FACIAL EMOTION RECOGNITION

Group 25: Ashwin Unnikrishnan, Abhishek Chandra, Shamekh Siddiqui

PROBLEM STATEMENT

- The ability to recognize emotional facial expressions is essential for good interpersonal communication.
- Facial Emotion Recognition has plenty of uses cases like
  - Smarter border control for screening
  - Healthcare – Prioritizing patients
  - Marketing – Analyzing focus groups
  - Alerting drivers
- Implemented CNN models to tackle this issue with comparison analysis
- On the frontend side, we have implemented a live image classification model based on ResNet.

Dataset

- This FER2013 kaggle dataset contains images with 7 labels categories:
  - {Angry, Fear, Disgust, Sad, Neutral, Surprise, Happy}}
- The images are 48x48(width, height) grayscale images, a sample batch shown.
- Preprocessing: horizontal flipping, reshape, normalizing, feature embeddings (toTensor)
- Live Video Data transformation includes:
  - capturing video stream and reshaping into cropped images of size 48x48
  - converting into grayscale

Sample Batch (64)

Training Dataset Summary

```
Count of label Neutral is 4965
Count of label Happy is 7215
Count of label Angry is 3995
Count of label Fear is 4097
Count of label Sad is 4830
Count of label Surprise is 3171
Count of label Disgust is 436
```

Testing Dataset Summary

```
Count of label Angry is 491
Count of label Disgust is 55
Count of label Fear is 528
Count of label Happy is 879
Count of label Sad is 594
Count of label Surprise is 416
Count of label Neutral is 626
```

Validation Dataset Summary

```
Count of label Angry is 467
Count of label Disgust is 56
Count of label Fear is 496
Count of label Happy is 895
Count of label Sad is 653
Count of label Surprise is 415
Count of label Neutral is 607
```

MODELS

- **LeNet**
  - First CNN model developed to classify handwritten digits, hence suitable for 7 labels classification. It uses Relu as the activation layers.
  - As our input is 48x48 we have used 3 Layered LeNet with two convolution layers followed by 1 fully connected layer. (large resolution issue)
- **AlexNet**
  - Deeper than LeNet, with more filters per layer, and with stacked convolutional layers. It consisted 11x11, 5x5,3x3 convolutions.
  - It deals with overfitting by using "dropout" rather than regularisation. Accepts 224x224, so we upscaled the image from 48x48 to the respective size.
- **VGG**
  - Instead of using large receptive fields like AlexNet (11x11 with a stride of 4), VGG uses very small receptive fields (3x3 with a stride of 1).
  - Because there are now three ReLU units instead of just one, the decision function is more discriminative
- **RESNET**
  - As we make the CNN deeper, the derivative when back-propagating to the initial layers becomes almost insignificant in value.
  - As a result the ResNet was developed to resolve the vanishing gradient problem. With ResNets, the gradients can flow directly through skip connections backwards from later layers to the initial filters.

	LeNet	**AlexNet**	**VGG**	**ResNet**
Layers	3	5	11	9
Learning Rate	0.01	0.01	0.05 (Decaying)	0.001(OneCycleLR)
Epochs	50	50	50	50
Batch Size	32	64	64	64
Optimizer	SGD	Adam	SGD	Adam, SGD
Loss Function	CrossEntropy	CrossEntropy	CrossEntropy	CrossEntropy

Results & Conclusion

LeNet

Test accuracy: 46%

	precision	recall	f1-score	support
0	0.40	0.24	0.30	491
1	0.00	0.00	0.00	55
2	0.34	0.17	0.23	528
3	0.63	0.72	0.67	879
4	0.32	0.40	0.36	594
5	0.70	0.43	0.53	416
6	0.38	0.62	0.47	626
accuracy			0.46	3589
macro avg	0.40	0.37	0.37	3589
weighted avg	0.46	0.46	0.44	3589

AlexNet

Test accuracy: 56%

	precision	recall	f1-score	support
0	0.51	0.41	0.45	491
1	0.81	0.47	0.60	55
2	0.47	0.42	0.44	528
3	0.69	0.80	0.74	879
4	0.37	0.43	0.40	594
5	0.74	0.66	0.70	416
6	0.52	0.50	0.51	626
accuracy			0.56	3589
macro avg	0.59	0.53	0.55	3589
weighted avg	0.56	0.56	0.55	3589

VGG

Test accuracy: 70%

	precision	recall	f1-score	support
0	0.59	0.69	0.64	467
1	0.87	0.59	0.70	56
2	0.66	0.49	0.56	496
3	0.87	0.89	0.88	895
4	0.61	0.52	0.56	653
5	0.84	0.82	0.83	415
6	0.56	0.71	0.62	607
accuracy			0.70	3589
macro avg	0.71	0.67	0.68	3589
weighted avg	0.70	0.70	0.69	3589

