ID	Contribution
ss4658	50%
ss4659	50%

Shared Linked for Model A: https://drive.google.com/file/d/1-gLXv79RuAi8L-n8MM3Gm7opcztHF1hw/view?usp=sharing

Shared Linked for Model B: https://drive.google.com/file/d/1-5qhi_KSY-1ZsMGFOx8TH5a-N8wgcZr5/view?usp=sharing

Do all the members agree with the above contributions? [Yes]

Section 1: Introduction

Convolutional neural networks have become central in image analysis jobs, specifically in processing the delicate face characteristics of a person. This work aims to build a customized CNN and fine-tune a pre-trained model to improve the accuracy of predicting gender and age in various human images.

Methodologically, our work begins with the precise design of the CNN, which is the systematic organization of layers and definition of hyperparameters. Further, the CNN is fine-tuned, that is the pre-trained model is being adjusted to enhance its capacity to generate predictions on our dataset. After that, we will explain its architecture, how it is trained, and its performance along with learning curves.

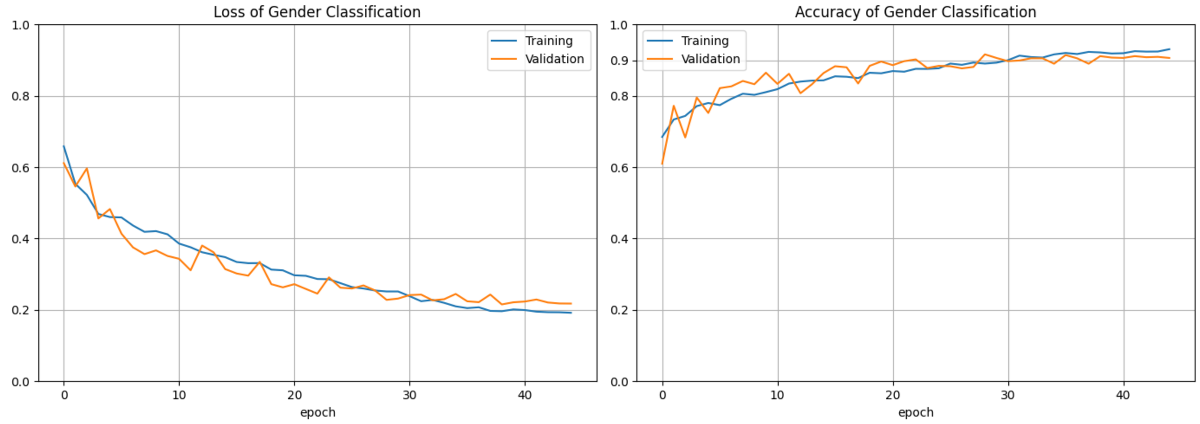
The project concludes by comparing the performance of these two models, exaggerating the information that we have realized during the training process, what is the problem we have faced, and how to improve the model performance in the future.

Section 2: My own CNN

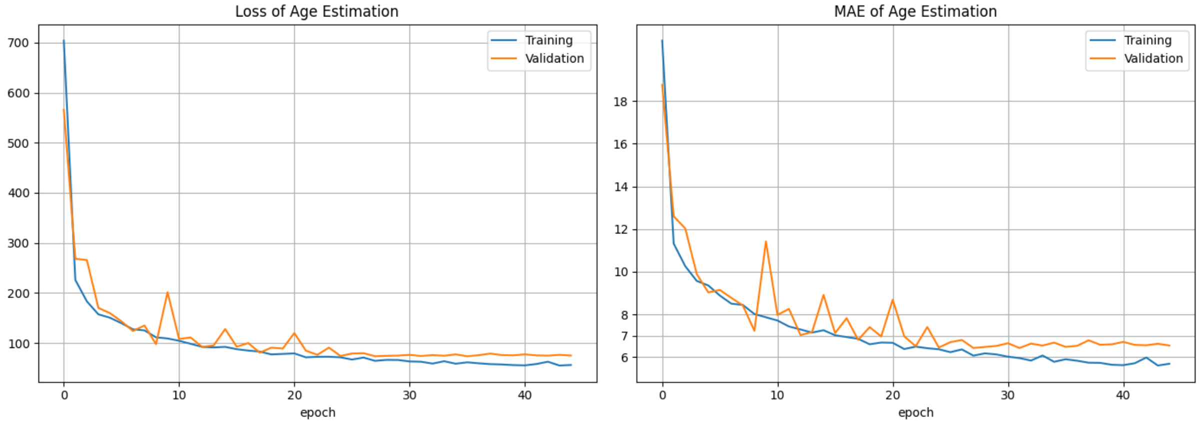
Our CNN model begins with 128 x 128 images with 3 color channels passing through 4 consecutive convolutional layers, each with increasingly growing sets of filters: 32, 64, 128, and 256. The filter size is kept consistent at 3 x 3 for pixel dimensionality and detecting features gradually from simple to complex. Each layer is fed with batch normalization and ReLU activation to facilitate training and prevent gradient vanishing. To avoid overfitting, L2 regularization is set to 0.005. MaxPooling 2 x 2 reduces spatial dimensionality following each convolutional layer. The subsequent flattening step, then, converts the dimensional data array into a one-dimensional array.

