INFT2060: Applied AI

Assessment Item 2

AI Project

Specification

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Overview

Assessment Type	Written Report
Weighting	30%
Involvement	Group
Due Date	Week 12 Friday 2025 (23:59)
Deliverables	- Report in pdf format - Completed cover sheet

Description

Student groups research a specific AI model and document their findings in form of a written report. The report is 20-25 pages long (excluding reference list and appendices) and includes:

- Title Page
- Table of Contents
- Executive Summary
- Introduction
- Background and Description
- Applications and Impact
- Experimental Evaluation
- Advantages and Limitations
- Future Directions and Conclusion
- List of references *
- Appendix A - Python Code *
- Appendix B - Supplemental Material *

Note: Sections marked * are excluded from page count

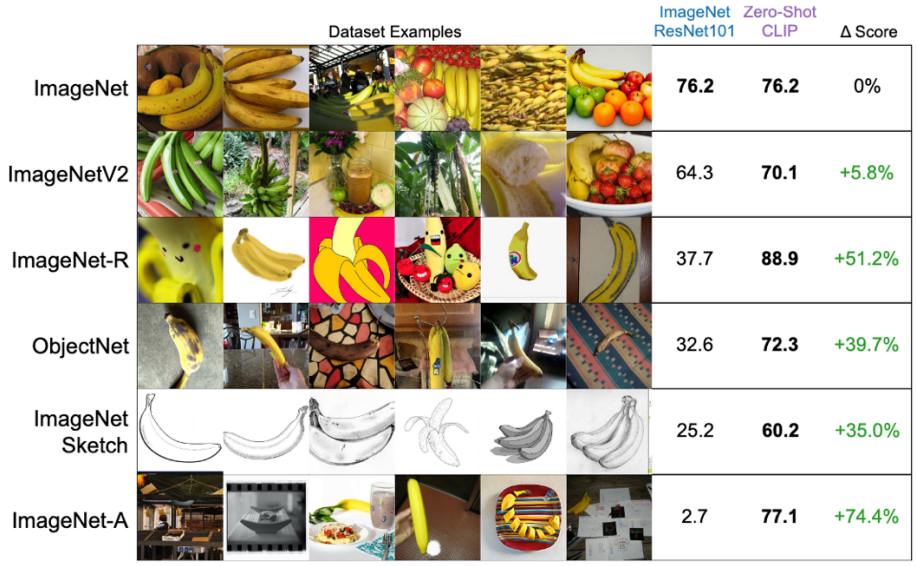
See the next pages for the *topic* and *marking criteria*. The marking criteria can also be used as a guide for the content and length of each section.

Submission

- The report and any attachments should be compiled to pdf format and submitted by one member of the group via Canvas, which also includes a Turnitin screen.
- Please attach a completed cover sheet on which all contributing group members have signed off the agreed group report.
- Optionally, a document that states the contributions of each group member can also be attached. In cases where member contributions within the group are very different and/or a conflict arises, the marks may be weighted according to this document.

Topic

Your group's topic for this report is the **CLIP (Contrastive Language–Image Pretraining)**, a neural network model trained on a variety of (image, text) pairs [1]. It learns visual concepts from natural language supervision. CLIP can understand images in the context of text and perform zero-shot classification, retrieval, and more.



The performance of the best zero-shot CLIP model, ViT-L/14@336px, VS the performance on the ImageNet validation set, ResNet-101. [1]

As part of *Applications and Impact*, you will identify several industries on which this model may have an impact. Your *Experimental Evaluation* task is to *experimentally evaluate the feasibility of using CLIP in at least two of the identified industries*. For each industry, you will need to

- Define a hypothetical scenario of how *CLIP* may be used in this industry
- Describe a methodology and performance metric(s) that experimentally determine how well the model performs in the hypothetical scenario. For example, you may decide to
  - Collect a small dataset of images and associated text prompts that may be used in this industry for various vision-language tasks.
  - Define tasks such as zero-shot classification, image-text retrieval, or caption matching
  - Choose metric(s) to evaluate how well the model does at this particular task
- Get access to a *CLIP* model for running inference
  - You will find a variety of *CLIP* models like *CLIPSeg*, *CLVP*, and *etc.* from https://huggingface.co/mlunar/clip-variants on Github or HuggingFace
  - To improve performance, you should also look into finetuning CLIP
- Run inference on your data with the *CLIP* model and calculate metric(s) using your labels
- Present the experimental results according to your methodology and performance metric(s)

Marking criteria

All group members listed on the cover sheet will normally get the same mark for the report, unless a contribution statement has been submitted and it is clear that equal marks would be unjust.

Criterion	Marks	Description
Executive summary	1	An executive summary provides a concise summary of the most essential information in the report.
Introduction	2	A clear and concise introduction to the given model.
Background and Description	3	Thorough explanation of broader context of the given model, including an overview of alternative models. Description of the components of the given model and how they work on a high level.
Applications and Impact	3	Exploration of practical applications in different domains. Building on this, a thoughtful analysis of the potential impact of the model on at least two industries and society as a whole.
Experimental Evaluation	12	According to the given task, a description of the hypothetical scenarios with associated evaluation methodology and performance metric(s). Accurate presentation of experimental results and performance analysis including tables, figures, and examples where appropriate.
Advantages and Limitations	2	Building on findings from the previous section, a well-rounded discussion of the strengths and limitations of the model. This also includes ethical considerations.
Future Directions and Conclusions	2	Identification of emerging trends, challenges, and future directions relating to the model. Coherent summarisation of key findings and well-supported concluding remarks.
Report Format	1	This report should be clearly structured and well organised. It should be designed in an appealing way (e.g., by using an eye catcher on the title page, including page numbers and by highlighting essential information as well as using footnotes and appendices to add supplementary information). The report length should be 20-25 pages. Be aware that any work after the maximum page limit may not be included in marking.
Appendix A (Python Code)	2	The Python code used to produce the findings reported in this report is to be included in Appendix A. Marks may be deducted if findings cannot be reproduced from code, or if code is poorly written, structured, or documented.
Writing and referencing	2	This criterion includes correctness of grammar and spelling, and appropriate use of references, direct quotes, paraphrasing etc. Students should note that there is a fine line between poor referencing and plagiarism, and reports that appear to be plagiarised will be referred to the Student Academic Conduct Officer, with possible outcomes such as a mark of zero. Students are strongly advised to repeat the University's Academic Integrity Module, and to be sure never to take text or ideas from anywhere without clearly noting the source.

References

[1] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021, July). *Learning transferable visual models from natural language supervision.* In International conference on machine learning (pp. 8748-8763). PmLR.

[2] Openai. (n.d.). *GitHub - openai/CLIP: CLIP (Contrastive Language-Image Pretraining), Predict the most relevant text snippet given an image*. GitHub. https://github.com/openai/CLIP