Comparative Profiling: Insights Into Latent Diffusion Model Training

Bradley Aldous

Queen Mary University of London
b.j.aldous@qmul.ac.uk

Ahmed M. Abdelmoniem

Queen Mary University of London
ahmed.sayed@qmul.ac.uk

AI + MUSIC

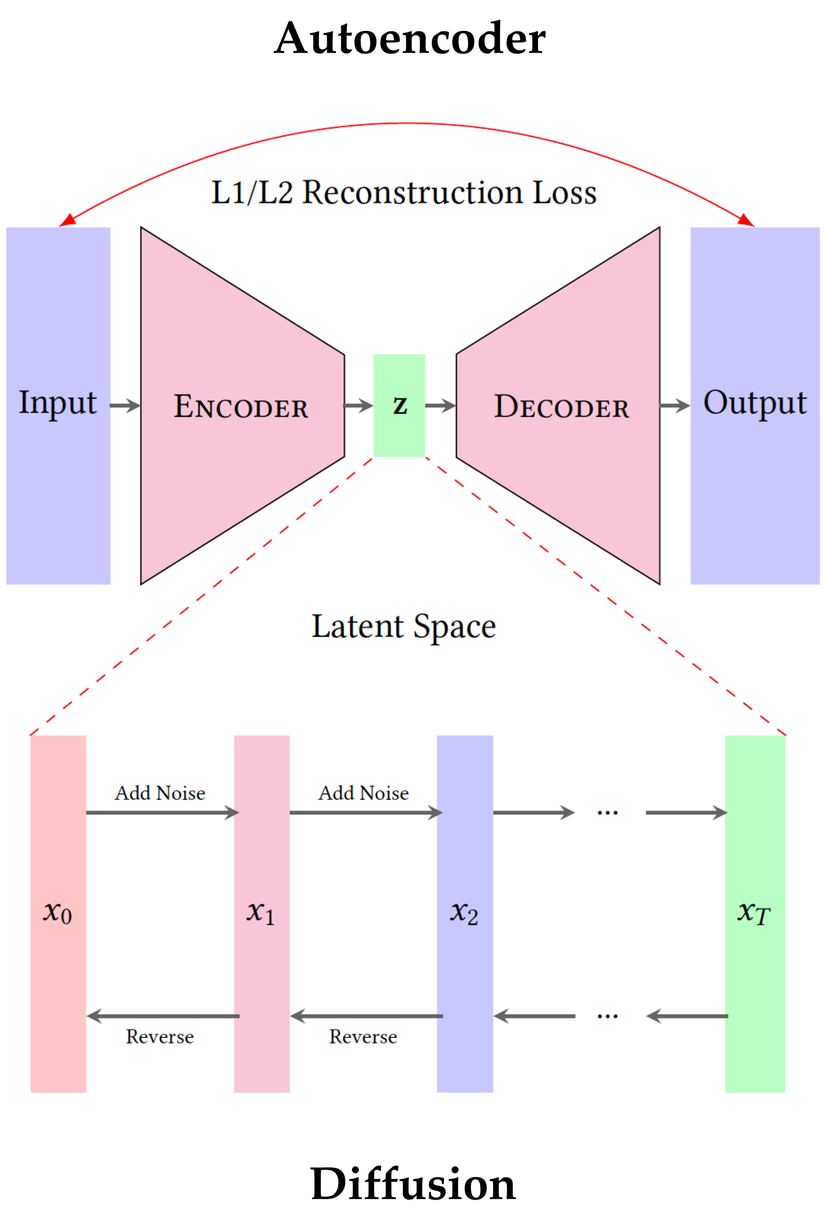
Queen Mary University of London

Background

- **Diffusion** models can generate realistic output (audio, image etc), by adding noise to training data and learning to reverse the process
- **Latent Diffusion Models** (LDMs) conduct the diffusion process in the **latent space** of an autoencoder

Motivation

- Require significant GPU resources to train on
- **AudioLDM** and **Stable Diffusion** (~700-800M params)
- Existing profiling studies usually use standard image benchmarks: no profiling studies focus on audio



Approach

Simple profiling approach:

- **Weights & Biases** used for monitoring GPU usage
- **PyTorch Profiler** used to profile operations

