy to use and remember.

NFR-04. The system shall the brief and simple language will be used on the website.

NFR-05. The system shall use meaningful images on the website.

#### 4.3.8.3 Performance

NFR-06. The system shall load the system webpages within 3 seconds on average.

NFR-07. The system shall implement server-side rendering for the interface due to the dynamic nature of the majority of the data on the website.

4.3.8.4 Design Constraints

NFR-08. The system shall be a web-based system.

## 4.4 Interfaces

The proposed system will have only one page with the following contents on it:

* Slideshow with images used as External Stimulus for the User.
* Webcam Live Feed.
* A Button for Initiating the Prediction Procedure.

## 4.5 Architecture Diagram

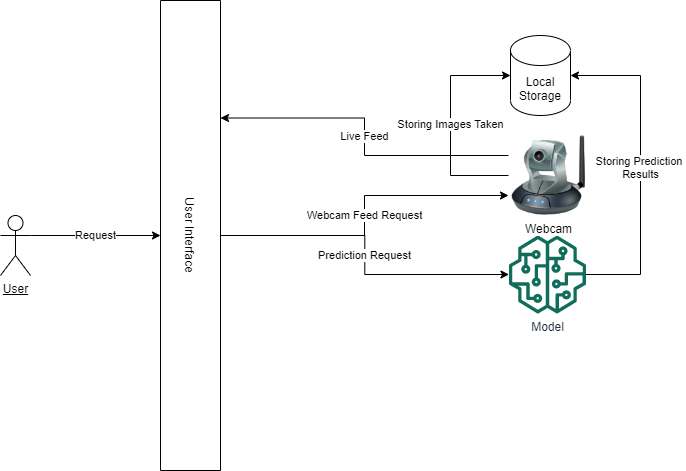


Figure 8 - Architecture Diagram

Figure 6 provides a bird’s eye view of the proposed system. There is only a single User Interface for the user. All the requests to different functions and routes of the project are accessed in this User Interface. The User Interface houses the slideshow that provides the external stimulus for the user. It also houses the live feed from the webcam and a button that acts as a trigger for the Machine Learning Model. The request for the live feed from the camera is continuously generated and handled by the User Interface and the business logic. The prediction button on the User Interface sends a trigger request to the Machine Learning model through routing and the Model starts predicting results by ingesting images taken by the camera in the local storage.

## 4.6 Use Cases

### 4.6.1 Use Case Diagram

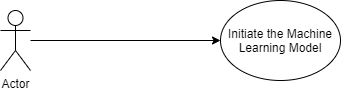


Figure 9 – Use Case Diagram

## 4.7 Use Case Descriptions

### 4.7.1 UC-01: Initiate the Machine Learning Model

|  |  |  |  |
| --- | --- | --- | --- |
| **Initiate the Machine Learning Model** | | | |
| **Actor: User** | | | |
| **Feature:** User wants to predict the emotions in the images of the user taken in the local storage. | | | |
| **Requirement ID** | | FR-02 | |
| **Use-Case ID** | | UC01 | |
| **Preconditions** | | The customer must be on the main page. | |
| **Step #** | **User Action** | | **System Response** |
| 1. | Navigate to main screen. | | System displays the main screen. |
| 5. | User waits for the Images to be Taken for each External Stimulus in the Slideshow. | |  |
| 6. | Click the ‘Prediction’ button. | | System produces the Prediction Result file in the Local Storage. |
| **Post Condition** | | | |
| The user will be able to access the prediction results. | | | |

