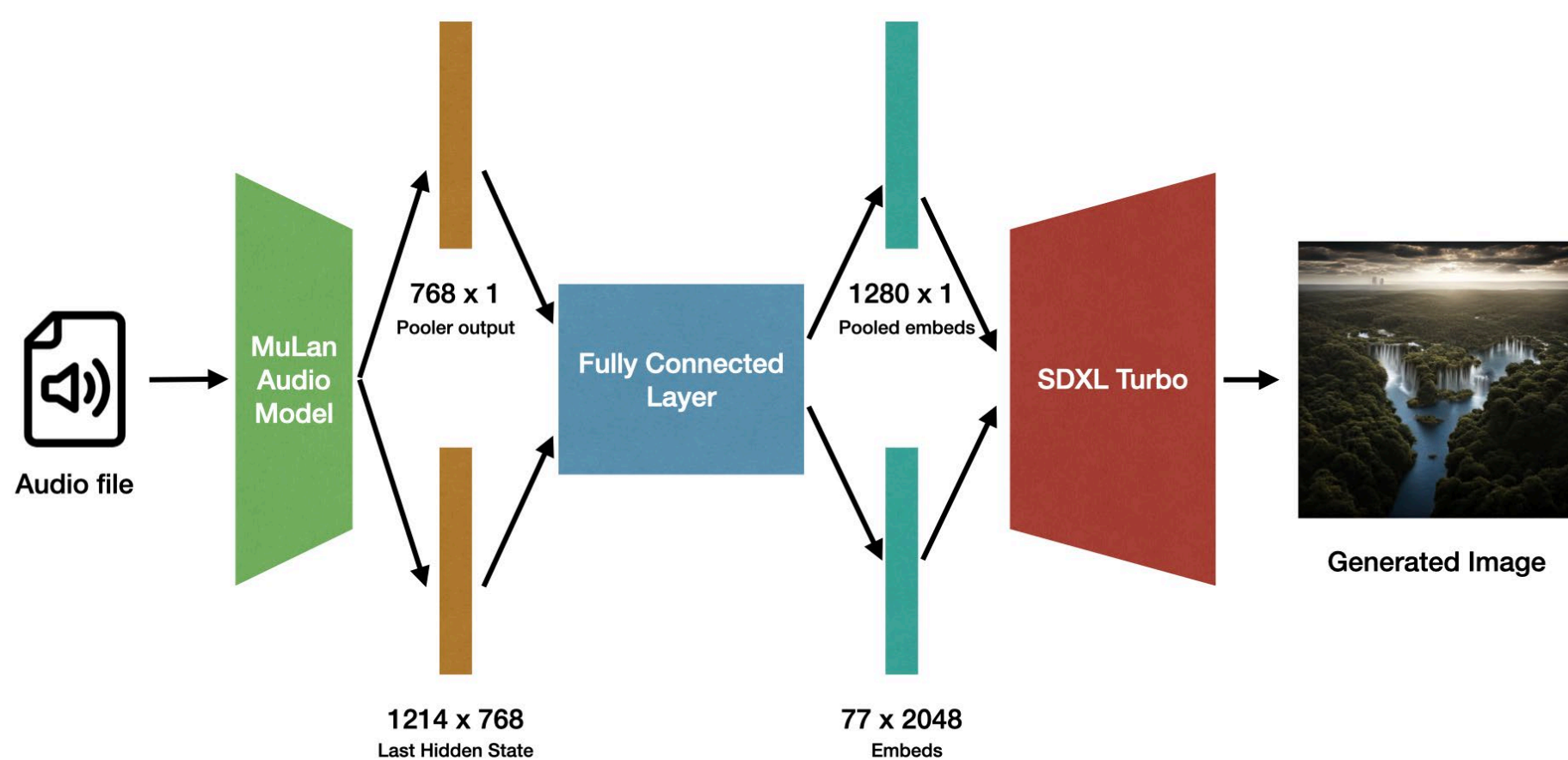


Music Visualisation With Deep Learning

Static Visualisations of Music Mood

Student: Ang Teng Hng Justin

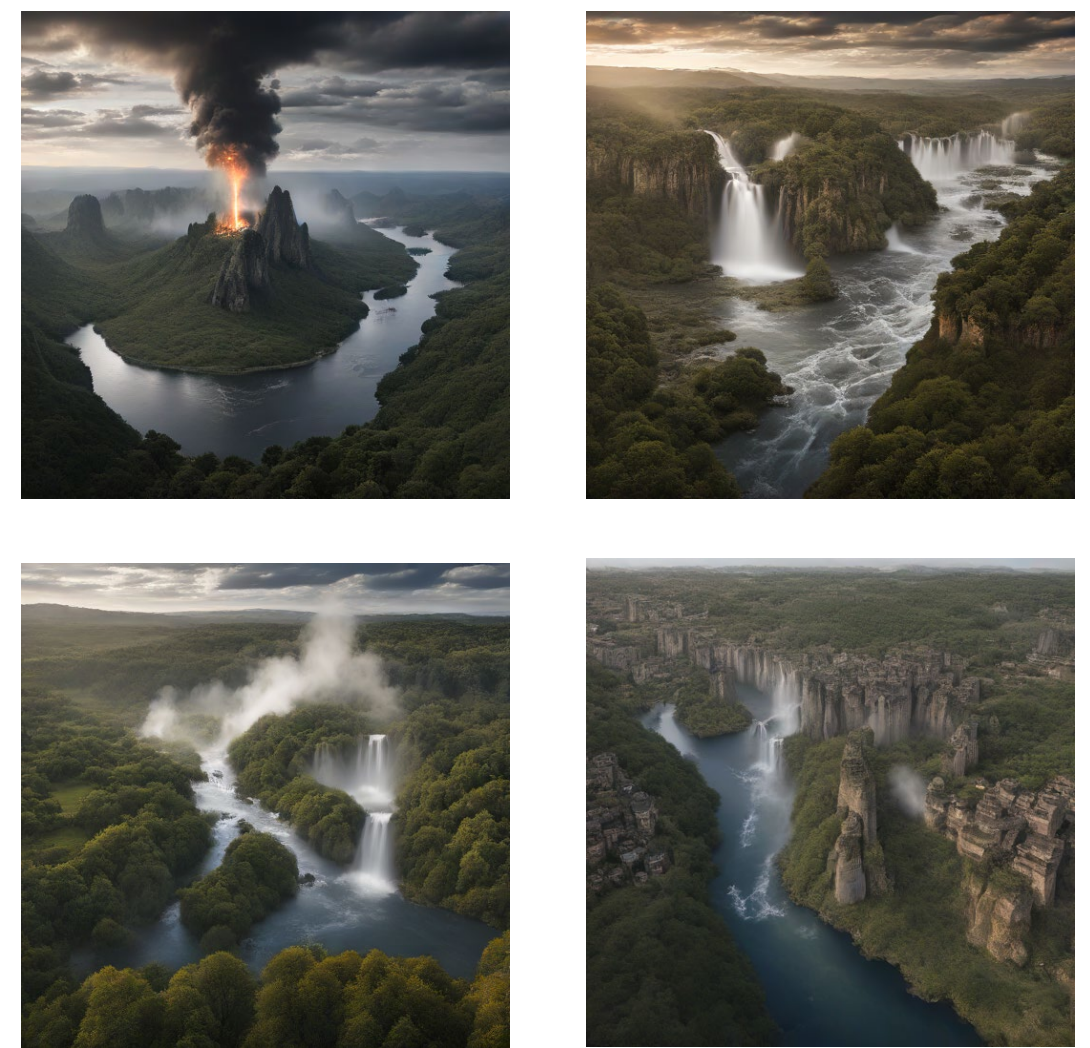
Supervisor: A/P Alexei Sourin



Proposed Network Design

Project Overview:

Mood is one of the least visualised aspects of music, and much of today's visualisation technology is animated, meant to be viewed while listening along. As such, a gap exists for static visualisations, which provide users with a quick overview of the overall ambience of a piece of music. This project proposes a model for visualising music mood through image generation. This model uses the MuLan audio model to employ a shared latent space between natural language and audio, as well as Stable Diffusion-XL Turbo for image generation. This model is trained using a dataset of classical music pieces and corresponding images generated using DALL-E. The generated images are subjected to analysis, and the model undergoes user testing to evaluate its effectiveness. The model can distinguish between different moods, which is validated by the user study.



Examples of Generated Images

| Question | Accuracy |
|----------|----------|
| 1 | 36.4% |
| 2 | 90.9% |
| 3 | 77.3% |
| 4 | 63.6% |

Table showing accuracy in user study