

Voicexer

Analysis Module

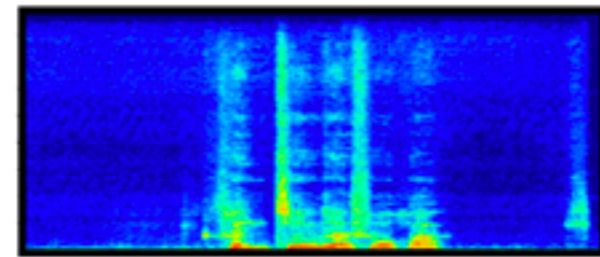
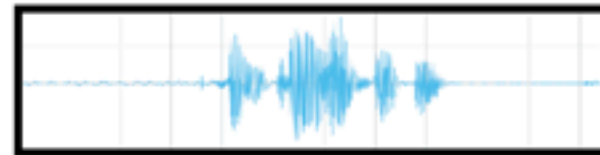
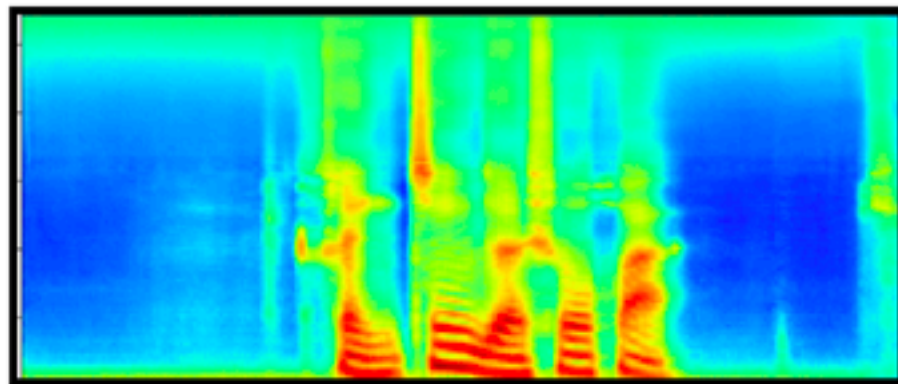
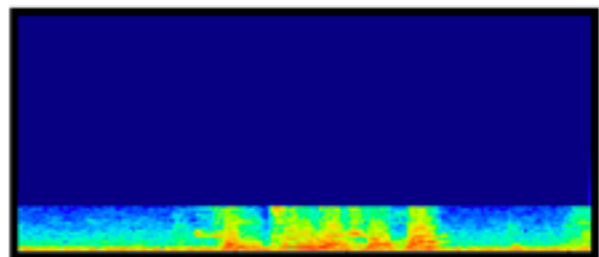
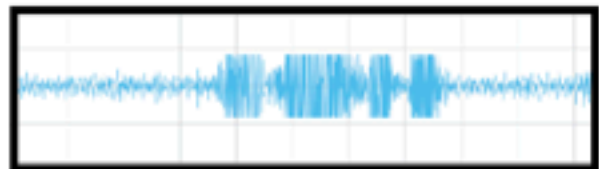
Low quality
speech

**Analysis
Module**

Mel
spectrogram
estimation

**Synthesis
Module**

High quality speech
estimation



• The prediction of high quality net spectrogram.

