Age and Gender Prediction from UTKFace Dataset

Shared links:

- Google Colab Link: https://colab.research.google.com/drive/1VJmJDPP216FOu-cWVbD37YF2rN-qMUho?usp=sharing
- age_gender_A.h5: https://drive.google.com/file/d/1BiJM6-jy7LKnlMjihRWwMxTI-STSvaLf/view?usp=sharing
- age_gender_b.h5: https://drive.google.com/file/d/1cfxiympMbeBDgtAdsIq8i7b6zYyTC7cU/view?usp=sharing

1. Introduction

The project aims to build and train two deep-learning models using images within the UTKFace Dataset. The dataset contains 5000 labeled face images, containing information including the subject’s age, gender, and race. We train two multi-output models to estimate the age and gender of the subject in the image. One model is a CNN model defined and trained from scratch, the other one involves fine-tuning the pre-trained VGG16 Model. We use mean absolute error (mae) to assess the performance of the age estimator, and accuracy to assess the performance of the gender classifier.

2. Model A

The CNN model was developed, it involves one base model and two branches for age estimation and gender classification respectively, built on top of the base model. The base model uses three sets of convolution-pooling architecture followed by a dropout layer with a probability of 0.2. The age and gender models both use two sets of convolution-convolution-pooling-dropout architecture, followed by their own fully connected layers with different depths and filter parameters. In total, there are 2295026 trainable parameters. The architecture of the whole model is as follows:

2.1. Model A Architecture

Model A Base Model:

Layer	Filters	Output Size	Kernel Size	Activation	Extra Info
Image	-	128x128x3	-	-	
Convolution	8	126x126x8	3x3	Relu	
MaxPooling	-	63x63x8	2x2	-	
Convolution	16	61x61x16	3x3	Relu	
MaxPooling	-	60x60x16	2x2	-	stride 1
Convolution	32	58x58x32	3x3	Relu	
MaxPooling	-	57x57x32	2x2	-	stride 1
Dropout	-	57x57x32	-	-	Dropout 0.2

Model A Age model - built on the base model

Layer	Filters	Output Size	Kernel Size	Activation	Extra Info
Convolution	32	55x55x32	3x3	Relu	
Convolution	64	54x54x64	2x2	Relu	
MaxPooling	-	27x27x64	2x2	-	
Dropout	-	27x27x64	-	-	Dropout 0.1
Convolution	128	24x24x128	4x4	Relu	
Convolution	128	22x22x128	3x3	Relu	
MaxPooling	-	7x7x128	3x3	-	
Dropout	-	7x7x128	-	-	Dropout 0.1
FC1	-	256	-	-	
Dropout	-	256	-	-	Dropout 0.2
FC2	-	32	-	-	
Output	-	1	-	Relu	

Model A Gender model - built on the base model

Layer	Filters	Output Size	Kernel Size	Activation	Extra Info
Convolution	32	55x55x32	3x3	Relu	
Convolution	32	54x54x32	2x2	Relu	
MaxPooling	-	18x18x32	3x3	-	
Dropout	-	18x18x32	-	-	Dropout 0.1
Convolution	64	15x15x64	4x4	Relu	
Convolution	64	13x13x64	3x3	Relu	
MaxPooling	-	4x4x64	3x3	-	
Dropout	-	4x4x64	-	-	Dropout 0.1
FC1	-	256	-	-	
Dropout	-	256	-	-	Dropout 0.2
FC2	-	128	-	-	
Dropout	-	128	-	-	Dropout 0.2
Output	-	1	-	Sigmoid	

