Expression Classification in Microexpressions

Sarah Houlton
Computer Science Department
Colorado State University
Fort Collins, Colorado, USA
shoulton@rams.colostate.edu

ABSTRACT

Microexpressions are very short facial expressions that are thought to portray the internal thoughts of the person displaying the microexpressions. It has been proposed that detection of these microexpressions could reveal when someone is lying and the emotions they are really feeling. Microexpressions also differ from normal expressions as they tend to be more subdued than regular expressions. The changes in the face are much smaller, which makes this an interesting classification problem. Deep convolutional neural networks have been shown to be successful in classifying regular emotions and so were chosen to classify microexpressions. In this paper, a deep convolutional neural network is used to classify microexpressions taken from the CASME II database [1].

CCS CONCEPTS

• Computing Methodologies • Machine Learning • Machine Learning Approaches

KEYWORDS

Artificial intelligence, Machine Learning, Classification, Convolutional Neural Networks, Emotion Classification, Microexpressions

1 Introduction

There is a common saying that up to 93% of human communication is nonverbal. This 93% is then made up of things like body language, eye contact, and expression. Whether or not this statistic is correct, it brings up a compelling point. Nonverbal communication is a vital part of interpersonal connection. Throughout the last year many of us have likely come to realize how truly important nonverbal communication is. From professors teaching to black screens to speaking to a friend via text because you cannot see them, we have all been deprived of these nonverbal cues and communication has suffered.

In the 1960's Ernest Haggard and Kenneth Isaacs proposed a form of nonverbal communication past facial expressions, body language, tone of voice, and eye contact. They deemed this form of communication microexpressions [2]. Microexpressions are small changes in expression that happen incredibly quickly, lasting only one-eight to one-fifth of a second. This is the equivalent of three to five frames of a video. The microexpressions themselves differ greatly from the expression prior to the observed microexpression, e.g., a smile to a grimace.

Haggard and Isaacs first isolated these microexpressions by reviewing hours of psychotherapy sessions [2]. They were investigating nonverbal communication between therapist and subject and found that when they watched the film frame by frame, they occasionally saw these fleeting shifts in expression. They came to the conclusion that these may be the inner emotions showing on the surface for a split second. For example, one subject was smiling and happy, talking about how he was doing better with the therapist when he displayed microexpressions of intense anguish. They later found out that the subject was experiencing suicidal ideation and thus was feeling this internal anguish which may have been projected through microexpressions [2]. Whether or not microexpressions are a reflection of inner thoughts is still debated, but this is the theory proposed by Haggard and Isaacs.

In the decades since the introduction of microexpressions, it has been proposed that there may be a use for isolating and detecting these microexpressions. Because they are thought to reflect inner, hidden thoughts, there is a possibility that microexpressions can reveal ill intent or an intent to deceive. Based on exactly this theory, some airports have even introduced special TSA agents called behavior detection officers [3]. Behavior detection officers are meant to be trained to identify microexpressions and screen individuals that they believe display ill intent through microexpressions. In reality, the chance of a human correctly detecting ill intent through microexpressions and being able to predict that a person is dangerous through these microexpressions is about the same as flipping a coin or making a random guess [2]. In fact, in studies where people were trained to detect microexpressions, they were not found to perform better than blindly guessing even after the training [4]. Jordan et. al conducted

an experiment to find the efficacy of this training [4]. In the experiment, people were instructed to watch a set of videos and try to detect microexpressions or intent to deceive, then they received training for microexpression detection, and their performance was compared before and after training [4]. The experimental group were trained on METT, the most common tool for micro-expression training [4]. The control group were either not trained at all or given a fake training that was not intended to improve detection of micro-expressions. The experimental group did not perform any better, once again performing only slightly worse than simple chance [4].

But, where humans are not very good at detecting these micro-expressions, computers may excel. In the field of artificial intelligence, classification is a commonly researched area. Of particular interest for this paper is the classification of human emotions. There have been many successful models for classifying human emotion. This paper takes the approaches of these well performing models and uses them to classify subtler microexpressions.