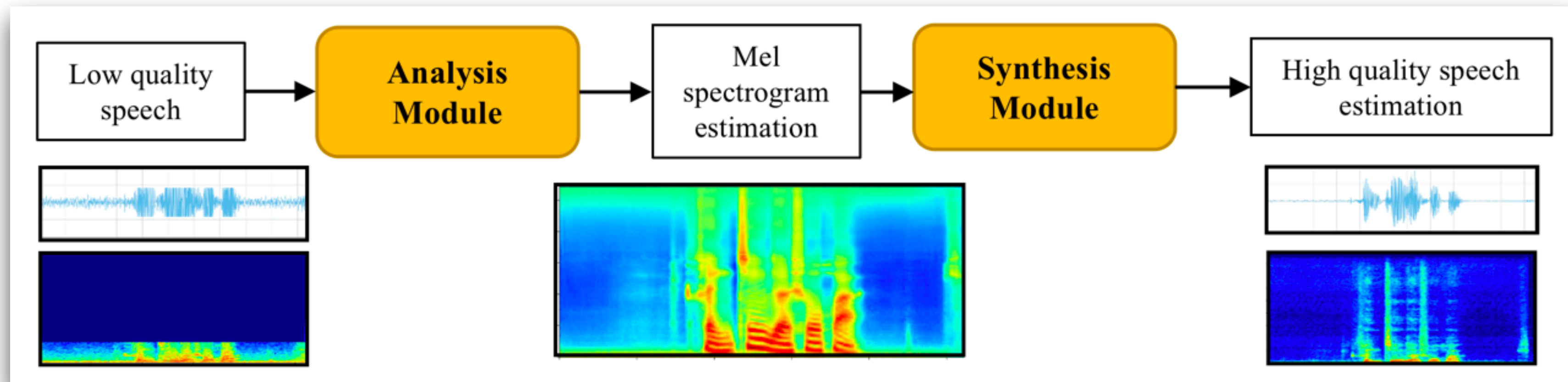
• Loss function is $\text{predictogram Mean Absolute Error}$.

$$\left| \hat{S} \right|_{mel} = F(\left| X \right|_{mel}) \odot \left| X \right|_{mel}$$

$$L_{ana} = \left\| \left| \hat{X} \right|_{mel} - \left| S \right|_{mel} \right\|_1$$