2.2. Model A Training Process and Performance

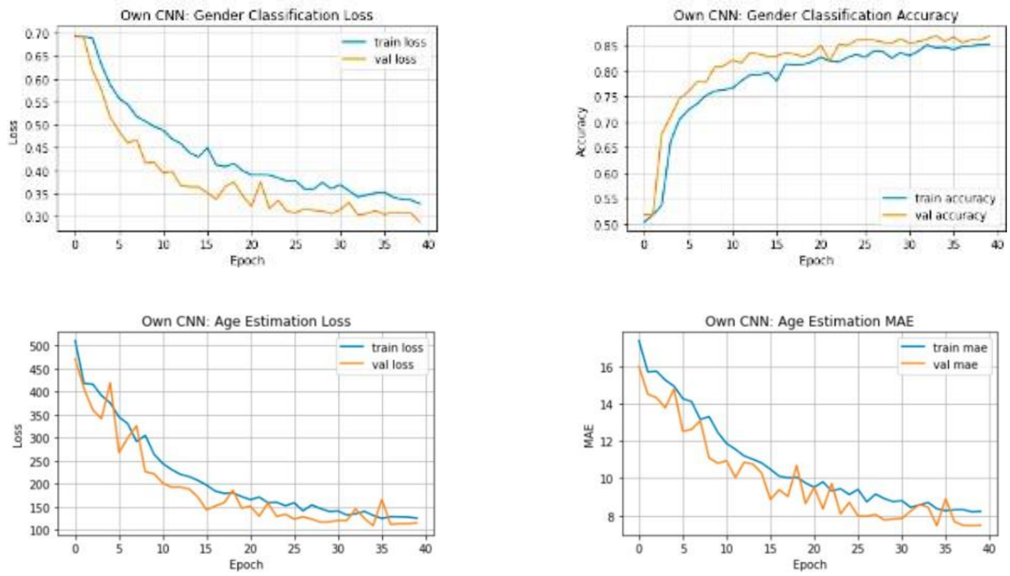
To prevent overfitting, the model used numerous dropout layers with probabilities of 0.1 and 0.2. Adam optimizer was used since it has a better general performance and converges faster according to research[1]. The learning rate 0f 0.001 was chosen as this produced the most stable results after experimenting. Furthermore, the training process used an early stopping with the patience of 5 epochs and monitoring the validation loss, ensuring the model does not overfit and saves on computational resources. An Epoch of 40 was chosen in the end because the performance started to stabilize around this value and further training would only lead to little improvements in performance with a large cost in computational resources. Since the model is trained from scratch, using these techniques above ensures the model would learn from most of the data available while also making sure it would be able to generalize with new data.

The model achieves good performance on the validation sets, with the final epoch age validation mae of 7.4766 and validation gender accuracy of 0.8690, there is also no significant overfitting nor underfitting present:

```
age_output_loss: 125.1706 - gender_output_loss: 0.3524 - age_output_mae: 8.2602 - gender_output_accuracy: 0.8428
age_output_loss: 128.7500 - gender_output_loss: 0.3419 - age_output_mae: 8.3155 - gender_output_accuracy: 0.8482
age_output_loss: 128.3801 - gender_output_loss: 0.3367 - age_output_mae: 8.3215 - gender_output_accuracy: 0.8500
age_output_loss: 127.9359 - gender_output_loss: 0.3361 - age_output_mae: 8.2010 - gender_output_accuracy: 0.8518
age_output_loss: 125.6119 - gender_output_loss: 0.3277 - age_output_mae: 8.2290 - gender_output_accuracy: 0.8525
```

```
- val_age_output_loss: 165.8938 - val_gender_output_loss: 0.3047 - val_age_output_mae: 8.9056 - val_gender_output_accuracy: 0.8669
- val_age_output_loss: 111.9710 - val_gender_output_loss: 0.3089 - val_age_output_mae: 7.6747 - val_gender_output_accuracy: 0.8558
- val_age_output_loss: 113.7129 - val_gender_output_loss: 0.3067 - val_age_output_mae: 7.4718 - val_gender_output_accuracy: 0.8619
- val_age_output_loss: 113.5995 - val_gender_output_loss: 0.3074 - val_age_output_mae: 7.4630 - val_gender_output_accuracy: 0.8609
- val_age_output_loss: 115.3775 - val_gender_output_loss: 0.2877 - val_age_output_mae: 7.4766 - val_gender_output_accuracy: 0.8690
```

Own CNN Model



---

[1] P. Kingma, Diederik, and Jimmy Ba. “Adam: A Method for Stochastic Optimization.” 2017, https://arxiv.org/abs/1412.6980.

