Institute for Cognitive Science
Emotive Computing Lab

From Emotion Recognition to Mind-Wandering Detection
A Comparative Analysis of Video-Based Emotion Foundation Models

Ekta Sood*, Sebastian Ricke*, Trisha Mittal, Sidney D'Mello
Univeristy of Colorado Boulder, Univeristy of Colorado Boulder, Dolby Labortories, Univeristy of Colordo Boulder

University of Colorado, Boulder

Motivation

• Predicting cognitive states such as mind wandering during learning is important for adaptive educational technologies

• Video-based mind-wandering detection is promising → continuous and non-intrusive, but progress is limited by data scarcity

• Prior work leveraged pretrained emotion recognition model – transfer learning to MW pred → ER models may provide useful visual representations for MW detection

• **Do more recent and stronger emotion-recognition foundation models improve frozen-transfer performance for video-based mind-wandering detection?**

Experiments

• Building on SOTA transfer-learning pipeline, keep downstream MW classifier fixed
→ swap encoder w. newer pretrained ER

• Baseline: AffectNet-pretrained ResNet50

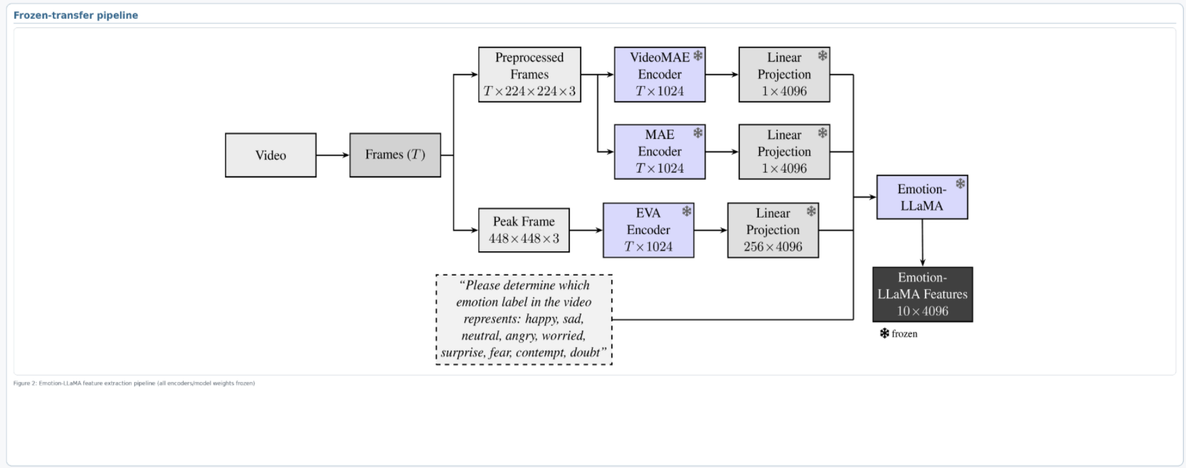
• Newer encoders: MAE, VideoMAE, Emotion-LLaMA

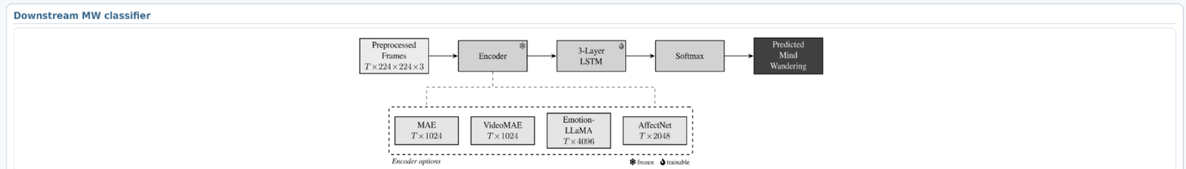
Evaluation

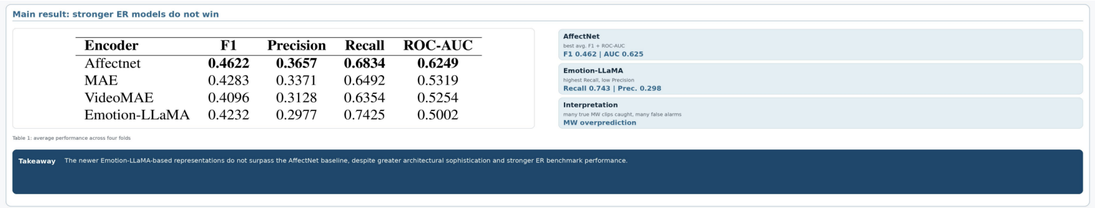
• Metrics: F1, Precision, Recall, and ROC-AUC

