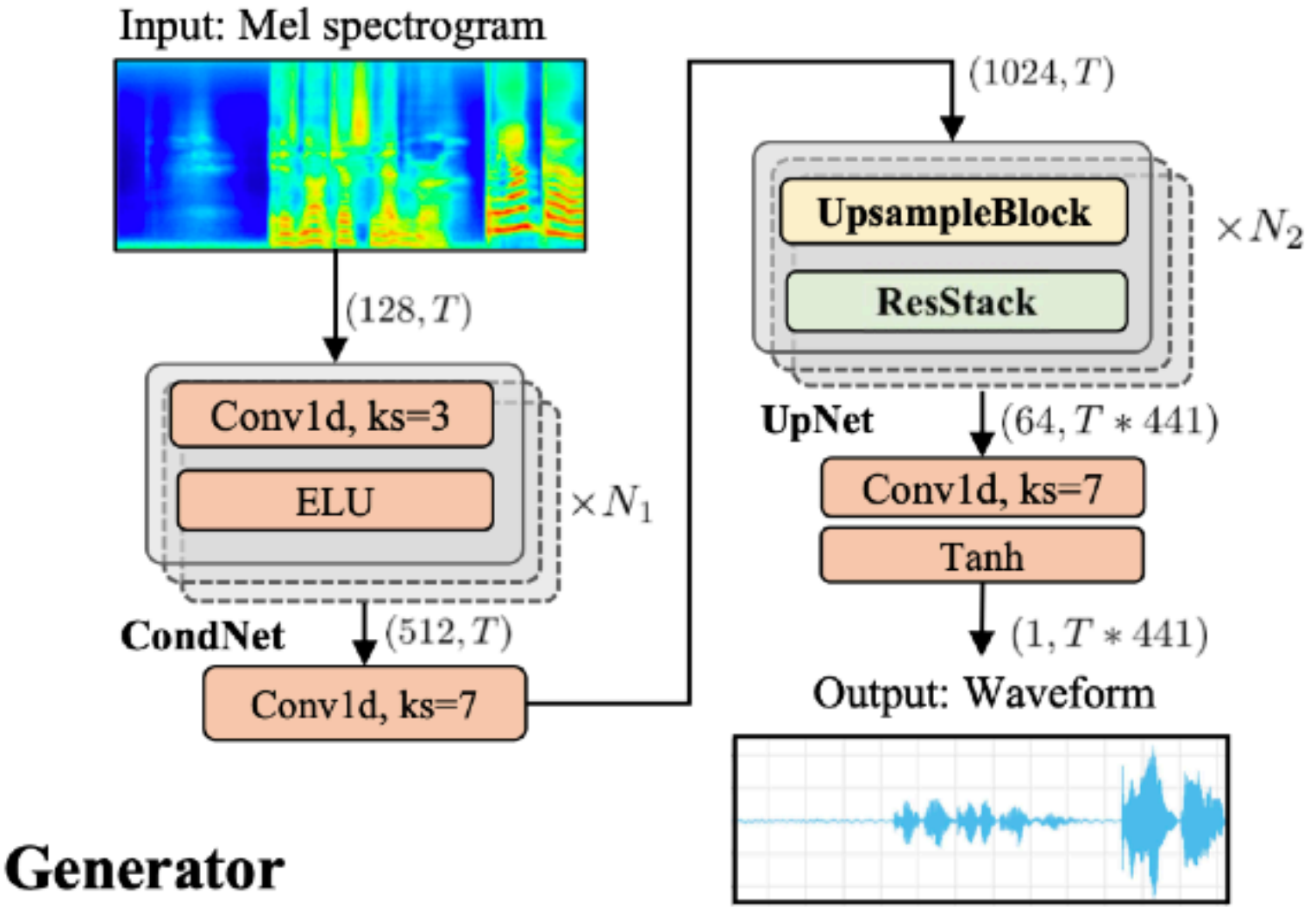
VoiceFixer

Synthesis Module - TFGAN Vocoder

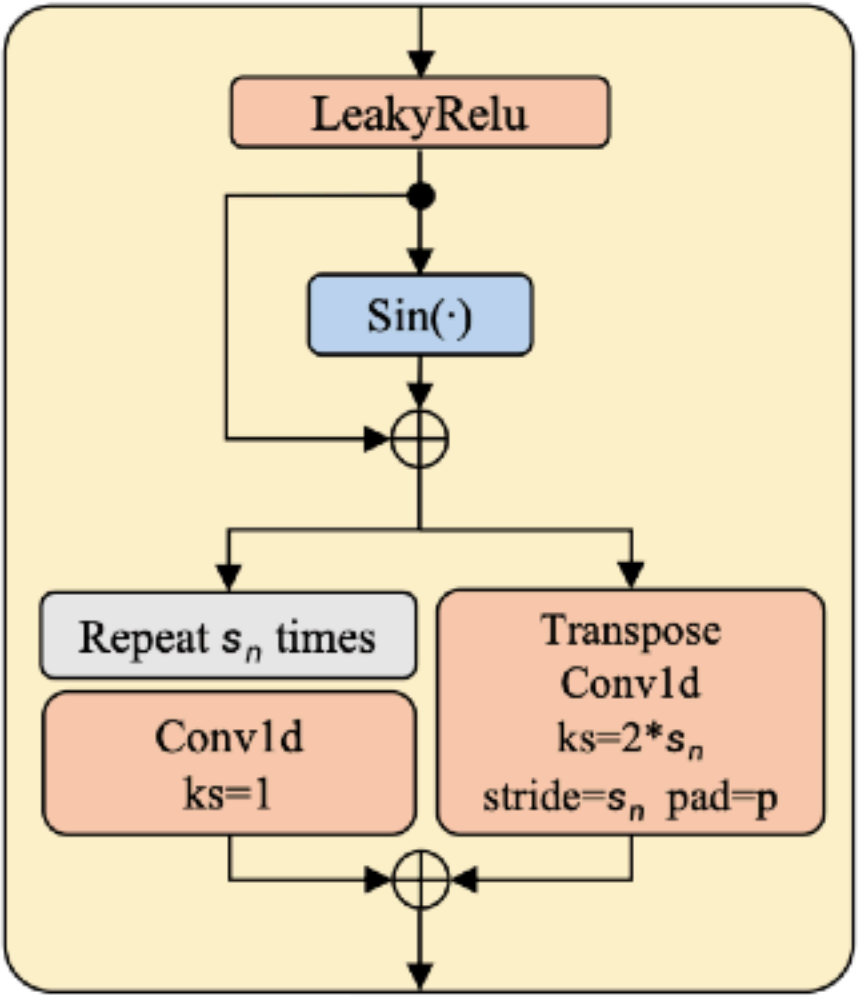
Input restored mel spectrogram and output waveform: $\hat{x} = V(\left|\hat{X}\right|_{mel})$.

The generator of TFGAN Vocoder

The module architecture inside generator

