VocalCycleGAN

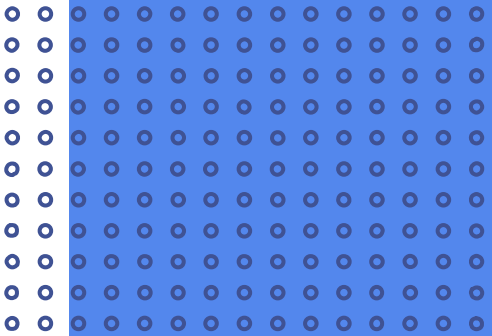
Speech to Vocal Synthesizer Effect Powered by Deep Learning

https://github.com/jjaw89/spring_2025_dl_audio_project/

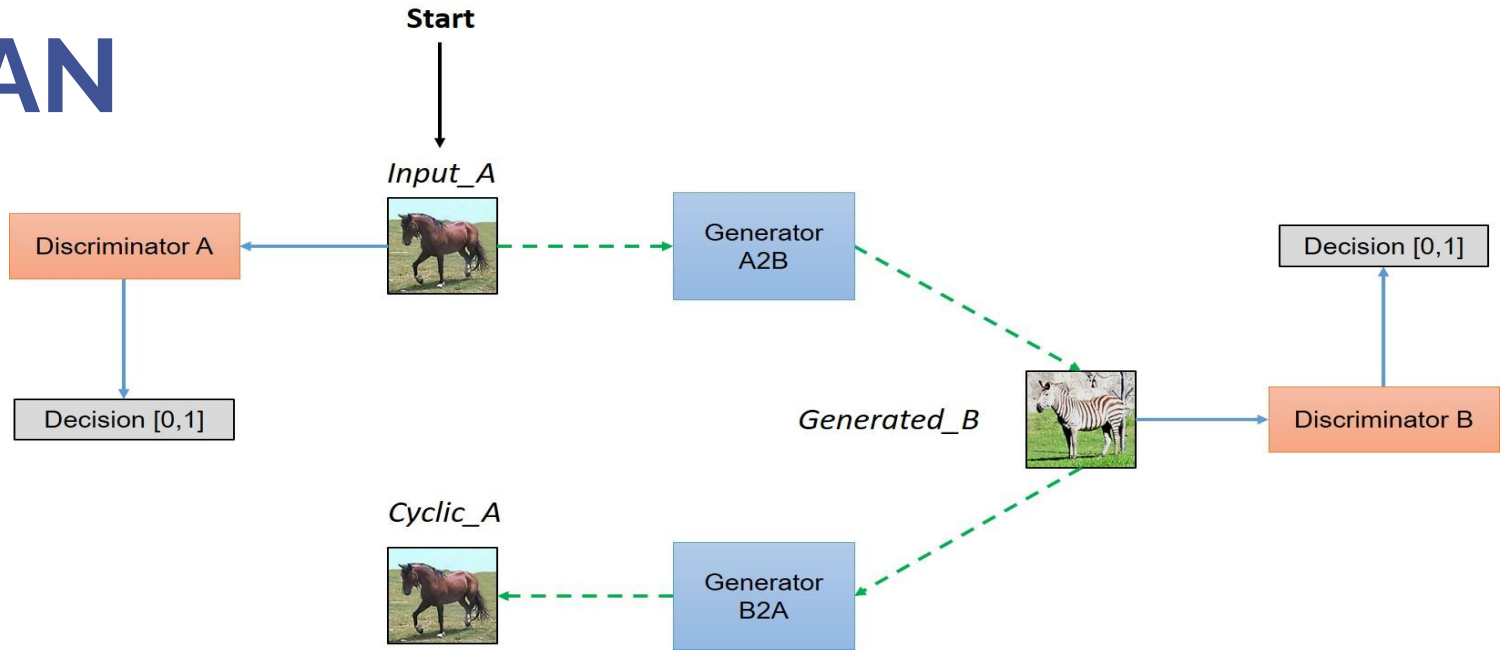
Chutian Ma, Greg Taylor, Jaspar Wiart
Erdös Institute Deep Learning Bootcamp
Spring 2025

Overview

- Datasets - LibriSpeech, MUSDB18
- CycleGAN (Generative Adversarial Network)
- Training behavior
- Result - an interesting “vocoder” effect



CycleGAN

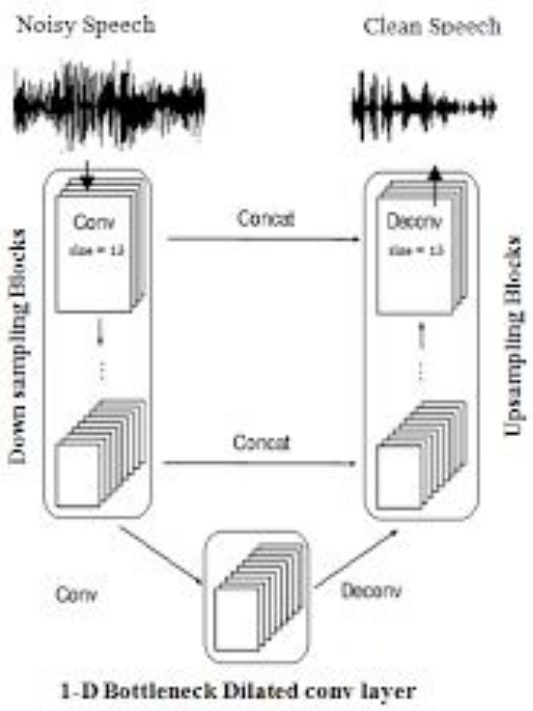


Losses:

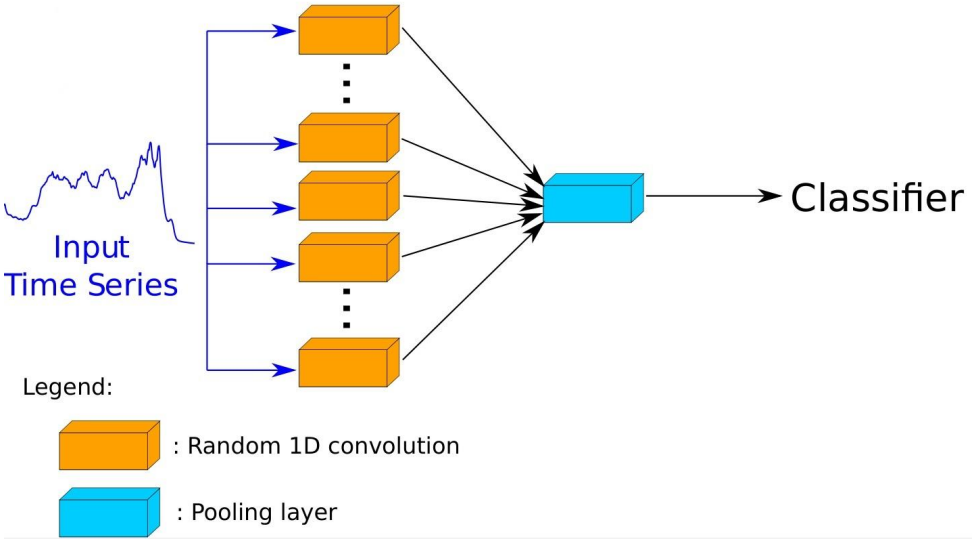
- Binary Cross Entropy (Discriminator)
- Adversarial Loss (Discriminator & Generator)
- Cycle Loss
- Identity Loss