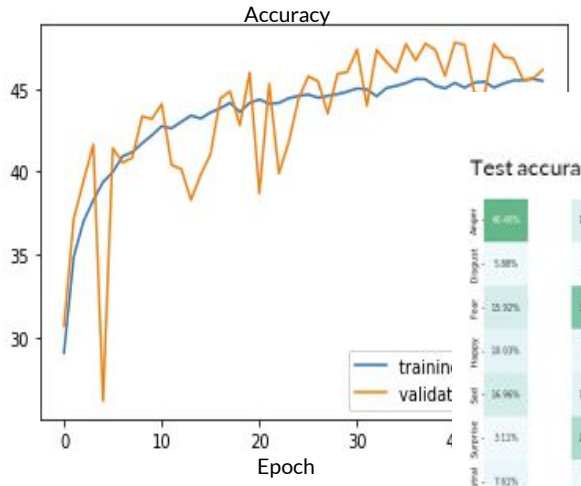
ResNet

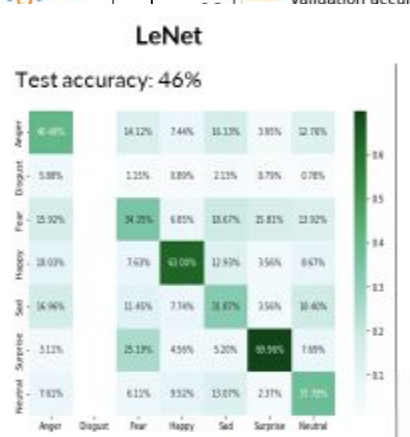
Test accuracy: 67%

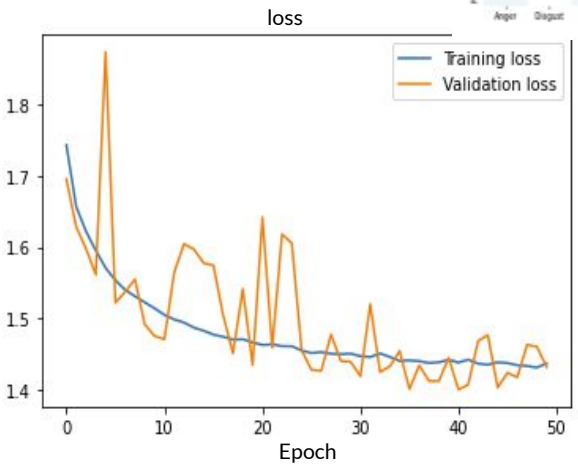
	precision	recall	f1-score	support
Angry	0.64	0.54	0.59	491
Disgust	0.59	0.47	0.53	55
Fear	0.59	0.35	0.44	528
Happy	0.84	0.91	0.87	879
Sad	0.51	0.59	0.55	594
Surprise	0.77	0.79	0.78	416
Neutral	0.61	0.73	0.66	626
accuracy			0.67	3589
macro avg	0.65	0.63	0.63	3589
weighted avg	0.67	0.67	0.66	3589