Figure 10 - Use Case 1 Description

## 4.8 Sequence Diagrams

### 4.8.1 Sequence Diagram 1 – Take Image and Save to Local Storage

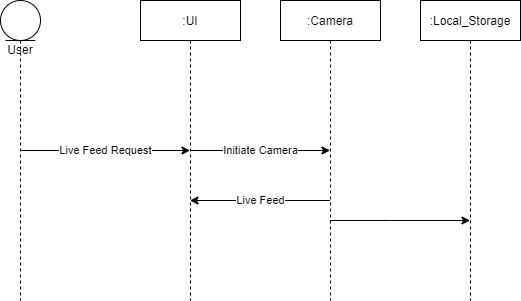


Figure 11 - Sequence Diagram 1

### 4.8.2 Sequence Diagram 2 – Machine Learning Model Prediction

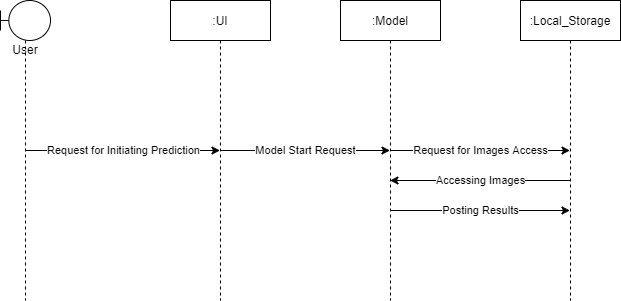


Figure 12 - Sequence Diagram 2

# Chapter 5

# Conclusions & Future Work

## 5.1 Conclusion

We can understand all facial emotional expressions as they are part of our evolutionary history and a biologically natural capacity. This is an ability we can achieve in our work and in our everyday lives. This is especially true when it comes to macros. On the other hand, most people are struggling to acknowledge micro or subtle expressions. The average accuracy rate in people in the Matsumoto & Hwang (on press) study was 48 percent; if happiness and surprise are omitted, the two expressions that are easiest to see decrease to 35 percent. And between them there are a lot of differences. Fortunately, methods were developed to help people improve their skills regardless of their original talents. There are tools if you work in an area where the understanding of facial emotional expressions can help you to be more efficient or accurate, in particular micro and subtle expressions.

However, it is only the first step to develop your ability to read nonverbal behaviour such as facial expressions. The decision on what to do with the data is the second phase of the interaction process. Being very sensitive to nonverbal behaviour, such as micro-expressions and other forms of nonverbal leakage, can have an adverse effect on interpersonal outcomes, as discussed in the literature (Blanck, Rosenthal, Snodgrass, DePaulo, & Zuckerman, 1981; Elfenbein & Ambady, 2002b; Rosenthal & DePaulo, 1979). Those who call out the feelings of others indiscriminately may be invasive, unpleasant, or dominant. It is also likely that the efficient handling of emotional information about others is an important element of the ability to interact well. Once emotions are read, skills such as when and how to intervene can all be implemented, modifying one's own behaviour and communication styles, as well as helping others.

In terms of results, experts anticipate that the structure will be accurate across six important outcomes, including "happiness," "fear," "sadness," "repression," "disgust," and "surprise." The three most valuable commercial and research measures of emotional well-being are pleasure, suffering, and fear. To extract the crucial fraction of a second of data required for micro expression feature extraction, our constructed network had to go through three stages: extraction, interpolation, and normalization. The majority of the macroscopic features have been recovered at this point. But if he had found out about it, his knowledge would have been extremely valuable, but he was unable to retrieve it because of this. From a broader perspective, each micro-level feature extraction area requires tuning. It has a large impact on feature learning if networks are trained intensively, but less intensively than required. This enabled the ResNet family of attention modules to go from the subatomic to the macroscopic scale, which included the creation of two different steps: the extraction of these steps from existing ResNet attention modules and various features, which ranged from the subatomic to the macroscopic scale. After going through these four steps, the process has transitioned to Stage 3, which is focused on larger-scale features. This is because the network was created using the "divide and rule" strategy to speed up calculations and learning. This groundbreaking approach to training and developing features was established as a result of this pioneering strategy for learning and training.

Despite hardware limitations, the model architecture claims that it can identify facial microexpressions. However, the accurate real-time results are accompanied by poor image quality when the distance between the camera and the person is greater than 10 metres. Findings from a recent study indicate that the results obtained using cameras that are more than 10 metres away are inaccurate. The definition of computing side-angle faces is the same regardless of the computing side-angle face that is being computed. Precise findings should be obtained by holding the face in the correct manner.