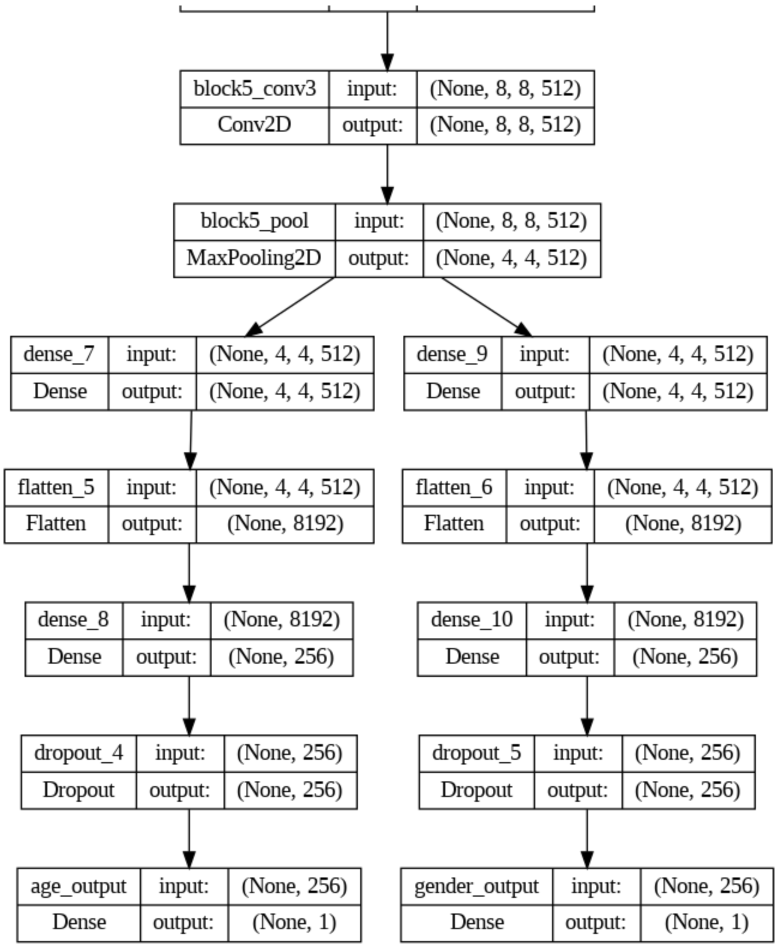
3. Model B CNN Built on Pre-trained VGG16 CNN

3.1. Fine-tuned VGG16 Architecture

The chosen pre-trained CNN model trained on an ImageNet dataset used for finetuning is VGG16 which had higher age estimation mean absolute error and gender classification accuracy on our UTKFace dataset. VGG16 model consists of 13 convolutional layers, 5 max pooling layers and 3 fully connected layers, with 16 learnable parameter layers.

The model was fine tuned with keras and Deep Learning by first loading the VGG16 model while leaving off the Fully Connected Layers and specifying the input size of image as 128 by 128 by 3. We then defined a 3 new fully connected layers for both the age estimation and gender classification consisting of a dense layer with 518 units and a relu activation function, reshaped the dimension of the layers into a single layer and added a second dense layer with 256 units and relu activation function. We added L2 weight regularizer of 0.0001 and a dropout layer with probability 0.5 to prevent the model from overfitting on the dataset.

Multi-output fully connected layers with units of 1 were defined for both age estimation and gender classification with relu and sigmoid activation functions respectively. The new Fully connected layers were placed on top of the VGG16 base model creating a fine-tuned multi-output model with 4,720,642 trainable parameters.