2 Motivation

Expression detection is a popular field that has a plethora of applications. In game development, emotion classification could be applied to a game similar to the popular game Phasmophobia. In Phasmophobia, groups of players work together to identify the type of ghost haunting a house. To do this, they utilize different ghost-hunting tools to find evidence. There is also an in-game voice feature. The game uses audio processing to listen in to that in-game voice feature, and it has a direct effect on the game. For instance, if a player says “fear”, “run”, “scared”, or the name of the ghost, the chance that the ghost becomes hostile increases. Using emotion detection, a game like this could access a web camera and monitor the face of the player. As the player expresses a certain emotion, thing in the game may change. For instance, if the player expresses a fear emotion, the ghost may become more hostile. Microexpression detection could also be added into a game as a sort of rudimentary lie detector.

On a more serious note, microexpression detection could be implemented to protect public safety. Where behavior detection agents failed, artificial intelligence detection algorithms may be able to accurately detect ill intent and select individuals for further screening. This would eliminate some of the inherent biases that humans have and better protect the public. In large public spaces

TABLE VI THE FOLLOWING ARE THE PARTICIPANTS'S RATINGS ON THE 17 VIDEO EPISODES, THE MAIN EMOTION FOR EACH EPISODE, THE NUMBER OF PARTICIPANTS WHO FEEL SUCH AN EMOTION AND THEIR CORRESPONDING MEAN SCORE (FROM 0 TO 6).

Episode NO.	Main emotions	Rate of selection	Mean score
1	amusement	0.69	3.27
2	amusement	0.71	3.6
3	amusement	0.7	3.14
4	amusement	0.64	4.43
5	disgust	0.81	4.15
6	disgust	0.69	4.18
7	disgust	0.78	4
8	disgust	0.81	3.23
9	fear	0.63	2.9
10	fear	0.67	2.83
11	/	/	/
12	disgust (fear)	0.60(0.33)	3.78(0.28)
13	sadness	0.71	4.08
14	sadness	1	5
15	anger (sadness)	0.69(0.61)	4.33(0.62)
16	anger	0.75	4.67
17	anger	0.94	4.93

Figure 1: Table VI from the “CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces” assigning an emotion to every video shown to participants and how strongly the emotion was felt by participants [1]

Figure 2: Subject 01 exhibiting a microexpression of disgust, original on left, after padding in the middle, after normalization on right [1].

like airports, bus stations, or concerts, this could help prevent terrorist attacks.

In police interrogations, having an AI microexpression detection system could help detectives to know when someone may be lying or feeling something that they are not outright saying. Implemented correctly, something like this could lead to quicker investigations. In investigations where time is limited, such as abductions, the ability to detect intent to deceive could help indicate if the person being interviewed had information beyond what they had disclosed. These are just a few possible applications of microexpression detection, there are many more places where the ability to detect ill intent or intent to deceive would be useful.

3 Dataset

The dataset for this research is the Chinese Academy of Sciences Micro-expression II dataset [1][6]. Access to this dataset was applied for and granted for use in this term paper by Professor Xiaolan Fu's lab at the Chinese Academy of Sciences [5]. For this project, 254 individual microexpression frames from this dataset were used.

This dataset was chosen because it contains truly spontaneous microexpressions [1][6]. These microexpressions were collected by recording individuals as they watched emotionally charged videos. They were told the purpose of the experiment was to see how well they could control their emotions and keep a neutral face, and that for every facial expression they displayed a small amount was deducted from their monetary reward [1][6]. After viewing the video, participants watched back the footage to clarify any facial movements that were not microexpressions, such as scratching the nose or coughing, so they could be removed from the data. They then also rated each video by the main emotion it elicited and how strongly they felt the emotion. Included in figure 1 is a table taken from section III of the "CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces" paper that shows for each video the main emotion the video was meant to elicit, the rate at which participants correctly identified the emotion, and the mean strength of emotion felt by participants on a scale of zero to six [1][6].