VoiceFixer

Analysis Module



- The prediction of high quality mel spectrogram.

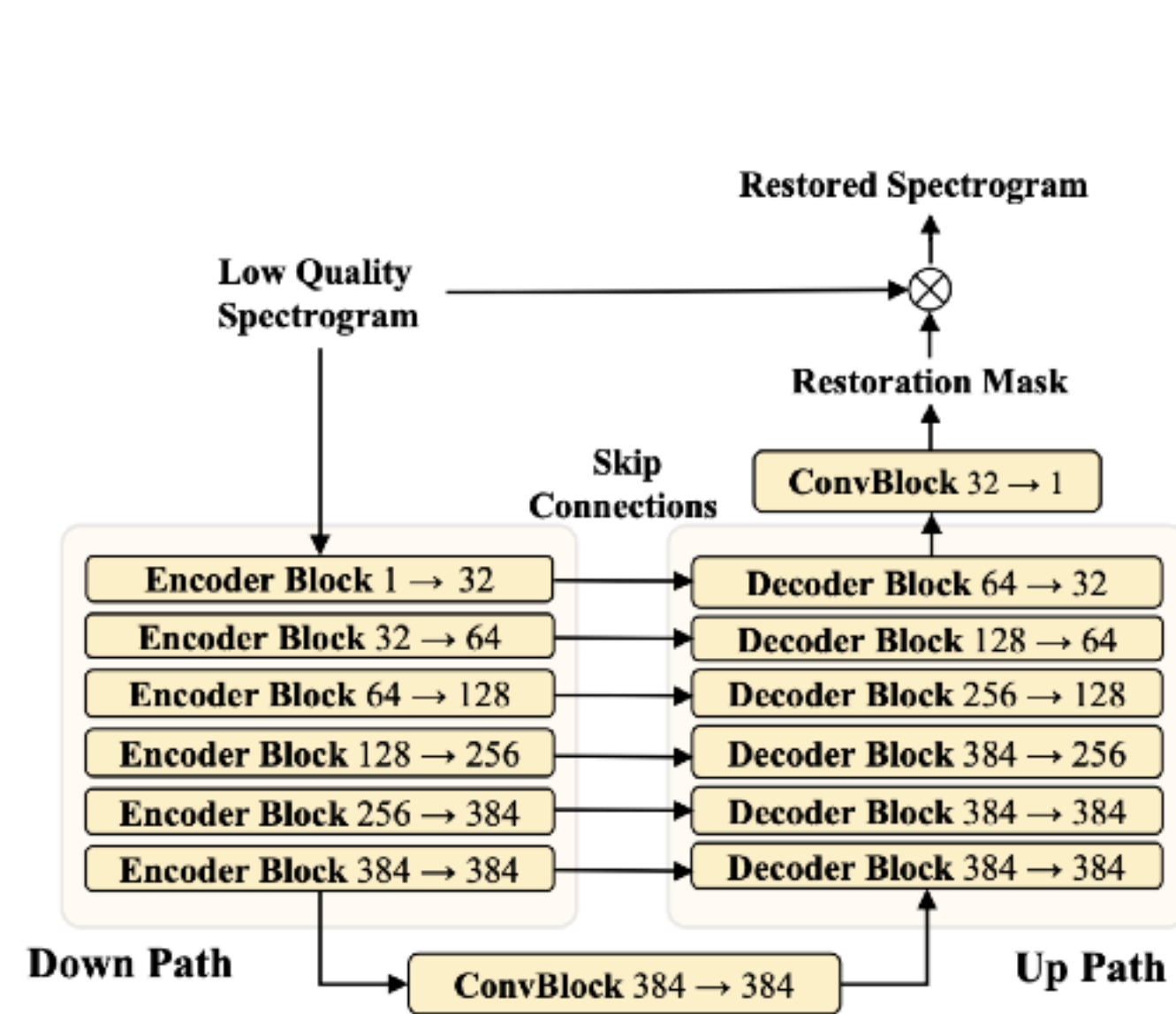
$$\left| \hat{S} \right|_{mel} = F(\left| X \right|_{mel}) \odot \left| X \right|_{mel}$$