Face microexpressions gathered in business districts are extremely significant because the information gained could be extremely relevant in some situations. If the scene understanding function is used while the user is looking at any type of merchandise, the proposed method would be more viable. To raise the quality of life for all members of society, all commercial complexes must implement scene understanding. The inclusion of a scenario helps all users understand their needs and helps improve products to meet those needs. If you take your face neutral, your results will be more accurate.

Although it was found that model had structural flaws, which are discussed in more detail below, these issues were not considered severe enough to interrupt service.

## 5.2 Limitations and Future Scope

To assess image quality, a camera capable of 200 fps or higher and of high resolution must be used. The majority of standard cameras are used for the aforementioned purposes. However, despite this, micro-level factors, such as the tiniest of characteristics, play a major role. Any image or movie that includes noise is prone to inaccuracy. The model is only trained for six unique classes, with the seventh classification, which serves as the "other" class, serving as a stand-in for all the others. Researching on the "bottom end of the emotional spectrum," as many as 20 different emotions are postulated. Because of this, it is difficult to compile a large dataset, making it challenging to adequately instruct all of the diverse skills needed to carry out each class. Medium-scale edge devices are the best hardware for running the model remotely because of the computational complexity of the model.

Future research will investigate microexpressions as lie detectors for the military and law enforcement. Research can continue when only a miniscule amount of facial expression is lost by a person telling a falsehood. We'll need to make the model smaller in terms of computation in order to run it on smaller edge devices.

# References

Bard, K. A. (2003). Development of emotional expressions in chimpanzees (Pan troglodytes). *Annals of the New York Academy of Sciences, 1000*, 88-90.

Blanck, P. D., Rosenthal, R., Snodgrass, S. E., DePaulo, B. M., & Zuckerman, M. (1981). Sex differences in eavesdropping on nonverbal cues: Developmental changes. *Journal of Personality and Social Psychology, 41(2)*, 391-396.

Burrows, A. M., Waller, B. M., Parr, L. A., & Bonar, C. J. (2006). Muscles of facial expression in the chimpanzee (Pan troglodytes): Descriptive, comparative, and phylogenetic contexts. *Journal of Anatomy, 208*, 153-167.

Cole, P. M., Jenkins, P. A., & Shott, C. T. (1989). Spontaneous expressive control in blind and sighted children. *Child Development, 60(3)*, 683-688.

Darwin, C. (1872). *The expression of emotion in man and animals*. New York: Oxford University Press.

de Waal, F. B. M. (2003). Darwin's legacy and the study of primate visual communication. In P. Ekman, J. Campos, R. J. Davidson & F. B. M. De Waal (Eds.), *Emotions inside out: 130 years after Darwin's The Expression of Emotion in Man and Animals* (pp. 7-31). New York: New York Academy of Sciences.

Ekman, P. (1972). Universal and cultural differences in facial expression of emotion. In J. R. Cole (Ed.), *Nebraska Symposium on Motivation, 1971* (Vol. 19, pp. 207-283). Lincoln, NE: Nebraska University Press.

Ekman, P. (2003). *Emotions revealed* (2nd ed.). New York: Times Books.

Ekman, P. (Ed.). (1973). *Darwin and facial expression; a century of research in review*. New York: Academic Press.

Ekman, P., & Friesen, W. V. (1971). Constants across culture in the face and emotion. *Journal of Personality and Social Psychology, 17*, 124-129.

Ekman, P., & Friesen, W. V. (1974). Nonverbal behavior and psychopathology. In R. J. Friedman & M. Katz (Eds.), *The psychology of depression: Contemporary theory and research* (pp. 3-31). Washington, D. C.: Winston and Sons.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). *Emotion in the human face: guide-lines for research and an integration of findings*. New York: Pergamon Press.

Ekman, P., & Oster, H. (1979). Facial expressions of emotion. *Annual Review of Psychology, 30*, 527-554.

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pancultural elements in facial displays of emotion. *Science, 164*(3875), 86-88.

Elfenbein, H. A., & Ambady, N. (2002a). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128*(2), 205-235.

Elfenbein, H. A., & Ambady, N. (2002b). Predicting workplace outcomes from the ability to eavesdrop on feelings. Journal *of Applied Psychology, 87*(5), 963-971.