Results & Conclusion

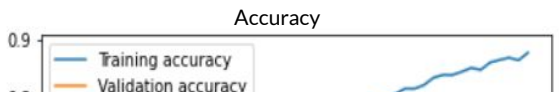
LeNet

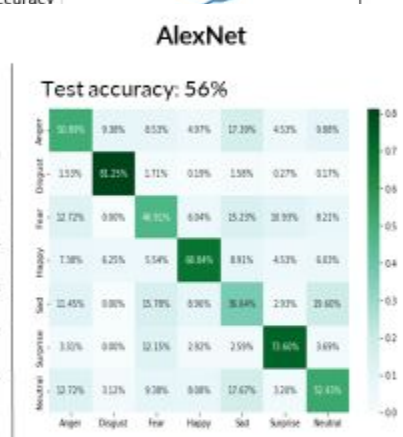


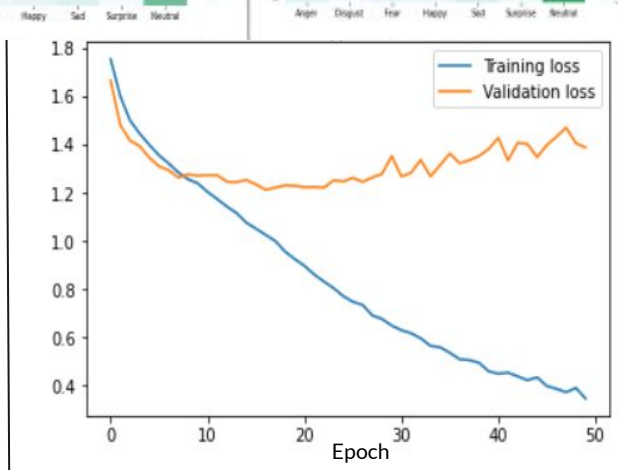




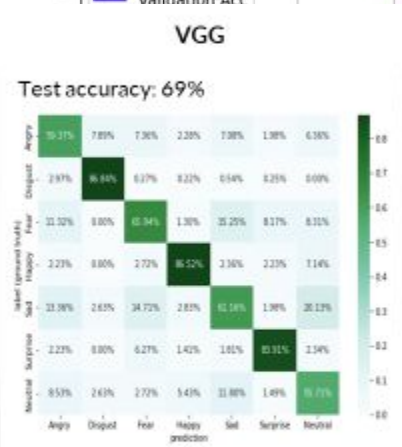
AlexNet

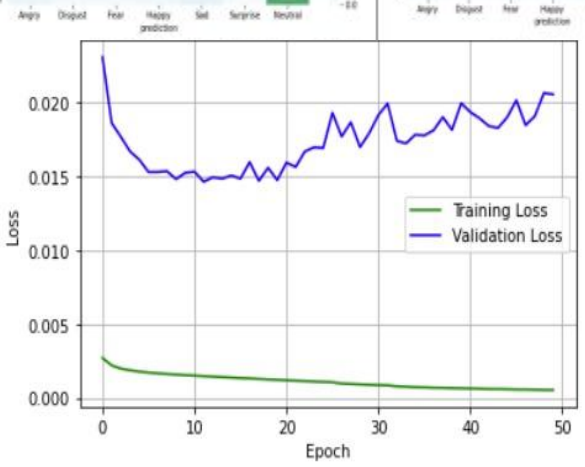




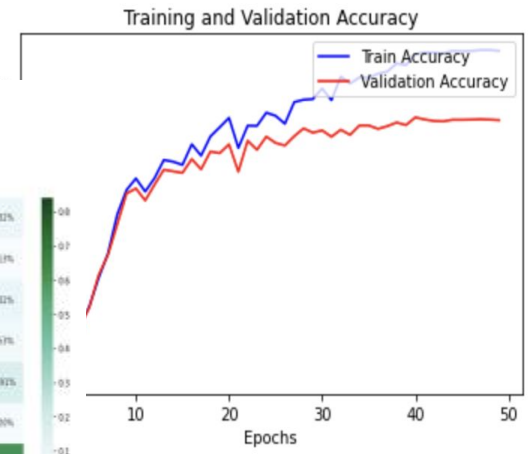


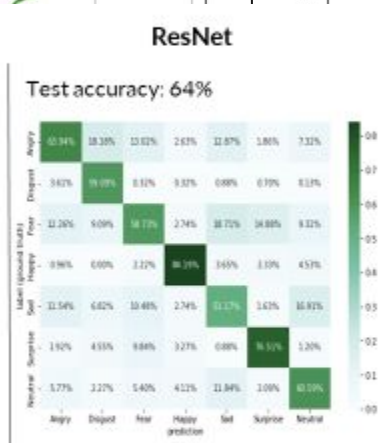
Vgg

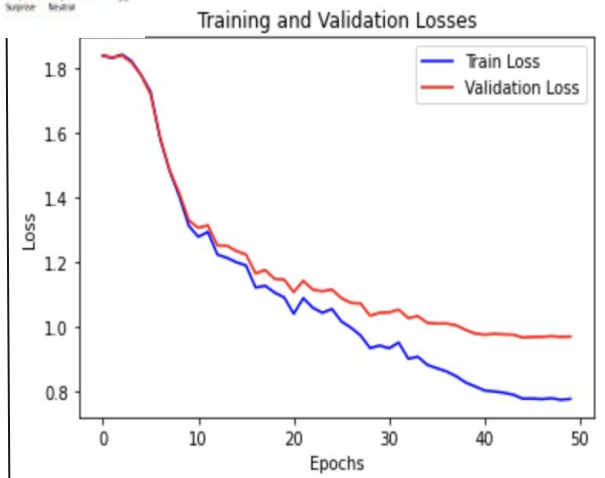




ResNet



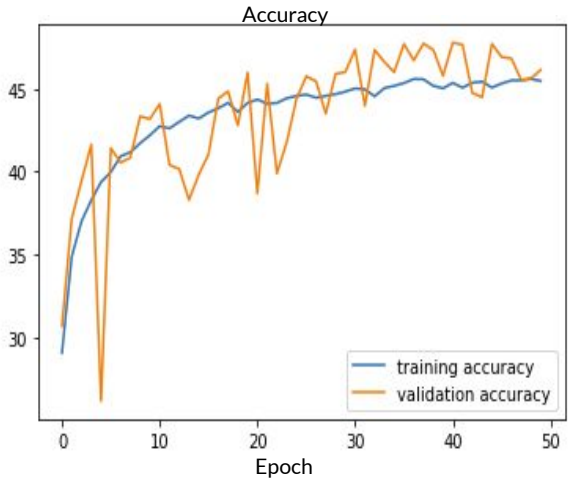


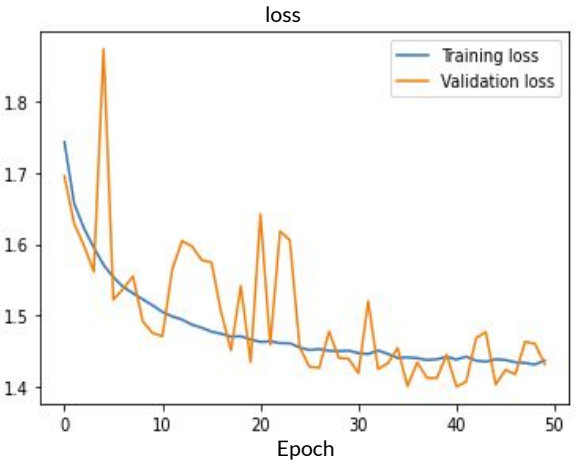


	LeNet	AlexNet	VGG	ResNet
Training Accuracy	44%	89%	87%	72%
Validation Accuracy	46%	58%	70%	65%
Testing Accuracy	46%	56%	70%	67%
Precision	0.40	0.59	0.71	0.65
Recall	0.37	0.53	0.67	0.63
F1-Score	0.37	0.55	0.68	0.63

