Stress Detection Utilizing Methods of Deep Learning

Mohammad Choudhury

June 11, 2021

Abstract

This paper details the use of deep learning techniques in training a Convolutional Neural Network with the task of detecting the seven basic emotions as proposed by Paul Ekman. I use these emotions to detect stress levels in real time and aim to utilize these predictions to give L3 Harris flight trainers a better understanding of how capable they are doing tasks with low levels of stress. I utilized a combination of data sets including FER2013, a data set containing over 30,000 images corresponding to one of the seven basic emotions. Stress is classified by Paul Ekman as a combination of disgust and anger and both emotions were used to solve our specific problem. By using detailed architecture and comprehensive data I achieve around a 66% accuracy, coming close to many cutting-edge papers.

1 Introduction

1 [**W**]e can recognize facial emotions as early as six years into our lives, such recognition being vital to our survival in both modern and primitive societies. Emotion recognition is vital in navigating our personal and professional lives, allowing us to act and react while considering the feelings of those around us. Such development has played a crucial role in our evolution as a species and understanding how to approach this problem using machine learning will undoubtedly lead us to a better understanding of the relationship between humanity and machine learning. In this paper, our specific problem of stress detection relies on solving the underlying issue of emotion detection. To solve this underlying problem, I undertook understanding all related works in the field, as well as the myriad of methods used to detect emotions. Methods include transfer learning, support vector machines, k-nearest neighbors, facial vector points, and Convolutional Neural Networks. The end product of this paper is a system that can in real-time, detect and quantify stress in L3 Harris flight simulators, allowing us to gauge the ability of a trainee and give useful feedback on how to improve. This method gives a non-invasive alternative to our current system of bio metrics as it relies on solely a camera and minicomputer to function.

2 Related Works

2 [**T**]here is a large array of papers that attempt to solve the problem of emotion detection, utilizing techniques ranging from Support Vector Machines to neural networks. Specifically utilizing the FER data set, Yinchuan Charlie Tang produced the best accuracy at 71.2% with the Goodfellow et al. designed dataset. Tang made several changes to a normal convolutional model, replacing the widely used SoftMax activation function with a linear SVM, and using the top layer of a Support Vector Machine as the loss function for training in addition to a L2-SVM loss function. Khanzada et al's work used the FER dataset as well as additional data sets along with transfer learning and several other techniques to achieve a 75% accuracy. These two works were instrumental in the development of our own, providing the idea for replacing the SoftMax activation function and for the use of transfer learning respectively. Several papers also

employed facial landmarks for the task of emotion detection. Classifying emotions with facial landmarks and support vector machines is done and described by Michel in his final year dissertation, providing valuable information on feature extraction. Many facial landmark techniques that employed support vector machines used Euclidean distances calculated from a neutral face and an expressed face through a uniform video sequence. As our data set does not include both a neutral and expressed face for the same image this implementation is not possible. However, many implementations extracted and used facial landmarks from single images and saw improved performance, hence providing the value for the technique's replication.

3 Data set

3 [**T**]he FER data set contains 35,887 48x48 gray scale images, all depicting a base emotion. The dataset contains the most images of angry faces and the fewest of disgust, making our underlying problem especially difficult. The specific distribution of the data includes Angry (4,953), Disgust (547), Fear (5,121), Happy (8,989), Sad (6,077), Surprise (4,002), and Neutral faces(6,198). I reconstructed all images using pixel values given in a CSV file found on kaggle.com and utilized a test training split of 20 and 80% respectively.

Figure 1: FER2013 Dataset

4 Initial Model

4 [**T**]he baseline model created to solve our underlying problem of stress detection contains 9,162,311 parameters and 12 convolutional layers. The model was staggered with batch normalization layers and 50% dropout layers. I experimented with both adam and stochastic gradient descent and chose the latter after extensive testing, along with a rectified linear unit for our activation function and a softmax function at the end of the model. Our loss function was categorical cross-entropy as our problem involves multi-class classification. The model takes a gray-scale image and outputs 7 percentages, each corresponding to a unique class or emotion. I used 150 epochs, a batch size of 64, and a learning rate of 0.001 to train the model, taking inspiration from several papers. I did no data preparation or augmentation of the data initially besides reconstructing and resizing the images from pixel values. Due to the imbalance in representation between classes, with anger and disgust being underrepresented, I experimented with the SMOTE technique. Synthetic Minority Over-sampling Technique details oversampling minority classes and under sampling majority classes to account for the disparity. While the idea is promising and has shown promise in other works, I was not able to get a reasonable difference in accuracy after this specific augmentation and due to lack of time did not explore further. Our final model involved no data augmentation but this is a large area to be improved on in future cases.

