MACHINE LEARNING 2 - COURSEWORK 1

AGE ESTIMATION AND GENDER CLASSIFICATION

ID	CONTRIBUTION
ID 1	50%
ID 2	50%

Do all the members agree with the above contributions? **Yes**

Specify two shared links for the two models:

age gender A.keras for the CNN model: age_gender_A.keras

age gender B.keras for the pre-trained CNN model: age_gender_B.keras

SECTION 1: INTRODUCTION

Age estimation and gender classification are tasks that humans perform effortlessly, yet they present a significant challenge for machines. With the rapid advancements in deep learning, particularly in the field of computer vision, it is now possible to train models to predict age and gender from facial images with impressive accuracy. This coursework focuses on building and training Convolutional Neural Network (CNN) models to classify gender and estimate age from facial images, developing one model from scratch and fine-tuning another from a pre-trained architecture.

The dataset used in this study is a subset of the UTKFace dataset, which contains 5,000 labelled face images. The UTKFace dataset is a large-scale collection of over 20,000 face images, spanning ages from 0 to 116 years, annotated with age, gender, and ethnicity, and captures diverse variations in pose, facial expression, illumination, occlusion, and resolution.

Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are deep learning algorithms designed to process input images by applying convolution operations (Chauhan et al., 2018).. CNNs consist of four main layers: the convolutional layer, pooling layer, fully connected layer, and activation layer.

The convolutional layer is responsible for detecting essential patterns and features by applying kernel filters to compute convolutions on the input image. The pooling layer reduces the dimensionality of feature maps by down-sampling, decreasing computational load, minimising overfitting, and retaining important information. It typically follows one or more convolutional layers to enhance feature extraction.(Purwono et al., 2023)

The fully connected layer called the convolutional output layer, is similar to a feedforward neural network. It receives flattened output from the final pooling of the convolutional layer and processes it for classification or prediction. The activation layer introduces nonlinearity into the network, enabling it to learn complex patterns. Common activation functions in CNNs include Sigmoid, ReLU, and Softmax (Purwono et al., 2023).

Pre-processing and data augmentation

Before pre-processing, the dataset was visualised to understand its structure and characteristics, as illustrated in Figure 1. The pre-processing step involves splitting the dataset into training and validation sets, loading image data along with gender and age labels, and normalizing pixel values to the range [0, 1] by dividing by 255. To improve dataset diversity and model generalization, TensorFlow's Sequential API applies data augmentation with rotations (15°), zoom (0.1), horizontal flips, and adjustments to brightness (0.001), contrast (0.1), and hue (0.1).

