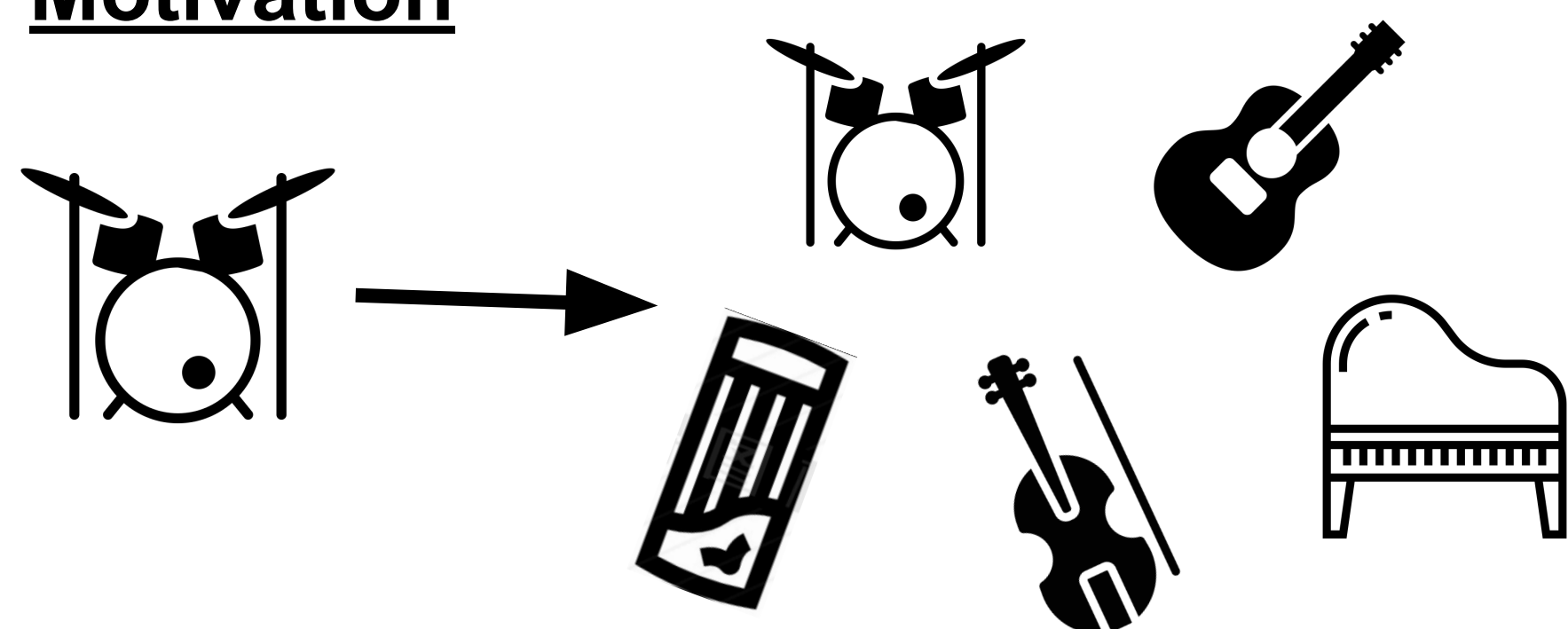