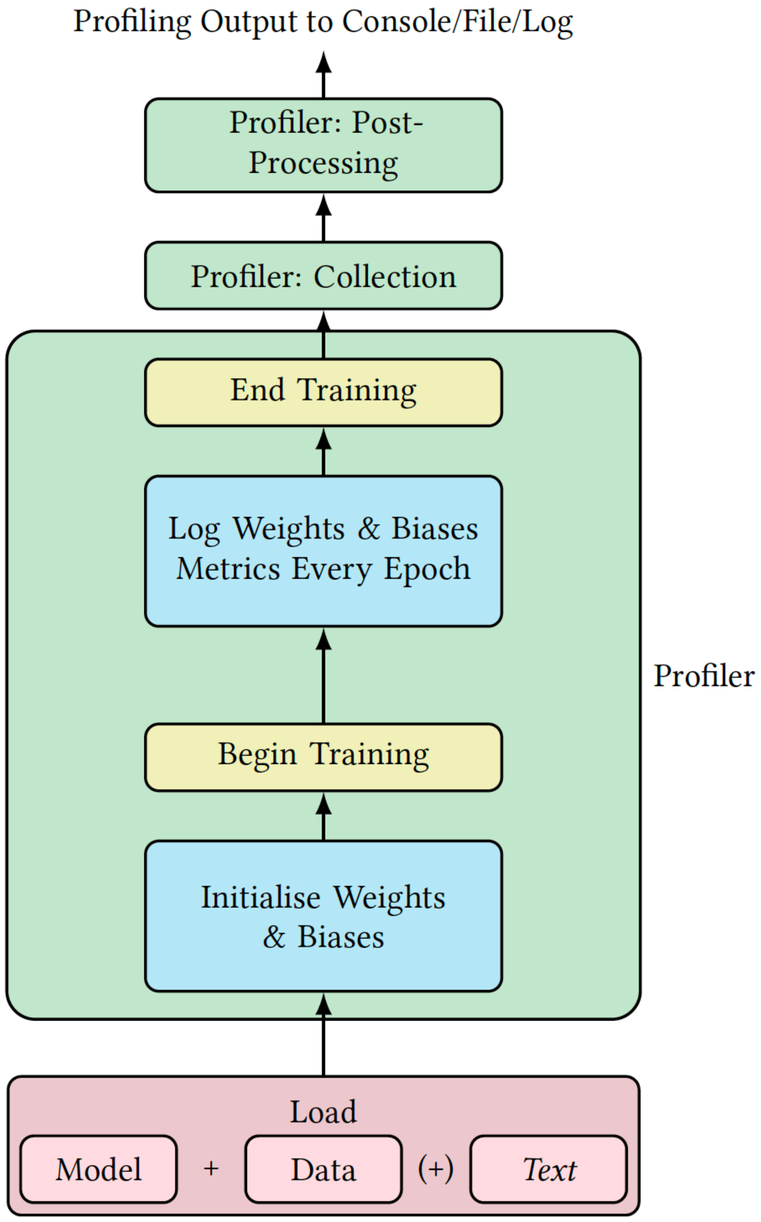
Metrics: Memory allocation, GPU power usage, GPU utilisation and profiling output all used in analysis

Further experiment:

- AudioLDM trained on single and dual GPU setups for the same number of epochs to observe acceleration gains from **data parallel distributed training**

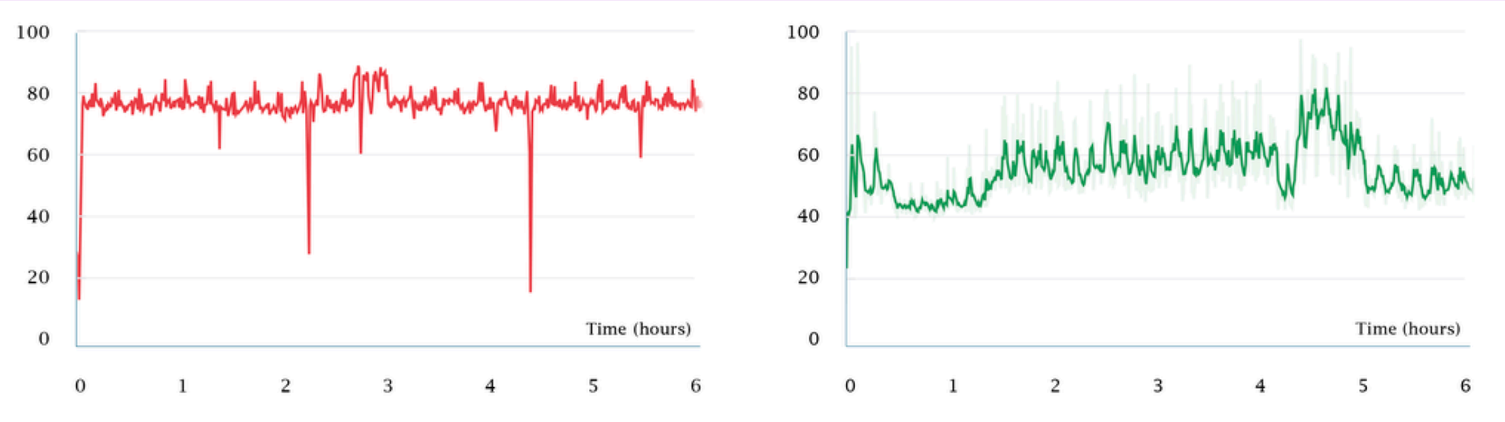
Limitations:

- Shared server leaves exposure to effects from other user's processes, profiling overheads vary between models

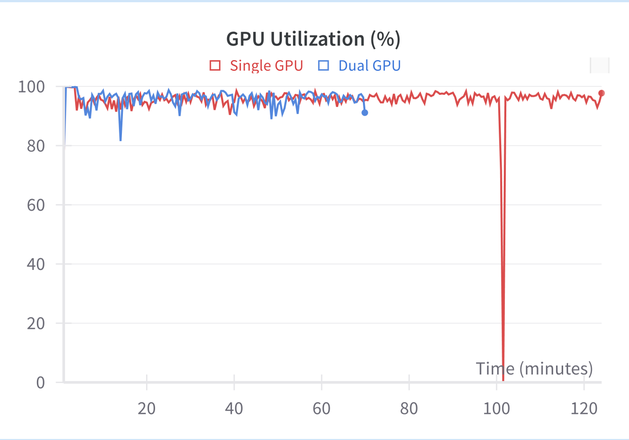


Insights & Conclusions

- GPU power usage was consistently higher for AudioLDM
- Indicating audio data is more resource intensive



- Distributing initially showed superlinear speedup, however isolating the system reduced the speedup to a factor of ~1.8
- Profiling showed the heavy nature of AdamW, this was shown to be more so in AudioLDM relative to other processes in the model



Need for research into optimisation in audio models to reduce computational costs and training times

Thanks for listening

AI + MUSIC

Queen Mary University of London