Neural Network Approach to Recognize Facial Emotions

1st Nathan Theng
Department of Computer Science
California State University, Fresno
Fresno, California
theng_nathan@mail.fresnostate.edu

1st Elvis Lor
Department of Computer Science
California State University, Fresno
Fresno, California
elvislor11@mail.fresnostate.edu

2nd Leo Llave
Department of Computer Science
California State University, Fresno
Fresno, California
compfresnostatestudent144@mail.fresnostate.edu

3rd Emily Barrett
Department of Computer Science
California State University, Fresno
Fresno, California
emkabar@mail.fresnostate.edu

***Abstract*—Facial emotion recognition received significant attention in recent years as a crucial component in human-computer interaction and biometric security. In this paper, a deep learning approach, specifically Convolutional Neural Networks (CNNs), was presented to automatically extract features from facial expressions for recognizing facial emotion.**

Two neural network models were proposed, designed, and trained on two large and diverse datasets of facial images with their corresponding emotion labels: CKPLUS and Natural Human Face Image for Emotion Recognition. Hierarchical representation from facial cues was learned to achieve accurate classification. The neural network architectures and hyper-parameters were optimized to enhance accuracy and performance.

The neural network was designed to learn hierarchical representations, capturing both global and subtle facial cues contributing to accurate emotion classification. Different neural network architectures' effectiveness and hyper-parameter optimization were explored to enhance model performance.

The results from the two neural network designs demonstrated a 44% and 49% accuracy in recognizing a diverse range of facial emotions: Anger, Contempt, Disgust, Fear, Happiness, Neutrality, Sadness, and Surprise. Since the majority of facial emotion recognition relies on static images, these models were implemented with live-feed images via a computer web camera.

I. Introduction

Emotion recognition has traditionally relied on machine learning models and algorithms that heavily depend on hand-crafted feature extraction. These methods require the identification and selection of specific features within static images or photos for emotional recognition. However, there is a shift towards utilizing various types of neural networks to enhance the analysis of facial emotion recognition, especially in real-time applications. This transition allows the processing and analysis of facial expressions as they are captured during the recognition phase, which enables the detection and recognition of diverse emotions as they unfold.

However, while it is difficult to recognize emotion in real-time from the live feed, some of the obstacles lie in the variability of how individuals express their emotions. Since emotion is completely individualist, and can vary from person to person, it poses a challenge for models such as our neural network to interpret and classify within these variations of emotions.

Models like these are essential when considering the emotional state of various individuals in different contexts. With the growing prevalence of emotional burnout and student mental health, there is a need to find a balance between academic life and personal mental and physical well-being. This technological advancement and model can help recognize these signs of exhaustion and stress within students. Being able to detect and respond to these emotional states, will provide valuable support and intervention to those in need.

Multiple of these facial features for emotional recognition will be extracted and processed through a deep learning model, and convolutional neural network. This is a regularized type of feed-forward network that has filter optimization and will perform a convolution operation to achieve an activation map and apply a pooling layer. It is then flattened into a linear vector that is then passed to an artificial neural network. The research will aim to improve the accuracy of detecting these facial emotions to contribute to the border understanding of the emotional well-being of students and professionals.

II. Related Works

A. Research Papers

Before proceeding with processing the data and creating the model and analysis of the neural networks, multiple article reviews were conducted to grasp a better understanding of emotion recognition and the current technology and research around the topic.

The first paper is *A Comparative Study on different approaches of Real-Time Human Emotion Recognition based on Facial Expression Detection* by Anurag De and Ashim Saha [2]. This paper focused on the different algorithms of recognition using a novel method [. It proposes a live video

input with frame extraction, face detection, facial feature point extraction, and emotion classification utilizing a Support Vector Machine. The second approach used an adaptive Canny operator edge detection with Active Appearance Model which video capture, image preprocessing, face detection, facial feature extraction, expression feature extraction, and classification and recognition. The paper goes into various other algorithms. The paper provided the team with different ideas and concepts on how to process and approach neural network deep learning.

The second paper is *Real-time Student Emotion and Performance Analysis* by Pramodini Metgud, Navaya Naik, et.al [4]. This paper focuses on student emotion and student burnout over the semester-long coursework. It utilized facial emotion recognition before and after students. It utilized an EifficientNet-B0 algorithm and received an accuracy of around 85.72% and was processed with the haar cascade and Ada Boost algorithm as face detectors. The student information and performance were later sent to faculty and parents. This paper gave the team a better insight into how emotion recognition technologies can be applied to student mental health.