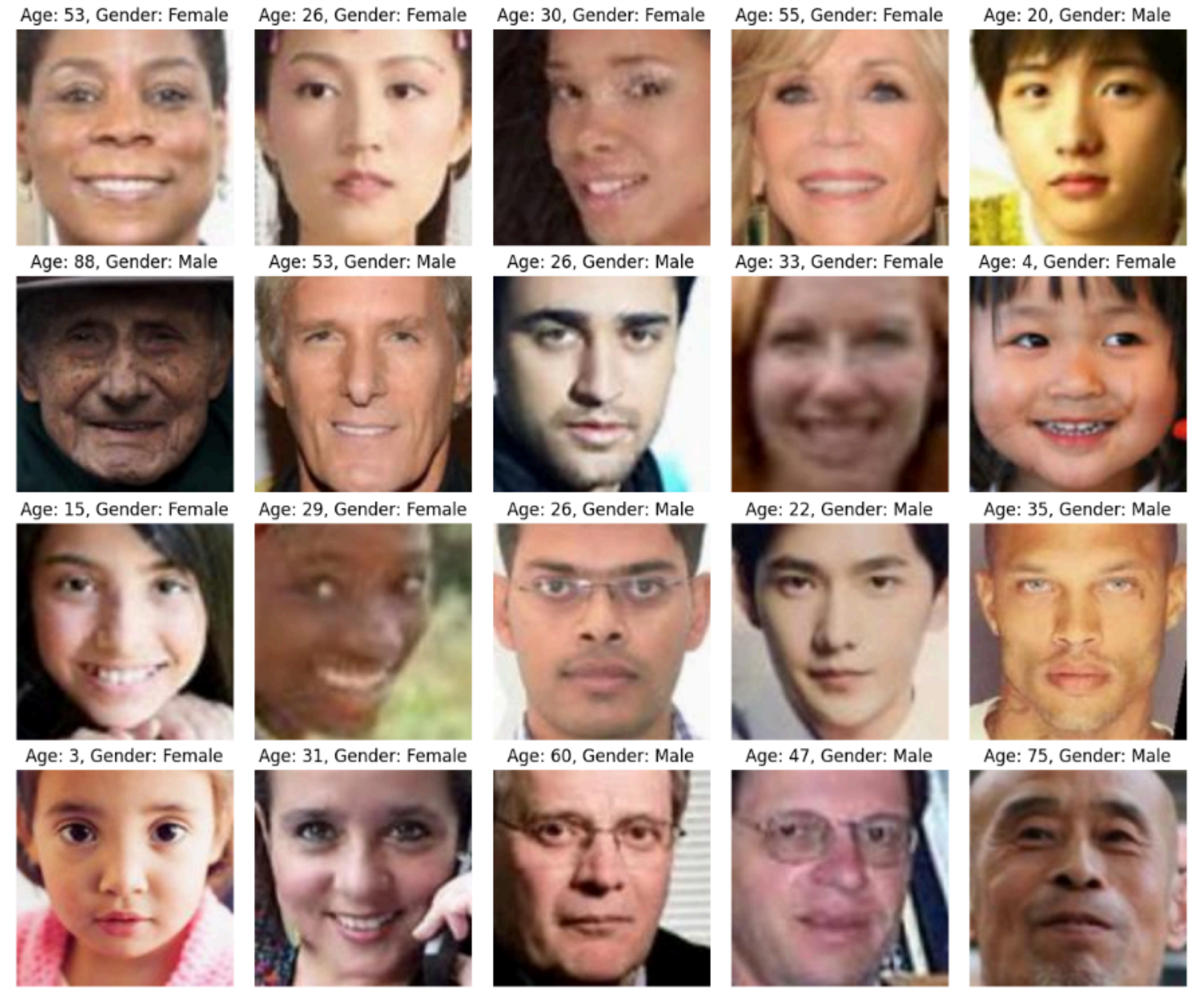


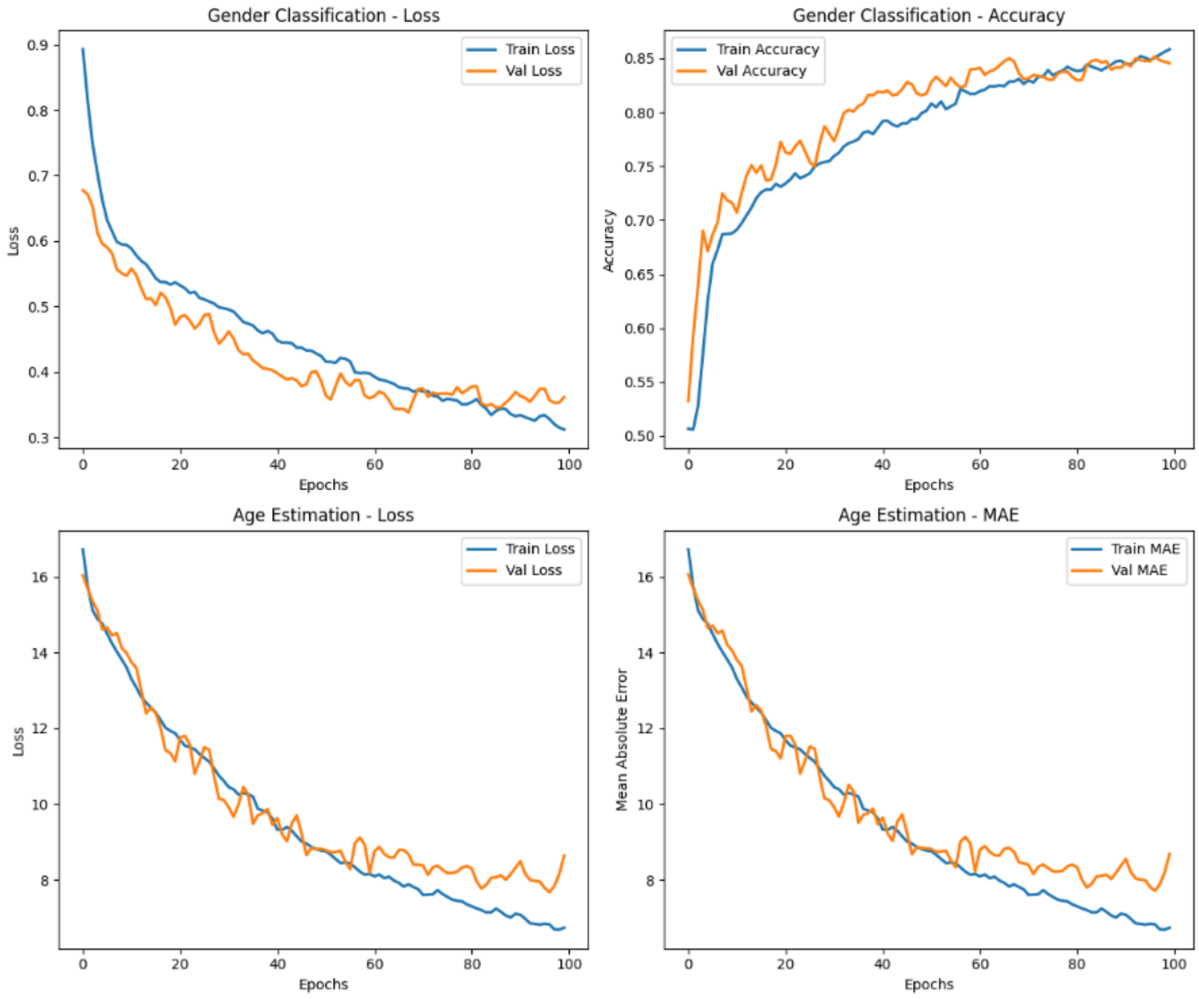
Figure 1: Visualization of the UTKFace dataset

SECTION 2: THE CUSTOM CNN

The proposed Convolutional Neural Network (CNN) architecture is designed to simultaneously predict age and gender from input images each of size 128x128x3. The model comprises four convolutional layers, each followed by max-pooling layers to reduce spatial dimensions while extracting hierarchical features. The input image is passed through the input layer, which serves as the starting point for feature extraction. Each convolutional layer uses Conv2D to perform convolution operations, where filters (or kernels) slide across the input image to detect patterns such as edges, corners, and textures.

The first convolution layer uses 32 filters of size 3x3, producing a feature map of size 128x128 that captures basic features. As the network deepens, the number of filters increases (64,128 and 256) to detect more complex patterns like textures and shapes. Each convolutional layer is followed by max-pooling to reduce spatial dimensions, ensuring feature maps are smaller than 10x10 before the fully connected layer. The final pooled feature map is flattened into a 1D vector and passed through a 128-unit fully connected layer with ReLU activation and dropout regularization, preventing overfitting. The model has two outputs: gender classification using a sigmoid function and age regression using a linear function.

The training process follows a structured approach to optimize both age estimation and gender classification. The model is compiled using the Adam optimizer with a default learning rate, ensuring efficient weight updates. It employs a dual-loss strategy: Mean Absolute Error (MAE) for age prediction and Binary Cross-Entropy for gender classification. Loss weights are set to 0.7 for age estimation and 1.5 for gender classification to balance the contribution of both tasks. The model is trained for 100 epochs with a batch size of 32, using training data. To enhance generalization and adapt the learning rate dynamically, the ReduceLROnPlateau callback is implemented, which reduces the learning rate by a factor of 0.5 if validation loss does not improve for 7 consecutive epochs, with a minimum learning rate threshold of 0.0001. The training process includes the validation dataset, ensuring that model performance is monitored throughout. The model's effective performance is tracked through MAE for age estimation and accuracy for gender classification.