- Loss function is spectrogram Mean Absolute Error.

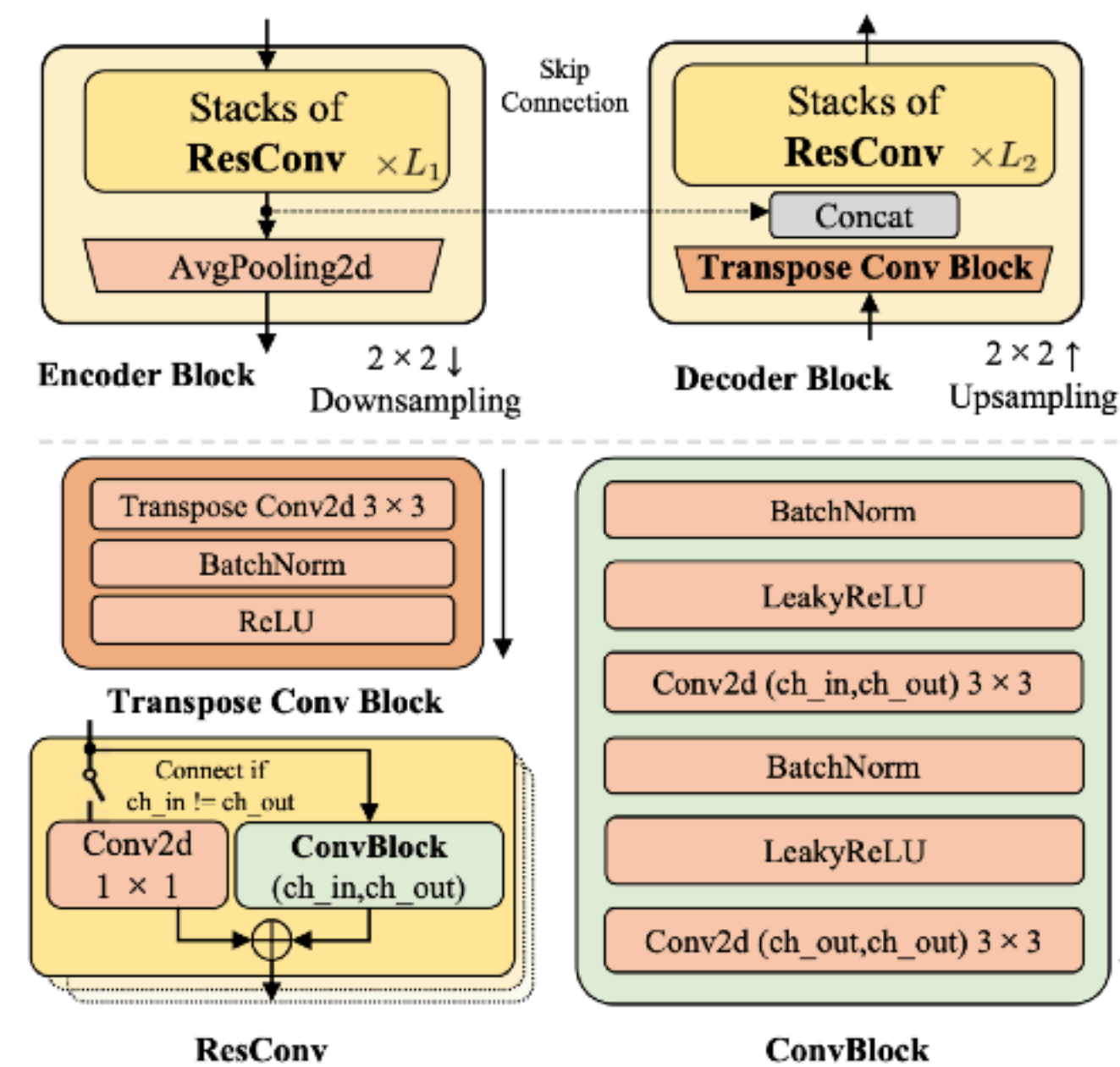
$$L_{ana} = \left\| \left| \hat{X} \right|_{mel} - \left| S \right|_{mel} \right\|_1$$

