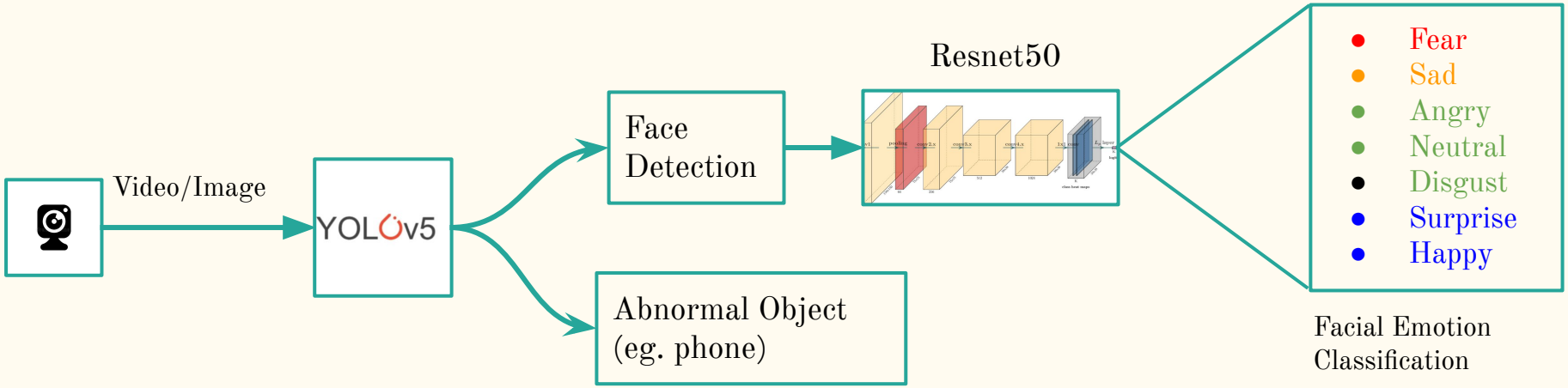
Proctor exams using AI

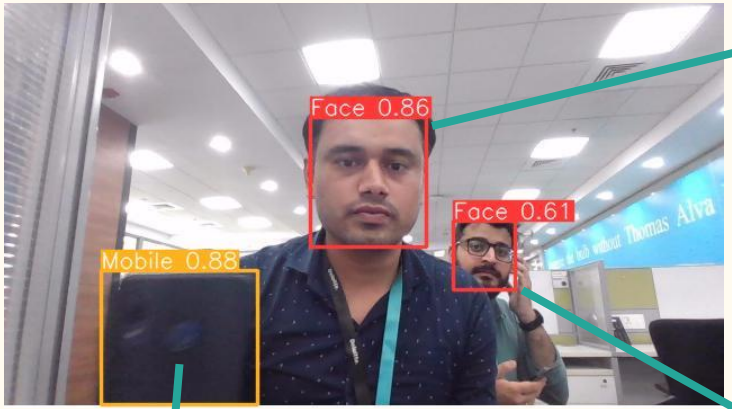
Fengwei Liu, Ruohua Li

Pipeline Overview

- Object Detection + Image Classification

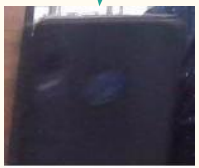


Demo



Student 0 emotion: Fear

Student 1 emotion: Neutral



Smartphone detected!

YOLOv5 (Object Detection)

- Why Yolo?
  - Performance:
    - High benchmark score(COCO)[1]
  - Complexity:
    - Easy to implement, train, and inference
  - Dataset:
    - Able to handle small dataset(around 1000 images)
    - Able to handle imbalanced classes
  - Time:
    - Less than 1 min for 1 epoch
    - Less than 0.5 second for inferencing one image
- Face Detection
  - (Online exam) count number of person shown in the camera
  - Facial emotion classification
- Phone Detection
  - Smartphone is not allow in most exams
- More objects could be added in the future