The network is divided into two branches. The gender branch is created by a 256-unit dense layer, batch normalization, and ReLU activation function, and makes use of a dropout with a rate of 0.4 to prevent overfitting. This use of dropout will also conclude with a sigmoid activation function for binary output. The age branch has the same architecture as the gender branch but with a 512-unit dense layer and a ReLU activation function for regression output. Furthermore, L2 regularization is set at 0.006.

we split the dataset into training and validation subsets at a rate of 80:20. To enhance robustness against the diverse input data, data augmentation was done by normalizing image pixels to the training data. To promote robustness against diverse input data, the picture then underwent normalization, random rotations by 30 degrees, and horizontal flips. We used the flow_from_dataframe method to group 32 images for each batch and resize images to 128×128. To achieve efficient learning and weight optimization, the model trained for 45 epochs and utilized the Adam optimizer with a 0.001 started learning rate and learning rate scheduler to dynamically polish its performance. After each epoch, we validated the model with the same steps, making sure that our binary gender classification and age estimation were reliable, evaluated accordingly by binary cross-entropy loss and mean squared error loss functions.



Loss and Accuracy for gender classification of ModelA

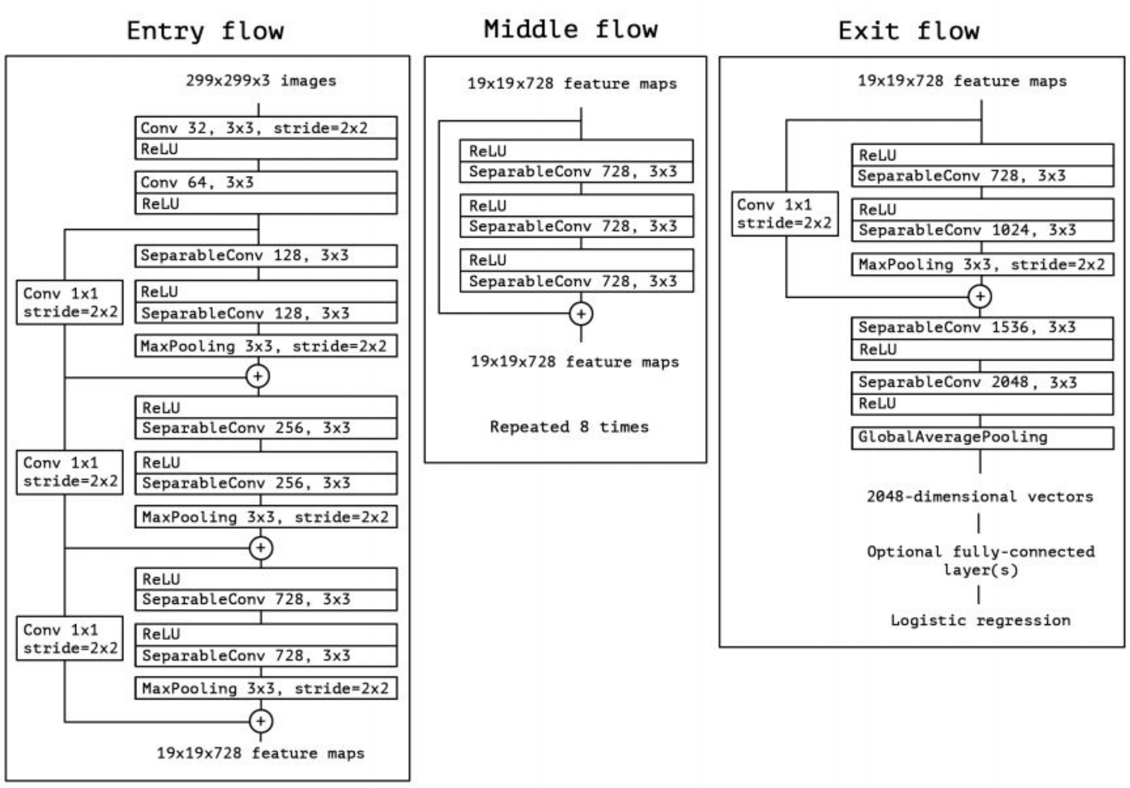


Loss and MAE for age estimation of ModelA