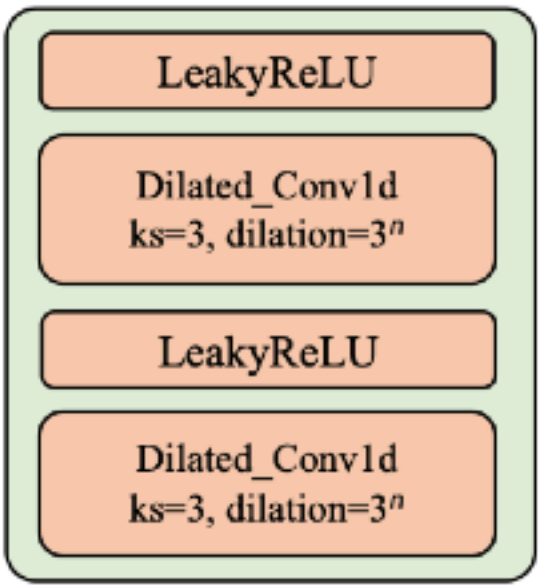

Input: Mel spectrogram
$(128, T)$
Conv1d, ks=3
ELU
$\times N_1$
CondNet
$(512, T)$
Conv1d, ks=7
$(1024, T)$
UpsampleBlock
ResStack
$\times N_2$
UpNet
$(64, T * 441)$
Conv1d, ks=7
Tanh
$(1, T * 441)$
Output: Waveform