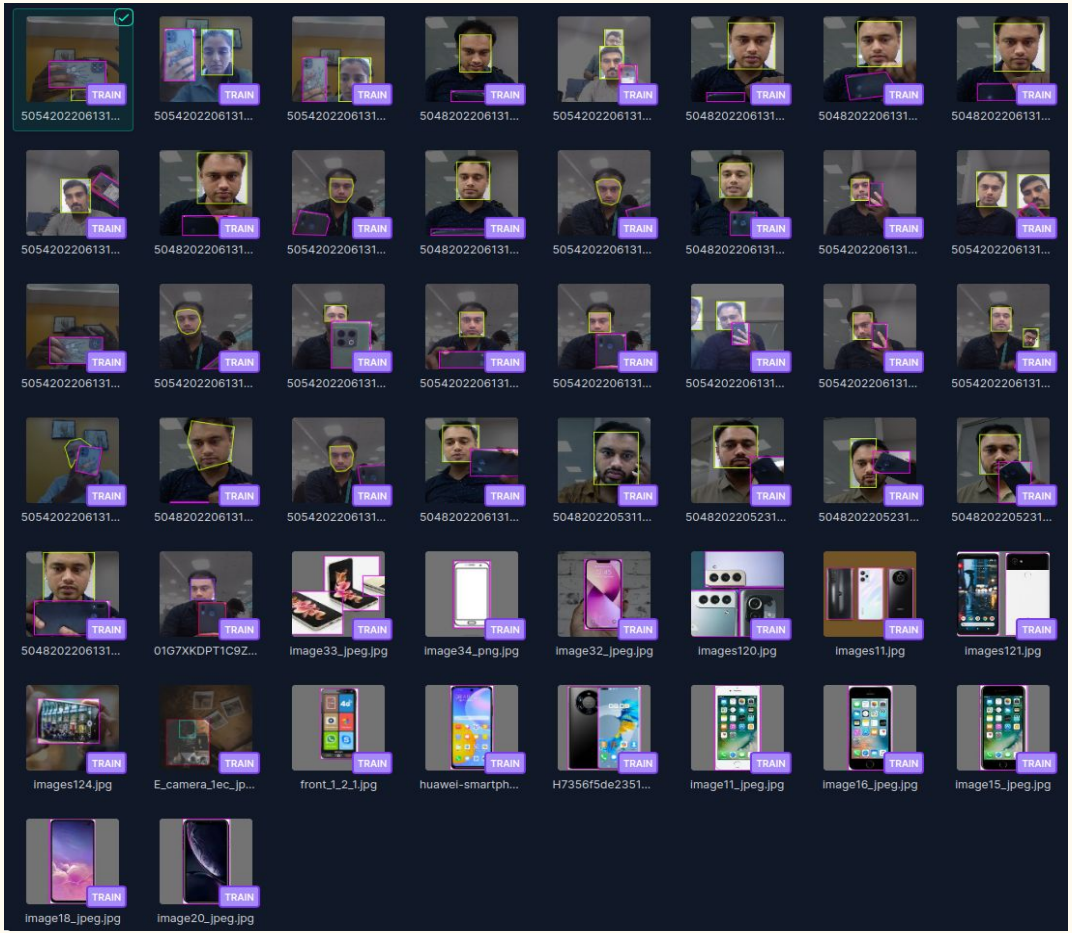
Method	Backbone	Size	FPS (V100)	AP (%)	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
YOLOv3 + ASFF* [18]	Darknet-53	608	45.5	42.4	63.0	47.4	25.5	45.7	52.3
YOLOv3 + ASFF* [18]	Darknet-53	800	29.4	43.9	64.1	49.2	27.0	46.6	53.4
EfficientDet-D0 [28]	Efficient-B0	512	98.0	33.8	52.2	35.8	12.0	38.3	51.2
EfficientDet-D1 [28]	Efficient-B1	640	74.1	39.6	58.6	42.3	17.9	44.3	56.0
EfficientDet-D2 [28]	Efficient-B2	768	56.5	43.0	62.3	46.2	22.5	47.0	58.4
EfficientDet-D3 [28]	Efficient-B3	896	34.5	45.8	65.0	49.3	26.6	49.4	59.8
PP-YOLOv2 [11]	ResNet50-vd-dcn	640	68.9	49.5	68.2	54.4	30.7	52.9	61.2
PP-YOLOv2 [11]	ResNet101-vd-dcn	640	50.3	50.3	69.0	55.3	31.6	53.9	62.4
YOLOv4 [1]	CSPDarknet-53	608	62.0	43.5	65.7	47.3	26.7	46.7	53.3
YOLOv4-CSP [30]	Modified CSP	640	73.0	47.5	66.2	51.7	28.2	51.2	59.8
YOLOv3-ultralytics[2]	Darknet-53	640	95.2	44.3	64.6	-	-	-	-
YOLOv5-M [7]	Modified CSP v5	640	90.1	44.5	63.1	-	-	-	-
YOLOv5-L [7]	Modified CSP v5	640	73.0	48.2	66.9	-	-	-	-
YOLOv5-X [7]	Modified CSP v5	640	62.5	50.4	68.8	-	-	-	-
YOLOX-DarkNet53	Darknet-53	640	90.1	47.4	67.3	52.1	27.5	51.5	60.9
YOLOX-M	Modified CSP v5	640	81.3	46.4	65.4	50.6	26.3	51.0	59.9
YOLOX-L	Modified CSP v5	640	69.0	50.0	68.5	54.5	29.8	54.5	64.4
YOLOX-X	Modified CSP v5	640	57.8	**51.2**	69.6	55.7	31.2	56.1	66.1