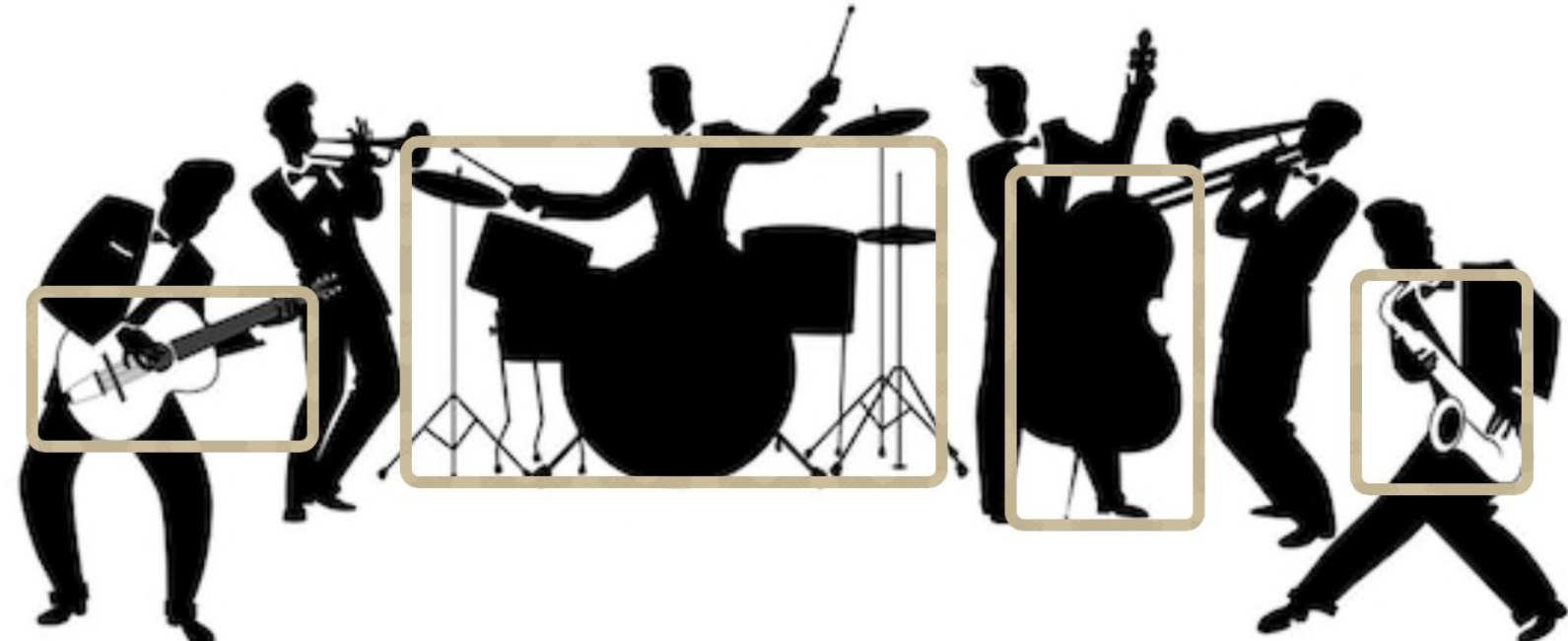


