Engagement Detection from Video Capture

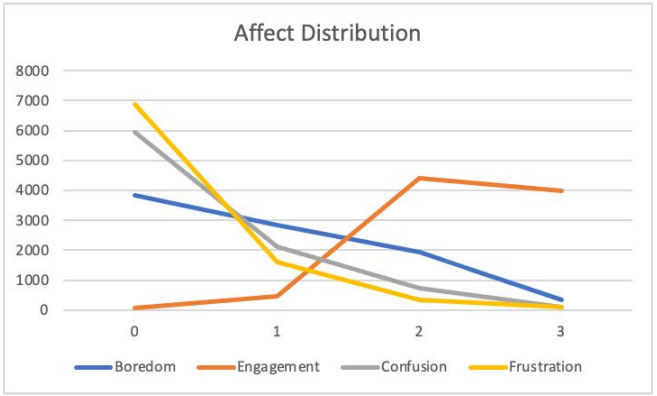
Chris Sexton, MIDS W251
December 2020

Problem Statement

Is it possible to assess whether a student is engaged in a virtual classroom, directly from video feed?

Data - DAiSEE

- 9068 multi-label video snippets captured from 112 users
- user affective states of boredom, confusion, engagement, and frustration.



	Boredom	Engagement	Confusion	Frustration
0	3822	61	5951	6887
1	2850	455	2133	1613
2	1923	4422	741	338
3	330	3987	100	87
total	8925	8925	8925	8925
average	0.86	2.38	0.44	0.29
% labeled 0	43%	1%	67%%%	77%