Friesen, W. V. (1972). *Cultural differences in facial expressions in a social situation: An experimental test of the concept of display rules*. Doctoral dissertation, University of California, San Francisco.

Galati, D., Miceli, R., & Sini, B. (2001). Judging and coding facial expression of emotions in congenitally blind children. *International  Journal of Behavioral Development, 25*(3), 268-278.

Galati, D., Sini, B., Schmidt, S., & Tinti, C. (2003). Spontaneous facial expressions in congenitally blind and sighted children aged 8-11. *Journal of Visual Impairment & Blindness, July*, 418-428.

Haggard, E. A., & Isaacs, K. S. (1966). Micro-momentary facial expressions as indicators of ego mechanisms in psychotherapy. In L. A. Gottschalk & A. H. Auerbach (Eds.), *Methods of research in psychotherapy* (pp. 154-165). New York: Appleton-Century-Crofts.

Hwang, H. S., & Matsumoto, D. (2010). *Training improves the ability to recognize subtle facial expressions of emotion*. Paper presented at the Annual Convention of the Association for Psychological Science, Boston, MA.

Izard, C. E. (1971). *The face of emotion*. East Norwalk, CT: Appleton-Century-Crofts.

Kendler, K. S., Halberstadt, L. J., Butera, F., Myers, J., Bouchard, T. J., & Ekman, P. (2008). The similarity of facial expressions in response to emotion-inducing films in reared apart twins. *Psychological Medicine, 38*(10), 1475-1483.

Matsumoto, D. (2001). Culture and Emotion. In D. Matsumoto (Ed.), *The handbook of culture and psychology* (pp. 171-194). New York: Oxford University Press.

Matsumoto, D., & Hwang, H. S. (in press). Training the ability to read microexpressions of emotion improves emotional competence on the job. *Motivation & Emotion*.

Matsumoto, D., Keltner, D., Shiota, M. N., Frank, M. G., & O'Sullivan, M. (2008). What's in a face? Facial expressions as signals of discrete emotions. In M. Lewis, J. M. Haviland & L. Feldman Barrett (Eds.), *Handbook of emotions* (pp. 211-234). New York: Guilford Press.

Matsumoto, D., & Willingham, B. (2009). Spontaneous facial expressions of emotion of congenitally and non-congenitally blind individuals. *Journal of Personality and Social Psychology, 96*(1), 1-10.

Peleg, G., Katzir, G., Peleg, O., Kamara, M., Brodsky, L., Hel-Or, H., . . . Nevo, E. (2006). Heriditary family signature of facial expression. *Proceedings of the National Academy of Sciences, 103*(43), 15921-15926.

Porter, S., & ten Brinke, L. (2008). Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions. *Psychological Science, 19*(5), 508-514.

Rinn, W. E. (1984). The neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expressions. *Psychological Bulletin, 95*, 52-77.

Rosenthal, R., & DePaulo, B. M. (1979). Sex differences in eavesdropping on nonverbal cues. *Journal of Personality and Social Psychology, 37*(2), 273-285.

Tomkins, S. S. (1962). *Affect, imagery, and consciousness* (Vol. 1: The positive affects). New York: Springer.

Tomkins, S. S. (1963). *Affect, imagery, and consciousness* (Vol. 2: The negative affects). New York: Springer.

Tomkins, S. S., & McCarter, R. (1964). What and where are the primary affects? Some evidence for a theory. *Perceptual and Motor Skills, 18*(1), 119-158.

Warren, G., Schertler, E., & Bull, P. (2009). Detecting deception from emotional and unemotional cues. *Journal of Nonverbal Behavior, 33*, 59-69.

Pei, W.; Dibeklio˘glu, H.; Baltrušaitis, T.; Tax, D.M. (2019). Attended end-to-end architecture for age estimation from facial expression videos. *IEEE Trans. Image Process. 29*, 1972–1984.

Goh, K.M.; Ng, C.H.; Lim, L.L.; Sheikh, (2020). U. Micro-expression recognition: An updated review of current trends, challenges and solutions. *Vis. Comput. 36*, 445–468.

Li, X. SMIC—Spontaneous Micro-Expression Database. Available online: https://www.oulu.fi/cmvs/node/41319 (accessed on 3 February 2020).