The third paper is *Real-time Algorithms for Facial Emotion Recognition: A Comparison of Different Approaches* by Aneta Kartali, et.al [3]. It utilized a conventional approach for a Histogram of Oriented gradient features that uses a support vector machine of HOG features and a Multi-Layer Perception (Neural Network). It also utilized three types of Convolutional Neural Networks: Alex Net CNN, Commercial Affdex CNN Solution, and Custom-made FER-CNN. This paper made the comparison of accuracy in recognizing the 4 basic emotions with an average frame-per-second prediction to predict the emotion classification. The Commercial Affdex CNN has the best overall accuracy with 85.05%. This paper provides the team with more insight into neural networks to predict the different eight emotions.

B. Datasets

The model utilized two different data sets: CK+ (extended Cohn-Kanade) and Natural Human Face Images for Emotion Recognition.

The CK+ data set had 980 images each with 48 by 48 pixels images and could be classified as one of seven different facial emotions: Happy, Sad, Angry, Afraid, Surprise, Disgust, and Contempt. The Natural Human Face Images for Emotion Recognition has 5500 images of size 224 by 224 pixels that would be classified as the following: Neutral, Happy, Angry, Sad, Fear, Surprise, Disgust, and Contempt. The research also utilized another data set which was identical to the Natural Human Face Images for Emotion Recognition, except with seven classifications as the Happy classification was removed in this data set.

C. Open-Source Programs

To build the neural network models, the designs of the model were based on an Open-Source Program through the MathWorks File Exchange: Live Emotion Detection Using CNN a Deep Learning Model. This was created by Akhilesh Kumar.

III. Methods

Deep learning is a sub-field of supervised machine learning that employs models to automatically learn how to categorize different kinds of data, such as text, audio, and images. Deep learning is usually implemented using neural networks, where ”deep” refers to the number of layers in the network. Convolutional Neural Networks is a branch of neural networks that take and process images. They are built using principles of neuronal organization discovered by connection ism in a biological neural network constituting animal/human brain t. Below are the different components to make a neural network

- Neurons: A crucial role that constitutes every layer (Input, Hidden, and Output layers). These neurons function similarly to the neurons in the brain cells with an activation function and connection. However, each neuron contains a bias parameter which is how the Neural Network learns and adjusts during the training process.
- Input Layer: The first layer of the neural network that represents the input data for the network, each neuron corresponds to one feature from the data set
- Hidden Layers: With. input and output pairs having complex relationships, hidden layers help to decode these relationships between the two. Hidden layers exist between the input and output layers and contain neurons that connect to every other neuron in adjacent layers
- Activation Function: Similarly to our brain, neurons get activated based on the signals from sensory organs. In the Neural Network, we have an activation function for neurons of every layer. There are Linear and Non-linear Activation Functions based on the input and output later. Some popular non-linear activation functions that are used are Sigmoid, ReLU, and Softwax.
- Parameters: Loss Function and Optimizer Algorithm. The Loss Function measures the difference between the predicted output of the model and the actual output. The Optimizer adjusts the Model's Parameters to Minimize the loss function.

Each neural network design creates an 'imageDatastore' from the image director. The data is then split into training and testing sets using the 'splitEachLabel' by an 80/20 ratio. 80% of the data is for training and 20% of the data is for testing.

Both designs will output seven different classes for the first data set and eight different classes for the other data set. The design uses the Stochastic Gradient Descent Optimizer, Learning rate = 0.001, 10 epochs, and shuffling of data. Both designs used the cross-entropy loss function which can be seen below:

$$L(y, \hat{y}) = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{1}$$

A. Design 1

For the first neural network design utilize **three fully connected convolution layers**, each i then followed by batch

normalization, the ReLU activation function, and the Max-Pooling Layers. The output layer utilized the softmax activation function which is used for the multi-classification (7-8 classes)

- Convolution Layer: Transform the input image to extract features from it. The image is convolved with a kernel filter which is a small matrix with height and width smaller than the image to be convolved. This is also called a convolution matrix.
- Batch Normalization: A method to make the neural networks faster and more stable through normalizing the layers' inputs by re-centering and re-scaling.
- ReLU Activation: Rectified Linear Unit. Returns 0 if it receives a negative input, but returns the value back if it is any positive value. it can be seen as below:

$$f(x) = max(0, x) \tag{2}$$

- Max Pooling Layer: A pooling operation that selects the maximum element from the region of the feature map covered by the kernel and filter.

B. Design 2

For the second neural network design utilize **six fully connected convolution layers**, each i then followed by batch normalization, the ReLU activation function, and the Max-Pooling Layers. The output layer utilized the softmax activation function which is used for the multi-classification (7-8 classes). This design runs with a batch size of 32.

