Team Number 18

Deep Age

Better Approaches for predicting age from images

By:

Srikanth Babu Mandru

Why Age from Images?

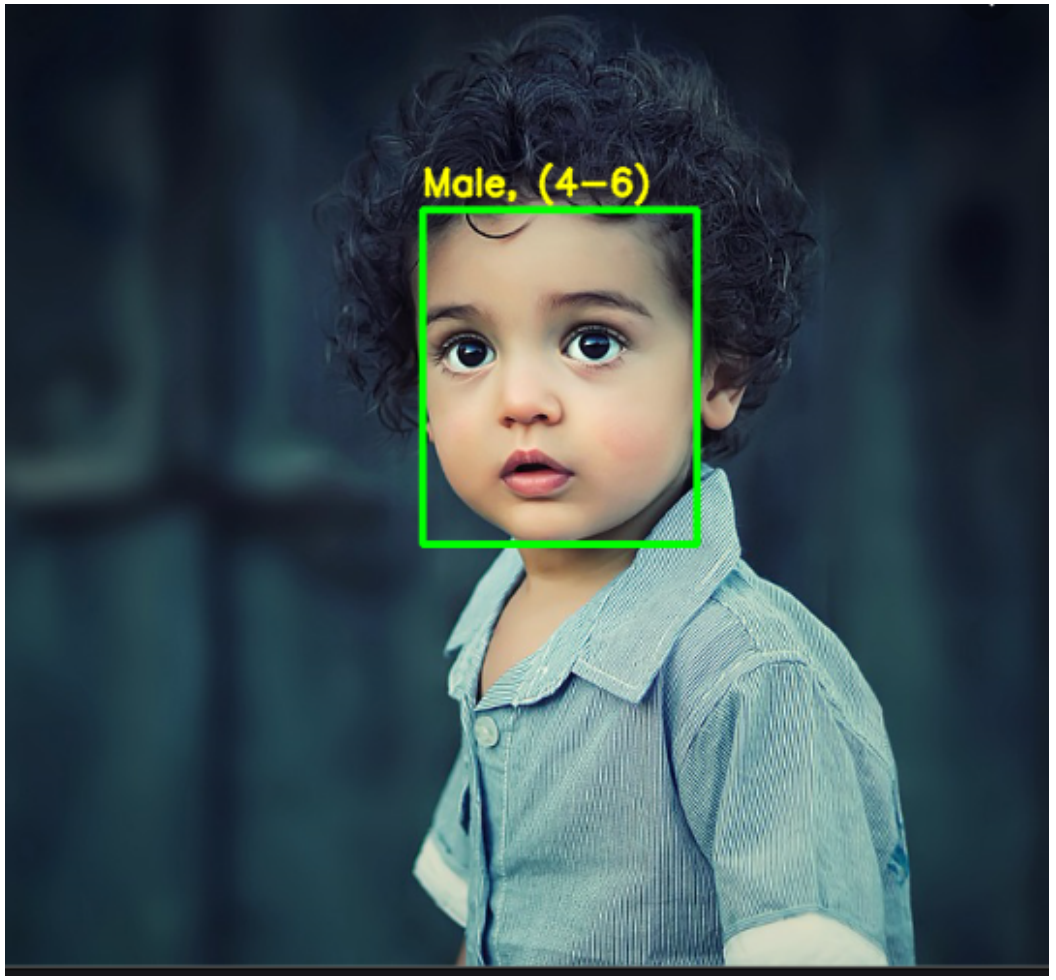
In the field of Computer Vision and Machine Learning, predicting the age using images is challenging and there is no one better machine learning model to predict accurately till now.

Applications:

- Child proof security systems
- Age based human computer interaction systems.
- Medical Diagnosis (Premature Facial Aging due to different causes)

Project Goals

- Predict age group from images
- Apply different State of the art neural network architectures



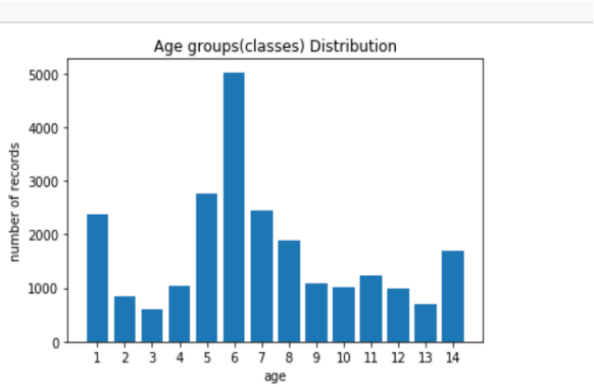
Data Set:

- Dataset Name: UTK Face
- Contains: 23708 images
- Ages range: 1 to 116 years
- Source: https://susanqq.github.io/UTKFace/

Data Pre-processing

Class labels encoding:

- Encoded ages from 1 to 116 years into 14 classes
  - ❑ How?
    - 1 to 65 years – 13 classes (Each Class covers intervals of 5 years)
    - 66 to 116 years – one class
  - ❑ Why dividing into classes works?
    - Features of face won't change drastically with each year
    - Predicting specific age might result in high variance among results
    - Even Humans are incapable in predicting exact age



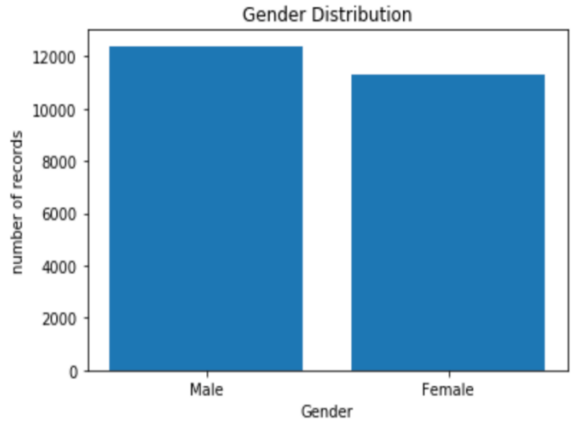
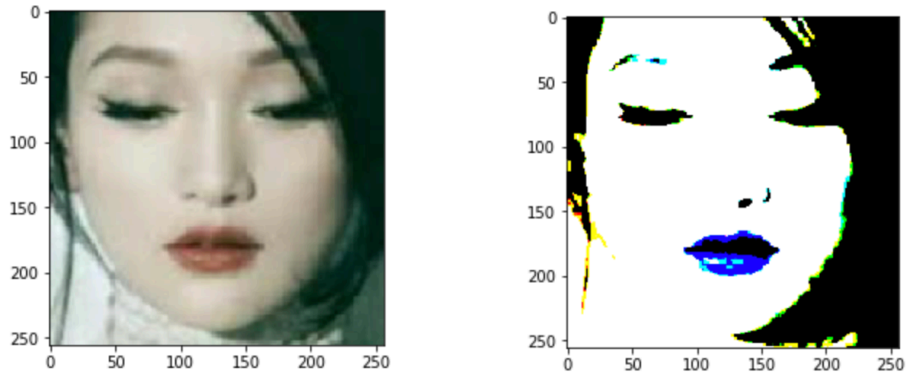


