Module Code
COM00024I

BSc, BEng and MEng Degree Examinations 2022–23

DEPARTMENT OF COMPUTER SCIENCE

Intelligent Systems 2

Open Group Assessment

Issued: 19 April 2023, 12:00 (noon)
Submission due: 10 May 2023, 12:00 (noon)
Feedback and marks due: 14 June 2023, 12:00 (noon)

All students should submit their answers through the electronic submission system: http://www.cs.york.ac.uk/student/assessment/submit/ by 10 May 2023, 12:00 (noon). An assessment that has been submitted after this deadline will be marked initially as if it had been handed in on time, but the Board of Examiners will normally apply a lateness penalty.

Your attention is drawn to the section about Academic Misconduct in your Departmental Handbook: https://www.cs.york.ac.uk/student/handbook/.

Any queries on this assessment should be addressed by email to cs-exams@york.ac.uk for forwarding to Dr Patrik Huber, or during the weekly group practical sessions in the assessment period. Queries will be discussed in these sessions, and answers that apply to all students will be posted on the VLE.

Rubric:

Answers must not exceed 2 A4 pages in total, using the IEEE Journal Paper Template with its default font and margins. This limit includes any diagrams, tables, and so on. Up to one separate page for only references (bibliography) is allowed. Excess pages will not be marked. Your report must address all sections outlined in the 'Report' section to gain full marks, and all the team members need to have completed the *self & peer assessment form* online.

The name of all your group members should be on the front cover of your assessment. Do not include exam numbers as these are private to individuals.

1 Scenario & Setting

The scenario for this assessment is the classification of images into specific classes. In machine learning, this is called *image classification*. The aim is to train a machine learning model from an image dataset, that can then, given a 2D input image, classify the content of that image into a set of pre-defined categories. The task might be to classify generic objects into broad categories, for example categories like 'airplane', 'cat' or 'horse', or more specific tasks like classifying digits.

In this assessment, you will construct and train a neural network classifier for the task of flower species categorisation, using the *flowers-102* dataset, collected and prepared by researchers from the University of Oxford. It is one of the well-known datasets for flower classification and contains a predefined set of training and test images.

This assessment requires you to construct a neural network classifier, that will take an image as its input, and predict a specific class as its output. Your task is to construct a suitable (deep) neural network architecture, train it on the flowers-102 training set, and then evaluate its classification accuracy on the official flowers-102 test set. You will then prepare a report in the style of a short academic research paper, presenting your method and results.

1.1 Group Assignment

The assessment is to be done in groups. You can find your group assignment on the CS Teaching Portal. Each group has a unique group number.

1.2 Software Framework

This assignment is to be done in Python 3, and you have to use one of the two most popular machine learning libraries/frameworks: PyTorch[1] or TensorFlow[2]. You are free to choose either PyTorch or TensorFlow, but your group has to agree on one framework. You have to use their most recent version: For PyTorch, that is either PyTorch 1.13.1 or 2.0 (or newer), and for TensorFlow, that is TensorFlow 2.12 or newer.

1.3 Dataset

The dataset we'll use is the Oxford 102 Category Flower Dataset (*flowers-102*). The details and data download are given on its website: https://www.robots.ox.ac.uk/~vgg/data/flowers/102/. However, instead of downloading the data manually we strongly recommend to use the dataset

[1] https://pytorch.org/
[2] https://www.tensorflow.org/

loaders for in PyTorch
(https://pytorch.org/vision/stable/generated/torchvision.datasets.Flowers102.html) or TensorFlow
(https://www.tensorflow.org/datasets/catalog/oxford_flowers102) to directly load the datasets into your environment, including the labels and train/validation/test splits.

You are not allowed to use images from any other dataset to train your network, and you have to use the official split into train, validation and test set used by original flowers-102 dataset for your experiments. The training set is to be used to train the model, the validation set is to be used to evaluate the model performance during the training and for hyper-parameter tuning, and the test set is to be used only for the final evaluation.

You may use data augmentation techniques, as long as the images are augmented from the flowers-102 training set, and no other external images or knowledge is used.

