Monash DeepNeuron | *CLIP Training*

Project Name: CLIP Training (HPC-DL Collab)	**Difficulty:** Easy **Intermediate** Hard

Project Lead: TBA
Asst. Project Lead: Prasanth Devireddy (pdev0010@student.monash.edu)

Overview of the project: This project is an exciting collaboration between the Deep Learning and High-Performance Computing branches. Parallel computing techniques will be utilised to train a deep learning model from scratch.

A CLIP model which efficiently learns visual concepts from natural language supervision and is used for connecting images with text will be architected and trained from scratch, as opposed to fine tuning a pre-trained model. Continuing the previous work of the team, which used a pre-trained NLP, the CLIP model will be trained using an NLP trained from scratch over the summer semester.

Core Deliverables: (HPC-specific goals)

1. Research and explore frameworks that allow parallel training of DL models on multiple CPU cores (for convenient local training)
2. Successful pre-trained parallel CNN execution on M3 MASSIVE across multiple nodes.
3. Successful parallel execution of CLIP model on multiple nodes.
4. Parallel CNN Training on M3 nodes.

Stretch Goals:

5. Parallel CLIP Training on multiple M3 nodes.
6. Positive relationship between performance and no. of nodes (parallel training scales well and doesn't result in worse performance as system overhead increases).
7. Ground-up re-implementation of framework used and comparison with existing parallel training frameworks (let's reinvent the wheel and maybe make it better)

External Collaboration: This is a joint HPC-DL collaboration project where half the team is from either branch. We plan on rotating between DL/HPC project leads every project season.

Skills Required: Basic knowledge of ML & DL, knowledge of M3, SLURM and distributed computing.	**Where this project could lead project members:** For HPC members, this gives you a chance to gain demonstrable skills in deep learning while you also get to apply HPC techniques and learn more about parallel & distributed computing.