Image pre-processing:

- Converted images to an input array of size (256,256,3)
- ImageNet Utils for Preprocessing images
- Normalized the images with a factor of 255
- Resized images to fit it as an input for different architectures



Training and Testing Procedure

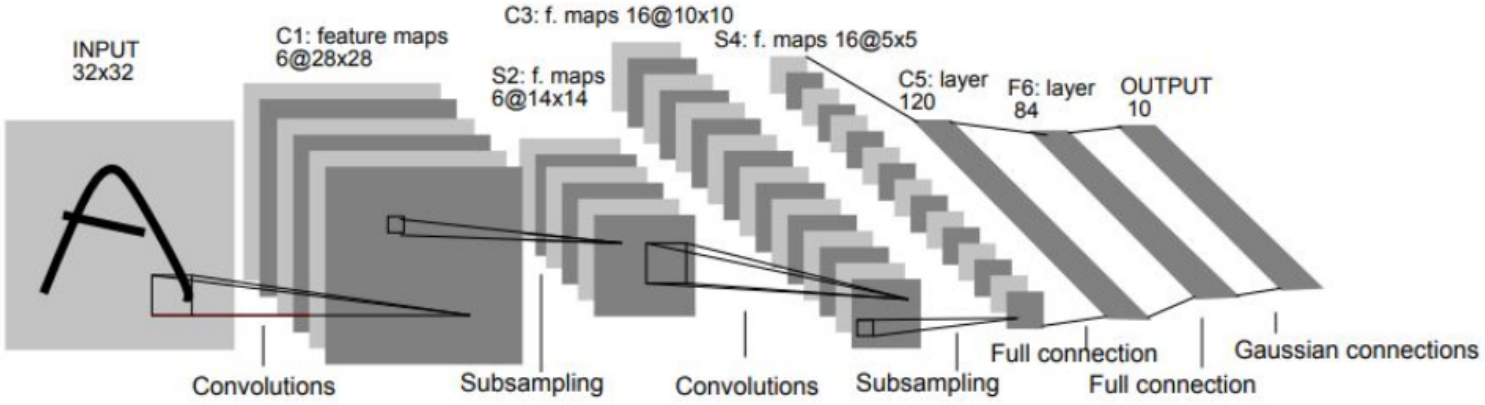
- **Train Test split:**
  - 80% Train data
  - 20% Test data

- **Hyperparameter Tuning (All Models)**
  - Epochs: 50
  - Batch Size: 64
  - Validation set size : 20% of train set (each epoch)
  - Optimizers: SGD (learning rate = 0.01)

Methods

- Soft max
- Alex Net
- LeNet - 5
- VGG - 19
- Inception – V3
- Resnet – 50(V2)
- 4C2F-NN
- AL Resnet