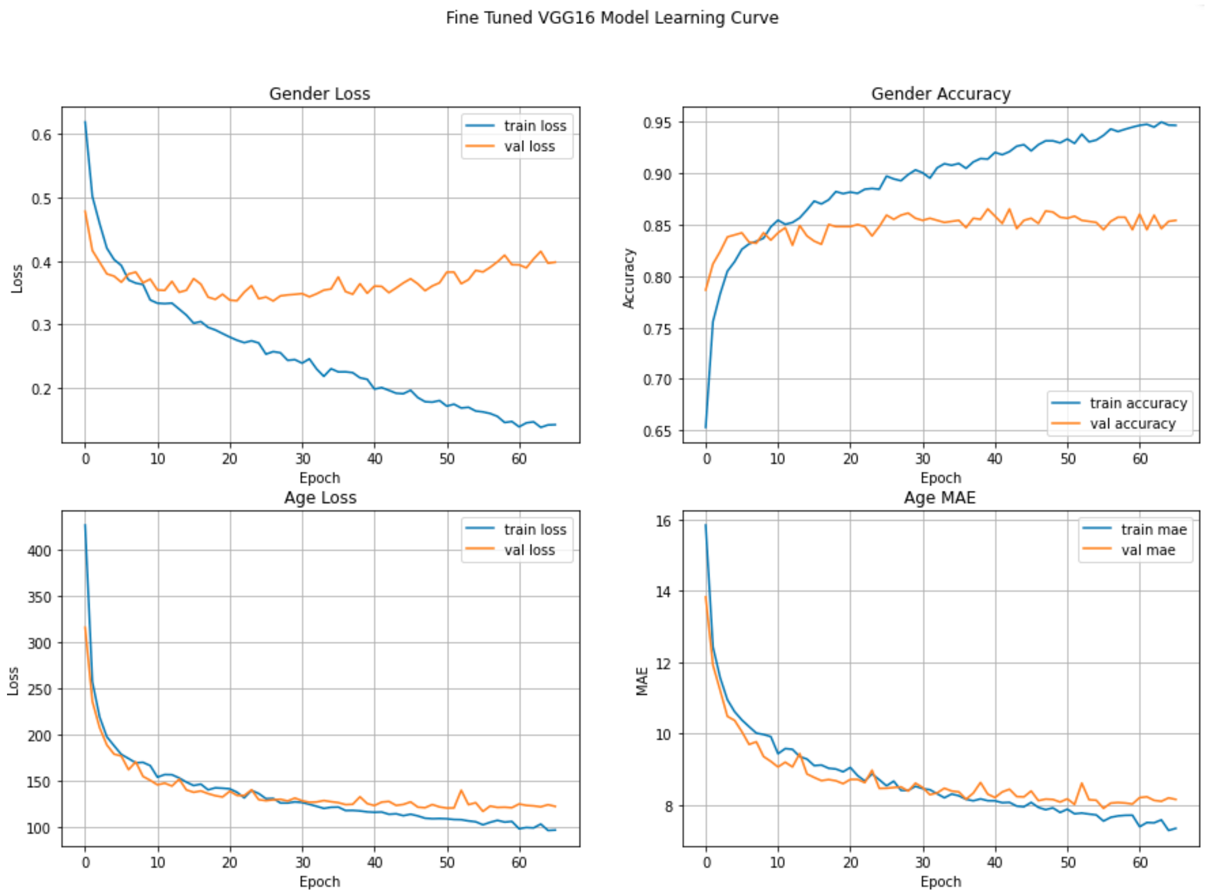
3.2. Fine-tuned VGG16 Training Process and Performance

We began the training process by freezing all the convolutional layers in the body of the VGG16 model, preventing weight updating during the training and ensuring only the fully connected layer weights were updated. Compilation of the model was done using Mean square error and Accuracy metrics for regression and classification with the loss functions of MAE and binary cross-entropy for age estimation and gender classification respectively Model optimization was experimented using different hyperparameters with SGD and Adam optimizer, with Adam optimizer with a learning rate of 0.0001 providing better results.