Gender Classification - Loss

Gender Classification - Accuracy

Age Estimation - Loss

Age Estimation - MAE

Figure 2: Learning curves for the custom CNN

Figure 2 depicts the loss and accuracy curves of age and gender. Over 100 epochs, the training accuracy for gender classification improved from 49.89% to 85.74%, while validation accuracy increased from 55.10% to 84.70%. Similarly, the training loss for gender classification decreased from 0.93 to 0.31, and the validation loss reduced from 0.67 to 0.35.

For age estimation, the training MAE decreased from 18.76 to 6.73, while the validation MAE improved from 16.12 to 8.65 by the 100th epoch, demonstrating significant performance enhancement.

SECTION 3: THE PRE-TRAINED CNN

The designed model uses VGG16Net, a widely used deep convolutional neural network (CNN) pre-trained on the ImageNet dataset (Thamina, 2019). The VGG16 network is initially imported without its fully connected layers (include_top=False), retaining only the convolutional base, which remains frozen (trainable=False) to maintain the learned feature representations and avoid overfitting during the fine-tuning. Subsequently, a customized fully connected was added. The output feature maps extracted from VGG16 are initially flattened, followed by a fully connected layer containing 256 units with ReLU activation through a dense layer. To improve training stability, a batch normalization layer along with dropout (0.4) is added to reduce the risk of overfitting. An additional dense layer with 128 units is implemented, which incorporates L2 regularization ($\lambda = 0.001$), to further refine the features, followed by batch normalization and dropout. The architecture splits into two distinct outputs: Age estimation, which utilizes a single neuron with a linear activation function for regression tasks; and Gender classification, which employs a single neuron with a sigmoid activation function, suitable for binary classification.

The training process involves two phases: initial training with a frozen VGG16 base model and fine-tuning by unfreezing the last few layers for further refinement. The model is compiled using the Adam optimizer with an initial learning rate of 0.001. The loss functions are defined separately as Mean Absolute Error (MAE) for age prediction and Binary Cross-Entropy for gender classification. During initial training, the model is trained for a maximum of 100 epochs with a batch size of 16. To optimize training, callbacks such as ReduceLROnPlateau are used, which reduces the learning rate by a factor of 0.5 if the validation loss stagnates for five consecutive epochs, with a minimum learning rate of 1e-6. ModelCheckpoint is implemented to save the best model based on validation loss, and EarlyStopping is used to halt training if the validation loss does not improve for 15 epochs while restoring the best weights. After the initial training, the model is fine-tuned by unfreezing the last 10 layers of the base model and reducing the learning rate to 0.0001 to ensure stable updates. Fine-tuning is performed for 50 epochs with the same batch size of 16 and defined callbacks as in the initial training. This process allows the model to adapt to more specific features in the data while maintaining stability.