BORED = 0

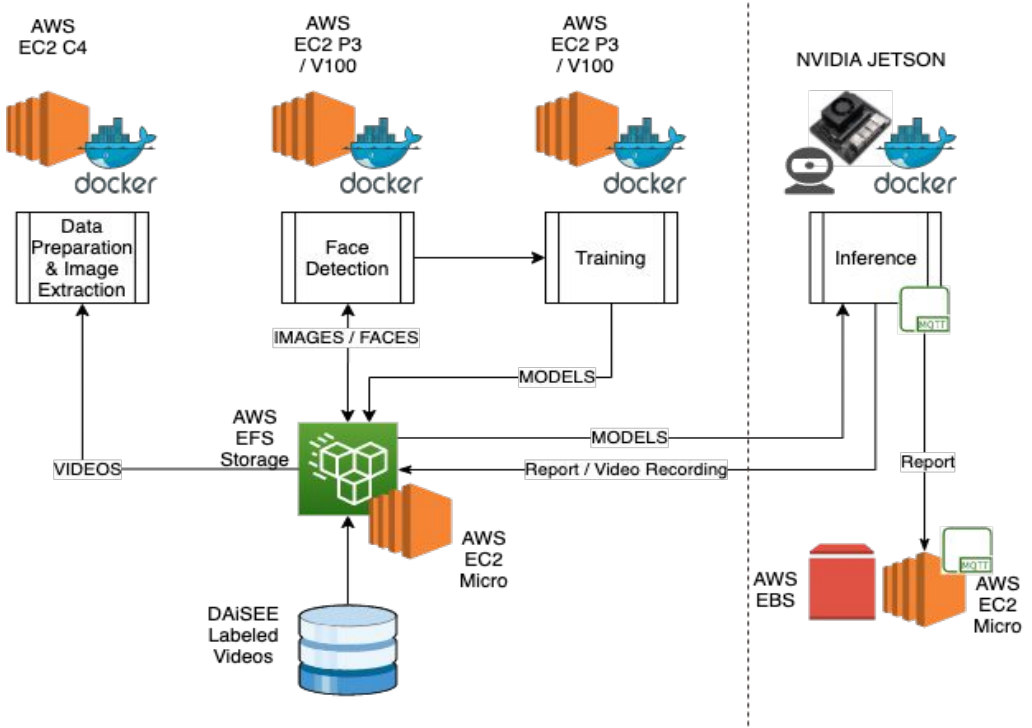
BORED = 1

BORED = 2

BORED = 3

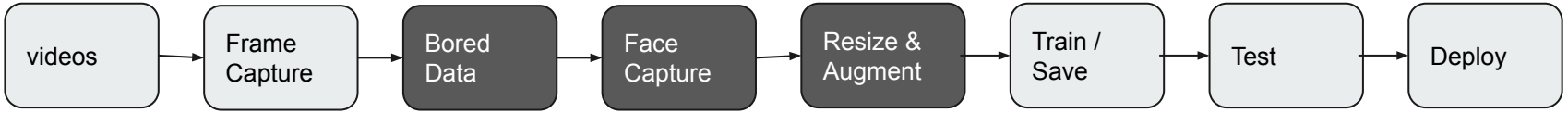
Environment

- TensorFlow 2
- OpenCV 4.4
- Python 3X

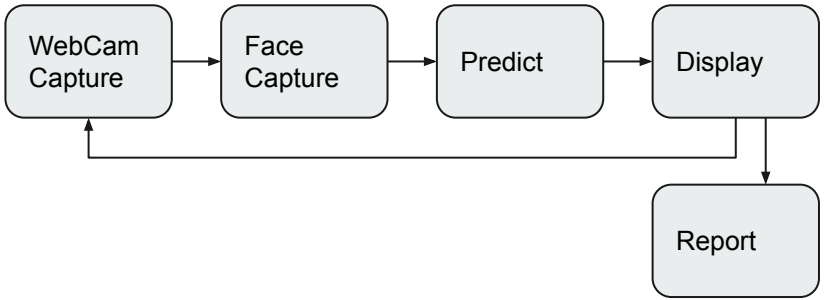


Pipeline

Training



Inference



	Original	haarcascade	dnn	difference
Test	17844	17443	17830	387
Validation	14294	14062	14289	227
Train	53584	53352	53566	214
Total	**85722**	**84857**	**85685**	**828**

Classification Model Approaches

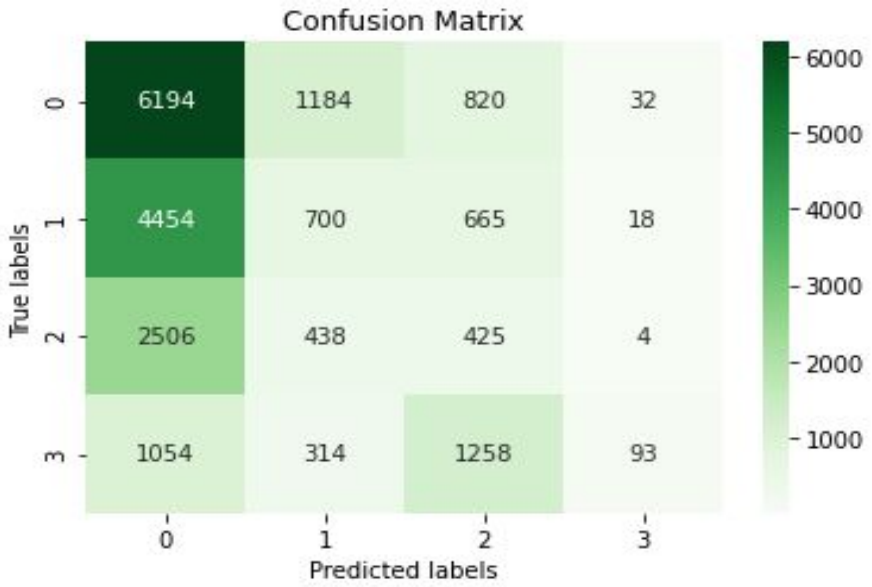
Model Family:	CNN / Transfer Learning	CNN->LSTM / CONVLSTM	Multi Task CNN
Image Extractions:	Extract Boredom Images at 1 FPS	Extract Boredom Images at 2 FPS	Extract All Images at 1 FPS
Whole Images or Faces:	Whole Images; Faces Only (DNN)	Whole Images	Whole Images

Best Model Results

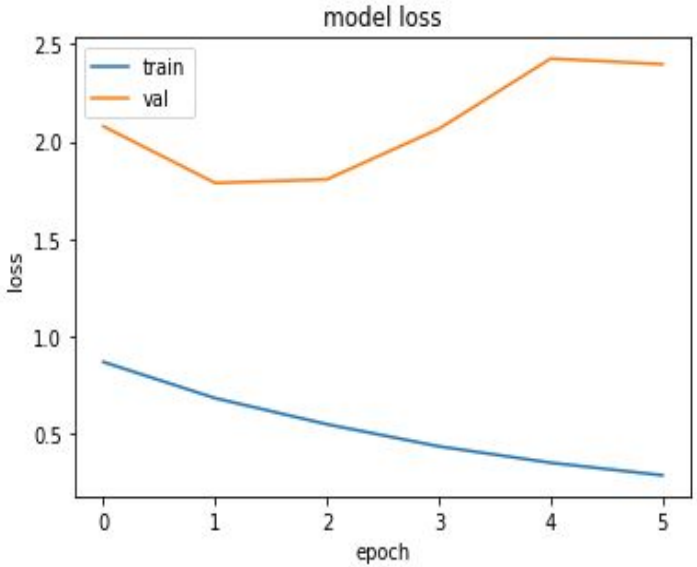
Model	data type	Base Model	frozen layers	epochs	batch size	learning rate	decay	accuracy	val Acc	test acc
CNN	Augmented Faces	MobileNetV2	[:126]	100	32	0.0001	1.00E-06	**0.885**	**0.341**	**0.367**
CONVLSTM	Whole Images	ConvLSTM2D	All	100	16	0.0001	1.00E-06	**0.960**	**0.321**	**0.344**
CNN -> LSTM	Whole Images	MobileNetV2 -> LSTM	All	100	32	0.0001	1.00E-06	**0.884**	**0.311**	**0.382**
Mult Task	Whole Images	Xception	All	10	16	0.0001	1.00E-06	**0.597, 0.658, 0.701, 0.597**	**0.393, 0.461, 0.645, 0.395**	**42.488, 56.013, 66.928, 42.777**

