Real-time Emotion Estimation

- Team 8: Daniel Paliulis, Gautam Pirthiani, Tyler Hinrichs, Vihaan Shah

1. Introduction

Emotions play an important role in human interactions. Understanding them is fundamental in establishing meaningful relationships and productive communication. For many, deciphering these emotional cues is challenging, especially for those with social or cognitive disorders. Leveraging the advancements in AI and machine learning can bridge this gap, providing real-time feedback on emotional states based on facial cues. This would provide more context to emotion during conversation or video calls for those that struggle reading such cues.

2. Objective

The objective is to employ image classification techniques to analyze facial expressions in real-time, providing instantaneous emotion estimations. The system will classify facial expressions into specific emotions and display the top predictions based on confidence scores. This will be presented in an intuitive and easily-interactable front end, providing immediate emotion estimation based on incoming video data from a real-time interaction.

3. Dataset

The dataset, titled FER2013, is sourced from Kaggle and contains labeled images representing various emotions.

Data Set: https://www.kaggle.com/datasets/msambare/fer2013

Specifications:
1. Features: The dataset comprises grayscale images of faces, each labeled with an emotion.
2. Number of images: 35,887 images.
3. Labels (Emotions covered): Angry, Disgust, Fear, Happy, Sad, Surprise, Neutral
4. Image Format: Grayscale, 48 by 48 pixels
5. Dataset Split: The dataset is categorized into training and testing validation.

4. Methodology

Data Processing:

a. Data Augmentation: We'll introduce varied transformations like rotations, zooms, horizontal flipping, and shear transformations to enhance the model and prevent overfitting (Lau 2017).
b. Data Rescaling: To normalize pixel values between 0 and 1, all images will be rescaled.

Model Architecture:

Our deep learning model will use a convolutional neural network (CNN) with multiple convolutional layers, max-pooling, dropout, flattening and dense layers.

To further improve the classification, we will use transfer learning techniques to leverage pre-trained deep convolutional neural networks like VGG16, ResNet, MobileNet, etc. These architectures have been trained on extensive datasets like ImageNet, extracting features that can be useful for our emotion estimation task (Baker 2022).

Metric Evaluation:

Accuracy: The proportion of correctly predicted emotions to total predictions. This will be a general metric of the overall performance of the model rather than focusing on any particular expression.
F1 Score: The F1 score will be utilized to evaluate the performance of the model in order to consider both precision and recall. It is important that the model both classifies a significant magnitude of the inputs and classifies them correctly. These features will ensure the practical use of the final application in a real-time setting.
Confusion Matrix: To observe the performance of the algorithm with respect to the possible classifications, a confusion matrix will be generated. This will allow generation of a visualization of the accuracy for each classification in comparison to the other potential classes. This will provide additional insight into what expressions are prone to be mistakenly classified as others. This also allows evaluation of the effectiveness in the model to predict each individual expression, rather than the general performance as a whole. This will be effective in showing if there are any expressions that the model struggles to predict.

Deployment and Real-time Implementation:

To ensure instantaneous emotion estimations, the model will be optimized for latency.

a. Model Optimization: Techniques such as quantization and pruning will be explored to reduce model size and inference time (Saturn 2023).
b. Hardware Acceleration: Different hardware may be utilized to ensure immediate feedback. This includes Utilizing GPUs to parallelize computations and edge devices, such as smartphones or the Nvidia Jetson, to decrease data transmission time (Shi 2016).
c. Data Stream Optimization: This will be performed through buffering data in batches and processing the batches in parallel using pipelines.

Final Testing of Application in Real-World Environment:

The final stage of the project will involve testing the application in its intended environment. This will involve having a conversation in front of a video and tracking which emotions the application predicts. This conversation will be controlled with a script where different emotions are scheduled throughout the conversation to test the full range of the tool.

5. Potential Applications

Deciphering emotional cues for people with cognitive disorders: The main application of our project revolves around enabling people with cognitive disabilities to receive information about the emotion of the people they are viewing on a screen. This could be in the context of a video, a video call, or other forms of digital communication or entertainment.

Conducting analysis of customer satisfaction during a video call: This tool could be used to assess current customer satisfaction, and later average customer satisfaction during video calls. It could be used as a tool to conduct larger studies on what brings success when interacting with customers.

Conducting sentiment analysis in intra-company calls: For a company that values how its employees feel about major decisions, it could be used during company-wide calls to assess average employee sentiment.

Monitor recruiter reactions during interviews: With many job interviews being conducted over video call, it would be advantageous for applicants to use this tool to receive insight into how their interviewer is reacting to responses in real-time.

6. Challenges

Real-time feedback - Ensuring instantaneous emotions estimations with no delay will be a primary technical challenge.

Interactive front-end - Along with the model, we will have the challenge of creating a front end to act as the interface of the tool between the end user and the model.

7. Conclusion

With the implementation of this real-time emotion estimation system, we aim to provide a valuable tool that offers immediate emotional insights, enhancing social interactions for those who use it. This will enable those who struggle evaluating emotional insights on their own and enhance conversational experiences conducted over video call.

References

Baker, N. A., Zengeler, N., & Handmann, U. (2022, January 14). *A Transfer Learning Evaluation of Deep Neural Networks for Image Classification*. Multidisciplinary Digital Publishing Institute. https://www.mdpi.com/2504-4990/4/1/2

Lau, S. (2017, July 10). *Image Augmentation for Deep Learning*. Medium. https://towardsdatascience.com/image-augmentation-for-deep-learning-histogram-equalization-a71387f609b2

Saturn Cloud. (2023, July 6). *Python Real-Time Image Classification Problems with Neural Networks*. https://saturncloud.io/blog/python-realtime-image-classification-problems-with-neural-networks/

Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016, June 9). *Edge Computing: Vision and Challenges*. IEEE Xplore. https://ieeexplore.ieee.org/document/7488250