CM50265: Machine Learning 2

Coursework 1: Age Estimation and Gender Classification

Feburary, 2025

Set: 8pm, 10 February (Mon.), 2025	
Due date for group signing-up: **8pm, 14 Feburary (Fri.), 2025**	
Due date for submission: **8pm, 14 March (Fri.), 2025**	
Percentage of overal unit mark: 20%	
Submission location: Moodle	
Submission components:	1. A notebook (.ipynb)
	2. A report (.pdf)
	3.* Two CNN models (.keras)
Submission format: See above	
Anonymous Marking: Yes	

*Note: *Due to the large size of models, they cannot be directly uploaded to Moodle. Instead, you are required to include their sharable links in your report. Make sure they are set to be "shared to anyone with the link".*

1 Overview

Humans can easily identify whether a person is a female or a male and estimate their age with reasonable precision. Now given a large labelled dataset, the computer can also perform gender classification and age estimation. Your task is to build and train deep learning models to predict gender and age on a given training face dataset with GPU in Google Colab. During the coursework evaluation, your models will be tested on a separate test set to assess their accuracy in both tasks.

2 Group of two

You are strongly encouraged to work in pairs (2 students per group), as teamwork is a vital skill for your future career. This assignment offeres a great opportunatity to develop that skill. You are free to form your own groups. Please ensure that you sign up a group on Moodle by the deadlind specified in the above table. Note that *each group only need have one submission. DO NOT include your names anywhere in your code or report.*

It is expected that all students participate in a group for this assignment. However, if you have a compelling reason to work individually, such as Individual

Learning Plan (ILP), please imform me before the group sign-up deadline. If you are allowed to work individually, you still need sign a group on Moodle. Please note that the assessment criteria will remain impartial, regardless of whether the assignment is completed individually or in a group.

3 The discription

This assignment consists of two compulsory components:

3.1 Part 1 – Coding

Train two CNN models:

- One is defined by you with a few restrictions and be trained from scratch, save it as `age_gender_A.keras`
- The other is to fine-tune a pre-trained model, save it as `age_gender_B.keras`.

Full details can be found in the provided notebook file. Your models will be subject to testing on unseen test data.

3.2 Part 2 - Report

You will write a report (max 1500 words). This limit does not include references, diagrams, tables and figures. In the report, please follow the following structure or similar.

- A **cover page** consists of the following:
  - A form to specify how much contribution each member made, based upon your agreed input. If each member took equal contribution, this would be all 50% (which is the general expectation for this assignment). More or less than 50% means higher or lower contribution for the project. The maximum contribution difference allowed is 70%: 30%.

ID	Contribution
ID1	50%
ID2	50%

    Reasons for different weights [Optional] :

    Do all the members agree with the above contributions? [Yes/No]
  - **Specify two shared links** for the two models:

    `age_gender_A.keras` for the CNN model you defined and trained from scratch;

    `age_gender_B.keras` for the pre-trained CNN model fine-tuned on this dataset.

- **Section 1: Introduction** Give a brief description of what this assignment is about.
- **Section 2: The custom CNN** Provide a description of your custom CNN architecture, detail the training process and include a disucssion of its performance.
- **Section 3: The pre-trained CNN** Provide a description of the pre-trained CNN architecture, detail the training process and include a disucssion of its performance.
- **Section 4: Summary and discussion** Compare and discuss the two models. Reflect on what you have achieved through this assignment and include any additional discussions or insights you have.

Both your code and report will be assessed jointly. Only submitting Part 1 or Part 2 will lead to a score of zero mark for the entire assignment.

4 Dataset downloading

You will use a subset of the UTKFace dataset. A `train_val` folder has been created by choosing a subset (5,000 face images) from the UTKFace dataset for you to train and validate your model. It is a shared google drive folder below.

`https://drive.google.com/drive/folders/1UjYRDyo10Fx-Rv91CQl5ZfwF85HiLUX8?usp=sharing`

- Click the link above. It must be in the "shared with me" in your Google Drive.
- Right-click this folder, choose "Organize" $\Rightarrow$ "Add shortcut to Drive", then specify the path where you would like to put it.
- Then you can access its path after you mount the drive.

UTKFace dataset is a large-scale face dataset with long age span (range from 0 to 116 years old). The dataset consists of over 20,000 face images with annotations of age, gender, and ethnicity. The images cover large variation in pose, facial expression, illumination, occlusion, resolution, etc. The labels of each image are embedded in the file name, formatted as `[age]_[gender]_[race]_[date&time].jpg`. You only need to use the first two labels for this assignment:

- **age** is an integer from 0 to 116, indicating the age.
- **gender** is either 0 (male) or 1 (female).

***Ethical disclaimer**: We acknowledge that some individuals may feel uncomfortable with aspects of this dataset, such as predicting a person's age and gender based solely on their appearance, or the limitation of gender labels to only*

'female' and 'male', etc. For this course, we are using the dataset purely as a machine learning exercise and are not addressing these moral concerns within this context. However, it is crucial that you carefully consider potential ethical implications when using any dataset in your future research or professional work.

5 Notebook file to start with

Download the attached notebook file (`age_gender_submit.ipynb`) for this assignment. You must follow the instructions provided within the notebook and add all your code as required. It is unlikely that I will run your code; therefore, it is ensential for you to retain all the output cells to demonstrate that your code runs correctly.

6 Marking Criteria

Your report and your code will be assessed jointly. Detailed marking criteria are listed in the attached `ML2-CW1-mark-criteria.pdf`. Please note that the criteria may be slightly adjusted during the marking process.

How to calculate individual marks in case of different contribution?

All members of a group are supposed to contribute equally, therefore will get the same mark. However, in the case that each member has different contribution, your mark consists of 60% of the common mark and 40% of your contribution:

*Group_mark * 0.6 + Group_mark * 0.4 * 2 * your_contribution*

For example, if one group's grade is 72, the contributions for the two members are 60% and 40%, then their marks would be:

Student 1 with 60% contribution: 78 (=72*0.6+72*0.4*2*0.6)

Student 2 with 40% contribution: 66 (=72*0.6+72*0.4*2*0.4)

7 Extension and Late submissions

Requests for extension should be made to the Director of Studies. Lecturers and tutors cannot approve extensions. Please make sure you are familiar with the department's coursework deadline extension policy.

The university policy will be followed on late submissions. If a piece of work is submitted after the submission date, the maximum possible mark will be 40% of the full mark. If work is submitted more than five days after the submission date, student will receive zero marks.

8 Academic Integrity and Use of Generative AI Tools

This coursework is classified as **Type A: GenAI is not permitted.** This category is applied to ensure that all work submitted reflects your own understanding, critical thinking, and original effort. By completing this work independently, you will gain hands-on experience with the fundamental concepts and techniques, strengthen your problem-solving abilities, and build a deeper understanding of ML models. Please do not use any GenAI tools to create any part of your submission.

All submissions will be checked with anti-plagiarism detection service provided by the University.

Please ensure that you write your code and your report independently of other groups. Copying is a serious issue as you all know. If you copy information directly from another source without attributing it through proper citation, your case will be referred to the Director of Studies for further investigation. For more details, please visit:

`http://www.bath.ac.uk/library/help/infoguides/plagiarism.html`
`http://www.bath.ac.uk/library/infoskills/referencing-plagiarism/`

9 Any questions about the assignment

Please do not email questions about this assignment. Instead, post all assignment-related questions in the **Moodle Discussion Forum (with a title starting with [CW1])**. This way, everyone can benefit from the answers.