This approach was taken for the generation of the dataset because microexpressions are meant to be spontaneous and uncontrollable, so by asking participants to keep a neutral face throughout the emotionally charged videos they captured natural microexpressions [1][6]. This differs from other datasets because many other datasets contain manufactured microexpressions, which are not as subtle as the natural microexpressions. For example, the USF-HD dataset contains both macro and micro expressions displayed by participants upon request. They were shown examples of microexpressions and asked to mimic them, which is a non-spontaneous microexpression [1]. While manufactured microexpressions can be useful for a proof of concept, in any real-world application spontaneous facial expressions are the ideal as they most closely mimic a real-world scenario.

Once the videos were taken, two trained coders looked through the footage to locate the microexpressions. Once the microexpressions were identified, the coders each individually selected onset, offset, and apex frames for each microexpression. If the coders disagreed, the frames were arbitrated and if a conclusion was not reached, the average of their two frames was recorded. They had an agreement rate of 73% on these frames [1][6]. They then assigned an emotion to the microexpression based off of the self-reported emotions from the participants. If an emotion could not be conclusively identified, the emotion was recorded as "others". In the dataset used for this project, there were 99 non-conclusively identified emotions out of 254 samples, making up about 40% of the dataset.

For this paper, only the apex frame was used as it should be the frame which displays the emotion most clearly. The dataset also offers full video of the sessions for microexpression recognition in video, just the video frames between onset and offset, or a cropped version of these frames that focuses on just the face. This paper uses the cropped version of the selected frames to avoid as much noise from the surroundings as possible. These cropped frames are then padded with a black border to all be the same dimensions of 400x400. This was done because the cropped images are of varying sizes and the neural net expects consistent sized input. The images are also normalized before input to avoid any interference from lighting differences. Examples of the types of images in the dataset are shown in figure 2. The expression of disgust is subtle, but there is indication in the eyebrows and around the mouth that categorizes disgust.

4 Related Works

The field of expression and microexpression classifiers is large, and ever-growing. This research leverages many of these publications in order to investigate the use of a deep convolutional neural network for classifying microexpressions.

A useful real-world application of expression detection is presented in "Student Emotion Recognition Using Computer Vision as an Assistive Technology for Education" from the 2019 Information Science and Applications conference [7]. It proposes an assistive technology for teachers who may be less able to empathize with students. The ability to empathize with students creates an environment which is open and accepting, and thus more conducive to learning [7]. Van der Haar et. al. created a prototype of a computer vision based emotion recognition system which was able to generate an emotion report in near-real time [7]. This would then be used by the teacher to supplement teaching. While this deals with expressions and computer vision rather than microexpressions, emotion recognition and microexpression classification are inherently closely linked and this research demonstrates an area where emotion recognition is being implemented in the real world.

In 2011, Qi Wu, Xunbing Shen, and Xiaolan Fu presented "The Machine Knows What You Are Hiding: An Automatic Micro-expression Recognition System" at the International Conference on Affective Computing and Intelligent Interaction [8]. In this research, the team was interested in detection of microexpressions and classification of microexpressions [8]. To select microexpression frames, the model analyzes each frame of video, searching for a threshold of action units. These action units are indications of different facial movements. If these facial movements only lasted a small time, this was recognized as a microexpression and would then be classified as one of seven emotions. These emotions are the six base human emotions that are used in emotion classification; namely happiness, sadness, disgust, anger, fear, and surprise, with an added neutral emotion for when the expression does not clearly fit any type. These expression frames were then classified using a combination of Gentleboost and an SVM, which they called GentleSVM. This gave an accuracy of 95.8% on Caucasian faces and 75% on Asian faces [8]. This shows how effectively artificial intelligence can be used for this type of work. My work differs from Wu et. al. because I am not attempting to extract microexpression frames from video, but rather just classifying static frames. By only classifying the static apex frames I do lose the context that Wu et al. get from the video, and in fact adding this context back in is a possible extension of this project. It also differs in that I have used a deep convolutional neural network to classify the images from the CASME II dataset.