LeNet- 5 and AlexNet

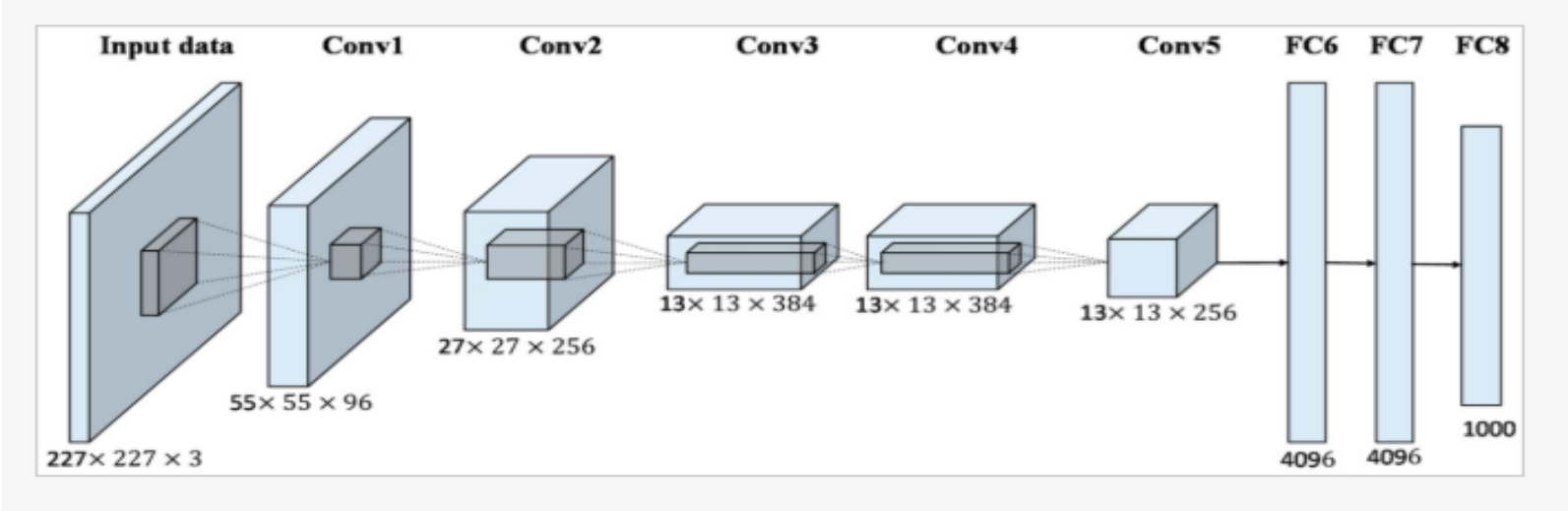


LeNet

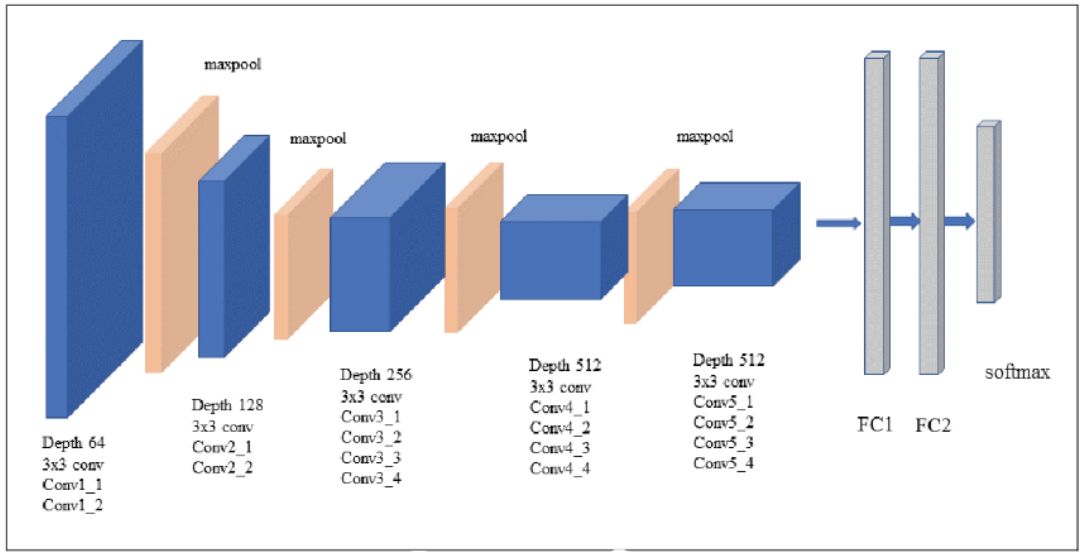
- Input Resized to 32*32 grayscale images for LeNet 5
- 3 convolutional layers
- Accuracy 32.13

AlexNet

- Input resized to (227,227,3)
- 5 convolutional Layers
- Accuracy 37.30



VGG - 19 and Inception (V3)

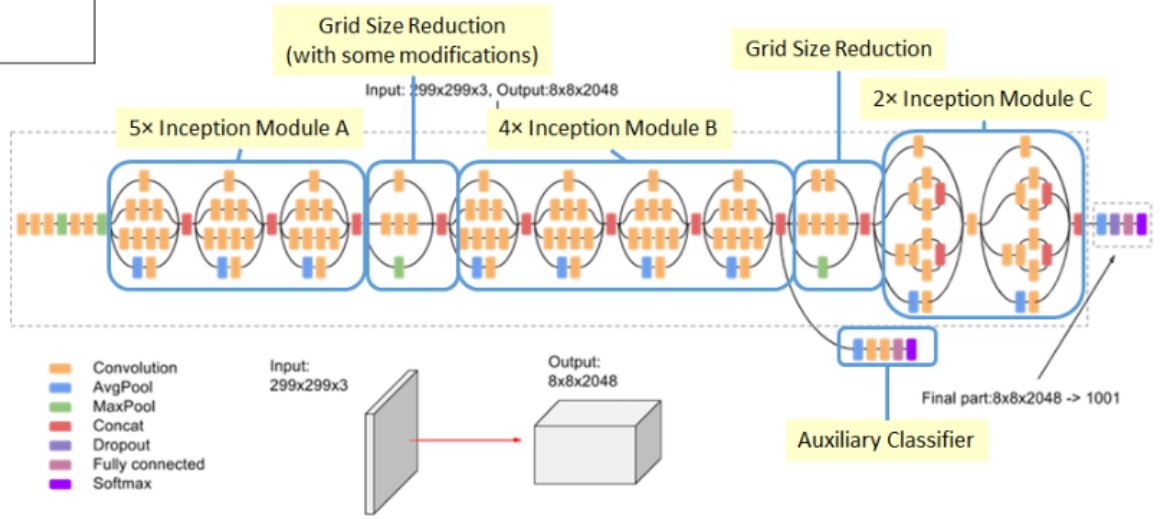


VGG 19

- Input image size (256,256,3)
- 16 Convolutional layers Layers
- Convolutional Filter size constant (3,3)
- Accuracy 37.937

Inception V3

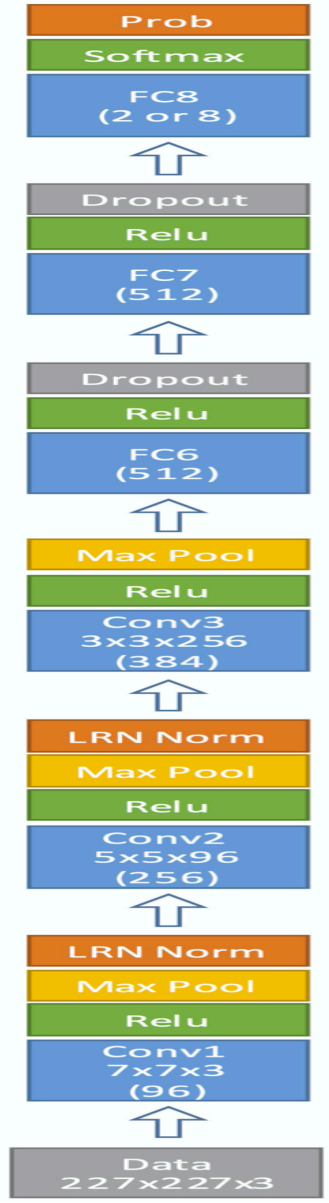
- Inception Blocks
- Input size : (256,256,3)
- 48 Layers
- Accuracy 35.765



4C2F-NN

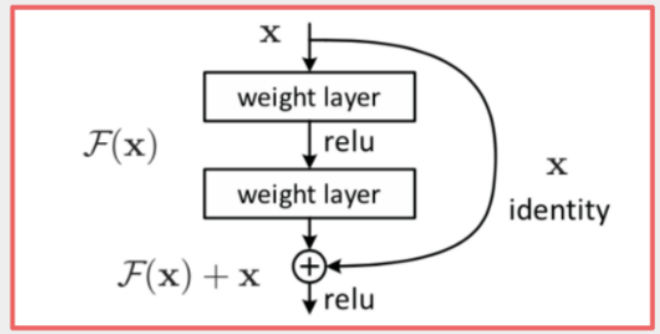
- A. Ekmekji, "Convolutional neural networks for age and gender classification", 2016.