C. Implementation

These two designs of the neural networks were implemented with MATLAB utilizing this workflow seen below:

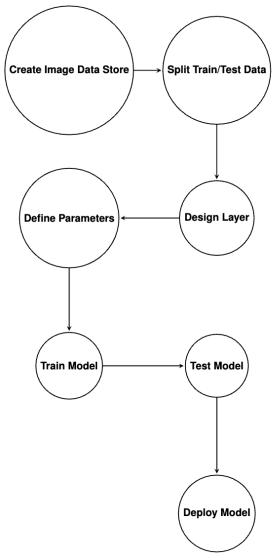

Fig. 1. Workflow Diagram

The project utilized the Webcam Support ToolBox as well when implementing the live-feed emotion recognition.

Below illustrate the various output of our web camera implementation, which demonstrated the algorithm classifying the emotion of the real-time image from the live feed.

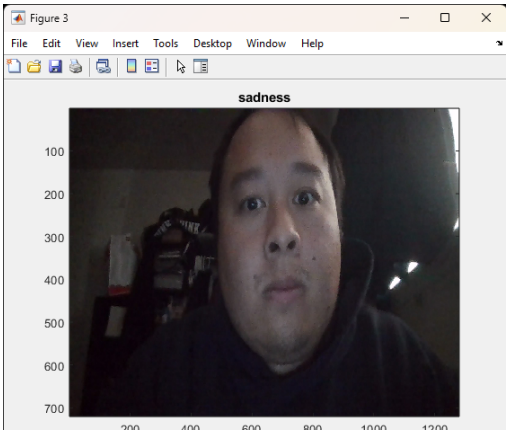

Fig. 2. Sadness Recognition

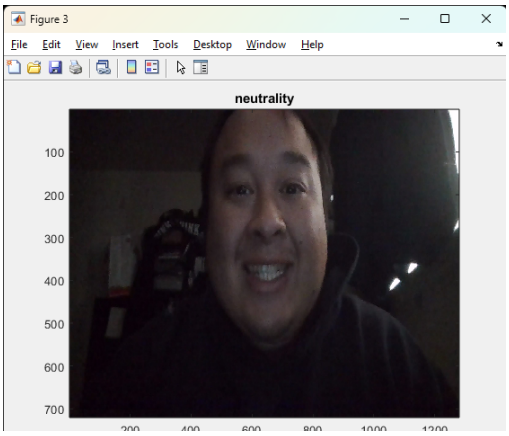

Fig. 3. Neutrality Recognition

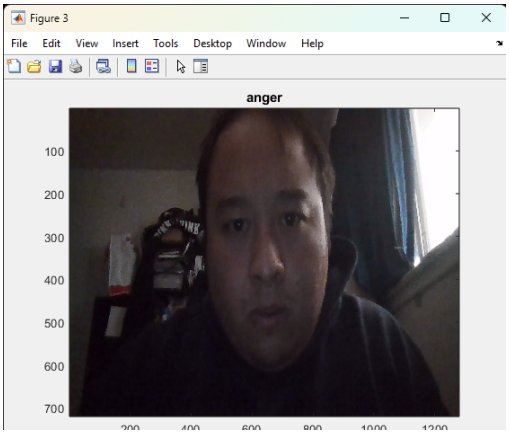

Fig. 4. Anger Recognition

The emotion classification web camera implementation is not perfect. In Figure 2, the individual's face demonstrates

that he is smiling, usually, humans would classify this as surprise or happiness. However, our algorithm classifies this emotion as neutrality. The web camera implementation is still a successful way to deploy the neural network models to a real-life application. Thus, by improving the models' accuracy, it should be expected that recognition from the real-time images from the live feed should improve as well for future research.

IV. Results

After running the both set of models the accuracy are listed below:

Design 1	Design 2
No MiniBatchSize	MiniBatchSize = 32
3 CNN Layers	6 CNN Layers
Accuracy = 44.2%	Accuracy = 49.2%

TABLE I
Accuracy Between Designs

By doubling the number of convolution layers within the model and adding a miniBatchSize of 32, the accuracy increase roughly by 5%.

Below is the Training Progress and Confusion Matrix for Design 1.

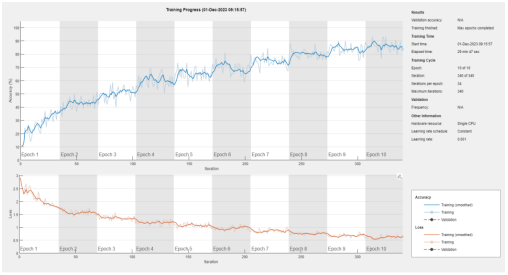

Fig. 5. Training Progress for Design 1

Fig. 6. Confusion Matrix for Design 1

As you can see the training accuracy increased throughout the the 10 epochs while the training loss decreased throughout the the 10 epochs.