1.4 Neural Network Architecture

You will have to construct a neural network classifier, that will take an image as its input, and predict a specific class as its output. You will have to use appropriate layers and loss function, and train your network using an optimiser on the dataset.

You may take inspiration from existing network architectures, but you need to design/write the architecture yourself, in your own code, using PyTorch or TensorFlow. Using existing models or existing model code from online sources (e.g. GitHub) is not allowed. If you take inspiration from existing code, you need to make it sufficiently clear that it is your own code, and not copy and pasted from the online source.

You need to train your own network from scratch, i.e. using a pretrained network or pretrained weights is not allowed.

1.5 Hardware Resources

To ensure fairness and equal opportunity for all, and to not make the assessment about who has the most computing power available, your network **must train within 12 hours**, on one GPU and/or CPU (it can be any model). If no person in your group has a suitable computer, you can request access to a departmental compute server via itsupport@york.ac.uk, or alternatively to a lab GPU via the Hardware Team. But note that it may likely take a few days from the day of request until access is granted, so plan for that.

1.6 Performance Evaluation

After any experimentation and testing, your report must present a final network for the classification task that can be used by others to classify new data. You must use the official flowers-102 test set and report the accuracy on the test set in your report.

2 Submission

Your final project mark will be made up of three components: A mark for a report (weighted at 40%), the reported classification performance of your network (weighted at 30%), and a mark for self/peer assessment (weighted at 30%). Each of the three components is described in the next three subsections.

2.1 Report [40% (total 40 marks)]

The report will present an overview to your method and results. You must evaluate thoroughly the performance of the network (in terms of accuracy). You must provide justification for the choices you make.

You are required to use the two-column *IEEE Journal Paper Template*. It is recommended that you use Overleaf[3], an online collaborative LaTeX editor, where you can work jointly on the report in a shared document. The IEEE template is available directly in Overleaf via this link: https://www.overleaf.com/latex/templates/ieee-journal-paper-template/jbbbdkztwxrd, by clicking "Open as Template" on that website. An introductory video to LaTeX and Overleaf in the context of this project is provided on VLE/Panopto.

If you are creating a new Overleaf account, log in with or link your University of York email address, so that you can benefit from the University's subscription, which allows to share documents with an unlimited number of participants.

Alternatively, you may use the equivalent official IEEE Word template, which can be obtained using the steps given on the VLE under the left-hand menu entry "Assessment".

Your report must be **no more than two A4 pages in length**, including any diagrams or tables. A third page for only references (bibliography) is allowed. At the top, it must indicate your group number, followed by your names: In the IEEE LaTeX template, find the `\author{...}` command, and put "Group X: Name 1, Name 2, etc.", where "X" is your group number. Your report must contain the following sections, in the order given below, and material outside the six prescribed sections will not be marked.

[3] https://www.overleaf.com

1. [4 marks] Abstract
   This section should give a brief overview of the highlights of the report: What the task is, what you've done, how, and what you've achieved. Each of these should be around one sentence.
   The last sentence of the abstract should clearly state your achieved classification accuracy on the flowers-102 test set.

2. [4 marks] Introduction
   This section should give a brief introduction to what the task is (in your own words), with a few references to the wider field of image classification, how other people have been tackling the task historically and currently, why it is an important machine learning problem, and then lead into how you are tackling the problem.

3. [8 marks] Method
   This section should introduce your approach: The network you've chosen to create, its layers, and the loss function, and how this works to train a model that learns to classify images. You should clearly describe and justify the type of architecture and/or layers that you've used. You can go into the full details of your approach here - for example, you could include a formula of the loss function.

4. [4 marks] A diagram of your network architecture
   Include a figure of the network that you've created to achieve your final result. The figure should exactly depict your network in terms of the layers used and their dimensions. Both TensorFlow and PyTorch include functions to visualise the graph of a network, but you may likely need to post-process that output to make a nice diagram out of it. Another option would be to draw the figure with a tool such as PowerPoint.

5. [8 marks] Results & Evaluation
   This section should describe your experiments, the evaluation metrics that you've used, your results, and draw initial conclusions from your results. If you've run multiple experiments, you could include a table with the results of the various experiments you ran. Your experiments should be described in a way such that they are fully reproducible, for example, describe (and justify) your chosen optimisation algorithm, and all the parameters and hyperparameters used for training. You should include here your classification test accuracy, and the setup of your evaluation to compute that. You also have to include the hardware and environment used for training, and how long it took to train your network.