The research I most heavily leveraged for this paper was "Extended deep neural network for facial emotion recognition" by D. K. Jain, P. Shamsolmoali, and P. Sehdev which discusses using a deep convolutional neural network to classify expressions [9]. This paper proposed using a deep convolutional neural network for the face detector and softmax for the classification of facial expressions. The network has six convolutional layers, and two deep residual learning layers. This network performed at 95% accuracy [9]. The success of a deep convolutional neural network on expressions is what inspired its use here. Given its success on regular expressions, I wanted to investigate its performance on microexpressions, given their subtleties.

There is much more published work worth reviewing on classifying expressions and microexpressions, but these were the most influential on this research [10][11][12][13].

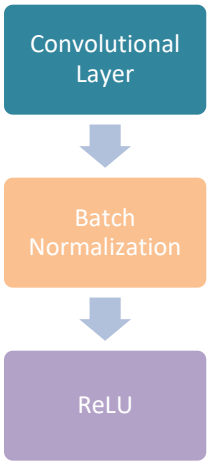

Figure 3: A representation of one convolutional module in the network.

5 Deep Convolutional Neural Network

Deep convolutional neural networks are a common choice for image classification because they extract features very well. A deep convolutional neural network is then ideal for a problem such as microexpression classification because it is able to detect fine-grained features of an image. This helps to classify the subtle changes that define microexpressions.

The convolutional neural network used here is made up of convolutional modules followed by pooling layers. The modules consist of a convolutional layer, followed by a batch normalization layer, followed by a ReLU layer. This structure is shown in figure 3. The final network is made up of 14 of these modules, with 3 intermittent pooling layers and a final fully connected layer, as is shown in figure 4. This gives a final 47 layers in the network.

6 Results

Initially, testing was run with a naïve 70/30 training and testing split. The images were shuffled to avoid grouping all images of a subject into either testing or training and 70% of the indices were moved into training and the remaining indices were moved to testing. The sets of images were then compared to be sure there was no overlap in the testing and training sets.

Because the dataset is small and somewhat repetitive, a naïve 70/30 training split was not effective. Figure 5 shows the results of using the naïve split. The model performed well on the training data, achieving about 95% accuracy after 50 epochs. The performance on the testing data was highly variable, however. The model either achieved 100% accuracy or 0% accuracy. This is indicative of a few things, but here I believe is a result of a small dataset and the subtlety of the data. Because 99 images were classified as others and 30% of the dataset is equivalent to 71 images, it is entirely