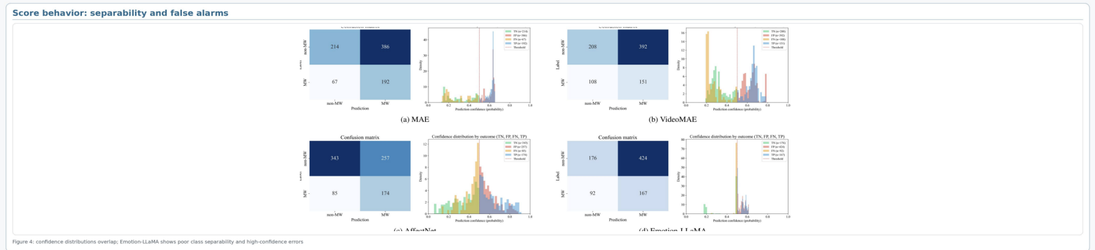
• Analysis: How each encoder fails
→ prediction confidence and score separability
→ FP vs. FN error profiles
→ shared vs. encoder-specific failures
→ predicted emotion labels vs. MW errors

Methodology

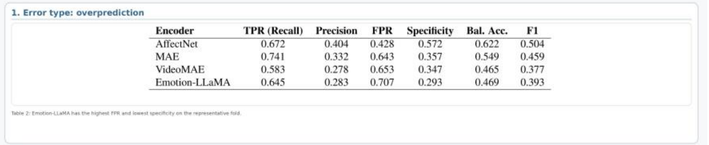


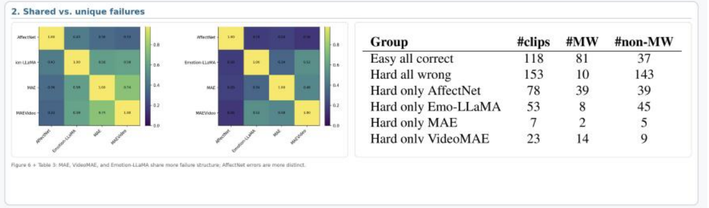


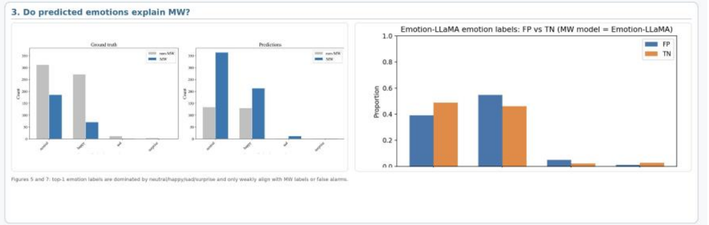


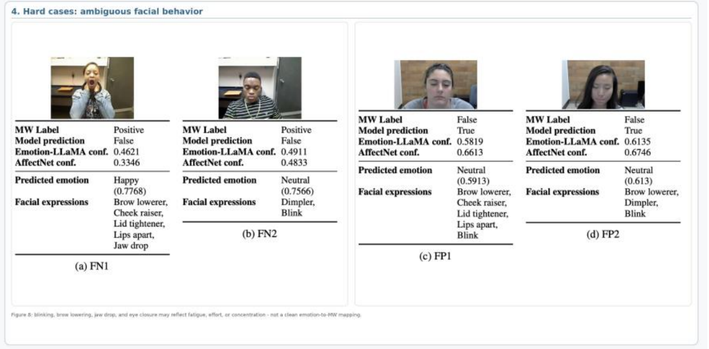


Analysis









Outlook

SOTA ER performance → **poorer MW** prediction
- emotion descriptors align weakly with MW-relevant behavior, weaker class separability, overpredict.
- Move beyond emotion-centric transfer → task-