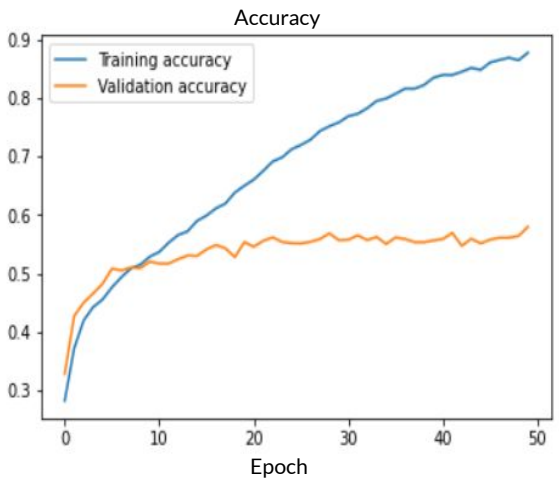
Results & Conclusion

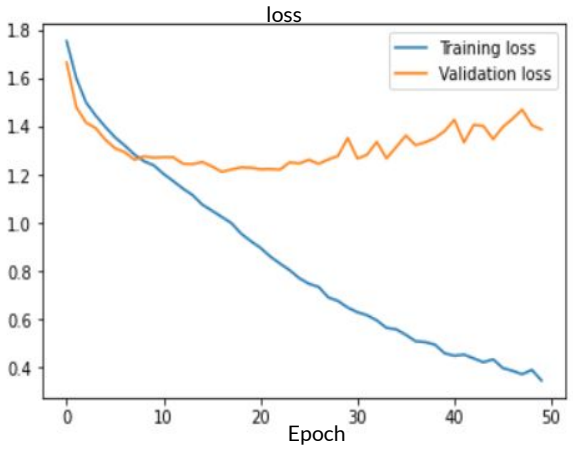
LeNet



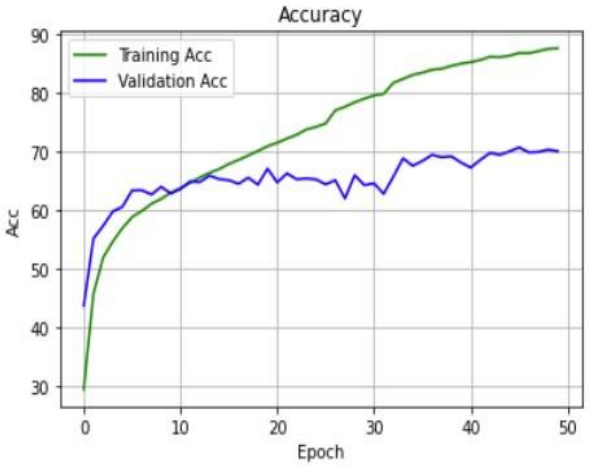


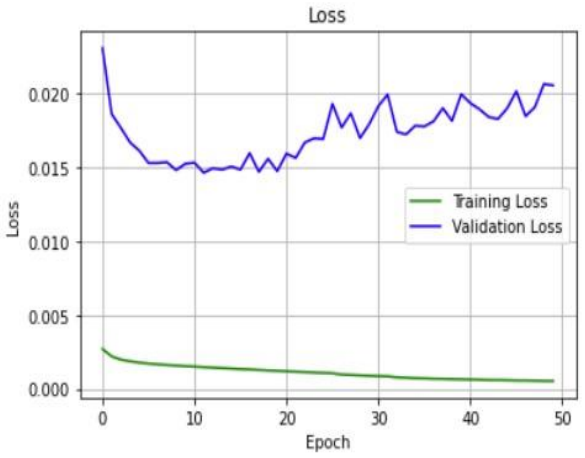
AlexNet



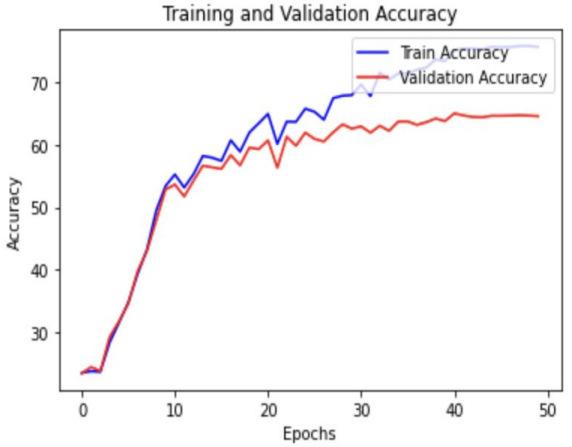