Datasets

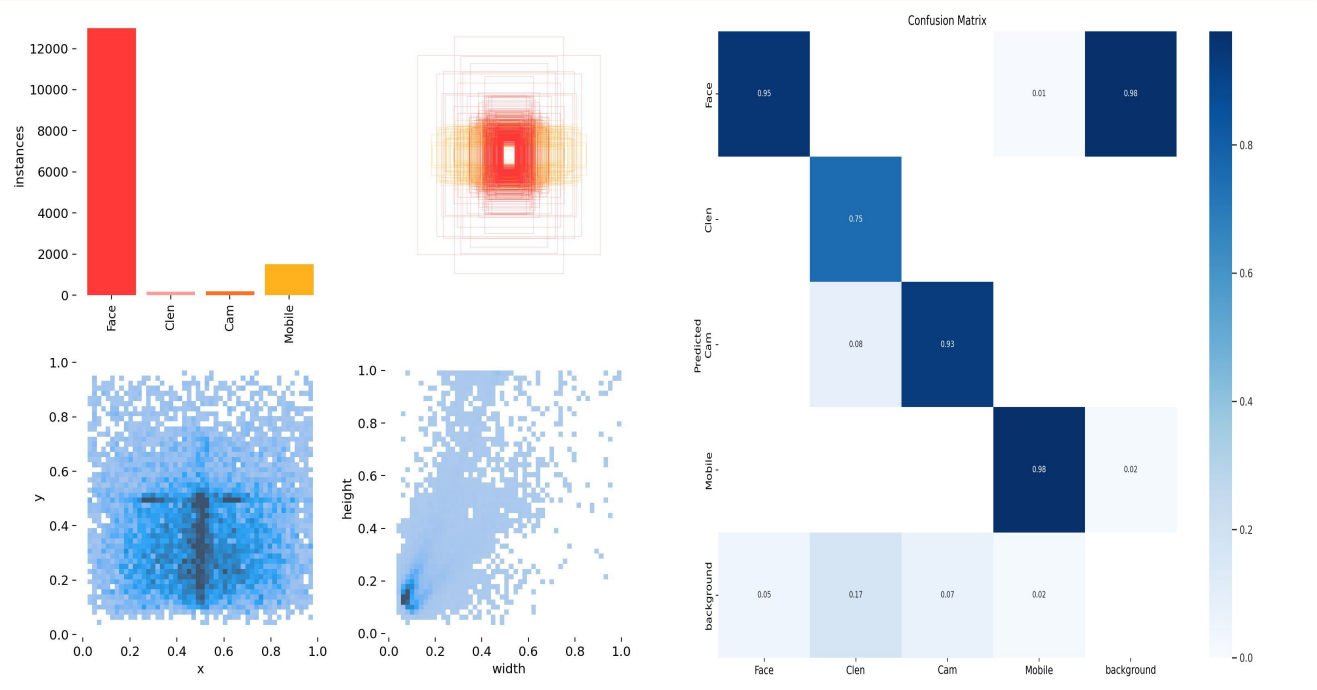
- WIDER FACE:
  - This is a face detection benchmark dataset, of which images are selected from the publicly available WIDER dataset. It has 32,203 images and label 393,703 faces with a high degree of variability in scale, pose and occlusion as depicted in the sample images. (7366 training images, 1856 validation images, 16098 test images after preprocessing)
- Tustrain Computer Vision Project(roboflow):
  - An open source Computer Vision Project with a dataset that has 1018 training images, 133 validation images, and 22 test images.