The Confusion also demonstrates how many times the model was able to classify the right emotion. For example, it was able to correctly classify happiness 207 times, which contributed 18.6% to the 44.2% overall accuracy. However, it was only able to correctly classify contempt three times, which contributed 0.3% to the 44.2% overall accuracy. The rest of the confusion matrix demonstrates how many times the model incorrectly classifies the emotions as well.

Below is the Training Progress and Confusion Matrix for Design 2.

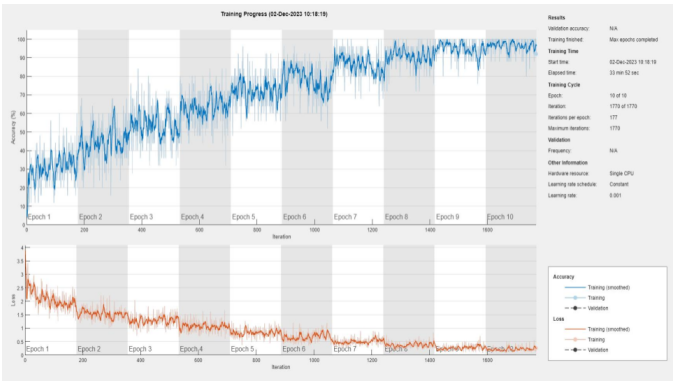

Fig. 7. Training Progress for Design 2

Fig. 8. Confusion Matrix for Design 2

As you can see the training accuracy increased throughout the the 10 epochs while the training loss was decreased throughout the the 10 epochs as well for Design 2.

The Confusion also demonstrates how many times the model was able to classify the right emotion. For example, it was able to correctly classify happiness 230 times, which contributed 20.7% to the 49.2% overall accuracy. However, it

was only able to correctly classify contempt three times, which contributed 0.5% to the 49.2% overall accuracy. The rest of the confusion matrix demonstrates how many times the model incorrectly classifies the emotions as well.

V. Discussion

Even, though the accuracy is below 50%, the model is still successful. Since there are eight different emotions, without the model, there is a 12.5% probability of correctly guessing the correct emotion. However, the model can correctly predict the correction emotion at 49.2% which is 3.936 more efficient in prediction. One of the various reasons why this would occur goes back to the discussion on how emotion is individualistic. Everyone has a unique way of demonstrating their emotion. No two people are the same. Another reason is that some of the emotions could overlap with another emotion. For example, some individual's emotions for neutrality could be mistaken and misclassified as contempt. This can be said for the same with happiness and surprise as well.

There were two potential flaws in the testing phase. One would be to increase epochs for the smaller data set so that there is a more accurate comparison between the smaller data set and the larger data set. This is because the one with a smaller data set has less amount of images so we can not gauge the accuracy between the two data sets The one drawback is that the is a very large and long execution time for each epoch. Another improvement would be to remove emotion with very low prediction. For example, contempt can correctly predict three and six times for Design 1 and Design 2 respectively. This emotion-low prediction could be classified as just noise within the data and decrease the overall accuracy of the models.

Finally, this research could be expanded by utilizing transformers and see how the accuracy and time execution compared to that of the convolutional neural network architecture.

Acknowledgment

This research project group thanks Professor Amith Kamath Belman, the Department of Computer Science at California State University, and CSci 158: Applied Biometric Security

References

[1] Akhilesh Kumar (2023). Live Emotion Detection using CNN a Deep Learning Model (https://www.mathworks.com/matlabcentral/fileexchange/75451-live-emotion-detection-using-cnn-a-deep-learning-model), MATLAB Central File Exchange. Retrieved December 10, 2023.

[2] A. De and A. Saha, "A comparative study on different approaches of real time human emotion recognition based on facial expression detection," 2015 International Conference on Advances in Computer Engineering and Applications, Ghaziabad, India, 2015, pp. 483-487, doi: 10.1109/ICACEA.2015.7164792.

[3] A. Kartali, M. Roglić, M. Barjaktarović, M. urić-Jovičić and M. M. Janković, "Real-time Algorithms for Facial Emotion Recognition: A Comparison of Different Approaches," 2018 14th Symposium on Neural Networks and Applications (NEUREL), Belgrade, Serbia, 2018, pp. 1-4, doi: 10.1109/NEUREL.2018.8587011.

[4] P. Metgud, N. D. Naik, S. M. S and A. S. Prasad, "Real-time Student Emotion and Performance Analysis," 2022 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2022, pp. 1-5, doi: 10.1109/CONECCT55679.2022.9865114.