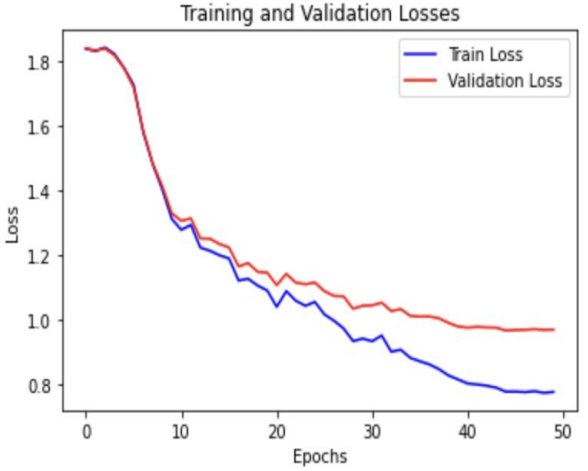
Vgg





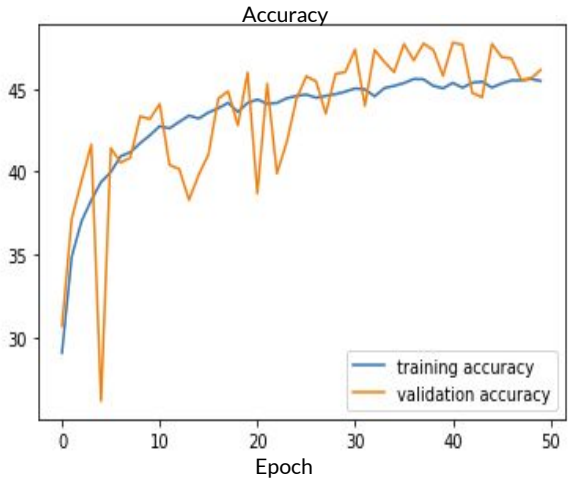
ResNet

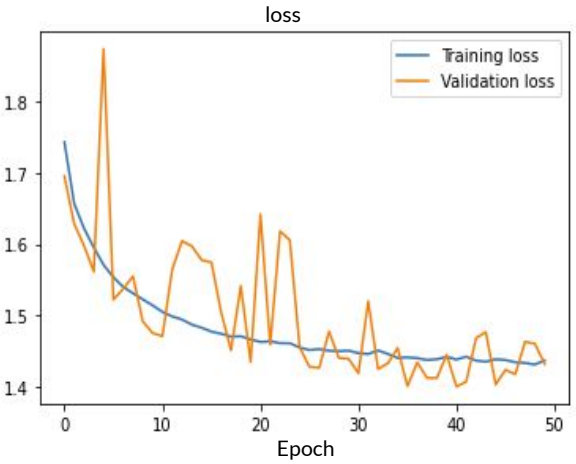




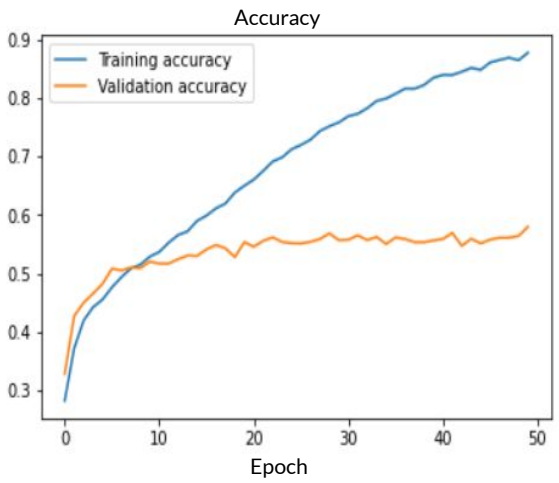
Results & Conclusion

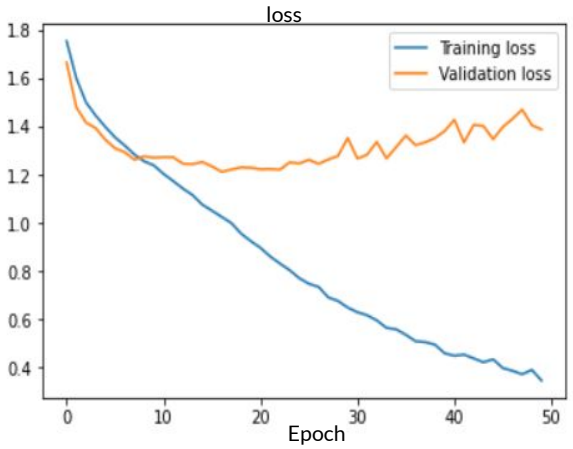
LeNet