- An RGB image input to the network is first scaled to 3x256x256 and then cropped to 3x227x227.
- There are 3 convolution layers, followed by 3 fully connected layers.
- This network design is intended to be relatively shallow so as to prevent over-fitting the data
- Used Batch Normalization layer instead of LRN
- Accuracy 42.619

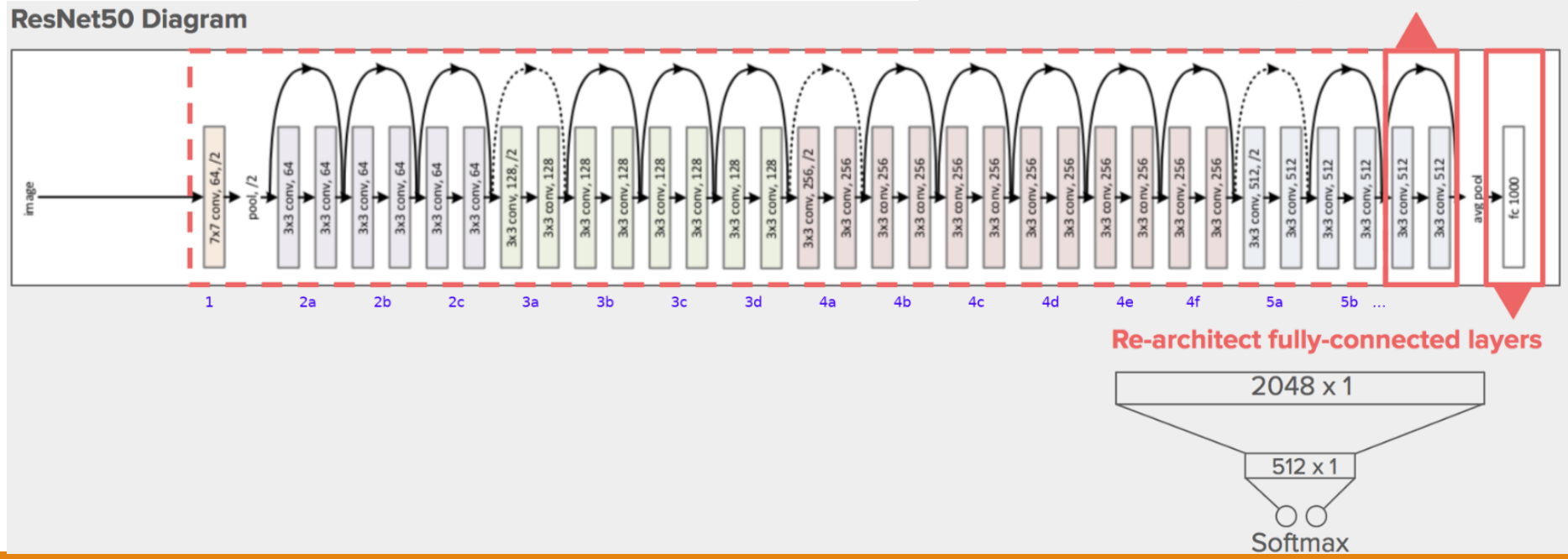


ResNet - 50

- Has residual blocks
- Input size (256,256,3)
- The **ResNet-50** model consists of 5 stages each with a convolution and Identity block.
- Each convolution block has 3 convolution **layers** and each identity block also has 3 convolution **layers**.
- 50 Layers
- Accuracy 43.10