Introduction

Motivation



- Prepare dataset to model individual instruments



- Separate each single part of a band or synthesize an orchestra using individual models learned.
- Variational-Autodecoder helps removing unrelated dimensions. Generalize better without the encoder.

Related Work



- MuseNet - takes MIDI file as input, mostly style changing instead of generation.



- Universal Music Translation Network - Converts whistling to orchestra music. Translating styles.
- Current work focuses on generating from MIDI file, which is not deterministic and needs extra musical editor to complete generation.

Novel Approaches

1. Two Channels Training and Conv-Transformer

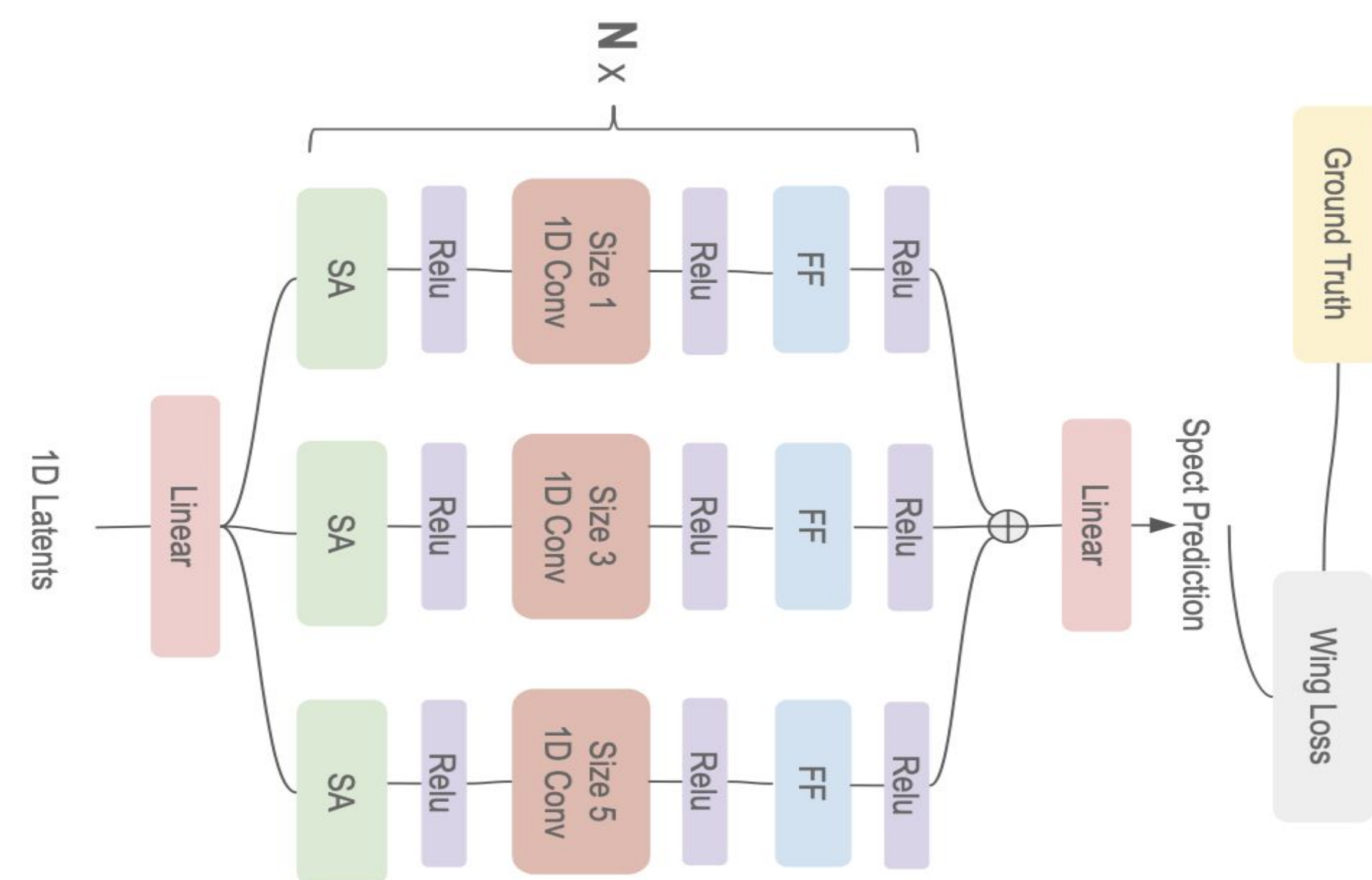


Fig 1. Convolutional Transformer with Convolutional layers of different dimensions

