DDA4220/MDS6224/MBI6011 Deep Learning

Final project

Due Date: 23:59, 15 May, 2023

This is a group project for 2 to 3 people. The final project will worth 40% of the final grade.

---

1 Project Topics

1.1 Object Detection Project

- **Background**
  Object detection is a computer vision task that involves detecting objects of interest within an image or video and localizing them by drawing bounding boxes around them. The development of object detection algorithms has been driven by the need for machines to perceive and understand the world in the same way as humans. Object detection has a wide range of applications, including autonomous vehicles, surveillance systems, robotics, and image retrieval.

  Object detection is an active area of research, and there are many challenges to overcome, such as detecting objects at different scales and orientations, dealing with occlusion and clutter, and detecting objects in complex scenes. However, with the continued advancement of deep learning techniques, it is expected that object detection algorithms will continue to improve and find more diverse and practical applications.
- **Task**
  You are required to finish the VOC2012 object detection competition. Specifically, there are 20 categories including: `Aeroplanes`, `Bicycles`, `Birds`, `Boats`, `Bottles`, `Buses`, `Cars`, `Cats`, `Chairs`, `Cows`, `Dining tables`, `Dogs`, `Horses`, `Motorbikes`, `People`, `Potted plants`, `Sheep`, `Sofas`, `Trains`, `TV/Monitors`. We provide the object detection data in `VOC2012_det.zip`. There are 5823 validation samples and 5717 training samples, please use the validation set as the test set (you can use a subset of the validation set). You can also access the full dataset containing other tasks on the competition home page: http://host.robots.ox.ac.uk/pascal/VOC/voc2012/. Please note that `/VOCdevkit/VOC2012/ImageSets/Main` contains the list for object detection task.

  Google Drive link for dataset:
  https://drive.google.com/file/d/1aLnY1PYFFBmR1NUEBdULXFxVx7iJQkzJ/view?usp=sharing
- **Suggestions**
  Since then, many other algorithms have been developed, such as Fast R-CNN, Faster R-CNN,

and YOLO, which have significantly improved the accuracy and speed of object detection. Now there are many models based on Transformer. Pascal VOC is a classic object recognition competition, you can find many materials online. Using some SOTA methods are OK, but you need to provide the reference, present your own understanding and do your own attempts (e.g. change some settings or do ablation experiments) to investigate the method you use.

1.2 Semantic Segmentation Project

- **Background**

  Semantic Segmentation is an important task in the field of computer vision, which aims to assign each pixel in an image to its corresponding semantic class. Unlike object detection, which only needs to determine the existence and position of objects in an image, semantic segmentation requires pixel-level classification of the image.

  Semantic segmentation has applications in various fields. For example, in autonomous driving, semantic segmentation can separate different objects such as roads, vehicles, and pedestrians, which helps the autonomous driving system make more accurate decisions. In medical image analysis, semantic segmentation can separate different tissue structures, which helps doctors make more accurate diagnoses.

- **Task**

  In this project, you need to use deep models to perform semantic segmentation on humans. Dataset is in `Pascal_seg.zip`. In specific, you should segment 7 semantic parts including `head, torso, upper-arms, lower-arms, upper-legs, lower-legs, and background`. The dataset is provided, which contains the training and validation sets.

  Google Drive link for dataset:
  https://drive.google.com/file/d/1dvQ6d7Bsf_esVEp74VIqTANIuYCln-C6/view?usp=sharing.

1.3 Machine Translation Project

- **Background**

  Machine translation is the use of computer algorithms to automatically translate text or speech from one language to another. It is a subfield of natural language processing (NLP) and has been an active area of research for several decades.

  Recently, with the development of deep learning and neural networks, neural machine translation (NMT) has emerged as a powerful approach to machine translation. NMT uses deep neural networks to learn a mapping from the source language to the target language. It has been shown to produce better translations than traditional statistical methods.

  Machine translation has many applications, including language learning, cross-cultural communication, and international business. However, machine translation is not perfect and can still produce errors and mistranslations, particularly for complex sentences and idiomatic expressions. Nonetheless, it continues to improve and has become an increasingly important tool for global communication.

- **Task**

  In this project, we will use neural networks to do **Chinese** to **English** translation. For example, if the user gives **'祝同学们最后的大作业顺利!'**, we can get something like **'Good luck with your final project!'**.

  The dataset is provided (`translation2019zh.zip`) in the format of two JSON files, one is for training *train.json*, and the other is for validation *valid.json*. You can distinguish them through the file names.

  Google Drive link for dataset:
  https://drive.google.com/file/d/1gGnpS0fQBNcTAwhpYNGjxJJwsHYQzKag/view?usp=sharing
- **Suggestions**

  Tokenizers from Spacy package are recommended, including Chinese version and English version.

  You can try some basic models, like RNN, transformer, as your baseline.

  You may know that ChatGPT and GPT4 are so hot nowadays. You can also try to use some open-source pre-trained large language model(LLM), like PaLM or LLaMA if you have enough GPUs.

2 Requirements

1. You should combine the content in the classroom and the content of the Internet, and use **Pytorch** to complete the experiment.

2. We do not provide any baselines, you should finish the investigations and report on your own. You are required to give a baseline and make enough attempts to improve the results. You can try any model architectures and any tricks, pretrained models are also allowed in this project. Creativity and novelty are encouraged in this project. It will be better if you can change something on your baseline based on your understanding of this topic. Off-the-shelf methods are OK, but you have to gives analysis to clarify these are useful to your result.

3. We do not grade your project based on performance, what we care about is your understanding of this project. You should really think about this project and try to get better results through experiments. Please refer to the detailed evaluation metrics on course site http://www.zhangruimao.site/DDA4220.html

4. When you investigate this topic, please pay attention to the metrics others are using. You should choose reasonable metrics to evaluate your performance. It will be better if you show both quantitative results and qualitative results in your experiments part.

5. If you buy Colab service for this project, we can cover the cost of the service for $10 **per team**. After the project finishes, you can provide your screenshot of buying Colab service, and we will pay this cost (up to $10).

6. Your need to write a report with this template: https://www.overleaf.com/read/dzrsxbcdwfpg, it contains Introduction, Related work, Method, Experiment, Conclusion, and Reference.

3 Submission requirements

1. You need to submit code, report and models to GitHubClassroom before May 15, 23:59. The link: https://classroom.github.com/a/7d6Hf_Nx

2. We will find a time slot for people to present their projects to the TAs and lecturer one by one. Slides are not necessary, but you have to explain your ideas, analyze your results, etc, we may asks you some questions. Plagiarism will not be tolerated; ChatGPT is not allowed.