CM50265: Machine Learning 2
Coursework 1: Age Estimation and Gender Classification

- Report -

ID	Contribution
RK957	50%
TA904	50%

Do all the members agree with the above contributions? [Yes / ~~No~~]

The shared links for the two models are as follows:
- age_gender_A.h5 :
  https://drive.google.com/file/d/1rFyADalhI6oXhxj4KkBBoWXIeNP87PcT/view?usp=share_link
- age_gender_B.h5 :
  https://drive.google.com/file/d/1-4yM0SN1wypCg3KG532bw3gYlG9gpX-J/view?usp=sharing

Section 1: Introduction

This assignment revolves around the task of building and training two Convolutional Neural Networks (CNNs) models for classifying gender and predicting age from facial images on a given training face dataset: a subset of the UTKFace dataset (5000 face images) with GPU on Google Colab. The first model, model A, is defined and trained from scratch. The second model, model B, involves selecting and fine-tuning pre-trained model to adapt it to this specific task. The aim is to build and train models that could achieve both tasks with high accuracy.

Section 2: My own CNN (model A)

Model architecture details:

An architecture of the model A begins with a series of convolutional layers, progressively deepening from 64 to 512 filters, each followed by batch normalisation and ReLU activation to ensure non-linearity and training stability. After every two convolutional layers, a max pooling layer reduces the spatial dimensions of feature maps, optimising the network for computational efficiency. Prior to entering the fully-connected layers, the feature map is reduced below 10x10 size, focusing on the most critical features. The network culminates in two distinct output layers: a sigmoid-activated unit for binary gender classification and a ReLU-activated unit for age regression. To prevent overfitting, batch normalisation is used to normalise the input layer by adjusting and scaling the activations; dropout is applied after pooling and dense layers, which randomly omits a fraction of the neurons during training. Data Augmentations including scaling pixel values, rotating, shifting, adjusting brightness, and flipping images horizontally enhances diversity of dataset to improve model generalization and robustness.

Training process:

The training process is characterised using the Adam optimiser with initial learning rate of 0.001, chosen for adaptive learning rate capabilities. Loss functions are specifically tailored to each output: binary crossentropy for gender (binary classification) and mean MAE for age (regression). Training is conducted over 100 epochs, with batch size of 32, ensuring balance between computational demand and learning efficacy.
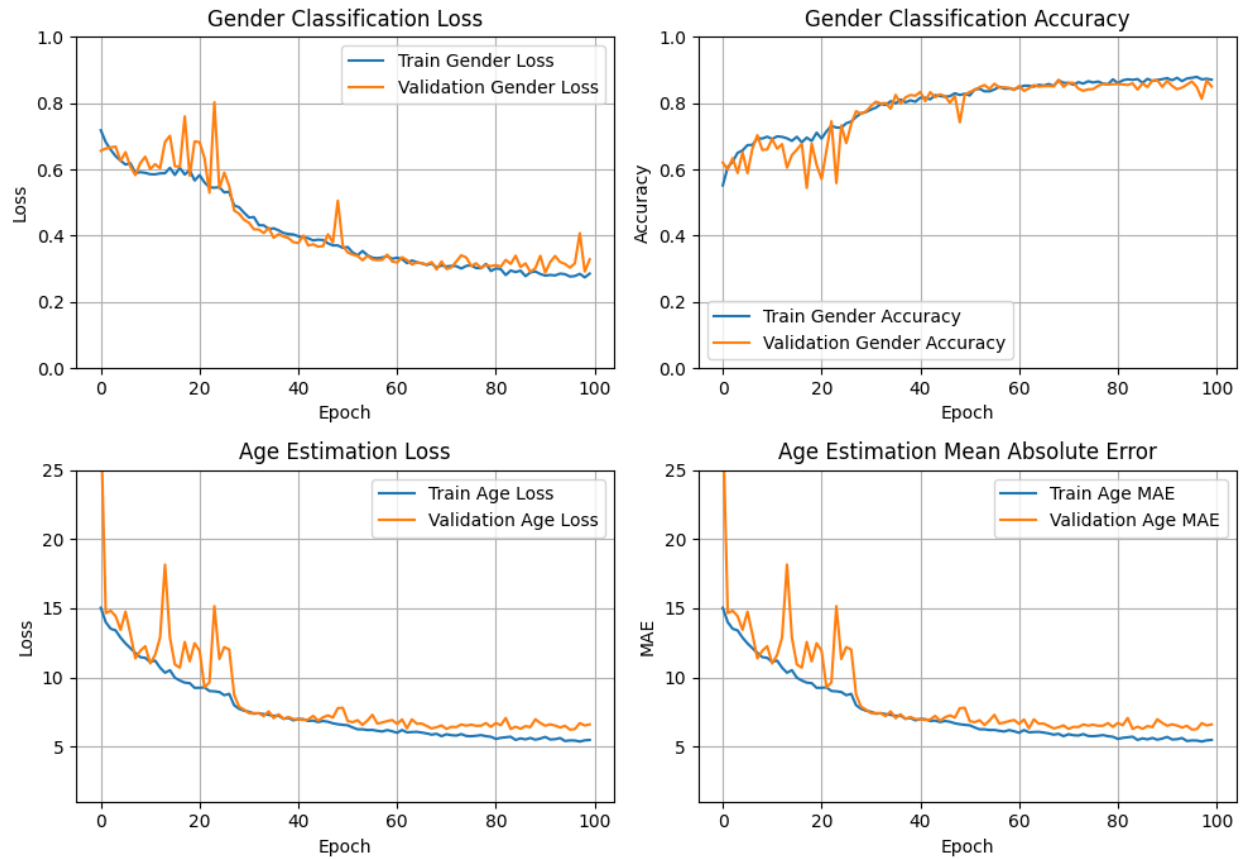
Fig. 1. Four learning curves of model A