Polikovsky, S. (2009). Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor. In Proceedings of the IET Conference, London, UK; p. 16.

Warren, G.; Schertler, E.; Bull, P. (2009). Detecting deception from emotional and unemotional cues. *J. Nonverbal Behav. 33*, 59–69.

Shreve, M.; Godavarthy, S.; Goldgof, D.; Sarkar, S. (2011). Macro-and micro-expression spotting in long videos using spatio-temporal strain. In Proceedings of the Face and Gesture 2011, Santa Barbara, CA, USA; pp. 51–56.

Yan, W.J.; Li, X.; Wang, S.J.; Zhao, G.; Liu, Y.J.; Chen, Y.H.; Fu, X. CASME II: (2014). An improved spontaneous micro-expression database and the baseline evaluation.

Fu, X. CASME II Database. Available online: <http://fu.psych.ac.cn/CASME/casme2-en.php>   
(accessed on 3 February 2020).

Ojala, T.; Pietikäinen, M.; Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern Recognit. 29*, 51–59.

Li, X.; Hong, X.; Moilanen, A.; Huang, X.; Pfister, T.; Zhao, G.; Pietikäinen, M. (2017). Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods. *IEEE Trans. Affect. Comput. 9*, 563–577.

Pfister, T.; Li, X.; Zhao, G.; Pietikäinen, M. (2011). Recognising spontaneous facial micro-expressions. In Proceedings of the International Conference on Computer Vision, Barcelona, Spain; pp. 1449–1456.

Guo, Y.; Tian, Y.; Gao, X.; Zhang, X. (2014). Micro-expression recognition based on local binary patterns from three orthogonal planes and nearest neighbour method. In Proceedings of the International Joint Conference on Neural Networks, Beijing, China; pp. 3473–3479.

Meyer, R.; House, C. (2015). *Preprocessing and Descriptor Features for Facial Micro-Expression Recognition*; Technical Report; Stanford University: Stanford, CA, USA.

Adegun, I.P.; Vadapalli, H.B. (2016). Automatic recognition of micro-expressions using local binary patterns on three orthogonal planes and extreme learning machine. In Proceedings of the Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech), Stellenbosch, South Africa; pp. 1–5.

Simonyan, K.; Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv*, arXiv:1409.1556.

Mayya, V.; Pai, R.M.; Pai, M.M. (2016). Combining temporal interpolation and DCNN for faster recognition of micro-expressions in video sequences. In Proceedings of the International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India; pp. 699–703.

Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R.; Guadarrama, S.; Darrell, T. Caffe: (2014). Convolutional architecture for fast feature embedding. In Proceedings of the 22nd ACM international conference on Multimedia, Mountain View, CA, USA; pp. 675–678.

He, K.; Zhang, X.; Ren, S.; Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA; pp. 770–778.

Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. (2016). Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA; pp. 1–9.

Huang, G.; Sun, Y.; Liu, Z.; Sedra, D.; Weinberger, K.Q. (2016). Deep networks with stochastic depth. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; Springer: Berlin/Heidelberg, Germany; pp. 646–661.

Ioffe, S.; Szegedy, C. Batch Normalization: (2015). Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the 32nd International Conference on International Conference on Machine Learning—ICML’15, Lille, France; Volume 37, pp. 448–456.

Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: (2014). A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res*, *15*, 1929–1958.

Viola, P.; Jones, (2001). M. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition—CVPR 2001, Kauai, HI, USA; Volume 1, p. I.

Ranjan, R.; Patel, V.M.; Chellappa, R. Hyperface: (2017). A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *IEEE Trans. Pattern Anal. Mach. Intell*, *41*, 121–135.

Cao, C.; Liu, X.; Yang, Y.; Yu, Y.; Wang, J.; Wang, Z.; Huang, Y.; Wang, L.; Huang, C.; Xu, W.; et al. Look and think twice: (2015). Capturing top-down visual attention with feedback convolutional neural networks. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile; pp. 2956–2964.

Stollenga, M.F.; Masci, J.; Gomez, F.; Schmidhuber, J. (2014). Deep networks with internal selective attention through feedback connections. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada; pp. 3545–3553.

