

Efficient training of Neural Networks

Adrien Aguilera-Multner

Enseirb-Matmeca & Inria

March 20, 2024

Who am I ?




Artificial
Intelligence

Optimisation

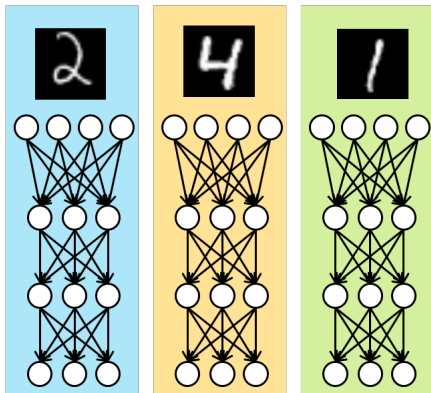
High
Performance
Computing

Internship topic(s)

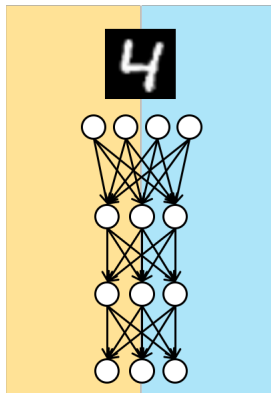
- Efficient training of neural networks
- Pipeline parallelism
- Nanotron 
HUGGING FACE
- Activation checkpointing (Rockmate)

Data Parallel vs Model Parallel

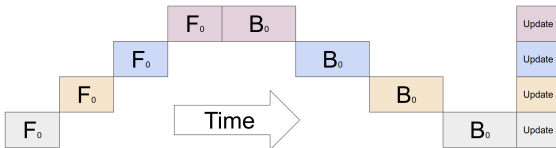
Data Parallel



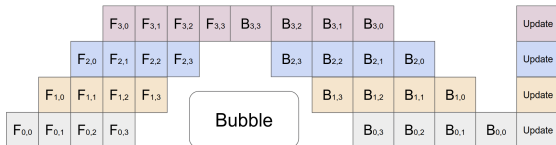
Model Parallel



Pipeline Parallel



Without pipelining



With pipelining

Torch distributed & Torchrun

Demo

Thank you !