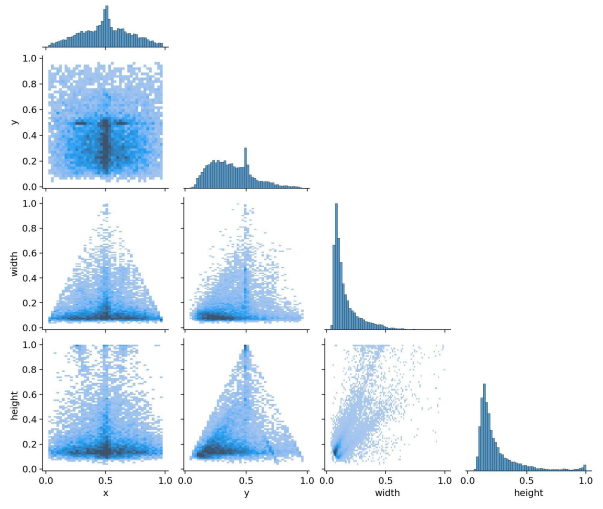
WIDER FACE

Tustrain Computer Vision Project

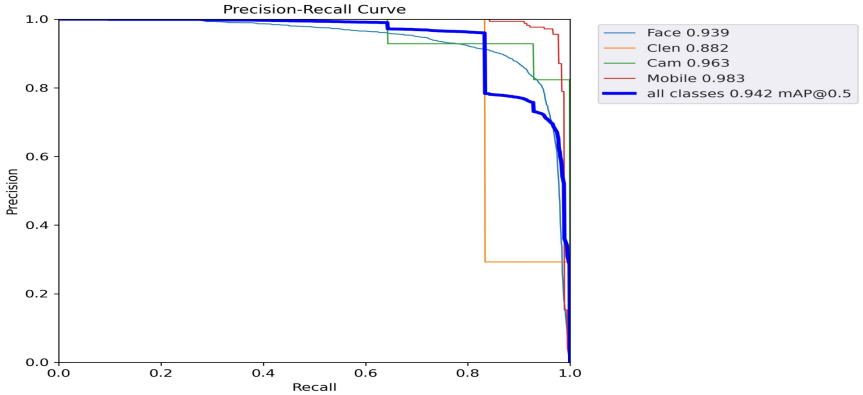


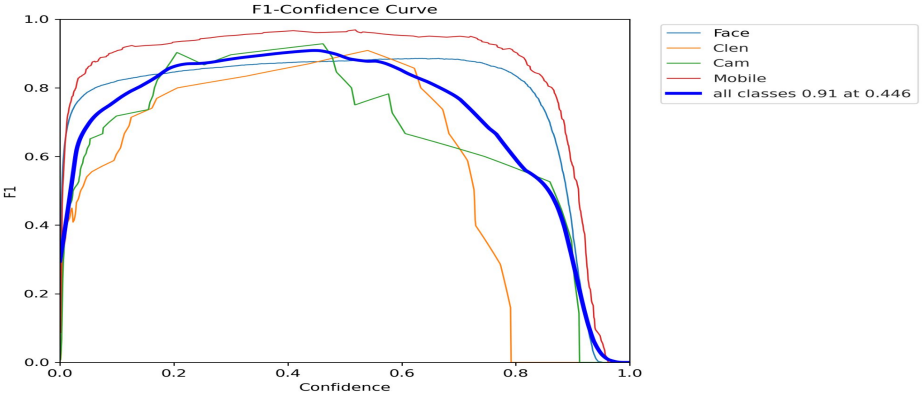
Labels and correlation map

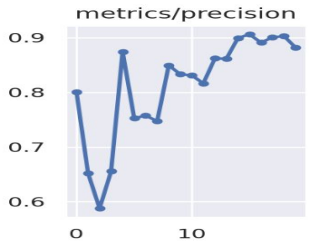


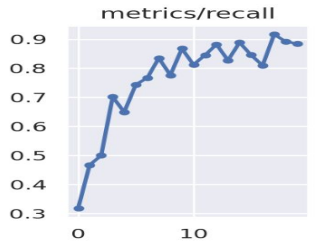


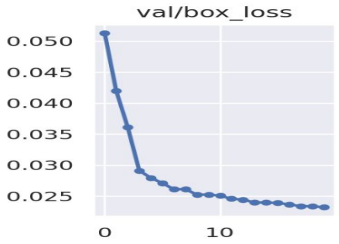
YOLOv5 Result

