Assignment 1B

CAB420, Machine Learning

This document sets out the two (2) questions you are to complete for CAB420 Assignment 1B.

The assignment is worth 15% of the overall subject grade. All questions are weighted equally. Students are to work individually. Students should submit their answers in a single document (either a PDF or word document), and upload this through the Canvas submission link.

Further Instructions:

1. Data required for this assessment is available on canvas alongside this document in *CAB420_Assessment_1B_Q1_Data.zip* for Question 1, and the provided template will download the required data for Question 2.
2. Juptyer Notebook python templates have been provided for each question on Cavnas.
3. Answers should be submitted via Cavnas. In the event that Cavnas is down, or you are unable to submit via Cavnas, please email your responses to `cab420query@qut.edu.au`.
4. For each question, **a concise written response** is required. Guidelines for the length of the written responses have been supplied in this assignment brief. These guides pertain to the **written** component of each section (ie. excluding figures), though figures should be included to enhance the response only. Figures without sufficient discussion show no evidence of understanding of the content or the task. Note that the CRA requires responses to be "clear and concise", so verbose and lengthy submissions that needlessly disregard these guides will be penalised according to the CRA. This similarly applies to the excessive use of figures without sufficient discussion or evaluation.
5. Responses should explain and justify the approach taken to address the question (including, if relevant, why the approach was selected over other possible methods), and include results, relevant figures, and analysis. **Python Notebooks, or similar materials will not on their own constitute a valid response to a question and will score a mark of 0.**
6. Responses should highlight where the modelling approach has been successful, and to highlight the limitations apparent. When limitations are encountered, students are encouraged to investigate and identify what the underlying cause of any failure or limitations are.

7. Python code, including live scripts or notebooks (or equivalent materials for other languages) may optionally be included as appendices. **Figures and outputs/results that are critical to question answers should be included in the main question response, and not appear only in an appendix**.

8. Students are encouraged to use **any code presented in lectures/tutorials/examples** from CAB420. Use of external code is also permitted, though requires to be correctly cited and referenced.

9. Students who require an extension should lodge their extension application with HiQ (see `http://external-apps.qut.edu.au/studentservices/concession/`). Please note that teaching staff (including the unit coordinator) cannot grant extensions.

Problem 1. Person Re-Identification. Person re-identification is the task of matching an image of a person (a *probe* sample) to a *gallery* of previously seen people. The problem can be seen as a ranking or retrieval task, in that the *gallery* samples are ranked based on their similarity to a given *probe* sample. Like many biometrics tasks, a common approach involves the use of dimension reduction techniques such as PCA and/or LDA, or deep learning and Siamese networks, to learn a compact sub-space in which samples can be compared. Ideally, this sub-space will be such that samples that belong to the same subject will lie close to one another.

Person re-identification (and performance for other retrieval tasks) is commonly evaluated using Top-N accuracy and Cumulative Match Characteristic (CMC) curves. Top-N accuracy refers to the percentage of queries where the correct match is within the closest N results, and is measured by ranking *gallery* samples based on their similarity to the *probe*, and determining the location of the true match within the ranked list. Ideally, the top result (i.e. the closest *gallery* sample to the *probe*) will be the same subject. A CMC curve plots the top-N accuracy for all possible values of N (from 1 to the number of unique IDs in the dataset).

You have been provided with a portion of the Market-1501 dataset [1]
(see `CAB420_Assessment_1B_Q1_Data.zip`, a widely used dataset for person re-identification. This data has been split into two segments:

- `Training`: consists of the first 300 identities from Market-1501. Each identity has several images. In total, there are $5,933$ colour images, each of size $128x64$.
- `Testing`: consists of a randomly selected pair of images from the final 301 identities. All images are colour, and of size $128x64$. These images have been divided into two directories, `Gallery` and `Probe`, with one image from each ID in each directory.

Using the `Training` dataset, a model to extract a compact representation of a sample can be trained. The resultant transform can then be applied to the `Testing` set to transform samples to a lower-dimensional representation, at which point samples can be matched. The testing set is broken into `Gallery` and `Probe` sets. To match samples, each sample in the `Probe` set can be compared to each image in the `Gallery` set, and based on the distance between pairs of probe and gallery samples the most similar instances can be identified.

Your Task: Using this data, you are to:

1. Develop and train a **non-deep learning** method using one of the dimension reduction methods covered in Week 7 for person re-identification. The method should be evaluated on the test set by considering Top-1, Top-5 and Top-10 performance. A CMC (cumulative match characteristic) curve should also be provided.
2. Develop and evaluate a **deep learning based** method for person re-identification, using one of the methods covered in Week 8. The method should be evaluated on the test set by considering Top-1, Top-5 and Top-10 performance. A CMC (cumulative match characteristic) curve should also be provided.

3. Compare the performance of the two methods. Are there instances where the non-deep learning method works better? Comment on the respective strengths and weaknesses of the two approaches.

