

Voicexer

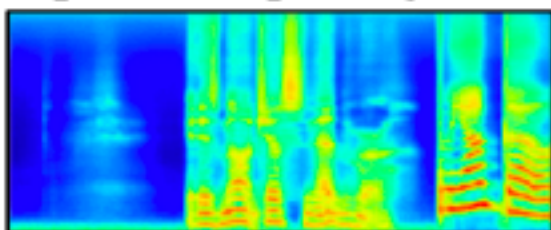
SynthesisModule - TFGAN Vocoder

Input restored mel spectrogram and output waveform: $\hat{x} = V(\left| \hat{X} \right|_{mel})$.

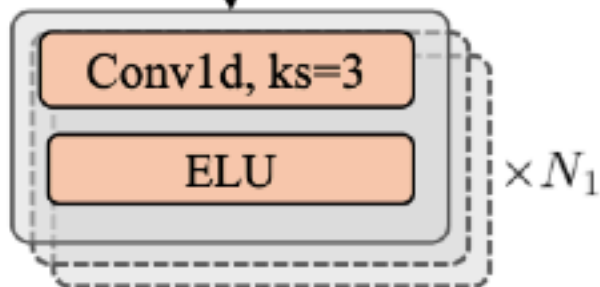
The generator of TFGAN Vocoder

The modular architecture inside generator

Input: Mel spectrogram



$(128, T)$



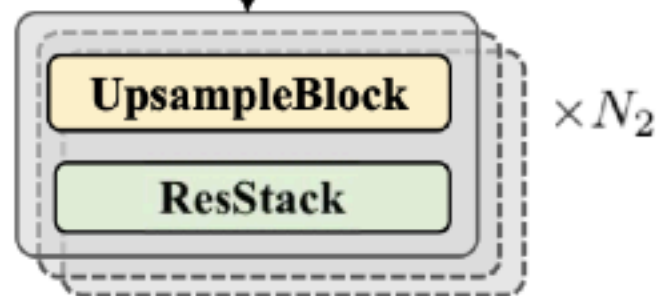
CondNet

$(512, T)$

Conv1d, ks=7

Generator

$(1024, T)$