Generator



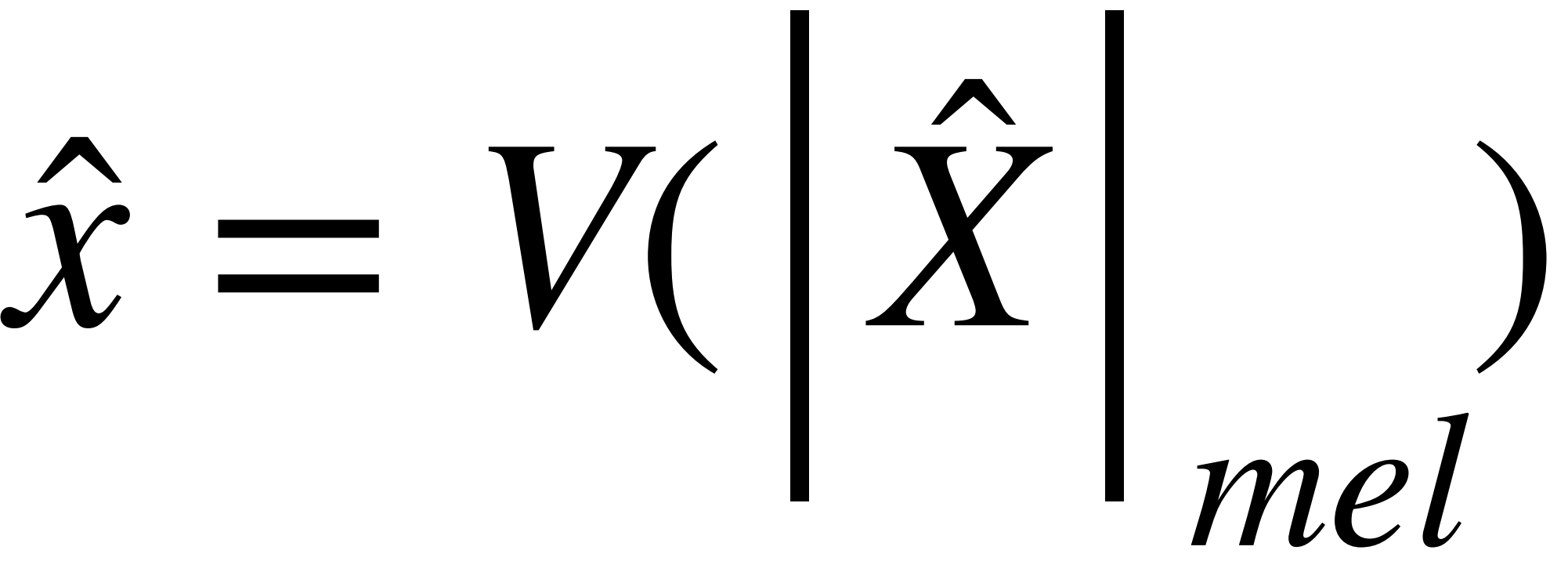
UpsampleBlock

$$s = [s_0, s_1, ..., s_{N_2-1}]$$
$$p = \lfloor \frac{s}{2} \rfloor + mod(s)$$
$$n = [0, 1, ..., N_2 - 1]$$

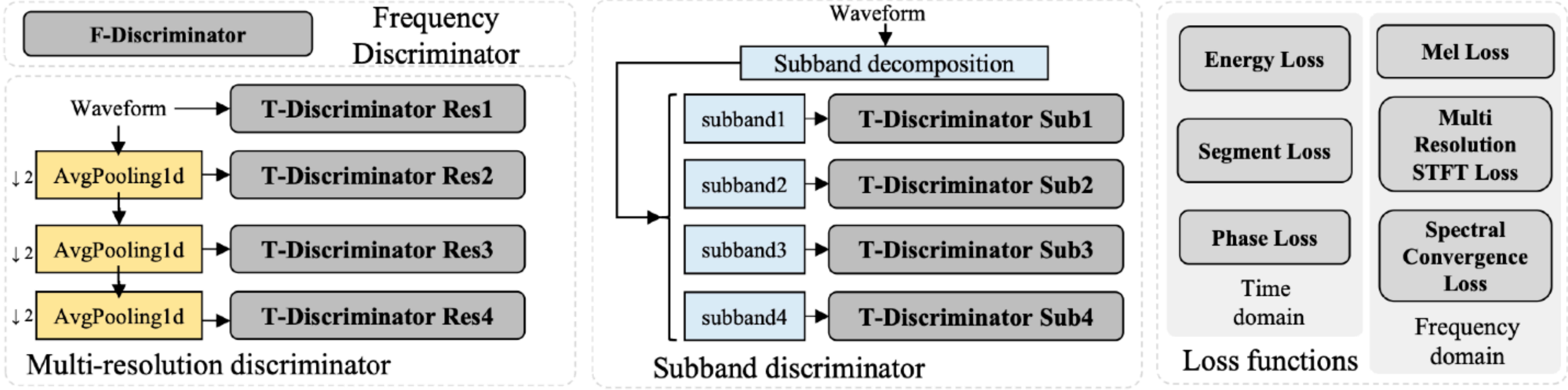


ResStack

Input restored mel spectrogram and output waveform:



Synthesis Module - TFGAN Vocoder - Training



Loss functions and discriminators