MiniProject 4: Reproducibility in ML

COMP 551, Fall 2021, McGill University

1 General information

- Lead TAs: Wesley Chung and Harley Wiltzer

- This mini-project is **due on Monday December 6, 2021 at 11:59pm (Montreal Time).** Submissions will be accepted without penalty until Friday December 10, 2021 at 11:59pm. Following this date, the usual late penalties will apply.

- This mini-project is to be completed in groups of three. All members of a group will receive the same grade except when a group member is not responding or contributing to the project. If this is the case and there are major conflicts, please reach out to the group TA for help and flag this in the submitted report. Please note that it is not expected that all team members will contribute equally. However every team member should make integral contributions to the project, be aware of the content of the submission and learn the full solution submitted. You should be in the same team with the same classmate more than two times.

- You will submit your assignment on MyCourses as a group. You must register your group on MyCourses and any group member can submit. See MyCourses for details.

- We recommend to use **Overleaf** for writing your report. You are encouraged to use existing open sourced python implementation.

- You should use Python for this mini-project. You are free to use any libraries as well as open sourced repositories or implement your own for anything you can't find an existing implementation.

- The TAs (Wesley and Harley) will hold an office hour on **Monday, November 30 at 1:00pm** during which can validate your project ideas (not mandatory, but recommended). As usual, you can always reach out to the TAs via Slack and email for any clarifications.

2 Background

One goal of publishing scientific work is to enable future readers build upon it. Reproducibility is the central theme to achieve this target, yet it is unfortunately one of the biggest challenges of Machine Learning research. Everyone is encouraged to follow the reproducibility checklist while publishing scientific research, to make the results reliable and reproducible. In addition, a challenge is organized every year to measure the progress of our reproducibility effort. The participants select a published paper from one of the listed conferences, and attempt to reproduce its central claims. The objective is to assess if the conclusions reached in the original paper are reproducible. The focus of this challenge is to follow the process described in the paper and attempt to reach the same conclusions. We have designed this miniproject in the spirit of the reproducibility challenge. Top projects can potentially be extended and submitted to the challenge in January. Note that in comparison to previous mini-projects, this is an open ended project meant to help you use the theoretical and applied knowledge from this course to experiment and tinker with actual, popular research work in the field.

3   Task

The goal of this assignment is to select a paper and investigate the main claims, to see if similar conclusions can be reached or the conclusions can be challenged. For this mini project, you are not expected to implement anything from scratch. You are encouraged to use any code repository published with the paper or any other implementation you might have found online.

It is up to you to define the experiments you would like to perform. The reasoning behind your choice of experiments is part of the evaluation criteria. Here are some possibilities:

1. Rerunning the models on reported datasets to see if you can reproduce the evaluation metrics reported in the paper, or alternative evaluation metrics.

2. Improving the baselines reported in the paper.

3. Applying the model to other datasets.

4. Investigating the effect of different choices in the proposed methodology/architecture (ablation study).

5. Improving the performance of the proposed method.

6. Comparison to a method that was not considered in the original paper.

3.1   Paper selection guidelines

You must choose a paper from the current pool of papers of the reproducibility challenge. If you have a good reason to work on another paper, check with the lead TAs for this project first.

Data   You should be able to access the data or environment you will need to reproduce the paper's experiments.

Code and Trained Model   In many cases a code might be available directly from the authors or another source. You should check whether you can work with the code before picking the paper. Similarly, there may be trained models available for use.

Computation   You should estimate the computational requirements for reproducing the paper and take into account the resources available to you for the project. Some authors might have had access to infrastructure unavailable to you; you might not want to choose such a paper. Alternatively, you can study the claims of the paper on smaller scales, due to limited time and computation.

You will undoubtedly find some papers that produce incredible demonstrations of deep learning feats. While it may be tempting to try to replicate tasks like image synthesis and text generation, note that these deep learning models tend to be quite large and consequently the experiments may demand extreme computational resources (and time!). Note that, this should not necessarily prevent you from choosing a paper as you can reduce the computational costs in many ways: running experiments on smaller datasets, reducing the size of the models or considering subsets of the experiments of the original paper. You can also attempt to design *new* experiments to investigate the same claims.

Several models also have **pretrained weights** available to download. Since these have been trained on huge datasets, you are encouraged to code up the models and directly import these weights instead of training from scratch. You can then use the pretrained model for experimentation as well as fine-tune the weights on new data. But make sure to add all the resources you have used in your references.

There are various ways to construct a very interesting project without requiring a massive amount of compute. For example:

- Averaging Weights Lead to Wider Optima and Better Generalization: This paper introduces a simple algorithm with some impressive results, which don't require enormous models to demonstrate. Some example experiments: Measuring flatness of minima, checking model performance, using different neural network architectures, etc.

- diffGrad: An Optimization Method for Convolutional Neural Networks: This paper introduces a novel optimization algorithm that claims to address shortcomings of existing optimization algorithms on training CNNs. Some example experiments include measuring the efficacy of the algorithm by different metrics (convergence rate, sensitivity to hyperparameters, variance, etc) or evaluating the algorithm on different datasets.

Note: You cannot choose these examples for your project. If you are unsure whether a paper would be suitable, you can contact the lead TAs for advice.

4   Deliverables

You must submit two separate files to MyCourses (using the exact filenames and file types outlined below):

1. code.zip: A collection of supporting code files. Please submit a README detailing the packages you used and providing instructions to replicate your results.

2. writeup.pdf: Your project write-up as a pdf (details below).

3. paper.pdf: Include a copy of the chosen paper (including the appendix).

4.1   Report guidelines

Write a report of no more than 3 pages (excluding reference) covering the below points. You can get inspiration from this latex template for your report, although not all the sections are necessary for this mini-project. You are allowed to have an additional appendix, but the main findings of the paper should be documented in the main body of the report (3 pages). Here is a suggestion for the content of your report. It is up to you to structure your report appropriately:

- Abstract and introduction defining the problem statement, experiments conducted and summarizing the results of your experiments.
- A quick summary of the paper under investigation.
- Specify which claims are being investigated from the main paper and the experiments you ran to assess them.
- Document the results of your experiments.
- From your experimental results, did you reach the same conclusion as the authors?
- Any necessary details for reproducing the results, but were not specified in the original paper.
- Challenges that you have faced and how did you solve them.
- Summarize the key takeaways from the project and possibly directions for future investigation.
- Optionally, state the breakdown of the workload across the team members.

5   Evaluation criteria

Your work will be graded based on the following criteria:

Understanding the paper (2.5 pts) does the report demonstrate understanding of the main claims and results of the paper?

Choice of experiments (2.5 pts) how did you choose the experiments you ran? Are the experiments motivated and interesting?

Results (2.5 pts) breadth and depth of the experiments and the conclusiveness of results. Did your experiments reveal something interesting about the ideas proposed in the paper? Did they confirm or refute some conclusions?

Report (2.5 pts) Quality of the report: does your report clearly describe the task you are working on (i.e., the paper you are reproducing),the experimental set-up, results, figures (e.g., don't forget axis labels and captions on figures, don't forget to explain figures in the text). Is your report well-organized and coherent? Is your report clear and free of grammatical errors and typos? Does your report include an adequate discussion of related work and citations?