UpNet

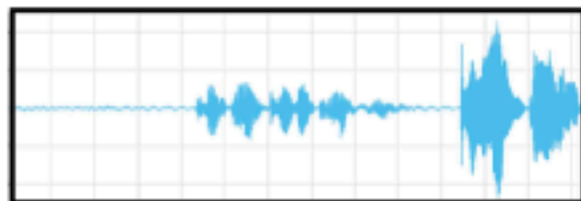
$(64, T * 441)$

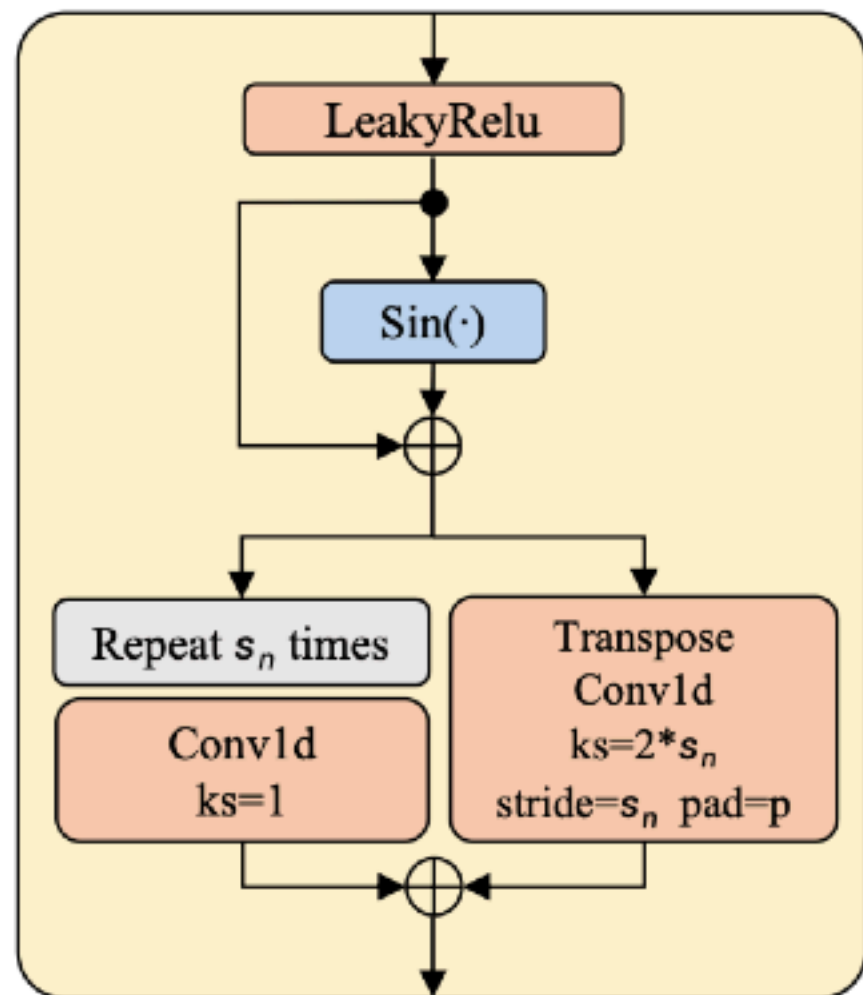
Conv1d, ks=7

Tanh

$(1, T * 441)$

Output: Waveform



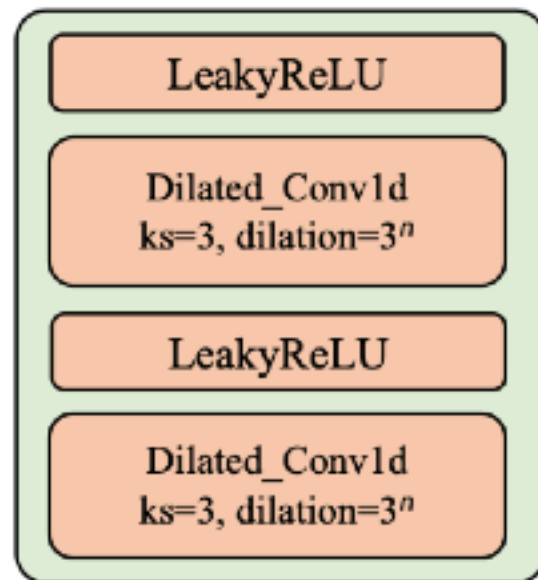


UpsampleBlock

$$s = [s_0, s_1, \dots, s_{N_2-1}]$$

$$p = \lfloor \frac{s}{2} \rfloor + \text{mod}(s)$$

$$n = [0, 1, \dots, N_2 - 1]$$



ResStack



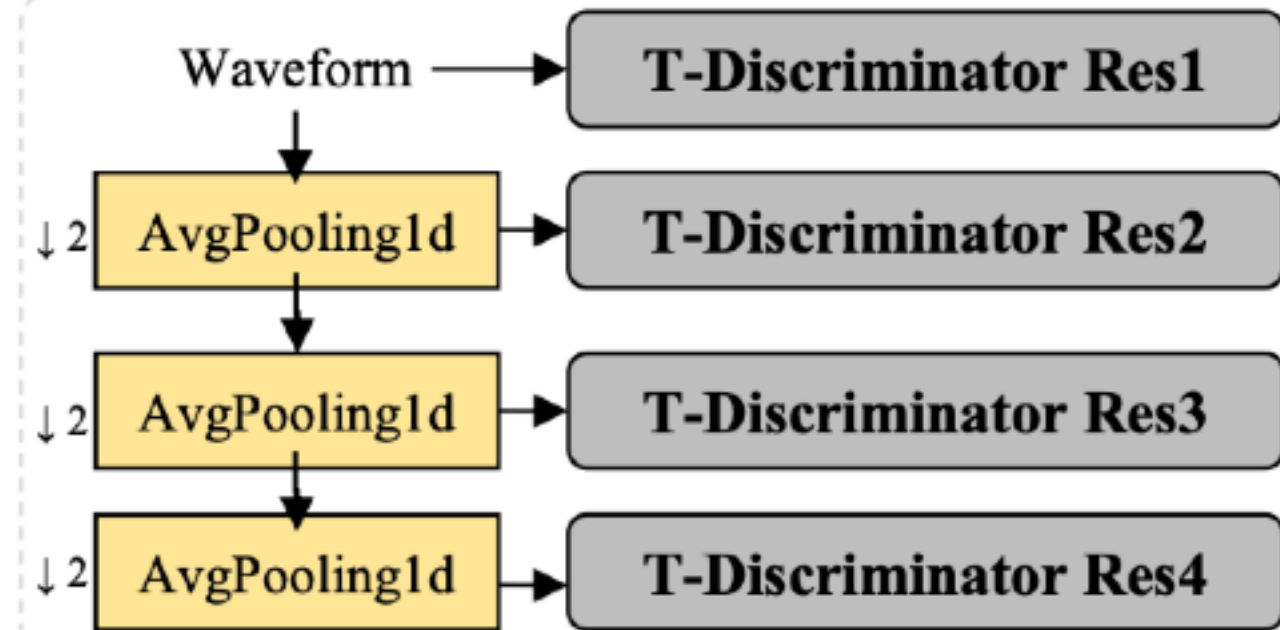
Input red and blue spectra and output waveform:

$$\hat{x} = V(|\hat{X}|_{mel})$$

SynthesisModule-TFGANVideo-Training

Frequency Discriminator

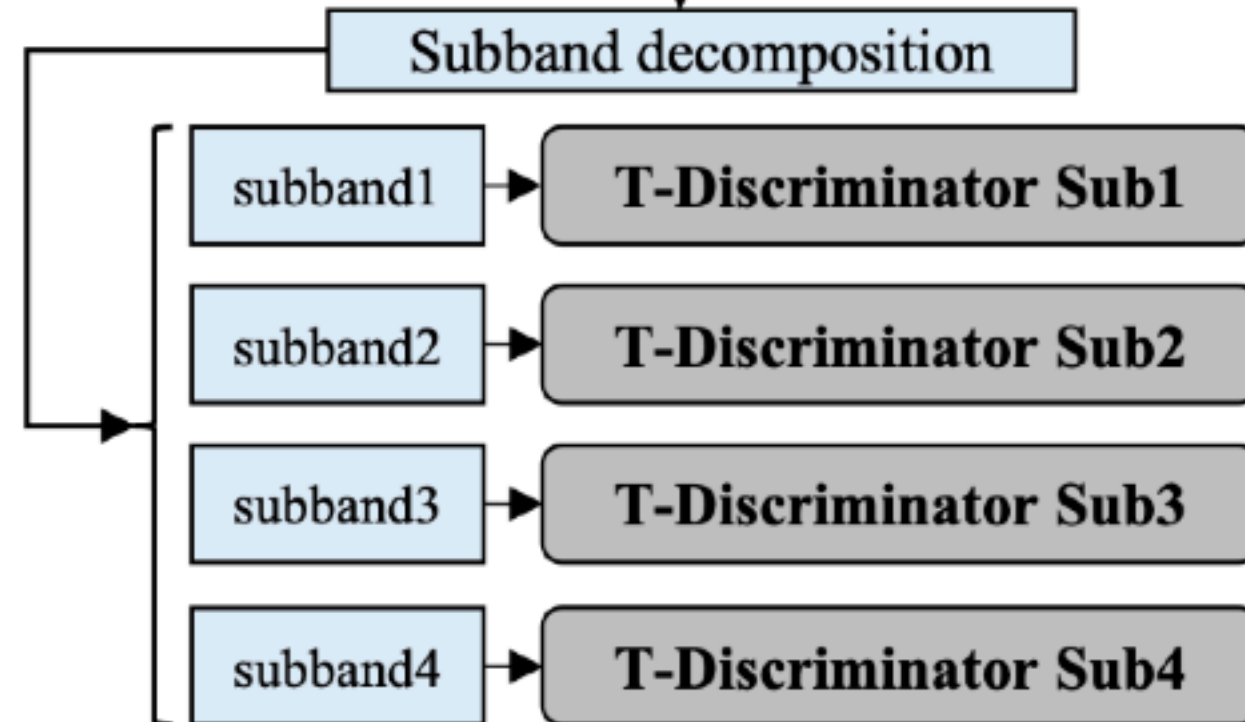
F-Discriminator



Multi-resolution discriminator

Loss functions and discriminators

Waveform



Subband discriminator

Energy Loss

Segment Loss

Phase Loss

Time
domain

Loss functions

Mel Loss

**Multi
Resolution
STFT Loss**

**Spectral
Convergence
Loss**

Frequency
domain