The graphs illustrate the positive trend of this model. It can generalize well for both gender classification and age regression tasks. The gender classification loss decreases rapidly in both training and validation sets, indicating good learning without overfitting. The accuracy is high and stable at between 90 and 91.5% at the end of the training period in the validation set, showing that the model classifies gender accurately. For age prediction, both training and validation loss decrease tremendously and level off, while the mean absolutes drop rapidly and steady at around 6.4-6.6 in the validation set. This suggests that age predictions are precise and consistent

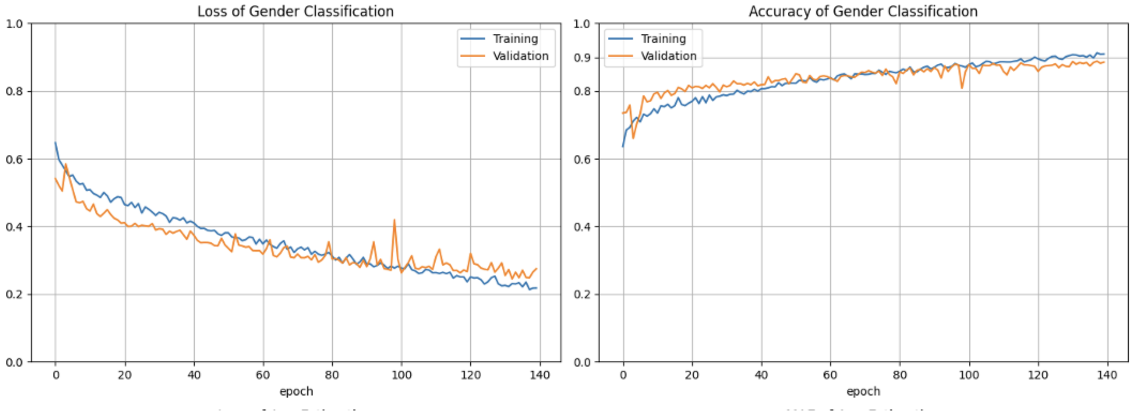
Section 3: Pre-trained CNN

For the second model, we build a model using the Xception architecture as a base model. Originally, the Xception architecture consisted of 36 convolutional layers mapped into 14 blocks with three different flows: entry, middle, and exit. We fine-tuned this model by freezing the first 50 layers, containing the entire entry flow and one block in the middle flow, maintaining the ability to extract generic features, and allowing the other seven blocks in the middle flow and the exit flow to be trainable. However, we excluded the top layer of the model and replaced it with a GlobalAveragePooling layer with a dropout rate of 0.3, which is then separated into two paths. Both routes start with a fully connected layer using the ReLU activation function with 256 neurons, L2 regularization and a dropout layer. While the L2 and dropout rates are set to be 0.002 and 0.2 for gender classification, the age estimation route applies 0.004 and 0.3, respectively. The output layer for both paths contain one unit neuron with sigmoid and ReLU activation functions for gender classification and age estimation.