Trained the network using our rescaled and augmented image data with Early stopping regularisation parameter monitoring the validation loss with a patience of 10. After Experimenting with different hyperparameters of epoch, optimizer, learning rate and batch size. The performance of the model for age validation mae is 8.19 and gender validation accuracy of 85%. There is no significant underfitting or overfitting in the learning curves which are presented below.

```
loss: 106.2123 - age_output_loss: 105.9917 - gender_output_loss: 0.1449 - age_output_mae: 7.7055 - gender_output_accuracy: 0.9423
loss: 106.9282 - age_output_loss: 106.7057 - gender_output_loss: 0.1465 - age_output_mae: 7.7092 - gender_output_accuracy: 0.9442
loss: 98.9912 - age_output_loss: 98.7768 - gender_output_loss: 0.1381 - age_output_mae: 7.3862 - gender_output_accuracy: 0.9460 -
loss: 100.3708 - age_output_loss: 100.1498 - gender_output_loss: 0.1443 - age_output_mae: 7.5031 - gender_output_accuracy: 0.9470
loss: 99.9041 - age_output_loss: 99.6811 - gender_output_loss: 0.1460 - age_output_mae: 7.4955 - gender_output_accuracy: 0.9442 -
loss: 104.0592 - age_output_loss: 103.8445 - gender_output_loss: 0.1372 - age_output_mae: 7.5786 - gender_output_accuracy: 0.9492
loss: 97.1501 - age_output_loss: 96.9314 - gender_output_loss: 0.1411 - age_output_mae: 7.2818 - gender_output_accuracy: 0.9463 -
```

```
- val_loss: 122.5564 - val_age_output_loss: 122.0715 - val_gender_output_loss: 0.4090 - val_age_output_mae: 8.0508 - val_gender_output_accuracy: 0.8569
- val_loss: 121.9985 - val_age_output_loss: 121.5283 - val_gender_output_loss: 0.3940 - val_age_output_mae: 8.0227 - val_gender_output_accuracy: 0.8448
val_loss: 126.0421 - val_age_output_loss: 125.5718 - val_gender_output_loss: 0.3938 - val_age_output_mae: 8.2019 - val_gender_output_accuracy: 0.8599
- val_loss: 124.4946 - val_age_output_loss: 124.0288 - val_gender_output_loss: 0.3889 - val_age_output_mae: 8.2251 - val_gender_output_accuracy: 0.8448
val_loss: 123.9640 - val_age_output_loss: 123.4838 - val_gender_output_loss: 0.4030 - val_age_output_mae: 8.1243 - val_gender_output_accuracy: 0.8589
- val_loss: 122.9795 - val_age_output_loss: 122.4868 - val_gender_output_loss: 0.4151 - val_age_output_mae: 8.0985 - val_gender_output_accuracy: 0.8458
val_loss: 125.2700 - val_age_output_loss: 124.7958 - val_gender_output_loss: 0.3963 - val_age_output_mae: 8.1943 - val_gender_output_accuracy: 0.8528
```



4. Summary and Discussion

In conclusion, Model A achieves a better performance with age validation mae of 7.4 and validation gender accuracy of 86%, compared to age validation mae of 8.19 and validation gender accuracy of 85% for Model B. In this assignment, both members gained valuable experiences in processing data, defining own CNN architectures, and fine-tuning a pre-trained model for it to be applied in a specific context. We learned that using dropouts and adjusting learning rates were good ways to resolve the issues with overfitting. Given more time and computation power, we could improve the models by experimenting with more variations of CNN architectures.

It should be noted, however, the approach of using face images to predict age and gender has moral issues as it relies on using the appearance of someone to determine their biological gender rather than the gender identity of the subject. In addition, the introduction of cosmetic procedures could mean that people in the future might look younger than their counterparts today, therefore, the model may not generalize well in a longer timespan and new training data would be needed for the model to keep up with its performance.