You have been provided with sample python code to:

- load all images, and to transform them into a vectorised representation for non-deep learning methods;
- resize images and transform the images to greyscale, which you may or may not wish to use;
- build image pairs or triplets for use with metric learning methods;
- plot a CMC curve given a set of distances between gallery and probe samples.

Your final response should include sections that address the following:

- What pre-processing (i.e. resizing, colour conversion) you apply to the data and why. Note that you do not need to pre-process the data, in which case you should explain why you are using the data 'as is'. For fair comparison, use the same pre-processing for both the methods (non-deep learning and deep-learning). **(1/2 page)**
- Details of the selected **non-deep-learning** approach(es) used, including justification for their selection; and of the **deep-learning** approaches used, including justification for their selection. Your justification should consider the nature of the problem, constraints imposed by the data, and computational considerations. **(1 page)**
- An evaluation that compares the two methods, reporting Top-1, Top-5 and Top-10 accuracy, and including CMC curves for the developed methods. Your evaluation should consider instances where performance differs between the two methods, and comment on the respective strengths and weaknesses (including computational efficiency/runtime) of the two approaches. **(1 page)**
- A **brief** discussion on ethical considerations of person re-identification. Research into person re-identification has enabled facial recognition for securing devices by Microsoft, Apple, and more[1]. There are also ethical considerations pertaining to person re-identification with respect to how data for training these models is collected, and how some of these models may be used in applications like surveillance. Briefly discuss the **ethical concerns** that need to be considered when developing person re-identification systems. This section will require additional research on your behalf, though to aid in starting discussing and investigating ethical considerations, we provide the following references.
  - Duke University removing the Duke-MTMC person re-identification dataset [2], following investigations from [3] and [4]

[1] Due to intellectual property reasons, Microsoft and Apple don't explicitly announce how they implement their the face authentication systems, though using more advanced methods of the models covered week 8 on metric learning would be sensible.

- A paper proposing a method for anonymous tracking and person re-identification: “REVAMP 2 T: real-time edge video analytics for multi-camera privacy-aware pedestrian tracking” [5]
- Given person re-identification is often used in video surveillance, it is worth considering some recent reviews of video surveillance with new legislation in different parts of the world [6].
- Amazon, Microsoft and IBM have been urged to stop selling facial recognition (enabling person re-identification) to law enforcement due to limitations in the systems and how they were being used [7].

These references are designed to serve as a starting point for an investigation into ethical considerations into person re-identification tasks. Each of these sources contain additional references which may be useful. In this section, please **briefly** discuss ethical considerations pertaining to,

- How data is collected to train these models
- How these models may be used (ie. authentication for digital devices, surveillance etc.)
- Limitations of these models

You may use the provided resources to form your investigation, though additional resources may be used (and is encouraged but not necessary). All sources must be cited correctly, using any well-defined referencing format. **(1/2 page, excluding citations)**

Problem 2. Multi-Task Learning and Fine Tuning. This task focuses on the Oxford-IIIT Pets Dataset [8] (and shown in Figure 1), which contains images of cats and dogs with labels corresponding to their breed (total of 37 classes), and semantic segmentation information labelling the foreground (cat or dog pixels) and the background.

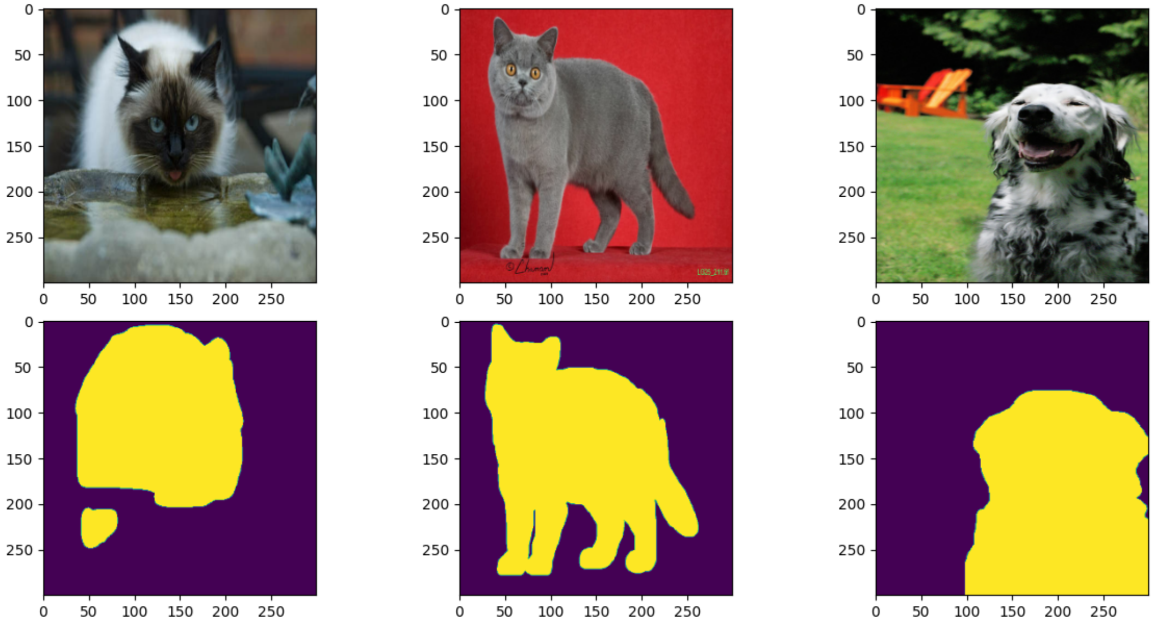

Figure 1: Sample Images from the Ofrod-IIIT Pets Dataset. The top row shows three sample images. The bottom row shows the corresponding semantic masks (foreground shown in yellow, background shown in purple).