5 Exploration into Detection Techniques

5 [**T**]he final methods used in this paper are a product of months of discovery and exploration into the vast world of machine learning. My personal journey involved several weeks spent exploring and documenting the large array of possible methods that could accomplish our underlying problem. I first started with the exploration of K Nearest Neighbors algorithms then from there different combinations of Support Vector Machines, the YOLO algorithm for object detection, Convolutional Neural Networks, and finally transfer learning. I spent the majority of my time exploring the YOLO method, pivoting af-

ter it was apparent such a technique was not worth exploring for emotion detection. Support Vector Machines were simply a poor way to detect stress and I abandoned them after my first implementation, although they are powerful for other classification problems. Throughout my time, I had the opportunity to delve into every viable method and take the time to understand the underlying mathematics and concepts in relation to my task. The exploration of these different methods were crucial in the arrival of our solution, both a CNN trained on facial vector points constructed from the FER data set and a CNN trained only on the FER data set.

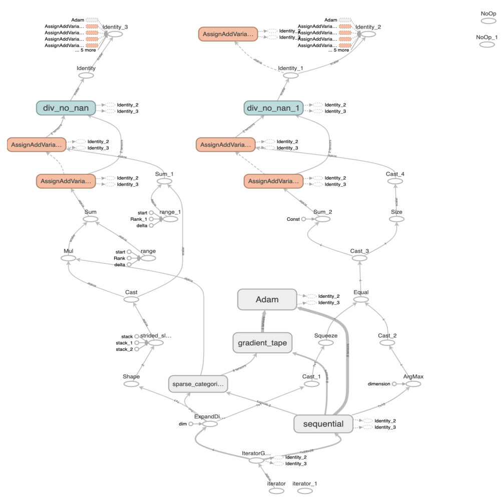

Figure 2: Conceptual Graph of Our Model

6 Real-Time Display

6 [**F**]or the end-use of the model, I used a laptop webcam to test the results. I utilized OpenCv's Haar cascade face recognizer as well as dlib's facial landmark library. Using the model that was without facial landmarks, I fed the model a face recognized by the haar cascade recognizer and displayed a bounding box on the same face. The bounding box was accompanied by a percentage level of stress, calculated by the parallel confidence in both anger and disgust. To provide an example, if anger and disgust were detected at a probability of 25% each and were the highest two probabilities, the stress level would be 50%. We normalize this percentage to 0.5, providing a relative measure of stress and particularly high levels of stress are recorded and stored, providing for future reference by the training pilot. The version that employed landmarks worked similarly, using landmark points calculated on a detected face, without the addition of a bounding box.

7 Conclusion

7 [**T**]he singular goal of this project involved many undertakings. The exploration and rigorous testing of numerous techniques ended in the utilization of a CNN, providing the means to detect stress. I utilized the publicly available FER data set to construct both a deep learning model capable of detecting stress using live images and facial landmarks. This paper hopes to give a starting point to those who explore the idea of stress detection at L3 in hopes of establishing a working non invasive way to improve our clients experience. While many explorations of the topic of stress detection resulted in no commercial value, I believe this final model can be optimized, improved, and put to use in L3's flight simulators.

8 Acknowledgements

This paper would not have been possible without the advisement and help of Brian Collen, who provided resources and approval for this project. I would also like to acknowledge Craig Bonner for his mathematical advising and expertise, especially in the field of optimization. Lastly, I thank James Hovey for instruction in basic computer science and continual help throughout my academic career.

9 References

1 Y. Tang, "Deep Learning using Support Vector Machines," in International Conference on Machine Learning (ICML) Workshops, 2013.

2 Redmon, J., Divvala, S., Girshick, R., amp; Farhadi, A. "You Only Look Once: Unified, Real-Time Object Detection." IEEE Conference on Computer Vision and Pattern Recognition, 2016.

3 P. Mihcel, "Support Vector Machines in Automated Emotion Classification." Churchill College, 2003.

4 I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee et al., "Challenges in representation learning: A report on three machine learning contests," in International Conference

5 Minaee S., Abdolrashidi A., "Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network", 2021

6 Y. Khaireddin, Z. Chen, "Facial Emotion Recognition: State of the Art Performance on FER2013", Computing Research Repository, 2021

7 K. Lawrence, R. Campbell, and D. Skuse, "Age, gender, and puberty influence the development of facial emotion recognition". Front. Psychol. 2015