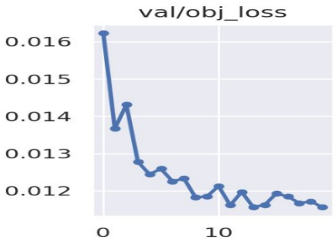


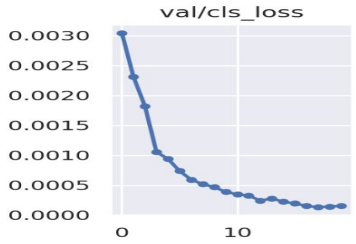


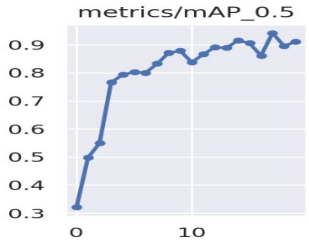


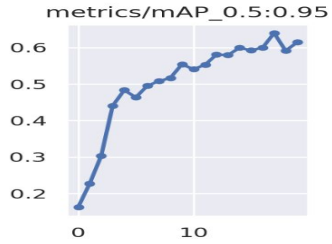








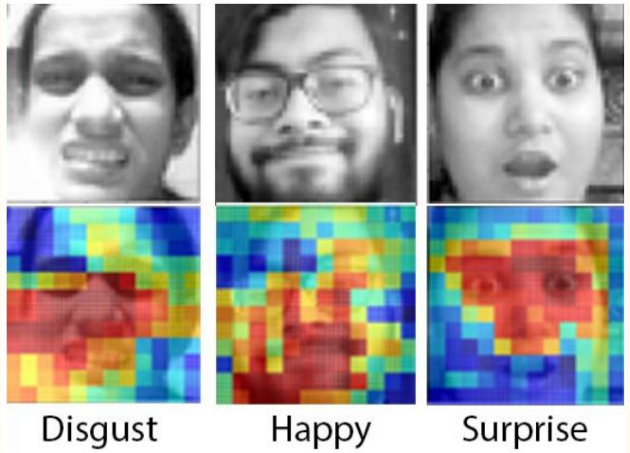




Facial Emotion Classification

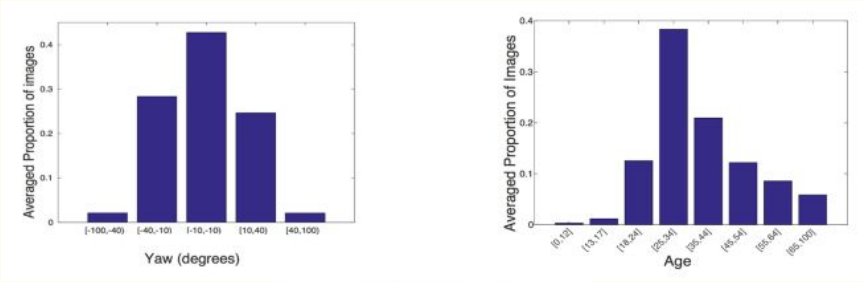
- Cheaters are more likely to show fearful emotion [2]