Original images from this dataset are of varying size, with some upwards of 400x400 pixels, though many are much smaller.

Your Task: Using this data, you are to:

1. Create two models suitable for performing image classification and semantic segmentation simultaneously. The first model you will design and train from scratch yourself, while for the second you will fine-tune an existing model. The fine-tuned model for this task will be MobileNetV3Small `https://www.tensorflow.org/api_docs/python/tf/keras/applications/MobileNetV3Small`.

2. Evaluate the models using the pre-defined training/testing splits for the dataset, and critically evaluate the performance of both methods (from-scratch and fine-tuned) for both tasks.

It is expected that the models developed for this task with the limited compute available will not perform well. This is intended, and we stress that a high accuracy is **NOT** required. Part of your task is to investigate where the models are failing, identify potential causes for this, and attempt to mitigate this. It is expected that your approach to addressing this problem will be iterative, in that you will you implement initial methods, evaluate these, and then attempt to improve performance through various strategies.

You have been provided with a sample python notebook which:

- Implements a data-loader for this task. This uses the built in data loader from `https://www.tensorflow.org/datasets/catalog/oxford_iiit_pet`, which will download the data for you and provide access to it through the `tf.data.Dataset` api. This dataset has been built upon for you to only include the data necessary (image input, class labels, segmentation masks). It is expected that you will need to modify this data loader to adjust image sizes, modify augmentation setting, etc.
- An example of and instructions for how to load the `MobileNetV3Small` model for fine-tuning.
- Links to additional resources for model fine-tuning within Keras.

Your final response should include sections that address the following:

- What pre-processing (i.e. resizing, colour conversion, augmentation, etc.) you apply to the data and why. Your pre-processing should consider aspects such as network size/complexity, and any compute restrictions imposed. For fair comparison, use the same pre-processing for both the methods (from-scratch and fine-tuned). **(1/3 page)**
- Details of two implemented methods. This should include a details of the final "from-scratch" approach and justification for the chosen design, and details of changes made to MobileNetV3Small for the "fine-tuned" approach. Details on how the models are trained are also to be provided. Justifications should refer to existing models for similar tasks, the design principles covered within the lectures, and compute restrictions imposed. **(1 page)**
- An evaluation that compares the two models for the two tasks (classification and semantic segmentation). Your evaluation should discuss overall model performance, how it differs between the two approaches, and include figures if/where necessary. **(2/3 page)**
- A discussion of methods that were explored to improve performance for both models and mitigate identified issues, and potentially other methods that were considered but not implemented due to computational constraints. We do not require your investigation to necessarily yield improved results, rather we require that your proposed methods to improve model performance are logical and consistent with difficulties encountered. If your investigation does not yield improved performance, reflect on potential reasons why this might have occurred. Include figures/tables as needed to support illustrate the impact on performance of the proposed methods. **(1 page)**

References

[1] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, ser. ICCV '15. USA: IEEE Computer Society, 2015, p. 1116–1124.

[2] J. Satisky. (2019) A duke study recorded thousands of students' faces. now they're being used all over the world. [Online]. Available: https://www.dukechronicle.com/article/2019/06/duke-university-facial-recognition-data-set-study-surveillance-video-students-china-uyghur

[3] J. Harvey, Adam. LaPlace. (2021) Exposing.ai. [Online]. Available: https://exposing.ai

[4] M. Murgia. (2019) Who's using your face? the ugly truth about facial recognition. [Online]. Available: https://www.ft.com/content/cf19b956-60a2-11e9-b285-3acd5d43599e

[5] C. Neff, M. Mendieta, S. Mohan, M. Baharani, S. Rogers, and H. Tabkhi, "Revamp 2 t: real-time edge video analytics for multicamera privacy-aware pedestrian tracking," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 2591–2602, 2019.

[6] D. Almeida, K. Shmarko, and E. Lomas, "The ethics of facial recognition technologies, surveillance, and accountability in an age of artificial intelligence: a comparative analysis of us, eu, and uk regulatory frameworks," *AI and Ethics*, vol. 2, no. 3, pp. 377–387, 2022.

[7] K. Hao. (2020) The two-year fight to stop amazon from selling face recognition to the police. [Online]. Available: https://www.technologyreview.com/2020/06/12/1003482/amazon-stopped-selling-police-face-recognition-fight/

[8] O. M. Parkhi, A. Vedaldi, A. Zisserman, and C. V. Jawahar, "Cats and dogs," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.