Dai, J.; He, K.; Sun, J. (2015). Convolutional feature masking for joint object and stuff segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA; pp. 3992–4000.

Erhan, D.; Szegedy, C.; Toshev, A.; Anguelov, D. (2014). Scalable object detection using deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA; pp. 2147–2154.

Hariharan, B.; Arbeláez, P.; Girshick, R.; Malik, J. (2014). Simultaneous detection and segmentation. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland; Springer: Berlin/Heidelberg, Germany; pp. 297–312.

Yang, S.; Luo, P.; Loy, C.C.; Tang, X. From facial parts responses to face detection: (2015). A deep learning approach. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile; pp. 3676–3684.

Shrivastava, A.; Gupta, A. (2016). Contextual priming and feedback for faster R-CNN. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; Springer: Berlin/Heidelberg, Germany; pp. 330–348.

Xiao, T.; Xu, Y.; Yang, K.; Zhang, J.; Peng, Y.; Zhang, Z. (2015). The application of two-level attention models in deep convolutional neural network for fine-grained image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA; pp. 842–850.

Peng, M.; Wang, C.; Chen, T.; Liu, G.; Fu, X. (2017). Dual temporal scale convolutional neural network for micro-expression recognition. *Front. Psychol.*  *8*, 1745.

Kim, D.H.; Baddar, W.J.; Jang, J.; Ro, Y.M. (2017). Multi-objective based spatio-temporal feature representation learning robust to expression intensity variations for facial expression recognition. *IEEE Trans. Affect. Comput.*  *10*, 223–236.

Choi, D.Y.; Song, B.C. (2020). Facial Micro-Expression Recognition Using Two-Dimensional Landmark Feature Maps. *IEEE Access*  *8*, 121549–121563.

Yu, Z.; Liu, Q.; Liu, G. (2018). Deeper cascaded peak-piloted network for weak expression recognition. *Vis. Comput.*  *34*, 1691–1699.

Li, Y.; Zeng, J.; Shan, S.; Chen, X. (2020). Occlusion aware facial expression recognition using cnn with attention mechanism. *IEEE Trans. Image Process.* *28*, 2439–2450.

Wang, K.; Peng, X.; Yang, J.; Meng, D.; Qiao, Y. (2020). Region attention networks for pose and occlusion robust facial expression recognition. *IEEE Trans. Image Process.* *29*, 4057–4069.

Chen, L.C.; Yang, Y.; Wang, J.; Xu, W.; Yuille, A.L. (2016). Attention to scale: Scale-aware semantic image segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA; pp. 3640–3649.

Jaderberg, M.; Simonyan, K.; Zisserman, A.; Kavukcuoglu, K. (2015). Spatial transformer networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada; pp. 2017–2025.

Wang, F.; Jiang, M.; Qian, C.; Yang, S.; Li, C.; Zhang, H.; Wang, X.; Tang, X. (2017). Residual attention network for image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA; pp. 3156–3164.

Long, J.; Shelhamer, E.; Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA; pp. 3431–3440.

Larochelle, H.; Hinton, G.E. (2010). Learning to combine foveal glimpses with a third-order Boltzmann machine. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada; pp. 1243–1251.

Godavarthy, S. Microexpression Spotting in Video Using Optical Strain. Available online: https://scholarcommons.usf.edu/etd/1642/  
(accessed on 3 February 2020).

Hashmi, M.F.; Ashish, B.K.K.; Keskar, A.G.; Bokde, N.D.; Yoon, J.H.; Geem, Z.W. (2020). An Exploratory Analysis on Visual Counterfeits using Conv-LSTM Hybrid Architecture. *IEEE Access 8*, 101293–101308.

Hashmi, M.F.; Ashish, B.K.K.; Keskar, A.G.; Bokde, N.D.; Geem, Z.W. FashionFit: (2020). Analysis of Mapping 3D Pose and Neural Body Fit for Custom Virtual Try-On. *IEEE Access**8*, 91603–91615.

Schroff, F.; Kalenichenko, D.; Philbin, J. Facenet: (2015). A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA; pp. 815–823.