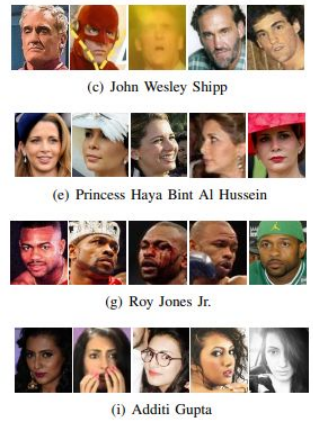
- Aim to spot micro movement on the face
  - Help proctor to make a decision
  - Pixels assigned warmer colors have greater importance in the prediction. (Figure on the right)[3]



Dataset

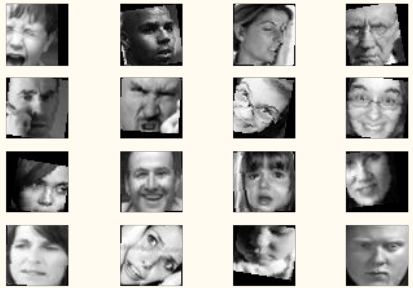
- VGGFace2 (Pre-trained)[4]
  - 3.31 million images of 9131 subjects
  - Wide range of ethnicities, professions, gender, pose, and ages
  - Some bias exist in the dataset
  - Pre-trained from scratch using Resnet50
- fer2013 (fine-tuned)[5]
  - 35886 images of 7 emotions (Angry, Disgust, Fear, Happy, Sad, Surprise, Neutral)
  - 48x48 pixel grayscale images of faces
  - 28709 training, 3589 validation, 3589 test



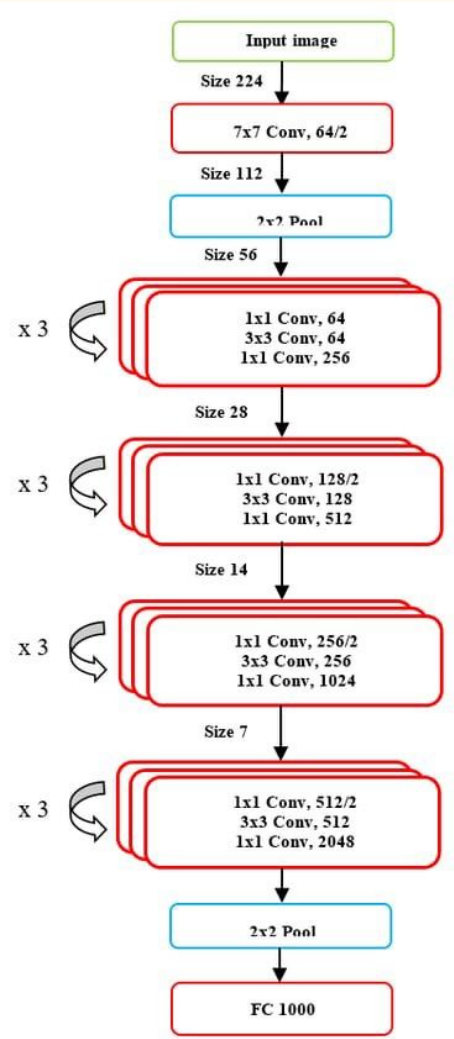


Model Selection/Training

- Resnet50 with some tweak
  - Change the first layer of conv into 3x3 with padding
  - Get rid of first maxpool layer
  - Change the last fc layer: add dropout to prevent overfitting
- Data Augmentation
  - Crop, Flip, Rotation

