

This project has laid the groundwork for future research into multimodal emotion recognition systems. In terms of development, there are 2 key areas for improvement:

1. Improvement of unimodal models.
  - a. For the visual based model this can be done by replacing the existing 1D CNN emotion classifier with a fine-tuned Visual Language Model capable of direct, continuous valence-arousal prediction. The fine-tuned VLM should be benchmarked against the current classifier on standardised datasets and on the pre-pilot pipeline. Additionally, Robustness of the new model can be improved via augmentation which could minimise the effect of pose, lighting, occlusion of the subject.
  - b. For the physiological signal based model this can be done by fully completing leave one out cross validation, and by fine tuning the CNN architecture. Alternatively, a remodel or alternative machine learning model could be developed. For fine-tuning of the model, CNN.py can be altered, and for adjustments of hyperparameters such as learning rate or epochs, Train.py can be altered.
2. Improvement of the fusion strategy. I got this TY
  - a. Edit late\_fusion\_module.py to move beyond fixed/ confidence-weighted averages to adaptive fusion that estimates each modality's reliability over time based on recent performances, or other specific conditions.
  - b. Optionally, future work could include training a small fusion model on synchronised unimodal outputs (and their uncertainties) to predict VA directly; include missing modality handling.