

Assignment 1B

CAB420, Machine Learning, Semester 1, 2021

This document sets out the two (2) questions you are to complete for CAB420 Assignment 1B. The assignment is worth 10% of the overall subject grade. All questions are weighted equally. Students are to work either individually, or in groups of two. Students should submit their answers in a single document (either a PDF or word document), and upload this to TurnItIn. If students work in a group of two, only one student should submit a copy of the report and both student names should be clearly written on the first page of the submission.

Further Instructions:

1. Data required for this assessment is available on blackboard alongside this document in *CAB420_Assessment_1B_Data.zip*. Please refer to individual questions regarding which data to use for which question.
2. Answers should be submitted via the TurnItIn submission system, linked to on Blackboard. In the event that TurnItIn is down, or you are unable to submit via TurnItIn, please email your responses to `cab420query@qut.edu.au`.
3. For each question, a short written response (approximately 2-5 pages depending on the nature of the question, approach taken, and number of figures included) is expected. This response 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.
4. MATLAB or 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.** Note that MATLAB Live Scripts, Python Notebooks, or similar materials will not on their own constitute a valid submission and a written response per question is expected as noted above.
5. 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. Training and Adapting Deep Networks. When training deep neural networks, the availability of data is a frequent challenge. Acquisition of additional data is often difficult, due to logistical and/or financial reasons. As such, methods including fine tuning and data augmentation are common practices to address the challenge of limited data.

You have been provided with two portions of data from the Street View House Numbers (SVHN) dataset. SVHN can be seen as a ‘real world’ MNIST, and although the target classes are the same, the data within SVHN is far more diverse. The two data portions are:

1. A training set, `Q1/q1_train.mat`, containing 1,000 samples total distributed across the 10 classes.
2. A testing set, `Q1/q1_test.mat`, 10,000 samples total distributed across the 10 classes.

These sets do no overlap, and have been extracted randomly from the original *SVHN* testing dataset. Note that the training set being significantly smaller than the test set is by design for this question, and is not an error.

Using these datasets you are to:

1. Train a model from scratch, using no data augmentation, on the provided abridged SVHN training set. *Batch size = 128, 64, 32, 16, 1*
Found to be best
2. Train a model from scratch, using the data augmentation of your choice, on the provided abridged SVHN training set.
3. Fine tune an existing model, trained on another dataset used in CAB420 (such as MNIST, KMINST or CIFAR), on the provided abridged SVHN training set. Data augmentation may also be used if you so choose.

All models should be evaluated on the provided SVHN test set, and their performance should be compared.

In addressing this question you should:

- Ensure that all choices (e.g. network design, type of augmentation) are explained and justified in your response.
- Consider computational constraints. You do not need to train the most complex model possible. It is acceptable to use a simpler architecture due to computational constraints, though efforts should still be made to achieve a good level of performance and details regarding these choices and the tradeoff between computational load and performance should be stated in the response.
- Your comparison should consider both raw performance and the different performance characteristics between methods. For example, do some methods work better on some classes than others, and why may this be? Are there cases that cause all methods to fail and if so, what are the characteristics of these?
- Include all relevant figures and/or tables to support the response.

Problem 2. Person Re-Identification. Person re-identification is the task of matching a detected person to a gallery of previously seen people, and determining their identity. In formulation, the problem is very similar to a typical biometrics task (where dimension reduction techniques such as PCA and/or LDA, or deep network methods using Siamese networks can be applied), however large changes in subject pose and their position relative to the camera, lighting, and occlusions make this a challenging task.

Person re-identification (and performance for other retrieval tasks) is typically 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 top N results. Ideally, the top result will always be the first (i.e. closest) returned match. 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 Q2/Q2.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.

In using these datasets, you should use the **Training** dataset to train your model, and determine any model hyper-parameters. You may wish to further divide the **Training** set into training and validation to do this. To evaluate your model using the **Testing** data, you should transform images in both the **Gallery** and **Probe** into your chosen representations, and then for each image in the **Probe** set, compare it to each image in the **Gallery** and determine the index of the correct match.

Your Task: Using this data, you are to:

1. Develop and evaluate a **non-deep learning** method 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. 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.

In completing your answer you may also wish to consider the following:

- You may wish to resize images to reduce computational burden. This is acceptable, but should be documented. Be mindful not to make images too small which will result in the loss of discriminative information and make identification very challenging.

- You may wish to fine-tune a pre-trained network. This is acceptable, but should be documented.
- A high level of accuracy alone will not guarantee a high mark for the question. The approach you choose and the rationale for it, and the quality of your evaluation and discussion are far more important.

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.