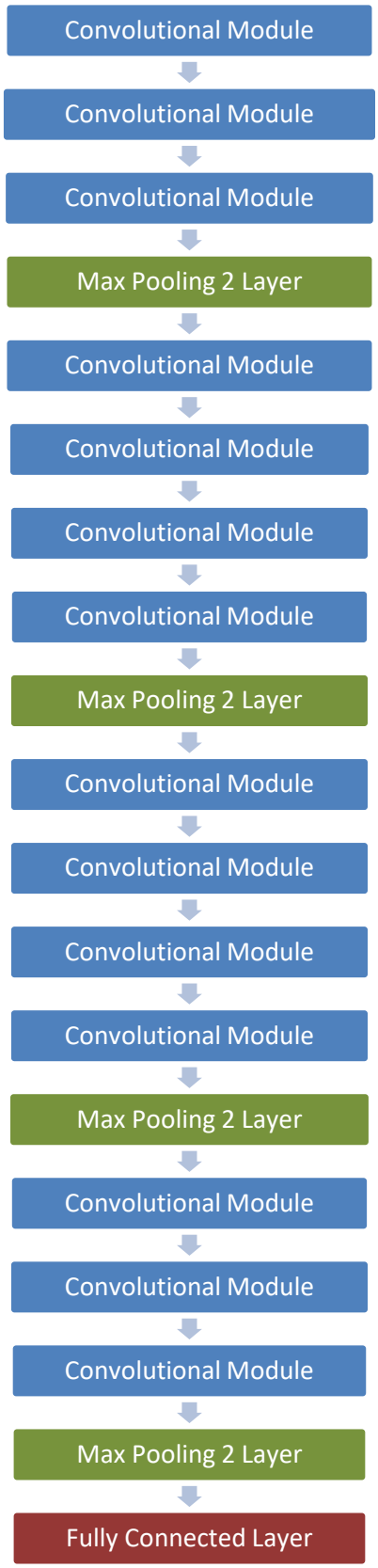

Figure 4: A representation of the full deep convolutional neural network.

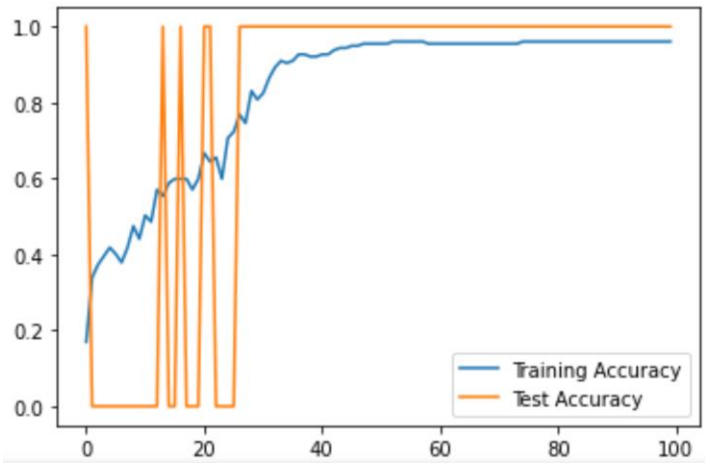

Figure 5: Performance of the model with naïve 70/30 split.

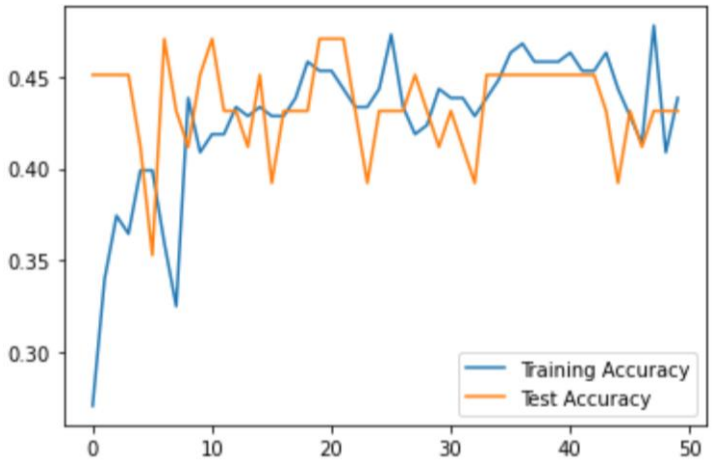

Figure 6: Performance of the model with 50 epochs and 5-fold cross validation.

possible the testing set was only made up of non-conclusive images and just got lucky in classifying these. It is also possible that some of the testing sets included every image of a particular subject and thus the model was never trained on this image. Nevertheless, I implemented k-cross fold validation in an attempt to improve the model.

For this research I ran the model with 50 epochs and 5-fold cross validation. I chose these parameters because in previous testing the accuracy seemed to plateau around 50 epochs and 5-fold cross validation fit within the time constraints I had. If I were to continue this project further, I would add more k-fold validations to see if it would improve performance. With these parameters, the performance was not good. At its best, the model was only able to achieve 45% test accuracy, as is shown in figure 6. This is similar to human accuracy according to Jordan et. al [4]. Figure 7 show shows the training loss over 50 epochs. While this is a disappointment, it was not entirely a surprise. I believe that a number of factors contributed to the failure of this model.

