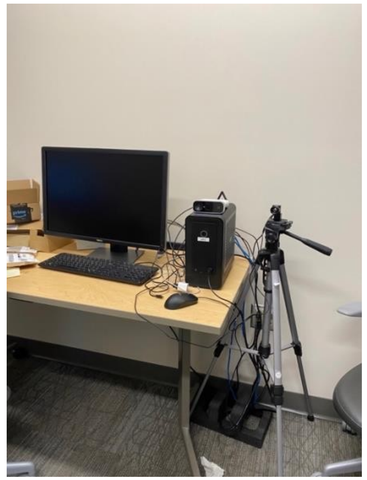
Detecting Human Engagement in Social Robots

Aditi Vijay Gode, Kowshik Selvam, Mihir Ravindra Patel | Luddy School of Informatics and Computing, Indiana University

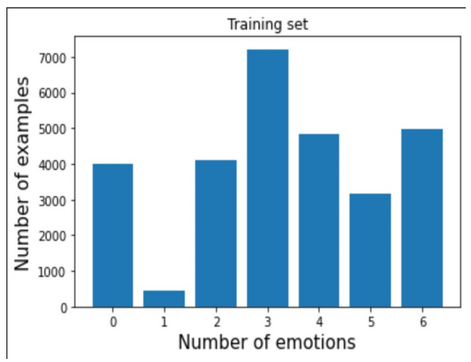
1. Motivation

- Our idea is to be a part of the larger social engagement algorithm in **"Detecting Human Engagement in Social Robots".**
- A model that detects emotions in toddlers while interacting with the socially assistive robot.
- Start with detecting discrete facial emotions but later move on to detect valence and arousal(more subtle aspects of emotions).



The Robot present in the Artificial Intelligence Lab at Indiana University which will be used in to test the existing model.

2. Challenges



High imbalance in fer2013 labels. Balanced it using data augmentation. The other main challenge was to determine the branching level in the ensemble network.

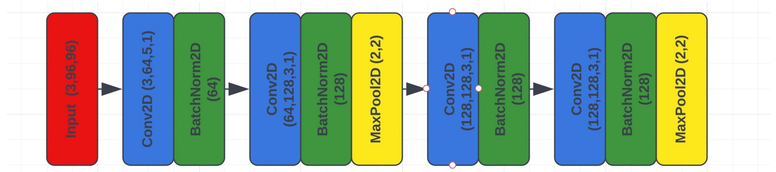
3. Experiments on FER2013

- Customized VGG16(with additional batch normalization and dropout layers)
- Compact Convolutional Transformer(CCT) to leverage generalization of transformers and locality of convolutional networks.
- Attentional CNN to concentrate learning on specific areas/features of the face with low computation.
- ResNet18 to experiment on transfer learning.

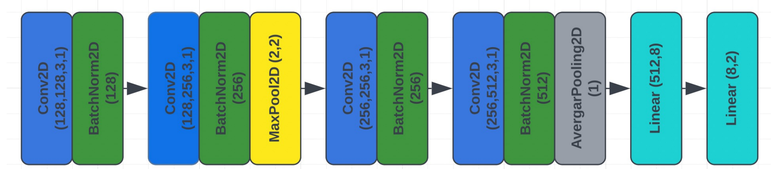
Model	Training Accuracy	Validation Accuracy	Testing Accuracy	# of Parameters
VGGNet16	79.71%	70.23%	68%	33M
CCT	46.81%	47.21%	46.81%	406k
Attentional CNN	71.671%	52.577%	54%	100K
ResNet18	50.45%	45.5%	40.23%	11M

4. Valence and Arousal Experiments

Valence – How positive/negative an event is.
Arousal – Event is exciting/soothing.
Lesr =$\Sigma b \Sigma i$ L[P(f(xi) = yi|xi, θshared, θb), yi]
Below are the 2 blocks of the wide ensemble model.

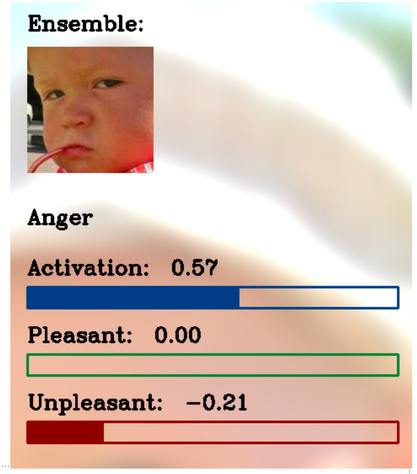


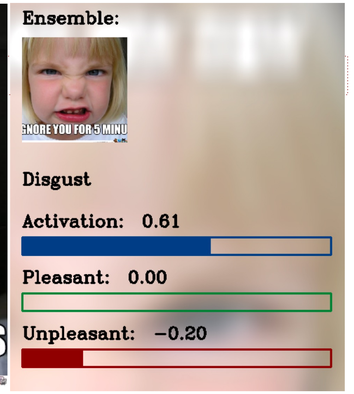
The Shared Base Network



Ensemble with the 9 convolution branches

5. Results





6. Future Work

- Integrating with thermal face data to determine gaze.
- Training on more complex video and web cam.
- Fine tuning with various other datasets (FER+).

).