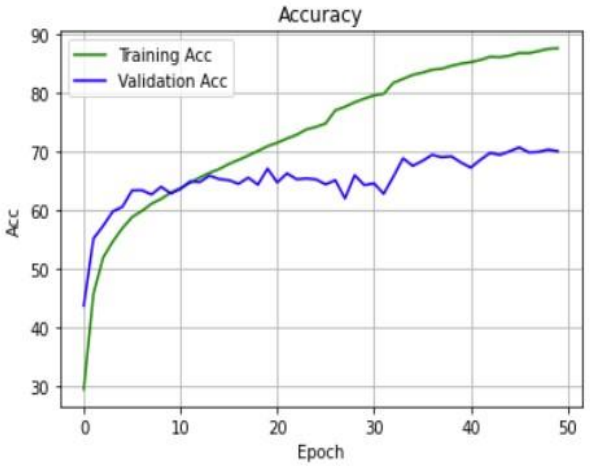


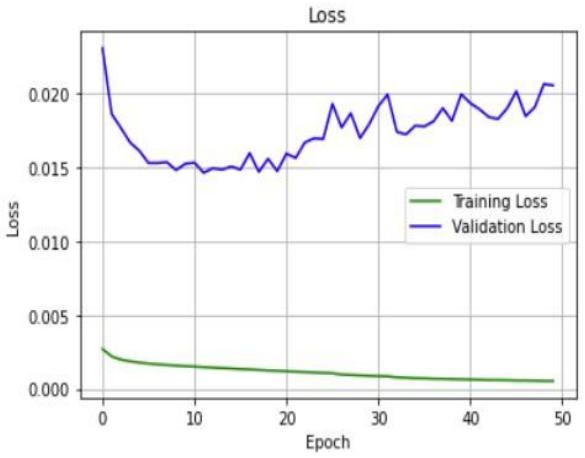
AlexNet



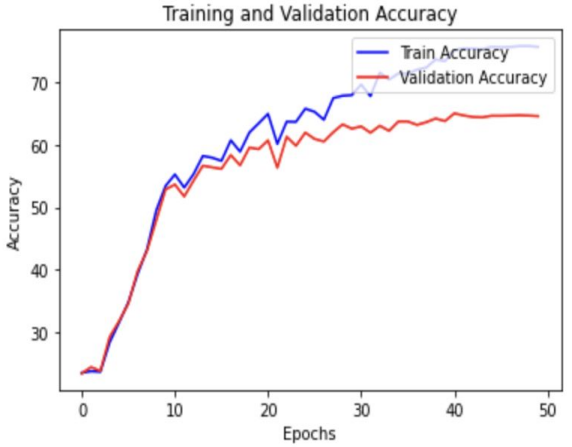


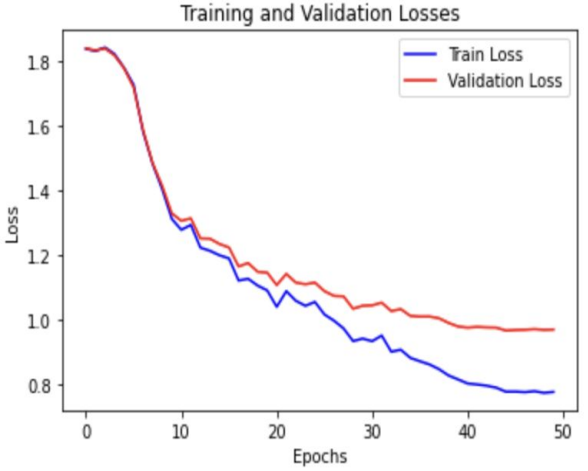
Vgg





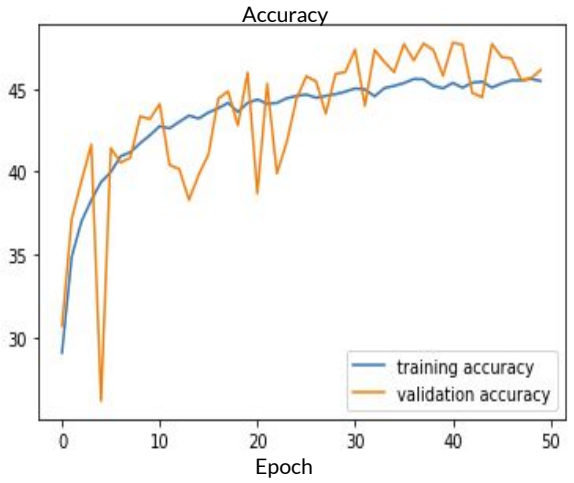
ResNet