The first factor is the variability in microexpressions. While they are all based on the same basic emotions, the expressions are subtle and present differently on different faces. Disgust may appear in the slightly furrowed brows of one subject, while it appears as flared nostrils on another. Since the CASME dataset only includes 26 subjects, it is possible the variability of expression was too great for the model to overcome. With a larger number of subjects, I believe there is a possibility of better generalization of microexpressions.

The second factor is the split of the dataset. With more folds in the k-fold cross validation, the model may improve its performance. Time and processing power constraints restricted my experimentation with these but with future research I believe this would help to improve the model. This would further randomize the testing and training sets and hopefully get a more representative sample to increase generalization of the model.

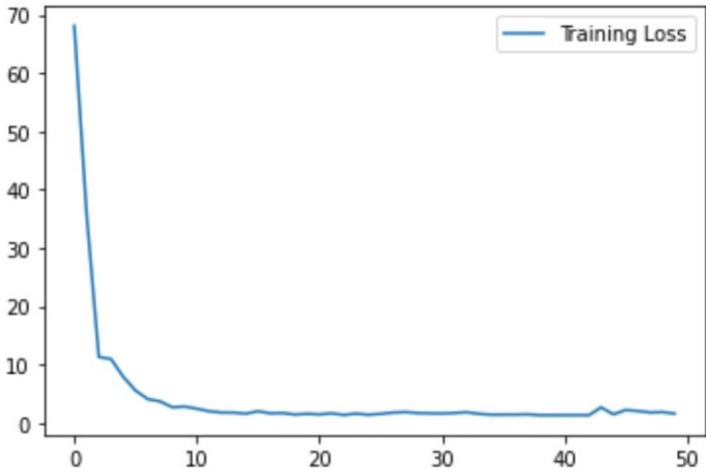

Figure 7: Training loss of the model over 50 epochs

7 Future Work

Despite the results of this research project, the field of microexpression recognition is still one that interests me. I believe there is a lot of room for future work on using deep convolutional neural networks for microexpression recognition.

The first extension of this project would be the addition of another dataset. It would be interesting to investigate whether the addition of another spontaneous microexpression dataset would increase the performance of the model by allowing for further generalization of expressions. It would also be interesting to add a more diverse dataset as the CASME dataset is made up of only Asian faces. Adding a more diverse dataset would avoid the problems encountered by Wu et. al. where the model is more accurate on one ethnicity than another [8]. This would make it more useable in real-world situations as well.