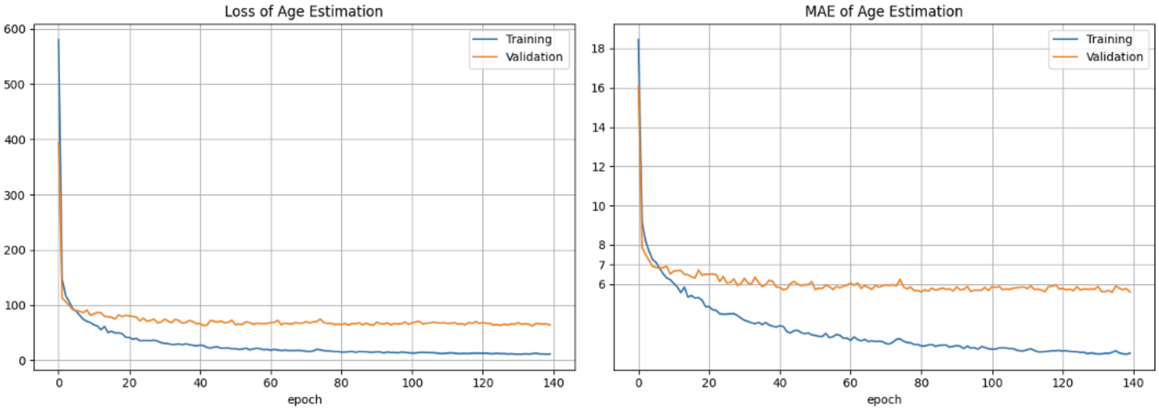


Architecture of the original Xception model [1]

After splitting the data into training and validation sets with the size of 4000 and 1000 images, respectively, the training process begins with augmentation. While the validation images are rescaled to 0 and 1, the training images are randomly augmented by various settings, including rescaling, rotating by 30 degrees, shifting both vertically and horizontally by 15%, and flipping horizontally. The input size is set to be 128x128 pixels along with two outputs for gender classification and age estimation, using binary cross entropy and MSE as loss metrics. We train the model using the Adam optimizer with a constant batch size of 32 images and a learning rate of 0.0001 for 140 epochs.



Loss and Accuracy for gender classification of ModelB



Loss and MAE for age estimation of ModelB

Despite the fluctuation, the fine-tuned model performs well on both training and validation sets, avoiding overfitting and achieving the highest accuracy of around 91% and 88.7% and the lowest loss of 0.2 and 0.3, respectively. Similarly, the model also performs impressively well on the age estimation task, reaching the lowest MAE of around 5.6 and 2.4 on validation and training sets.

Section 4: Summary and Discussion

Summary:

Overall, both ModelA and ModelB perform well on both gender classification and age estimation, not overfitting and sharing similar patterns of performance on both tasks on the validation set. For gender classification, the performance of ModelA is better in terms of how quickly to converge and the final accuracy, reaching the maximum of 91% within 40 epochs. Meanwhile, ModelB requires at least 108 epochs to reach its highest accuracy at 88.5%. Conversely, on the age estimation task, ModelB outperforms by achieving the lowest mean absolute error of 5.6 within 81 epochs with a rapid drop during the first 10 epochs. Although a similar pattern of error can be seen in the ModelA, it levels off at around 7 after the 40th epoch.

Discussion:

ModelA with only four convolutional layers performs surprisingly better than ModelB with more complex architectures on gender classification. This indicates that weights on the freezing layers are not suitable for this dataset, leading to a lower accuracy. This can be summarised that a fully customised model might be better for some problems.

Despite the impressive performance of both models on both tasks, there is still potential for improvement. Enhancing the dataset size and image resolution might allow the models to learn much more on facial data and make a more accurate prediction. However, these models demonstrate the ability to predict age accurately, which exceeds human's capability, indicating that they can be applied to solve more complex problems in various industries, such as detecting diseases in medical images.

Reference

[1] Chollet, F., 2017. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251-1258).