VoiceFixer

Three implementations of analysis module

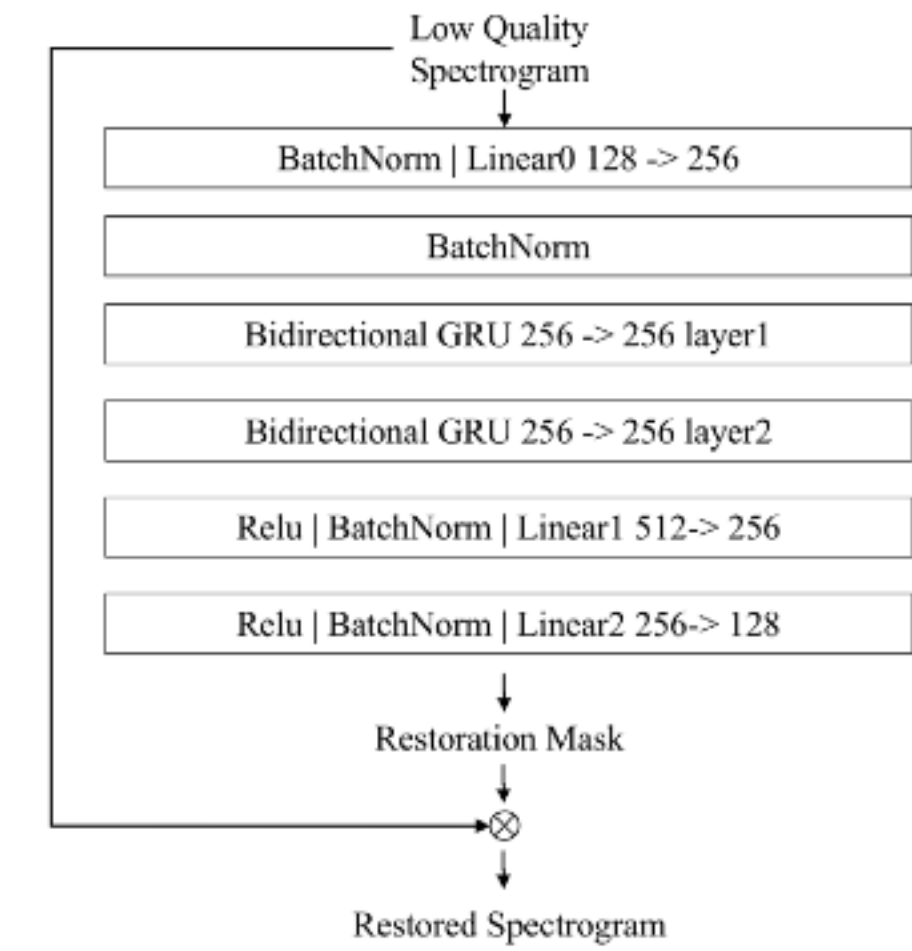


(a) The architecture of ResUNet

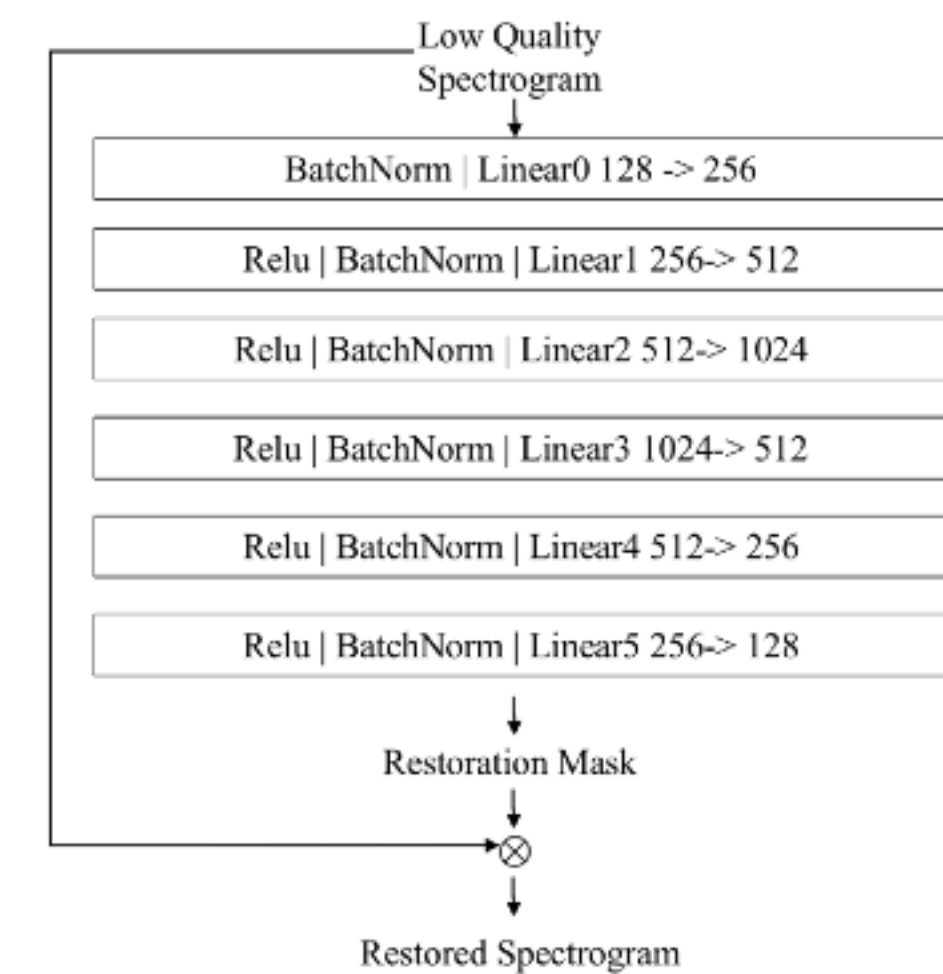


(b) Details of the encoder and decoder blocks of the ResUNet

Architecture1: ResUNet



Architecture2: BiGRU



Architecture3: DNN