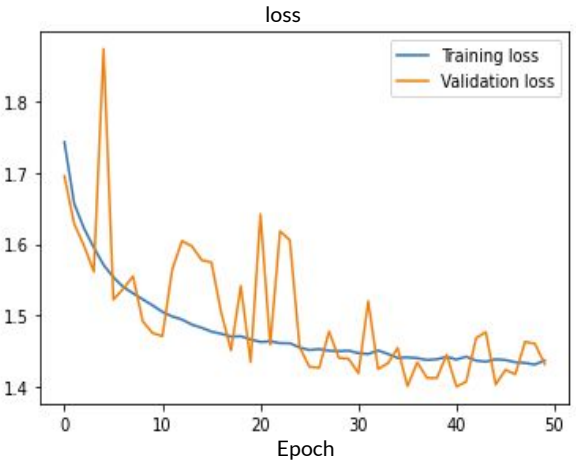




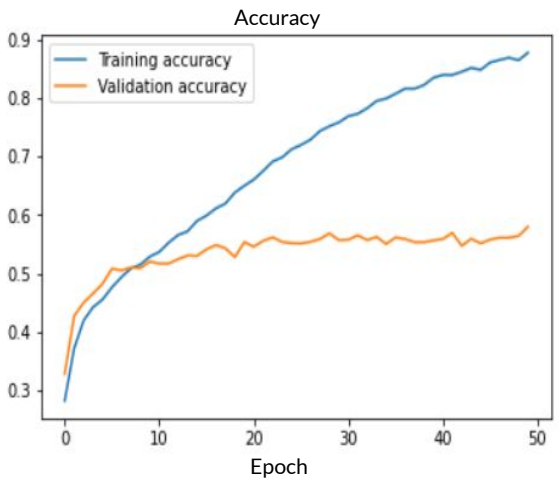
Results & Conclusion

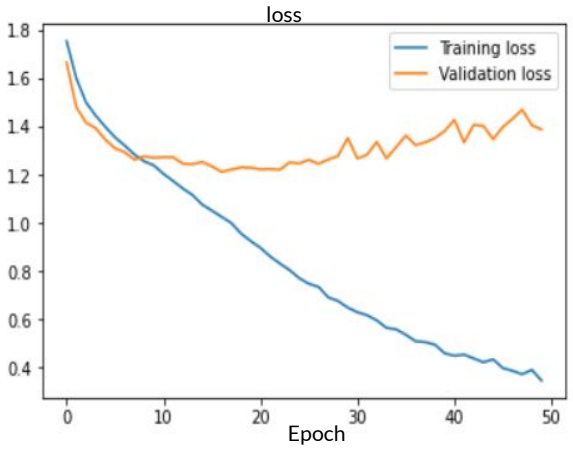
LeNet



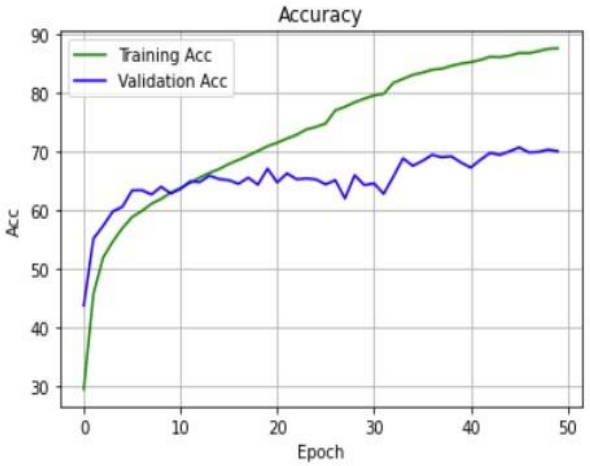


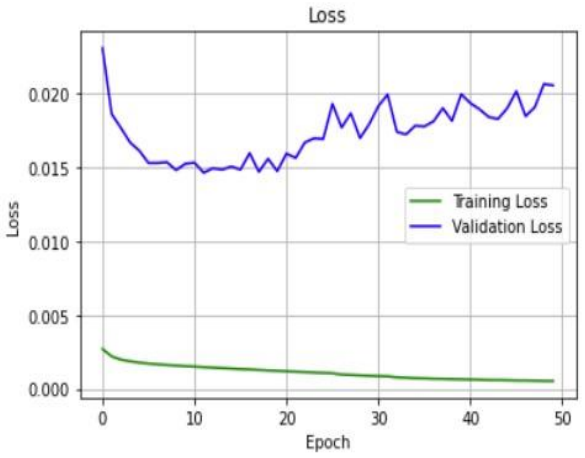
AlexNet