6. [4 marks] Conclusion & Further work
   This section should draw conclusions from the whole project. Did you achieve a good performance? Is your architecture a good choice? What went well, what did not go very well? This section should also identify at least one item for further work.

Additional marks will be given for:

- [4 marks] Appropriate and correct use of the IEEE double-column template.

- [4 marks] References to appropriate literature. Your text should show that you have understood the literature you have selected.

The 4 marks for each section (or 8 marks for Method and Results & Evaluation) are awarded using stepped marking[4] according to the following scheme:

- 0 marks: Not present or not meeting the minimum standard
- 1.5 marks (3 marks for Method/Results): *Low*: Meeting the minimum standard
- 2.5 marks (5 marks for Method/Results): *Mid*: Meets most of the standards well
- 4 marks (8 marks for Method/Results): *High*: Fully meets, and, in some aspects, exceeds the standards

You are free with regards to the length of the individual sections of the report (as long as the total is below two A4 pages, without references). On the VLE in the "Assessment" section, there is an example report provided with dummy text, to help you gauge the format, structure and length of individual sections.

You need to be able to clearly explain each component of the network that you've built, and justify the overall architecture. If it becomes apparent that you do not understand the architecture that you've used, it will result in zero marks for the respective section of the report as well as for the classification performance (section 2.2).

The final report must be submitted as a single PDF file.

2.2 Classification performance [30% (total 30 marks)]

The report submission has to clearly state the achieved classification accuracy on the official flowers-102 test set in the abstract and in the *Results & Evaluation* section. Marks are awarded according to the achieved test accuracy of the trained model. It has to be clear from reading your report that you have trained your network only on the flowers-102 training set, and that the test accuracy has been computed only from images in the test set - otherwise you will not be awarded any marks in this section.

Your code needs to be able to reproduce your reported accuracy within a margin of 3% when run/re-run (as there is some randomness involved when training most neural networks). The assessors may ask students to demonstrate their training and accuracy live to the assessor, to verify its performance.

[4] University of York Guide to Assessment (2020/21), Appendix R

2.3 Self- & peer assessment [30% (total 30 marks)]

You will be asked to fill out an online form to rate both yourself and each of your team members. You **must** complete this form. If you do not complete this form, you will be awarded a mark of zero for this part of the assessment.

The mark awarded to each student in this part is the average of the marks that they received from their peers.

The assessors may ask students to provide supporting evidence for their contributions as well as for the marks they awarded to their peers. As such, team members should only award peer-assessment marks that they are comfortable to justify if they are asked to do so.

Where there is a significant difference between a self-assessment score and scores given to that person by their team members (in both directions), the assessors may talk to the group and ask for supporting evidence. This evidence will be used by the assessors to determine an appropriate peer assessment score.

If there is sufficient evidence that a team member has not engaged with the assessment, the assessors may award a mark of zero to that individual for the entire assessment.

Note that individuals or teams are strongly encouraged to contact one of the module staff, if and as soon as a problem within the group becomes apparent.

2.4 Code repository & submission

You **must** submit your code and model as part of your submission. This will not be marked, but will be used to validate the reported classification performance. You are expected to use git and either GitHub[5] or GitLab[6] (or a similar platform) throughout your project. All team members are expected to work on that joint repository and contribute via *commits*. In case of unequal peer evaluation, the markers may ask the team for access to the repository and its full commit history, and the team has to be able to provide such.

If multiple team members work together on the same code at a single computer, or multiple team members have contributed to a single commit, then one team member should commit the change, and the names of the team members that contributed to that particular commit must be clearly listed in the commit message.

If a team member makes a significant contribution to the team outside of the code base, then the team is expected to either note this in a related commit message, or keep a separate note of such contributions (for example in a shared Google Docs document). If the team is asked for

[5] https://github.com
[6] https://gitlab.com

access to their code, an URL to such a log should then be submitted to the assessor together with the URL / access to the code.

End of examination paper