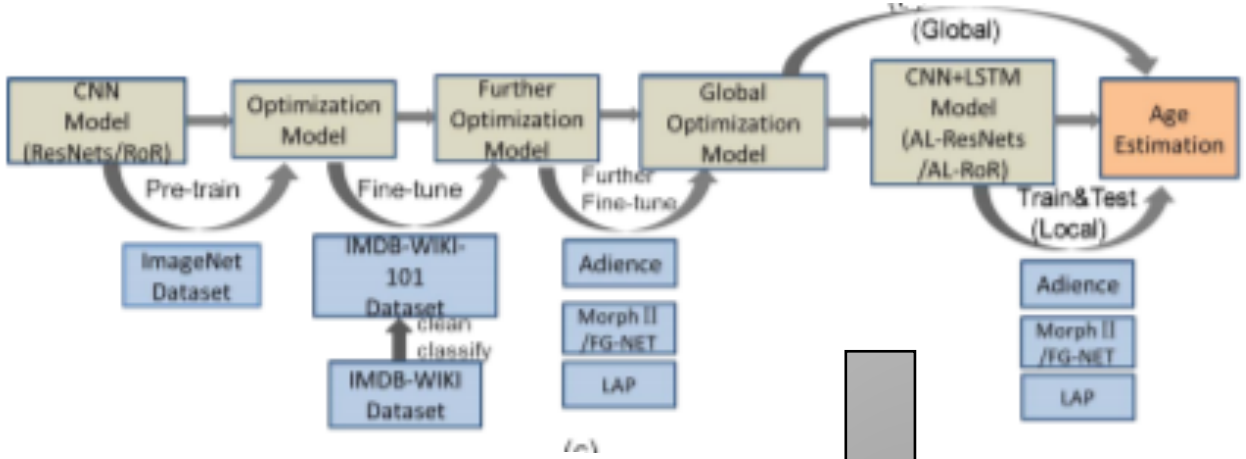


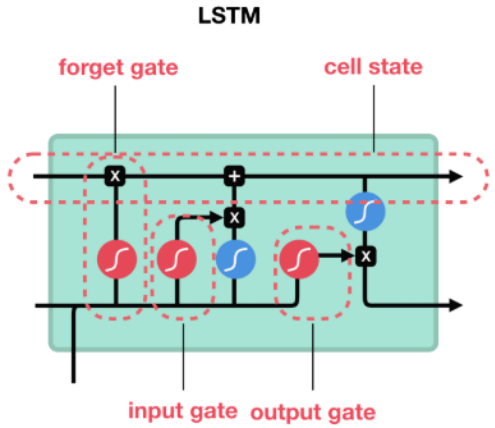
Residual Learning Block



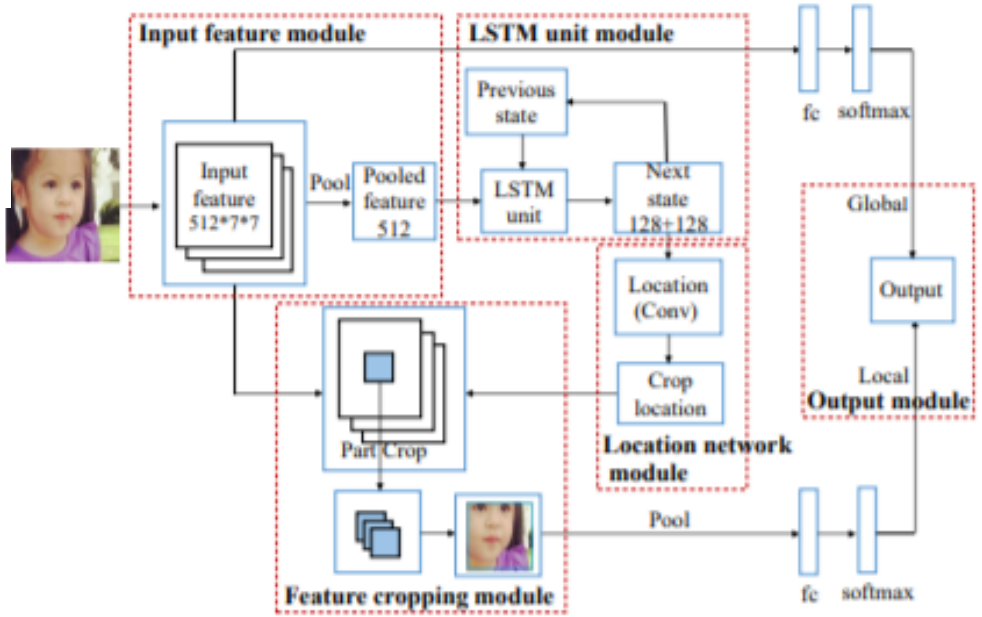
AL ResNet

-K. Zhang, N. Liu, X. Yuan, X. Guo, C. Gao, and Z. Zhao. Fine-grained age estimation in the wild with attention LSTMnetworks. arXiv preprint arXiv:1805.10445v2, 2019.