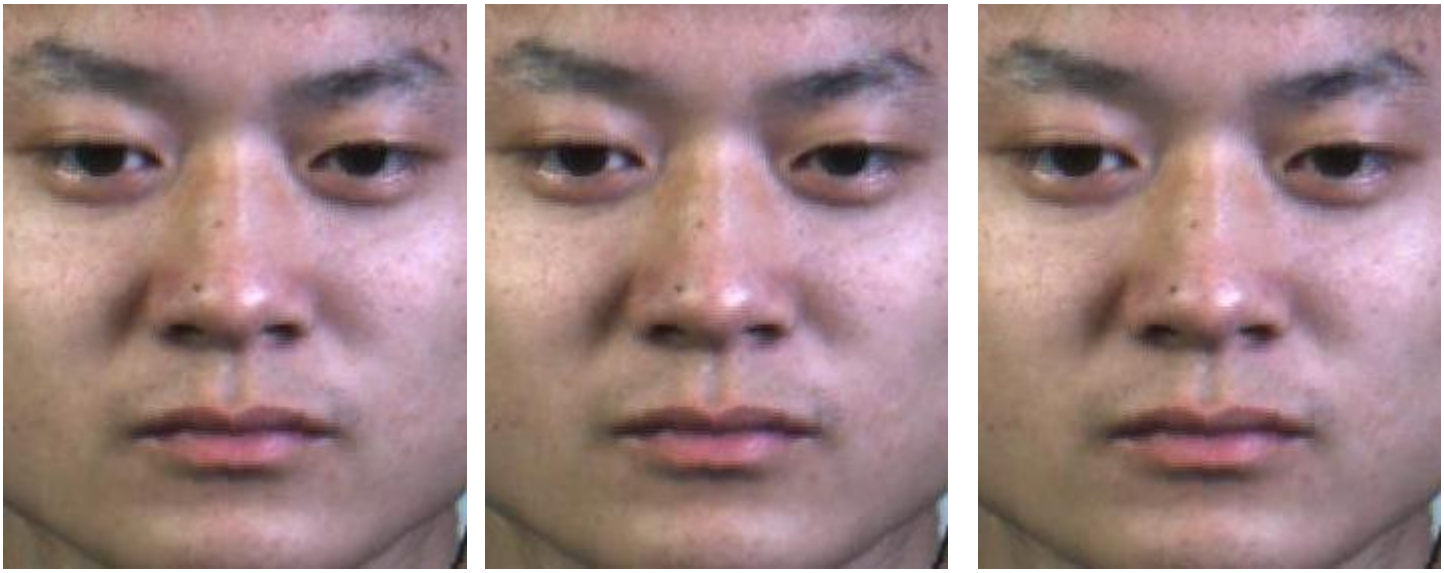

Figure 8: From left, onset, apex, and offset frames for a microexpression of happiness in subject 1 [1].

Additionally, it would be an interesting extension to test this on a dataset of non-spontaneous microexpressions. The USF-HD dataset is made up of microexpressions generated by subjects mimicking microexpressions and would be a fascinating test for this model [1]. Presumably, generated microexpressions would be less subtle than spontaneous microexpressions because it is difficult for a person to perfectly mimic the very quick and small movements of a microexpression. If the expressions were less subtle, it is possible the model would perform better because generalization would be easier.

Outside of just experimenting with different datasets, a worthwhile extension of this project would be adding in some time consideration through the inclusion of onset and offset frames. In extracting microexpressions, human coders pick out the small differences by watching a video and going through frame by frame as they notice a microexpression. Even so, when just comparing frames it can be hard even for a human to identify the markers of a microexpression. For example, figure 8 contains several frames from the CASME dataset.

These frames are the onset, offset, and apex frames of a happiness microexpression. Viewing the images like this, it is not immediately obvious that there is an expression of happiness. In reviewing the video, it is more obvious that there is a slight facial movement, however. This leads to including time consideration in the model. There are examples of using the surrounding frames to influence the classification of microexpressions, such as in Wu et. al [8]. One possible idea for adding time consideration to this model would be including the onset and offset frames as well as the apex frames and trying to classify the set of onset, apex, and offset frames.

Additionally, this project could be broadened from classifying microexpressions to simply identifying the existence of one. It would be simpler to scan the frames of a video and identify movements that could be a microexpression without trying to classify the specific emotion. A model could then be used to signal the existence of a microexpression and a person could review and classify it manually, should it be used in the real world.

Finally, a possible extension on this project would be simply adjusting the deep convolutional neural network. Would a deeper network result in better results because it can classify more features? Would a different activation function act differently to ReLU? Or is a deep convolutional neural network simply not the correct choice for classification of microexpressions?

8 Conclusion

Microexpression recognition is a small facet of the larger idea of expression recognition and computer vision. The inherent subtleties in microexpressions are what set it apart from general emotion classification. Because a deep convolutional neural network was so successful in classifying emotions, I was hopeful that its success would carry over into microexpression classification [9]. It seems, however, that a CNN struggles to classify microexpressions without context in the same way that humans struggle to classify microexpressions without context.

The final deep convolutional neural network included 47 layers, made up of 14 convolutional layers, 14 batch normalization layers, 14 ReLU layers, 4 max pooling layers, and a final fully connected layer. Running with 50 epochs and 5-fold cross validation, this model was able to achieve 45% accuracy on the testing set. This

accuracy is low, but the problem is very complex. It is a difficult task to classify the emotions being displayed in these frames, and it is probable that this amount of accuracy is higher than most people looking at these images could achieve.