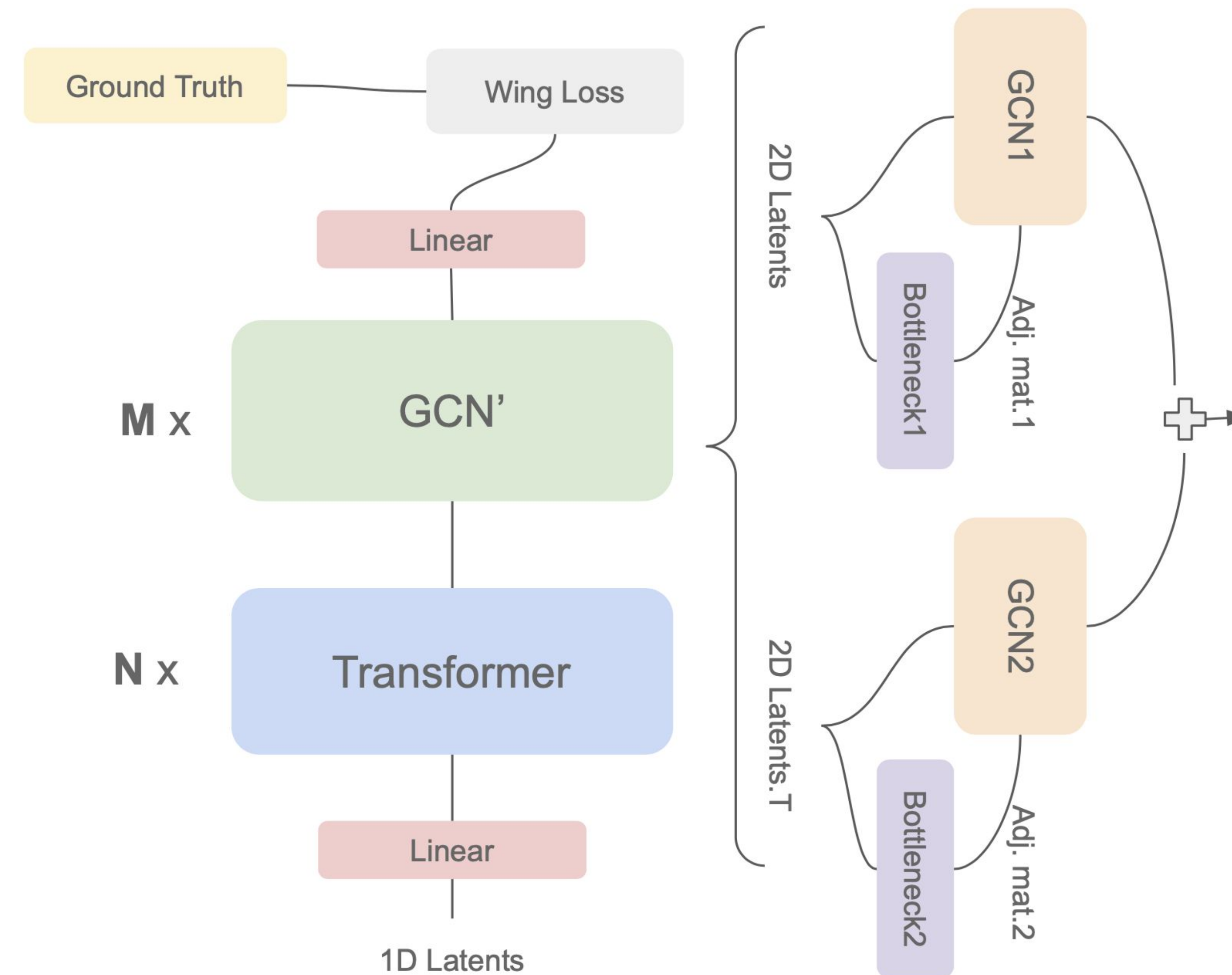


Fig 2. Bi-dimensional Trans-GCN

2. Generation

- **Interpolation** among sampled 1D latents
- **Random mutation** added to 1D latents
- Using **GCN** to order latents