Discussion of the performance:

The custom CNN model, model A, achieves an accuracy of 86-87% (validation dataset) in gender prediction, and age Mean Absolute Error (MAE) of 6-6.5 (validation dataset) in age prediction, demonstrating the model's adeptness at both gender and age relevant features extraction. The model's learning curves for both loss and accuracy indicate that the model converges well. While gender classification shows high accuracy, age estimation accuracy suggests slight room for improvement, highlighting the need for further optimisation strategies to enhance overall model performance and generalisation.

Section 3: Pre-trained CNN (model B)

Model architecture details:

An architecture of the model B leverages MobileNetV2, pre-trained on the ImageNet dataset, a decision driven by its compelling blend of efficiency and adaptability, and proven performance in visual recognition tasks. Initially, all layers in the MobileNetV2 base model are frozen to retain the comprehensive image features learnt from ImageNet, ensuring these foundational

insights remain intact during the initial adaptation phase. Building on pre-trained base, custom layers are introduced to address the dual objectives: age prediction (regression task) and gender prediction (binary classification task). Key enhancements include data augmentation, a GlobalAveragePooling2D layer for feature map compression and dense layers equipped with L2 regularisation and batch normalisation to mitigate overfitting. Additionally, an additive layer strengthens the model's predictive capacity. The model concludes in two distinct output layers: a sigmoid-activated unit for binary gender classification and a ReLU-activated unit for age regression.

Training process:

The model training begins with freezing all layers of the base model. Post an initial 5 epochs, progressively unfreezing the layers for fine-tuning is executed to gently adjust the mode to the dataset's specifics. The training process is characterised by use of Adam optimiser, aiming for efficient convergence, and minimising the loss functions assigned for each output: binary crossentropy for gender and mean absolute error for age. A ReduceLROnPlateau callback is applied to adjust learning rate in response to validation performance, ensuring continual progress even when facing learning plateaus. The mode trains over 90 epochs, with data batch-processing facilitating efficient learning. This gradual unfreezing, combined with dynamic learning rate adjustments, allows model to refine its pre-learned pattern for age and gender prediction.
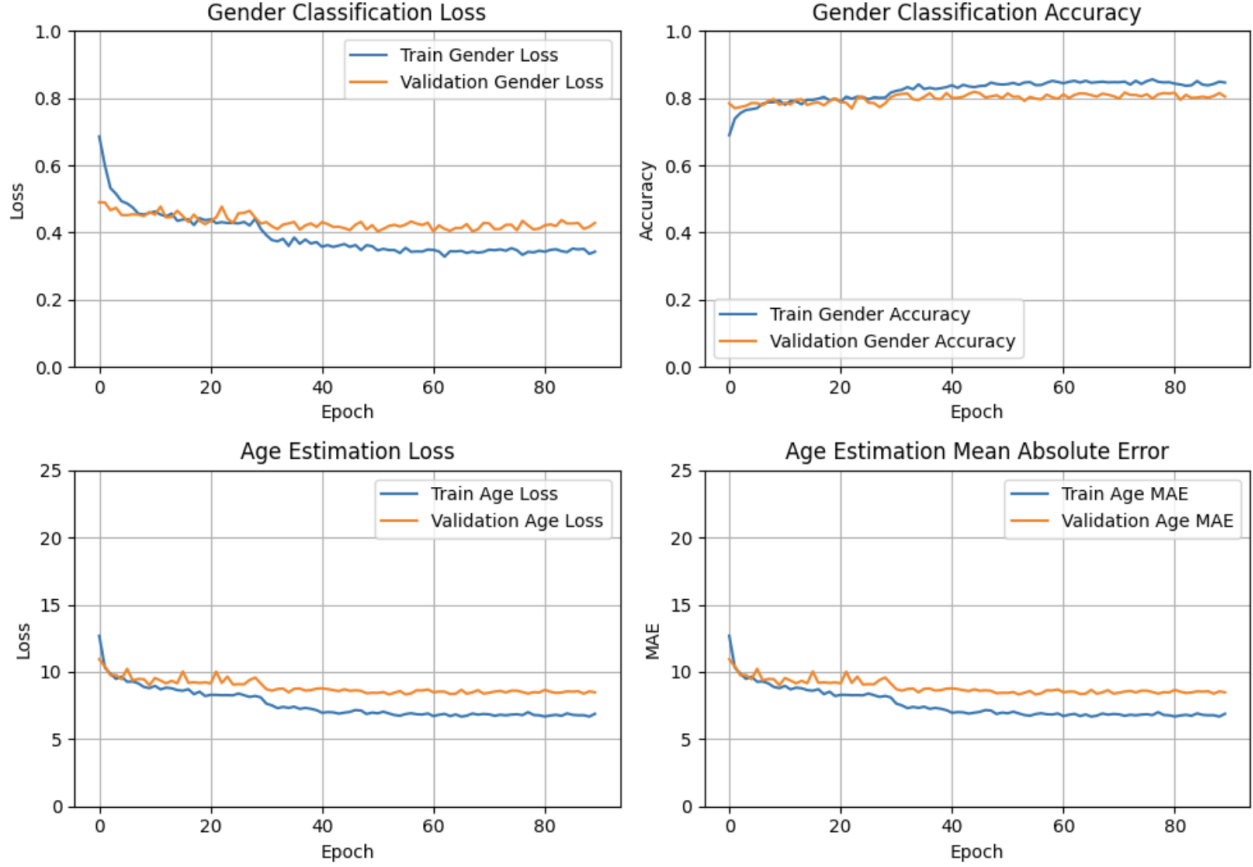
Fig. 2. Four learning curves of model B

Discussion of the performance:

The pre-trained model, model B, shows promising results in gender prediction with an accuracy of 82-83% (validation dataset), which highlights its capability to effectively discern gender-specific features. However, the age prediction, indicated by the Mean Absolute Error (MAE) of 8-8.5 (validation dataset), suggests room for refinement. Age prediction, a regression task, presents a more complex challenge than binary classification, requiring the model to capture more nuanced, age-relevant features from the facial images. The strategic gradual unfreezing of layers post-initial epochs likely contributes positively to the model's adaptation to these specific tasks, allowing for refined learning without significant overfitting, as evidenced by the balance between training and validation loss. Overall, results demonstrate the model's capability to adapt pre-learnt features to new, specific tasks efficiently.

Section 4: Summary and Discussion

The comparative analysis between Model A (custom CNN model) and Model B (pre-trained model) on validation dataset reveals distinct outcomes: Model A achieves a gender prediction accuracy of approximately 86-87% and an age prediction MAE of 6-6.5, whereas Model B records a slightly lower performance with 82-83% accuracy for gender and an age MAE of 8-8.5.

Model A's superior performance demonstrates the value of a unique approach in machine learning, where models are carefully designed and rigorously optimised for the task at hand. The architecture of Model A, tailored specifically for age and gender prediction, adeptly captures and interprets relevant facial features, benefiting from targeted optimization and regularization strategies. This demonstrates that with careful design and a focused training regimen, custom models can achieve and even surpass efficiency of pre-trained models in some scenarios.

On the other hand, Model B's performance, while slightly lagging behind Model A, offers a valuable perspective on limitations of transfer learning. The slight shortfall in accuracy and precision suggests that while pre-trained models provide a robust starting point, their effectiveness depends upon the extent and specificity of fine-tuning employed. This aligns with understanding that pre-trained models require meticulous alignment with the specific features and nuances of the new dataset and task.

This assignment not only showcases the mechanics of CNN architectures and the strategic implementation of deep learning solutions but also illustrates the broader implications of model choice in real-life problem-solving. Furthermore, this comparison emphasises how important it is to keep an iterative and informed approach to model development and optimization. It shows the importance of understanding when to invest and develop custom models and when to leverage complexity of pre trained models. The insight is invaluable for strategic decision-making and effective problem solving in developing deep learning models.