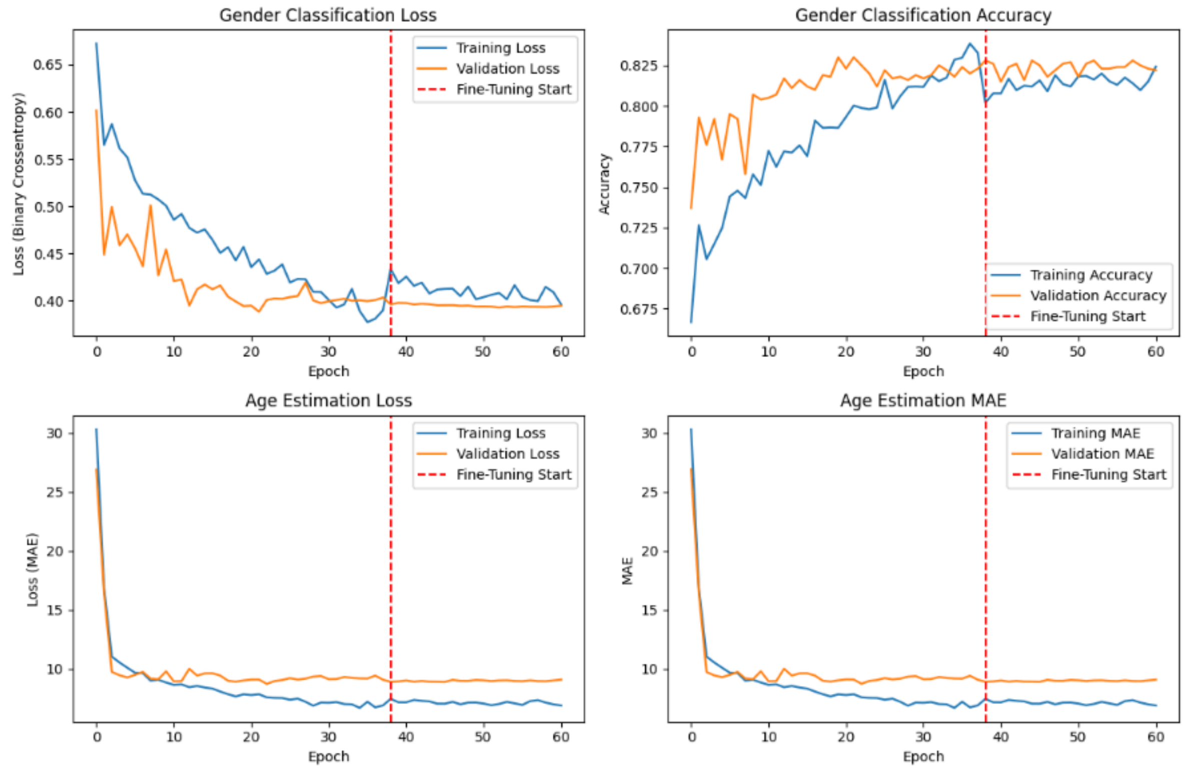


Figure 3: Learning curves for the pre-trained CNN

Figure 3 illustrates the training and validation performance for gender classification and age estimation. For gender classification, the loss for training and validation decreases from 0.73 to 0.43 before fine-tuning. After fine-tuning, training loss then further stabilizes with little fluctuations till 0.41 while validation loss stabilizes steadily to 0.39. Whereas gender accuracy for training and validation increases from around 63.20% to 83.0% before fine-tuning. Afterwards, with a slight drop, the training accuracy improves from 80.19% to 81.76% and validation accuracy fluctuates around 82%. For age estimation, before fine-tuning, the MAE gradually decreases from 32.46 to around 7-8 for training, and from 26.93 to 9.08 for validation. After fine-tuning, both training and validation MAE stabilize at these values with only minor fluctuations.

SECTION 4: SUMMARY AND DISCUSSION

Table 1 presents the performance comparison between the custom CNN and the pre-trained model for gender classification and age estimation. To sum up, the custom CNN outperforms the pre-trained model in both tasks.

MODEL	GENDER ACCURACY		AGE MAE	
	TRAIN (%)	VALIDATION(%)	TRAIN	VALIDATION
THE CUSTOM CNN	85.74	84.70	6.73	8.65
THE PRE-TRAINED MODEL	81.76	82.20	7.17	8.89

Table 1: Summarization of the model

To conclude. this assignment provided insights into designing and training CNNs for multi-task learning in age and gender prediction.A key challenge was balancing the dual-task nature of the problem,, as improving age estimation could hinder gender classification. The choice of hyperparameters, such as learning rate, batch size, dropout rate, and loss function weighting, played a crucial role in achieving this balance. The custom CNN, designed from scratch, provided control over feature extraction through carefully chosen convolutional layers, max pooling, and dropout regularization. On the other hand, the VGG16-based model enabled efficient transfer learning but required careful fine-tuning to prevent overfitting.

Ultimately, a combination of data preprocessing, feature extraction, and optimization techniques proved essential for effective simultaneous age and gender prediction.

REFERENCES

Chauhan, R., Ghanshala, K. K., and Joshi, R. (2018) ‘Convolutional Neural Network (CNN) for image detection and recognition’, *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, 278-282.

https://doi.org/10.1109/icsccc.2018.8703316

Purwono, P. et al. (2023) ‘Understanding of Convolutional Neural Network (CNN): a review’, *International Journal of Robotics and Control Systems*, 2(4), pp. 739–748. https://doi.org/10.31763/ijrcs.v2i4.888.

Tammina, S. (2019) ‘Transfer learning using VGG-16 with deep convolutional neural network for classifying images’, *International Journal of Scientific and Research Publications (IJSRP),* 9(10), pp.143-150. http://dx.doi.org/10.29322/IJSRP.9.10.2019.p9420