Results

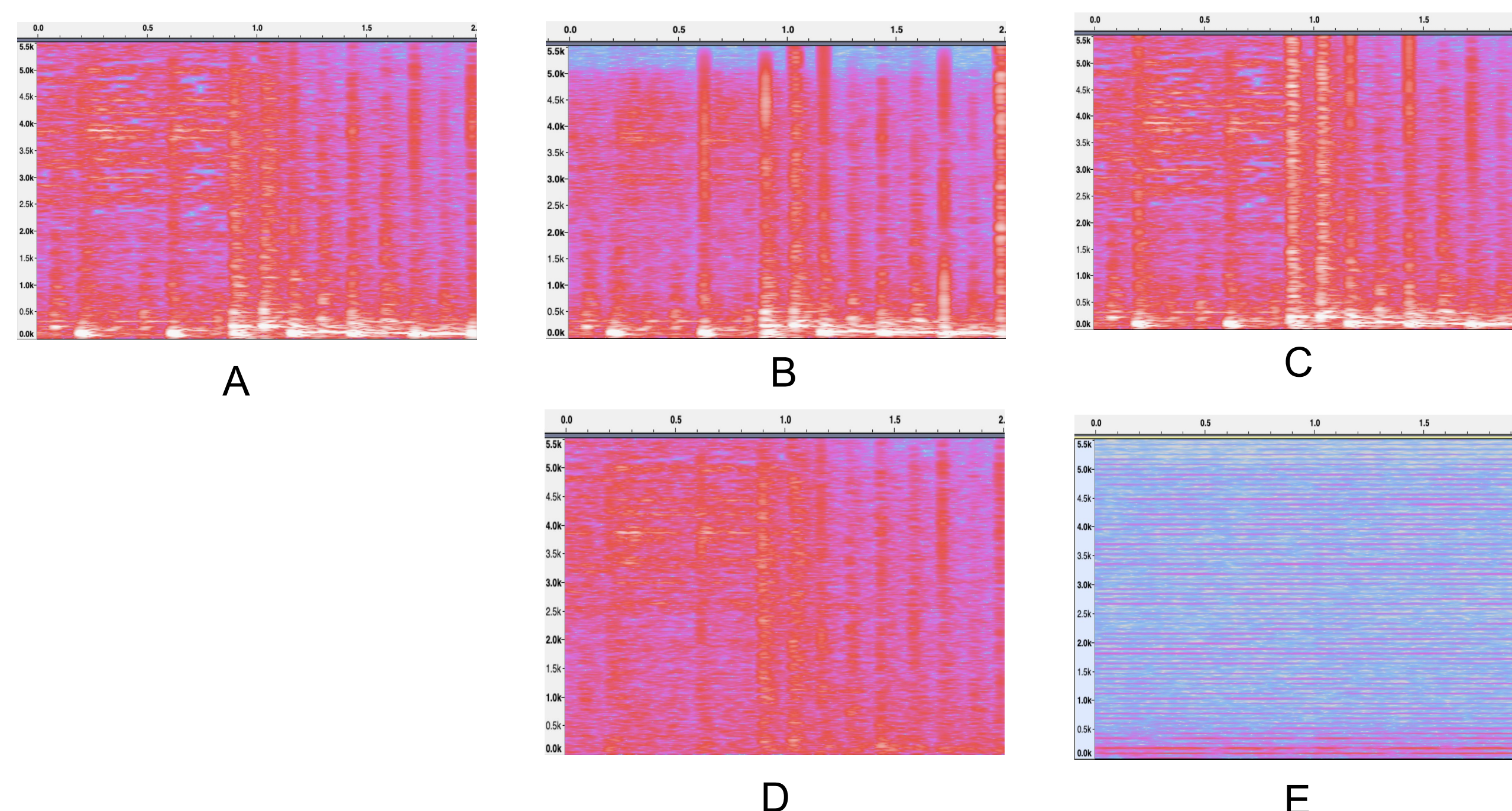
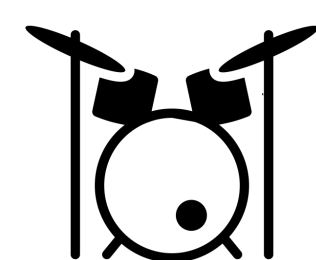


Fig 3. A: Ground Truth; B: Output w/o channels and WL; C: Output w/ channels and WL; D: Loss w/o channels and WL; E: Loss w/ channels & WL

Results (Cont'd)

- **Wing Loss** helped training to **intensify hits** in the audio.
- **Splitting** to different channels forces model to **focus** and **prevents vanishing gradients** problem at high frequencies.
- **Hearing is believing.**

Dataset



Goal: 30 hours
Current: 30 hours



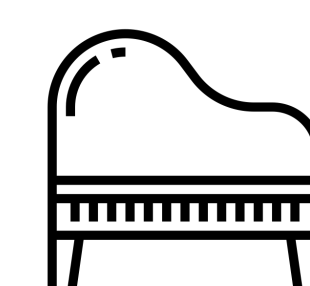
Goal: 30 hours
Current: 30 hours



Goal: 30 hours
Current: 0 hour



Goal: 30 hours
Current: 10 hours



Goal: 30 hours
Current: 0 hours

- 6000 training samples, 1000 test samples. Spectrograms of 10s segments extracted.
- Audio files splitted into segments, with each segment labeled with key, genre, tempo and detailed time in the file

Future Work

- Conduct research on generation methods that integrates semantic meanings of music such as pitch, rhythm, style, etc. Explore different ways to generate new samples.
- **Complete the dataset to make it more suitable for research purposes.**
- **Generalizing model to datasets of other instruments and finally synthesize an orchestra using learned models.**