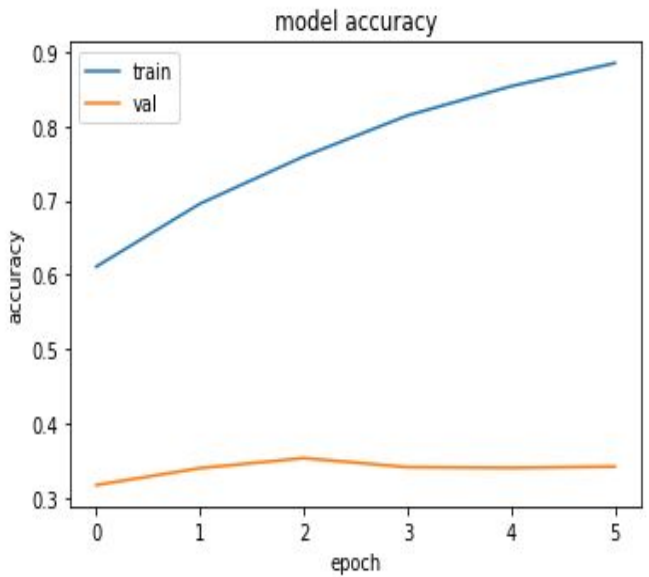
Results CNN B53

- Model: MobileNetV2
- FPS: 1
- Initial Weights: Imagenet
- Face Detect: Yes
- Image Augmentation: B3
- Balanced Data Set: Yes
- Optimizer: Adam
- Batch: 32
- LR: 0.0001
- Decay: 0.000001
- Early Stopping: Yes
- Unfreeze Layers: 126
- **Accuracy: .885**
- **Val Accuracy: .341**
- **Test Accuracy: .267**



Loss and Accuracy





Results MultiTask

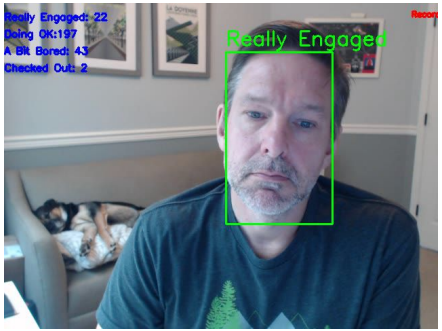
- Model: MobileNetV2
- FPS: 0.7
- Initial Weights: Imagenet
- Face Detect: No
- Image Augmentation: No
- Balanced Data Set: No
- Batch: 16
- LR: 0.0001
- Decay: 0.000001
- Early Stopping: No
- Unfreeze Layers: 126

Top 1 Accuracy:

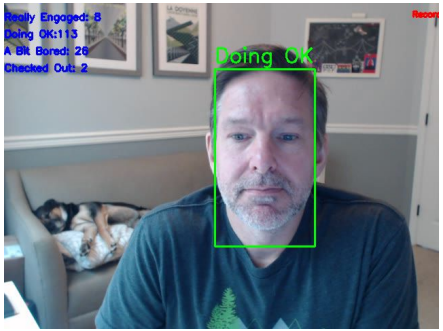
Boredom: 42.4888
Engagement: 56.014
Confusion: 66.928
Frustration: 42.777