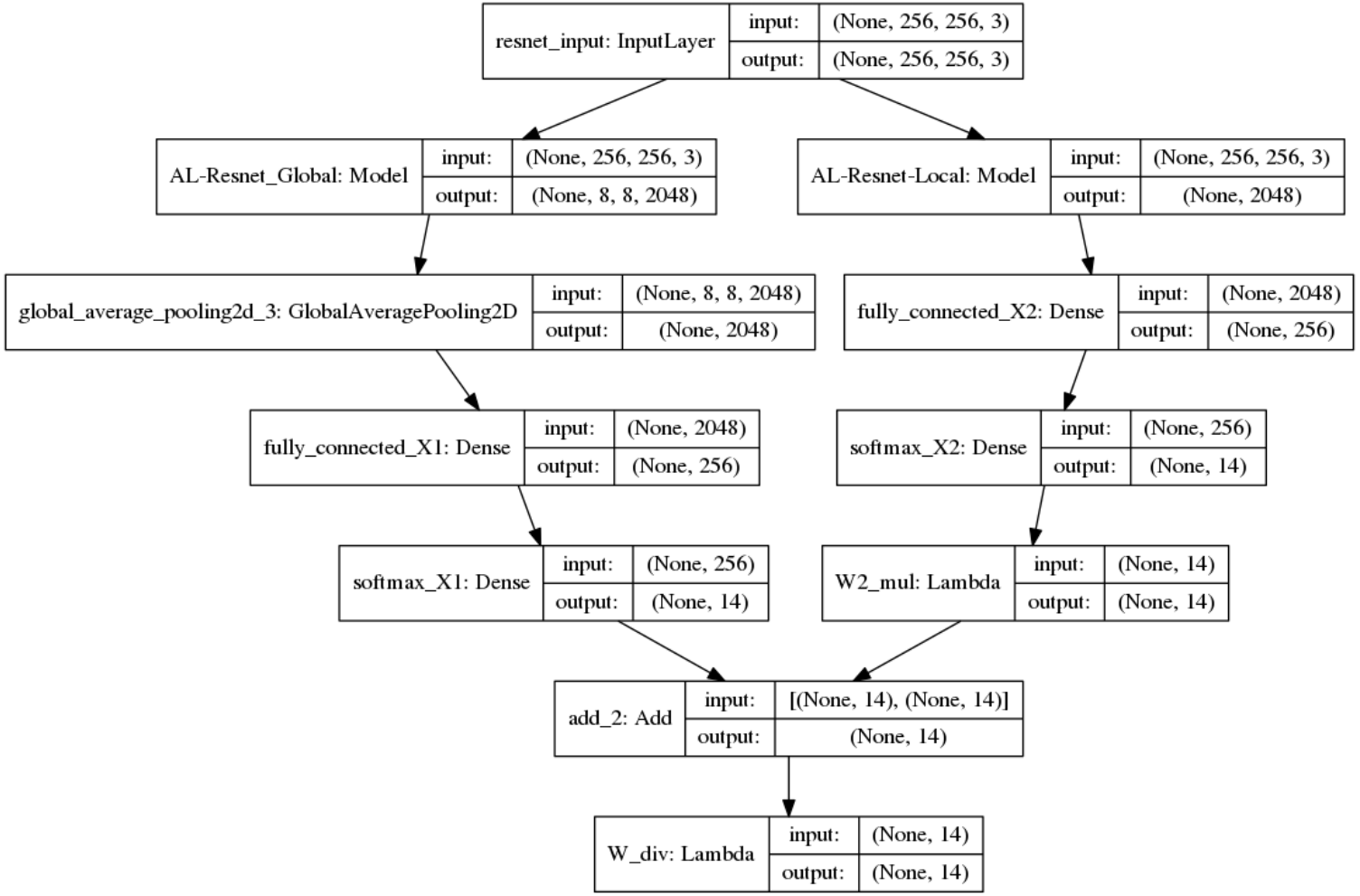


Accuracy 14 Classes: 43.16
Accuracy 8 Classes: 59.04



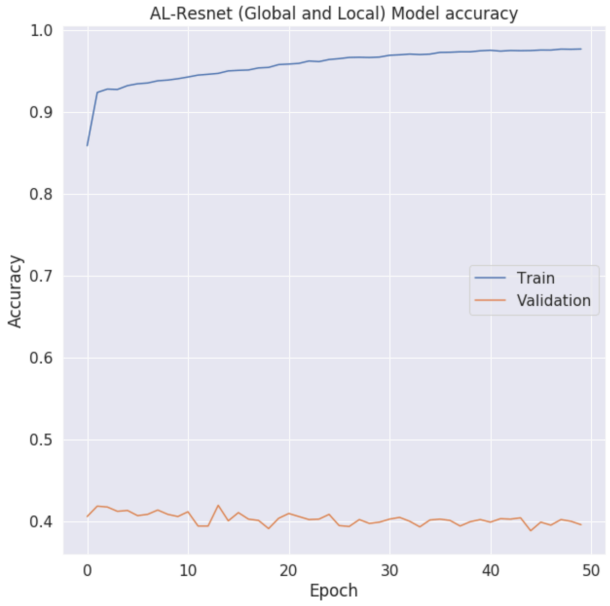
AL ResNet Implementation

-K. Zhang, N. Liu, X. Yuan, X. Guo, C. Gao, and Z. Zhao. Fine-grained age estimation in the wild with attention LSTMnetworks. arXiv preprint arXiv:1805.10445v2, 2019.

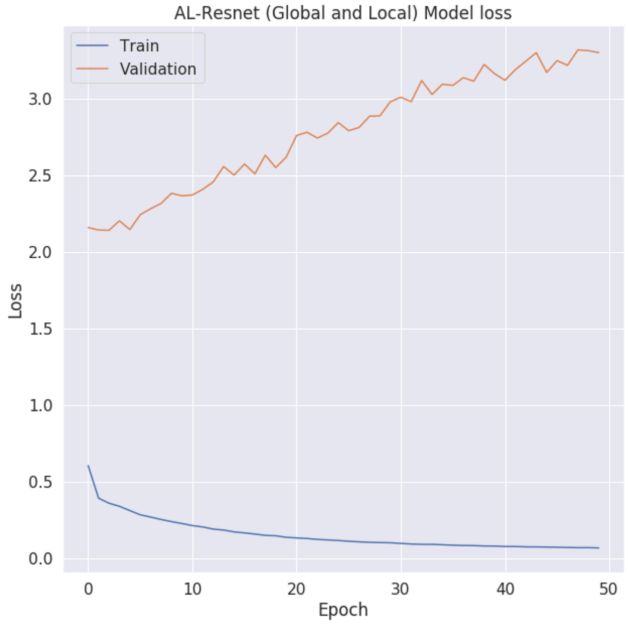


Results

Accuracy



Loss

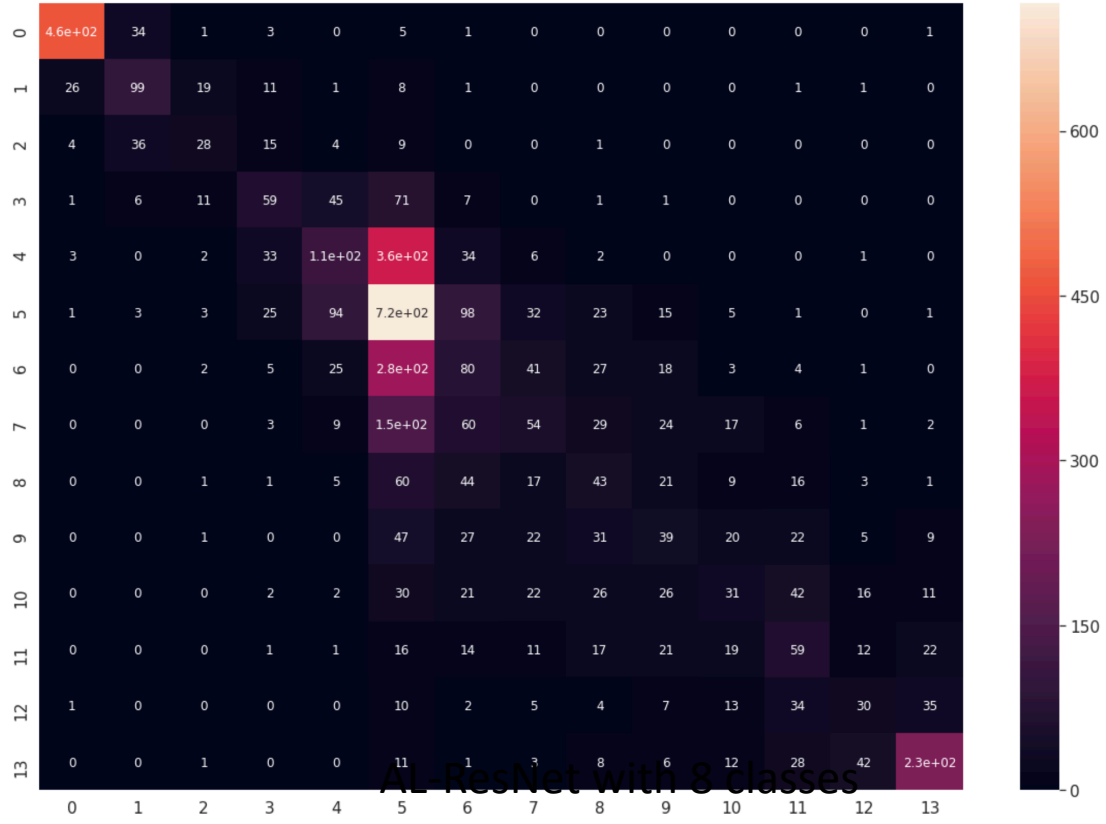


AL-ResNet with 14 classes

#	Models	Accuracy
1	SoftMax	31.99
2	LeNet 5	32.13
3	AlexNet	37.30
4	Inception	35.765
5	VGG 19	37.937
6	4C2F-NN	42.619
7	ResNet – 50 (imagenet pre trained weights)	43.10
8	Al - ResNet	43.16
9	AL – ResNet (8 class)	59.04