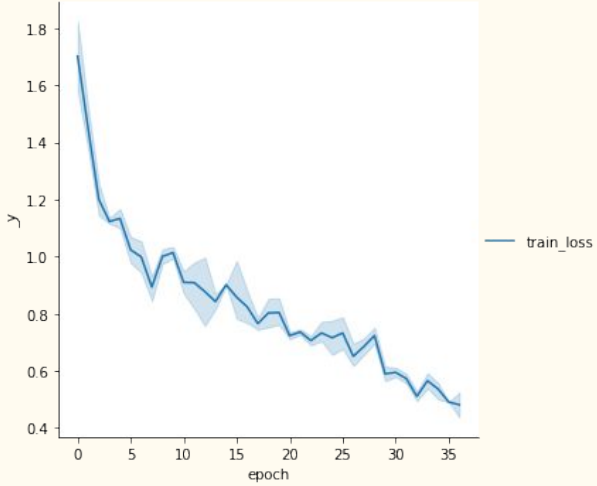


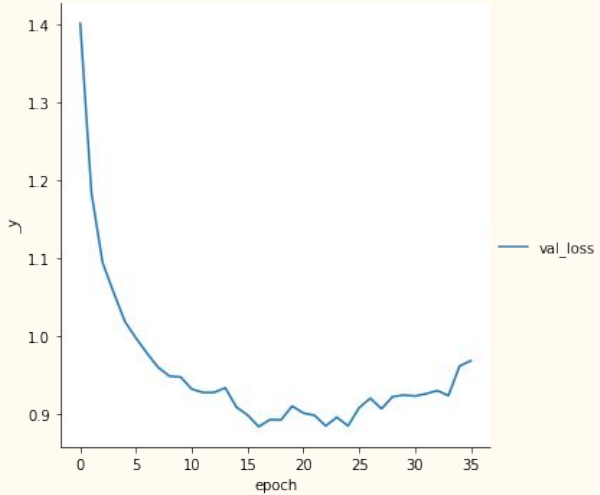
- SGD + ReduceLROnPlateau

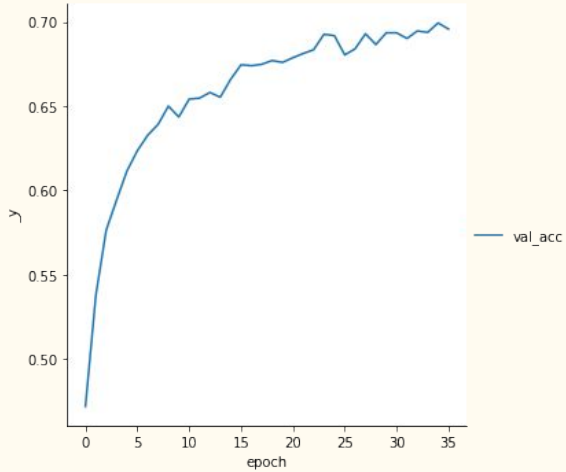


Result

- Human performance on fer2013: 65±5% accuracy[3]
- Our model: 71±1% accuracy (test)
- Some overfitting issue

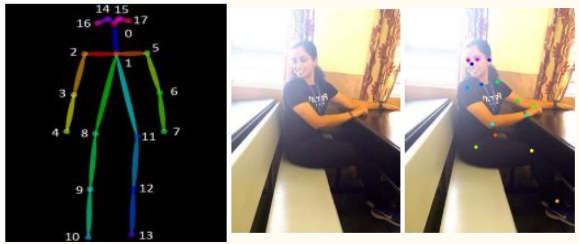


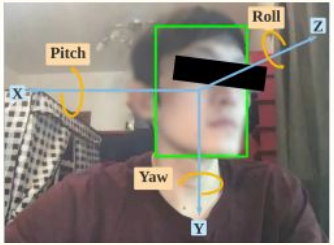


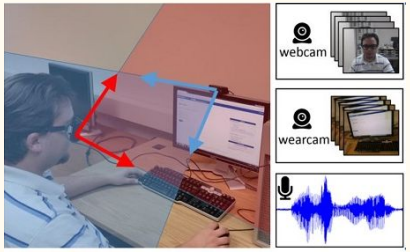


Limitation/Future Work

- Lack of “real-world” dataset
  - Students’ privacy issue
- Emotion detection is not a foolproof
  - Other checks are needed
- More than just image
  - Posture Analysis[2], Abnormal Head Movement[6], Multi-model integration [7]







Reference

[1]https://doi.org/10.48550/arXiv.2107.08430

[2] J. Nishchal, S. Reddy and P. N. Navya, "Automated Cheating Detection in Exams using Posture and Emotion Analysis," *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, 2020, pp. 1-6, doi: 10.1109/CONECCT50063.2020.9198691.

[3] http://cs230.stanford.edu/projects_winter_2020/reports/32610274.pdf

[4] https://arxiv.org/pdf/1710.08092.pdf

[5] https://www.kaggle.com/datasets/deadskull7/fer2013

[6]https://arxiv.org/pdf/2101.07990.pdf

[7] http://cvlab.cse.msu.edu/project-OEP.html