Inference - CNN with DNN

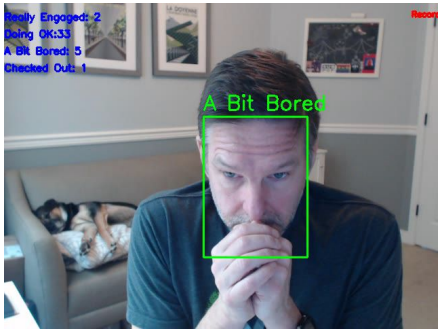
BORED = 0



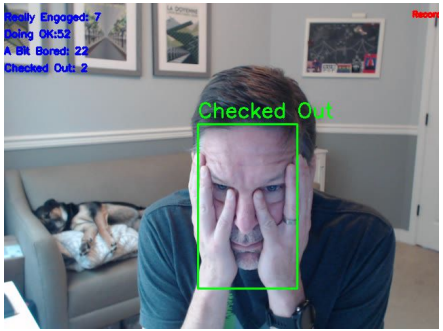
BORED = 1



BORED = 2



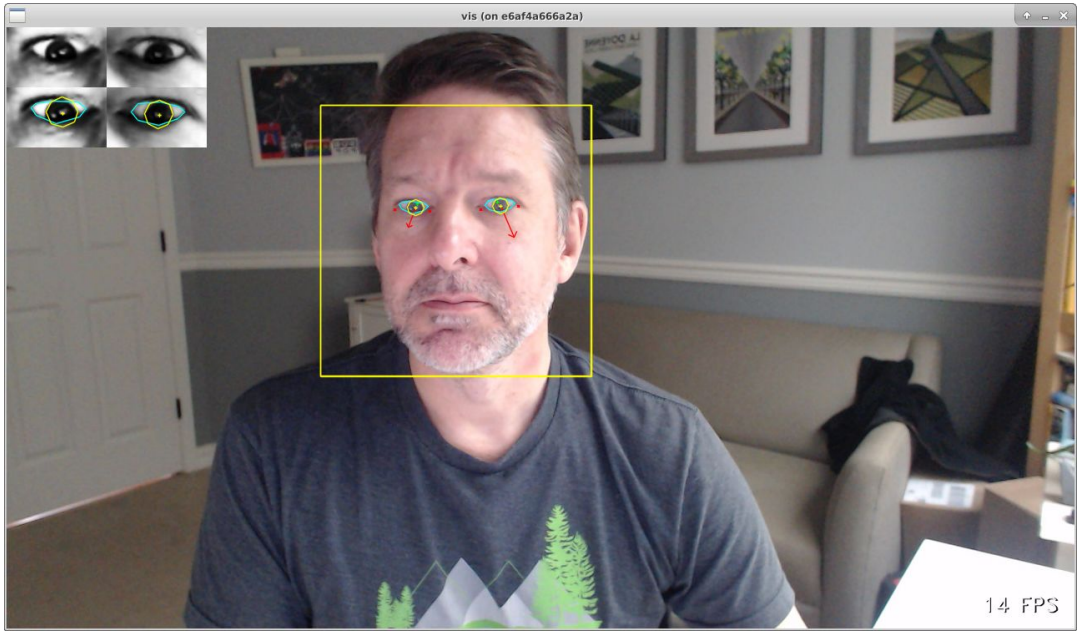
BORED = 3




```
Really Engaged: 27
Doing OK:198
A Bit Bored: 43
Checked Out: 2
```

Inference - Multi Task CNN

Addendum - Gaze Detection



https://github.com/swook/GazeML

Improvements

- Better Models
- Human In the Middle feedback / model update
- Integration of GazeML
- Better Reporting
- Multi-person capture from screen
- Anonymization / Ethical improvements
    - Do not capture student video / id information
    - Align report with teach recorded view

Questions ...

Live Demo!

References

[1] Automatic Recognition of Student Engagement using Deep Learning and Facial Expression, Omid Mohamad Nezami1,2 (□), Mark Dras1, Len Hamey1, Deborah Richards1 Stephen Wan2, and Ce´cile Paris2, 2018, https://arxiv.org/abs/1808.02324

[2] Prediction and Localization of Student Engagement in the Wild, Amanjot kaur, Aamir Mustafa, Love Mehta, Abhinav Dhall, 2018, https://arxiv.org/abs/1804.00858

[3] DAiSEE: Towards User Engagement Recognition in the Wild, Abhay Gupta, Arjun D'Cunha, Kamal Awasthi, Vineeth Balasubramanian, 2016, https://arxiv.org/abs/1609.01885

[4] Gaze360: Physically Unconstrained Gaze Estimation in the Wild, Petr Kellnhofer, Adria` Recasens, Simon Stent2, Wojciech Matusik, and Antonio Torralb, http://gaze360.csail.mit.edu/iccv2019_gaze360.pdf

[5] Learning to Find Eye Region Landmarks for Remote Gaze Estimation in Unconstrained Settings, Seonwook Park, Xucong Zhang, Andreas Bulling, Otmar Hilliges, 2018, https://ait.ethz.ch/projects/2018/landmarks-gaze/

Challenges

Issue	Mitigation
No docker image support for DNN	Compile and install OpenCV from source
No docker image support for Tensorflow 2	Compile and install Tensorflow 2 from source
Model Sizes too big for Jetson	Use MobileNetV2
General Memory Errors	Configure Tensorflow to grab more memory as needed
Unbalanced training set	Augment Images for class three Reduce # of images for class zero
Base models not designed for engagement	Unfreeze later layers for training
RNN Inference	Loop over frame capture for 20 frames
Gaze ML - inconsistent, TF1, complicated inference code (demo only)	Convert to TF2