COSE474-2024F: Final Project Proposal
“Emotion Recognition”

DINIE (2022320110)

1. Introduction

The motivation behind this project is to develop a robust system for emotion recognition using artificial intelligence. With the increasing reliance on AI in human-computer interactions, accurately detecting and responding to human emotions can significantly enhance user experiences in various domains such as customer service, healthcare, and entertainment. The aim is to explore how AI can be trained to recognize emotional cues from visual and audio inputs to improve interactive systems.

2. Problem Definition & Challenges

The problem lies in the inherent complexity of human emotions and the subtle variations in emotional expression across different individuals and cultures. Recognizing emotions from various input modalities, such as facial expressions, voice tone, and body language, poses a significant challenge. One of the main difficulties is achieving high accuracy in emotion detection across diverse datasets while avoiding biases that might arise from limited training data.

3. Related Works

Research on emotion recognition has explored models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to analyze facial expressions and voice patterns. Previous works include models like FER+ for facial emotion recognition and audio-based emotion classifiers such as Emo-DB. However, many existing models struggle with real-time processing and multi-modal emotion recognition.

4. Datasets

For this project, publicly available datasets like FER2013 for facial emotion recognition and RAVDESS for audio emotion recognition will be used. These datasets provide labeled emotional expressions and speech samples that cover a wide range of emotions. The combination of these datasets will allow for a multi-modal approach to emotion recognition, enhancing the model’s ability to generalize.

5. State-of-the-Art Methods and Baselines

The current state-of-the-art for emotion recognition relies heavily on CNNs for image-based emotion recognition and RNNs or Transformer models for audio-based emotion detection. Baselines for comparison will include standard models like VGGNet and ResNet for facial recognition and Long Short-Term Memory (LSTM) networks for analyzing vocal emotion patterns.

6. Schedule & Roles

The project will follow the following tentative schedule:

- **Week 1-2**: Literature review and dataset preparation
- **Week 3-4**: Model design and selection of algorithms
- **Week 5-6**: Model training and testing on multi-modal datasets
- **Week 7**: Fine-tuning and performance evaluation
- **Week 8**: Final testing and report writing

7. References

- Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1), 39-58.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- 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 (CVPR)*, 770-778.
- Livingstone, S. R., & Russo, F. A. (2018). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PLOS ONE*, 13(5), e0196391.