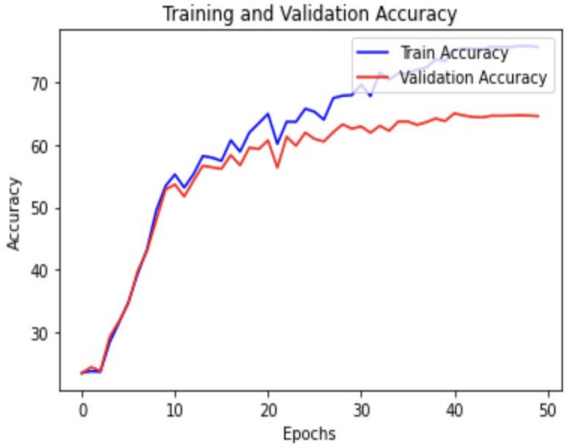


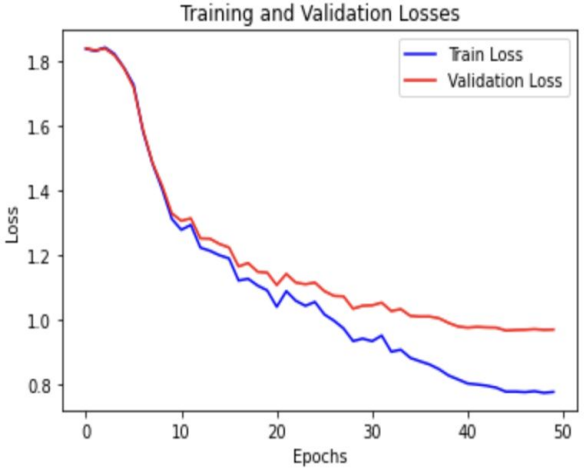
Vgg



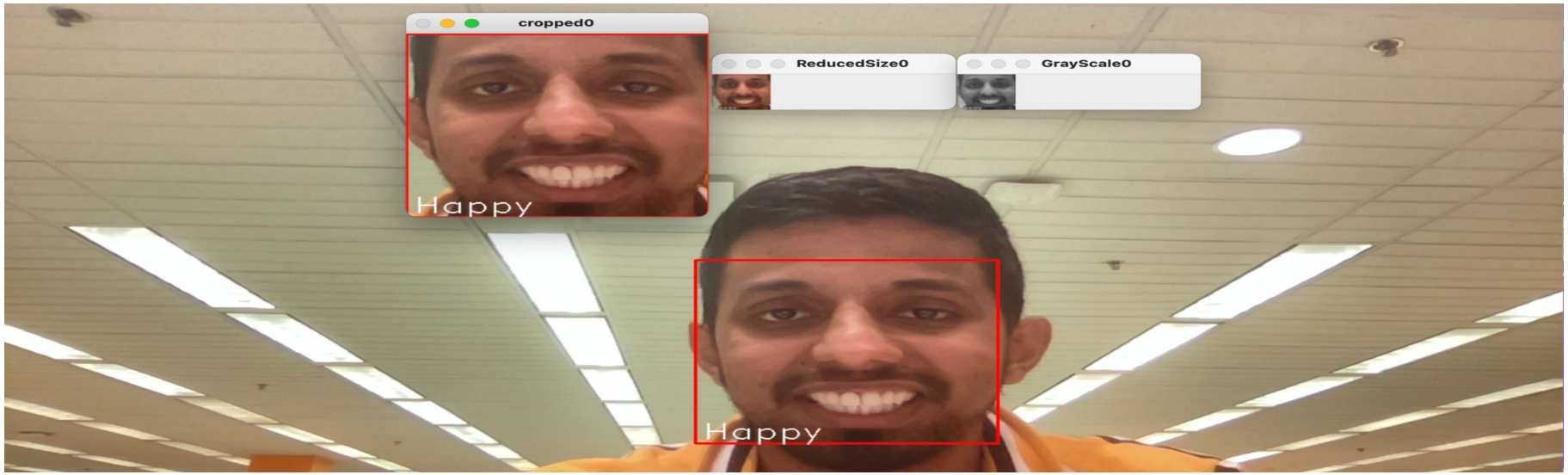


ResNet

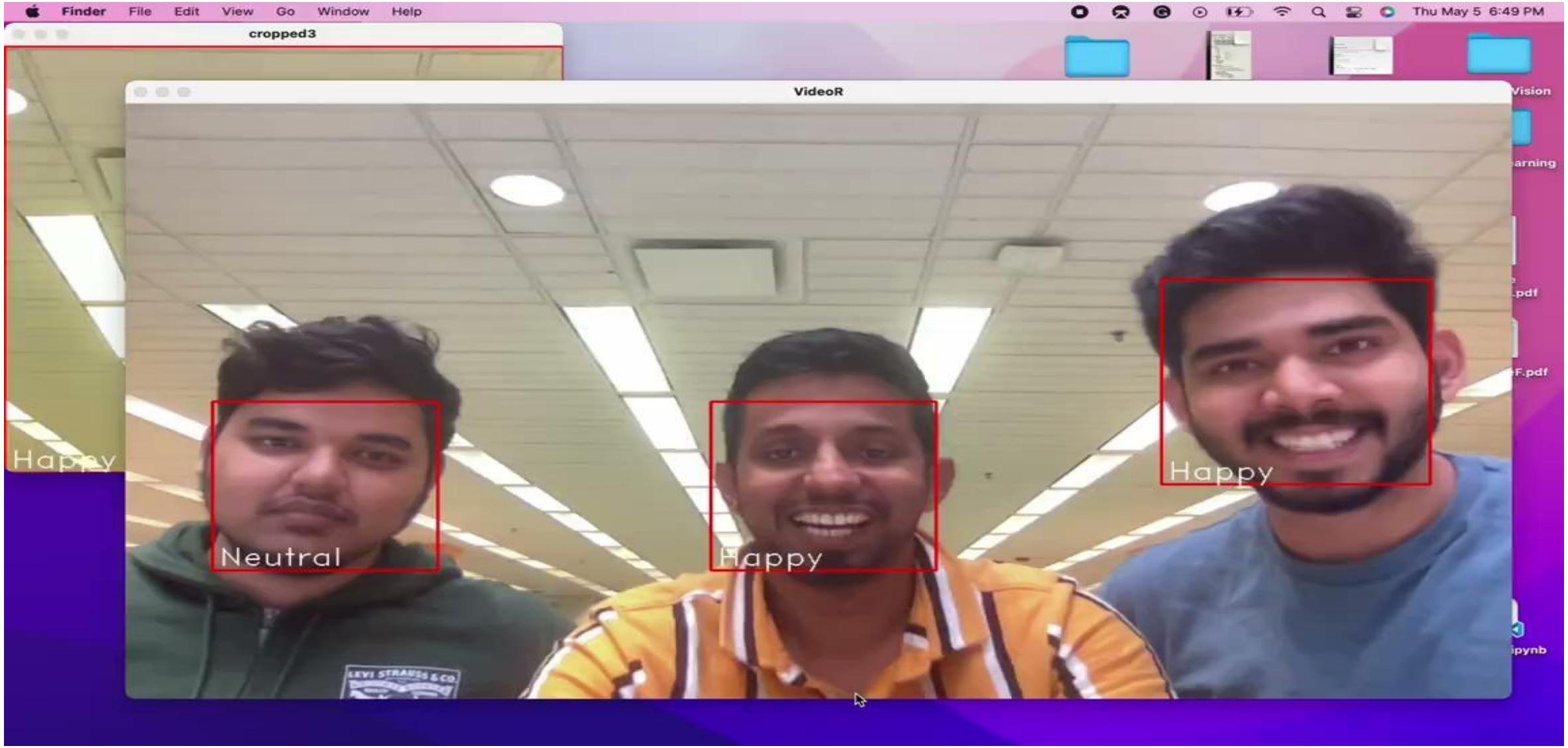




Live Video Transformation



Live Facial Expression Detection



THANK YOU