Deep Learning Audio

Lecture 7

Pavel Severilov

Moscow Institute of Physics and Technology

2022

Outline

1. Text-to-speech

Outline

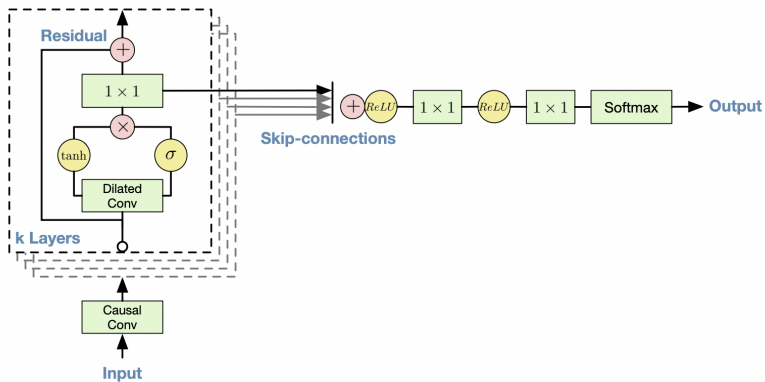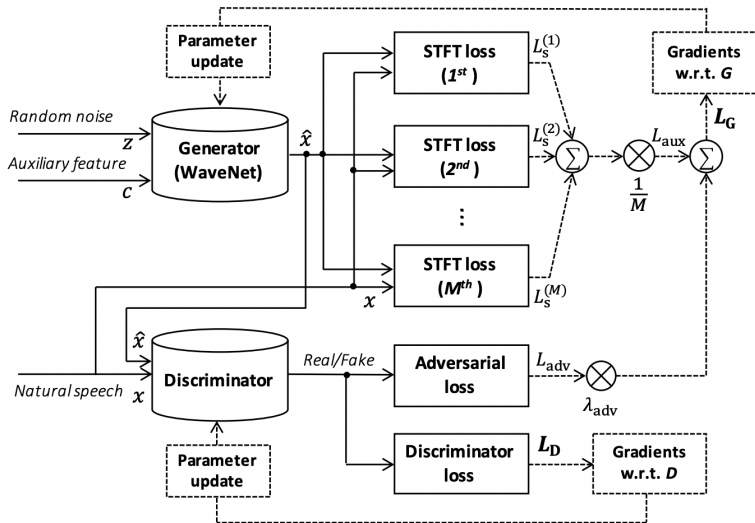1. Text-to-speech

WaveNet

Figure: Wavenet

Oord et al., WaveNet: A Generative Model for Raw Audio, 2016

Mu Law Encoding

Parallel WaveGAN

Figure: Parallel WaveGAN

Yamamoto et al., Parallel Wavegan: A Fast Waveform Generation Model Based on Generative Adversarial Networks with Multi-Resolution Spectrogram, IEEE ICASSP, 2020

STFT Loss

Takaki et al., STFT Spectral Loss for Training a Neural Speech Waveform Model, IEEE ICASSP, 2019