The low accuracy result may be due in part to how small the facial changes that make up a microexpression are. Because microexpressions happen so fast, the markers of them remain small and variable. This is to say that the same emotion may appear differently in different people. This is true of emotion recognition as well, but with general emotion recognition the facial changes are much more accentuated, which allows for easier generalization. With a larger dataset and more folds in the k-fold cross validation, I believe performance may improve.

While peak performance was not achieved, many avenues for further research were opened up. Combining the ideas of a deep convolutional neural network and time consideration that was presented in Wu et. al.'s research is just one of many possible extensions of the work presented in this paper that may increase performance.

Ultimately, I consider this research to be a success. As a proof of concept, the deep convolutional neural network did better than simple guessing, considering that there are seven emotions in the dataset. This shows that, despite the failure here, there is a possibility that with further improvements, a deep convolutional neural network could be used for microexpression classification.

REFERENCES

[1] Wen-Jing Yan, Q. Wu, Yong-Jin Liu, Su-Jing Wang, and X. Fu, "CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces," 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Jul. 2013.

[2] H. Auerbach, L. A. Gottschalk, E. A. Haggard, and K. S. Isaacs, "Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy," in Methods of research in psychotherapy, 1st ed., Boston, MA: Appleton-Century-Crofts, 1966, pp. 154–165.

[3] S. Weinberger, "Airport security: Intent to deceive?," Nature, vol. 465, no. 7297, pp. 412–415, May 2010.

[4] S. Jordan, L. Brimbal, D. B. Wallace, S. M. Kassin, M. Hartwig, and C. N. H. Street, "A test of the micro-expressions training tool: Does it improve lie detection?," Journal of Investigative Psychology and Offender Profiling, vol. 16, no. 3, pp. 222–235, 2019.

[5] X. Fu, Welcome to Professor Fu's Lab, 2006. [Online]. Available: http://fu.psych.ac.cn/CASME/casme2-en.php. [Accessed: 16-Mar-2021].

[6] W.-J. Yan, S.-J. Wang, Y.-J. Liu, Q. Wu, and X. Fu, "For micro-expression recognition: Database and suggestions," Neurocomputing, vol. 136, pp. 82–87, Feb. 2014.

[7] D. van der Haar, "Student Emotion Recognition Using Computer Vision as an Assistive Technology for Education," in INFORMATION SCIENCE AND APPLICATIONS: icisa 2019, S.l.: SPRINGER VERLAG, SINGAPORE, 2020, pp. 183–192.

[8] Q. Wu, X. Shen, and X. Fu, in International Conference on Affective Computing and Intelligent Interaction, 2011, pp. 152–162.

[9] D. K. Jain, P. Shamsolmoali, and P. Sehdev, "Extended deep neural network for facial emotion recognition," Pattern Recognition Letters, vol. 120, pp. 69–74, Apr. 2019.

[10] Saxena, A. Khanna, and D. Gupta, "Emotion Recognition and Detection Methods: A Comprehensive Survey," Journal of Artificial Intelligence and Systems, vol. 2, no. 1, pp. 53–79, 2020.

[11] Y.-H. Oh, J. See, A. C. Le Ngo, R. C. Phan, and V. M. Baskaran, "A Survey of Automatic Facial Micro-Expression Analysis: Databases, Methods, and Challenges," Frontiers in Psychology, vol. 9, Jul. 2018.

[12] T. Pfister, Xiaobai Li, G. Zhao, and M. Pietikainen, "Recognising spontaneous facial micro-expressions," 2011 International Conference on Computer Vision, Jan. 2012.

[13] S. Koelstra, M. Pantic, and I. Patras, "A Dynamic Texture-Based Approach to Recognition of Facial Actions and Their Temporal Models," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 11, pp. 1940–1954, Mar. 2010.