Photo from https://hardikbansal.github.io/CycleGANBlog/

Wave-U-Net

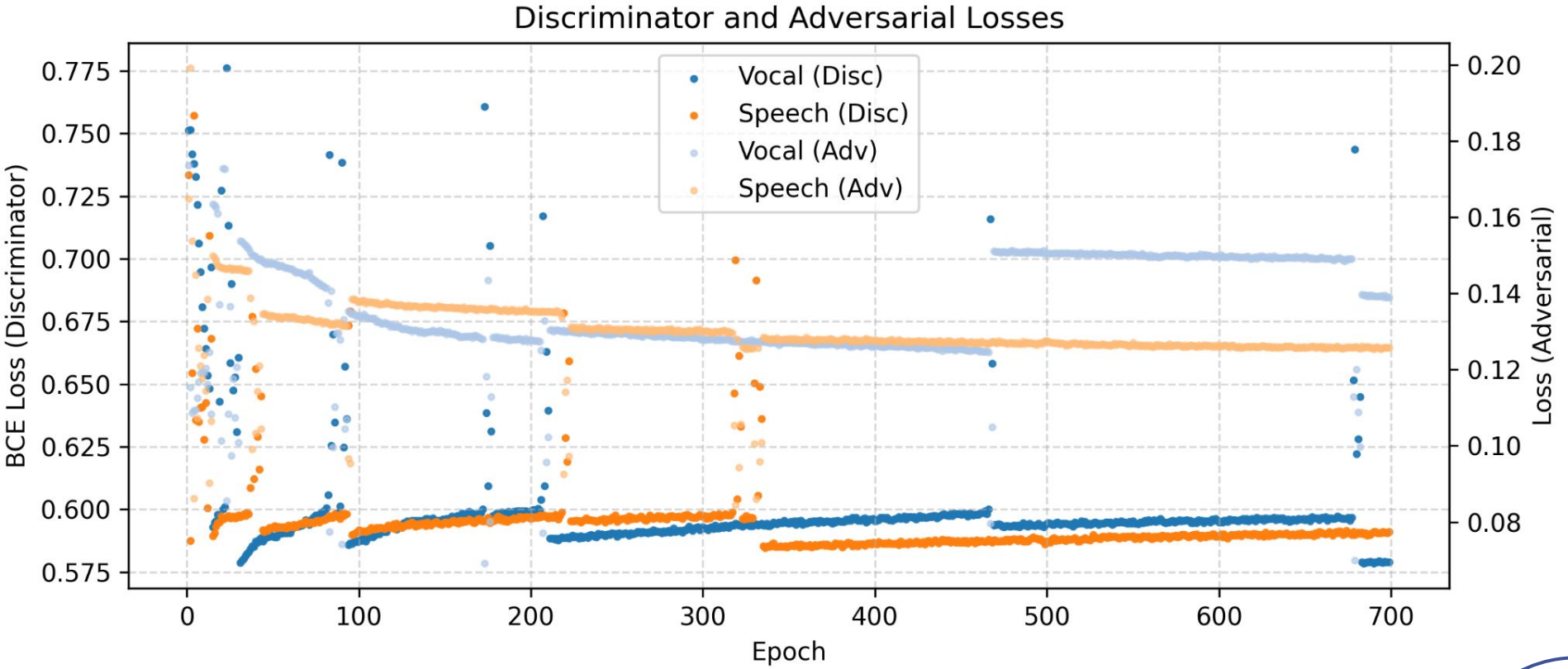


MiniRocket Classifier

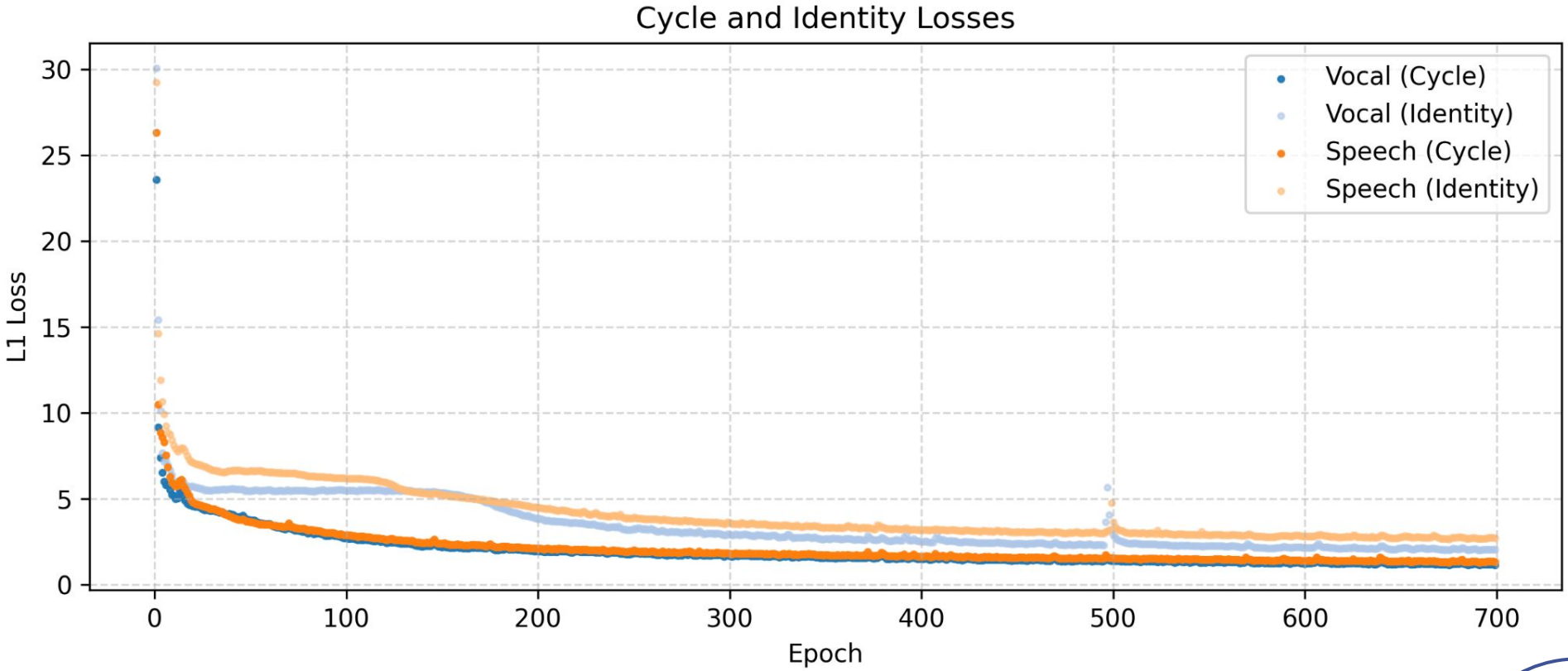


Loss Tuning

Identity Loss	Cycle Loss: 1	Cycle Loss: 0.001
0	**Generated Vocals:** silent	**Generated Vocals:** melodic tones, no words
0.001	No data	**Generated Vocals:** identical to input
0.00001	No data	**Generated Vocals:** words with tone changed



- The discriminators improved rapidly.
- We froze the discriminators while the generators caught up.
- Adversarial loss steadily improved over the course of training.



- Cycle and identity losses decreased steadily.
- As expected: cycle loss < identity loss