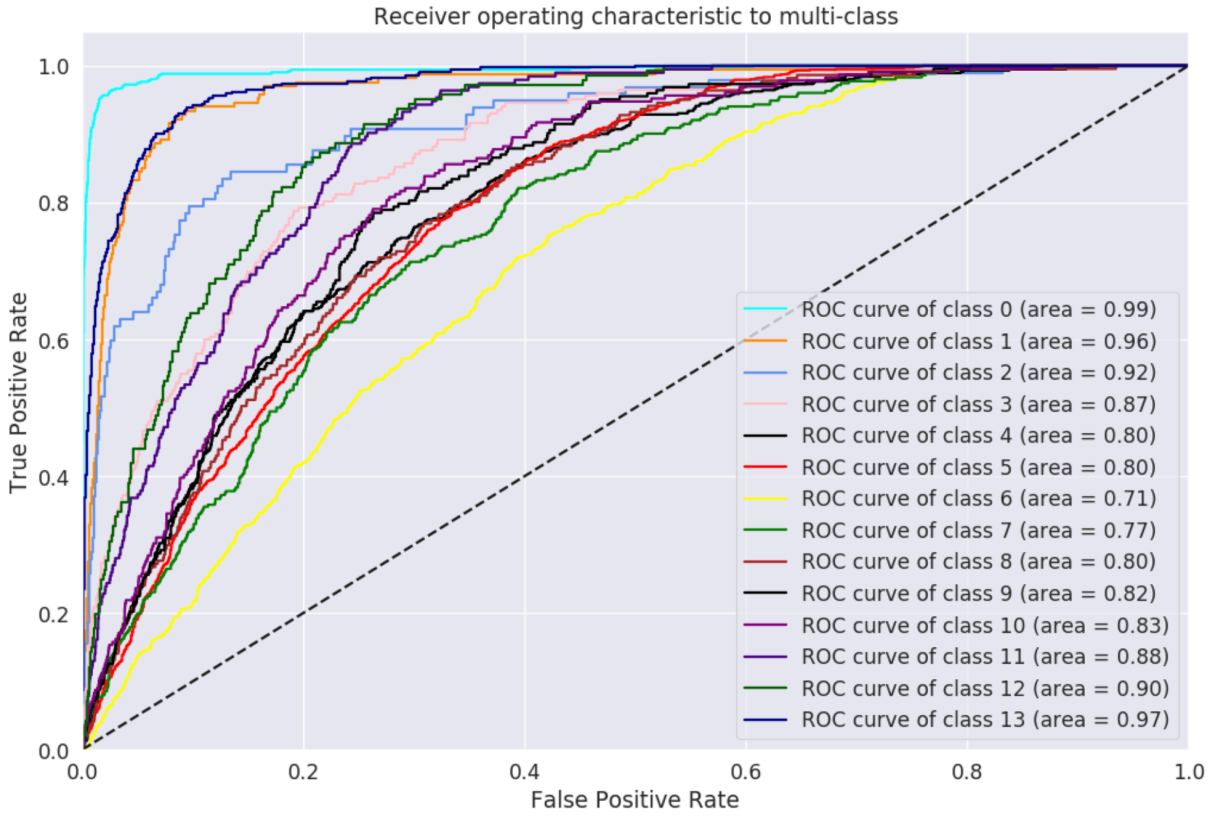
Results

model accuracy on test data is : 0.43167439103126526



AL-ResNet with 8 classes

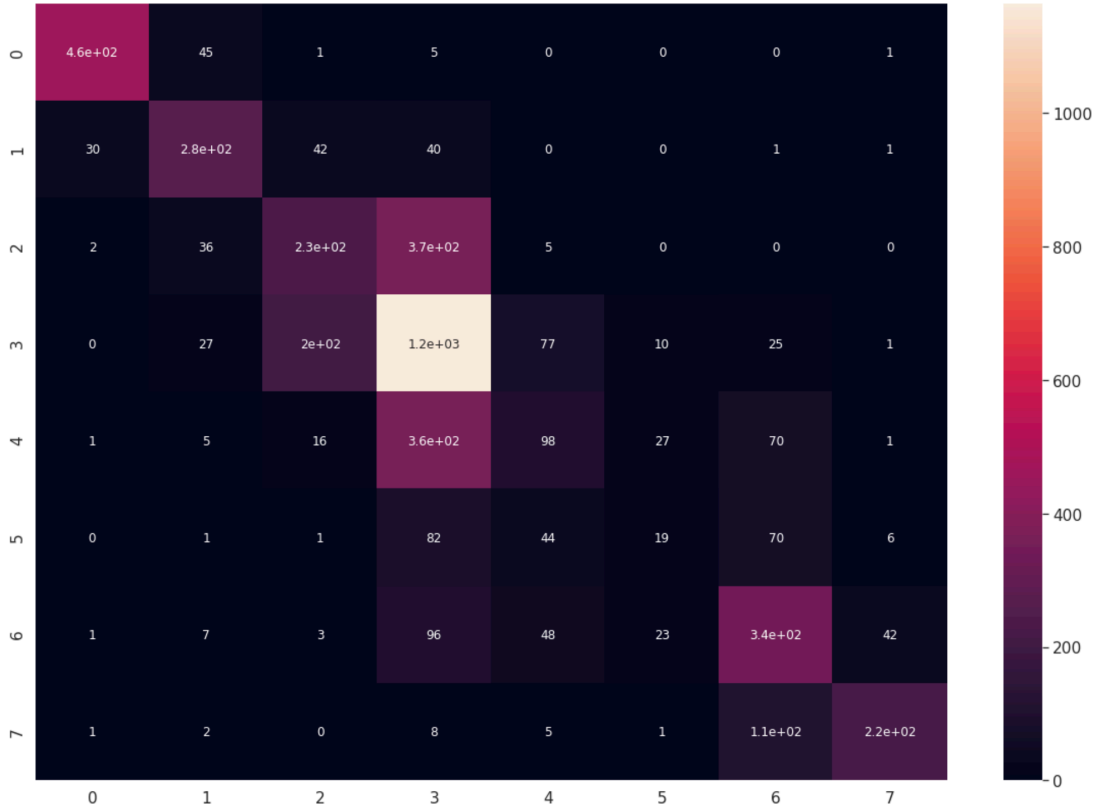
AL-ResNet with 14 classes



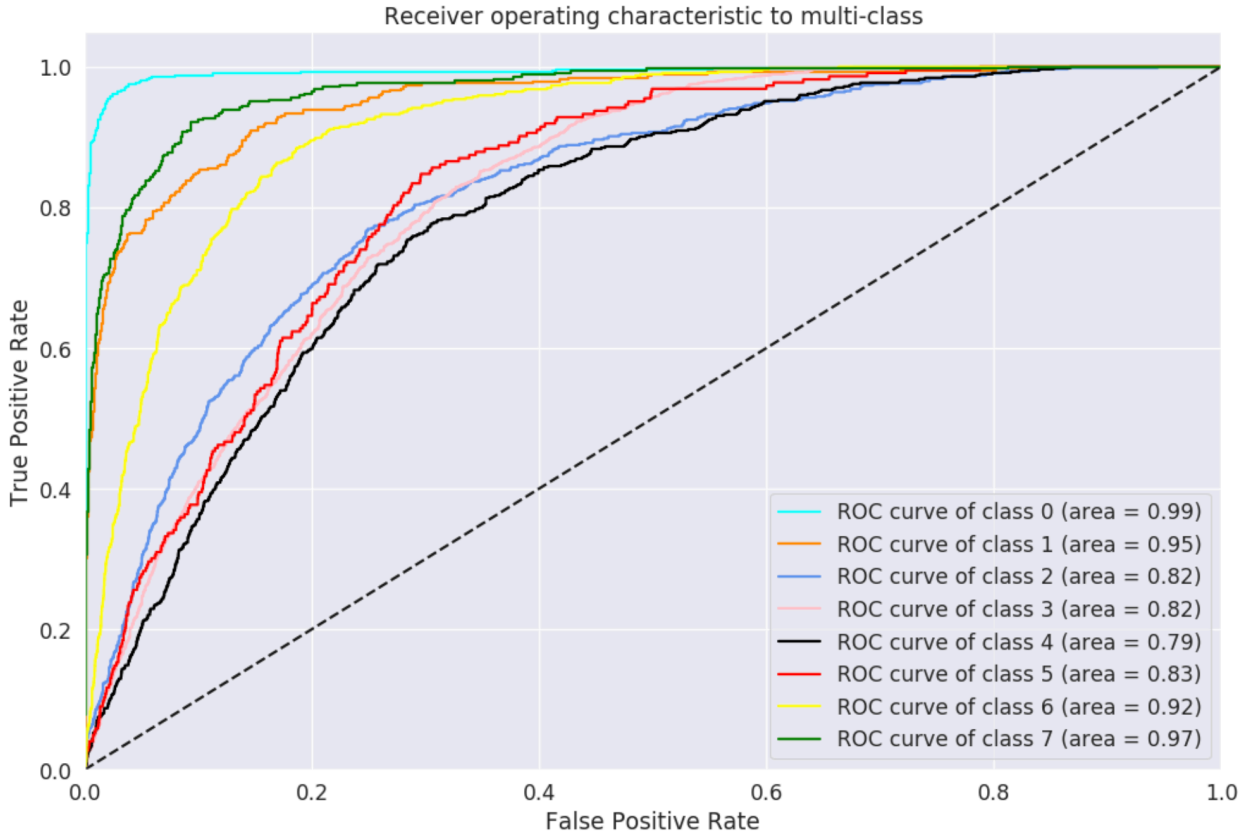
AL-ResNet with 14 classes

Results

model accuracy on test data is : 0.590468168258667



Confusion matrix (8 classes)



ROC (8 classes)

Conclusions

- AL ResNet architecture provides us with the highest accuracy
- AL RestNet Model shows promising improvement in accuracies
  - The AL-ResNet model is set up to automatically find age-sensitive regions and discriminative local features
  - We use face global features to update the internal state of the LSTM unit and extract age-sensitive position
  - With fine tuning on other data sets this model proves to be very effective.
- 4C2F-NN model has a rather shallow architecture but shows good performance in classifying the age

Future work

- Attention Layers can be added to AL ResNet Model to focus on specific regions of images for better generalization of data
- Fine tuning the model on multiple data sets to increase predictive accuracies
- Trying out other RNN's instead of LSTMS for extracting local features from the images
- Extending to gender and age classification

Thank you

References

[1] Gil Levi and Tal Hassner.Age and Gender Classification Using Convolutional Neural Networks. IEEE Workshop on Analysis and Modeling of Faces and Gestures (AMFG), at the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Boston, 2015.

[2] K. Zhang, N. Liu, X. Yuan, X. Guo, C. Gao, and Z. Zhao. Fine-grained age estimation in the wild with attention LSTMnetworks. arXiv preprint arXiv:1805.10445v2, 2019.

[3] Karen Simonyan, Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556v6, 2015.

[4] Krizhevsky, Alex; Sutskever, Ilya; Hinton, Geoffrey E. (2017-05-24). "ImageNet classification with deep convolutional neural networks" (PDF). Communications of the ACM. 60 (6): 84–90. doi:10.1145/3065386. ISSN 0001-0782.

[5] LeCun, Yann; Léon Bottou; Yoshua Bengio; Patrick Haffner (1998). "Gradient-based learning applied to document recognition" (PDF). Proceedings of the IEEE. 86 (11): 2278–2324. CiteSeerX 10.1.1.32.9552. doi:10.1109/5.726791. Retrieved October 7, 2016.