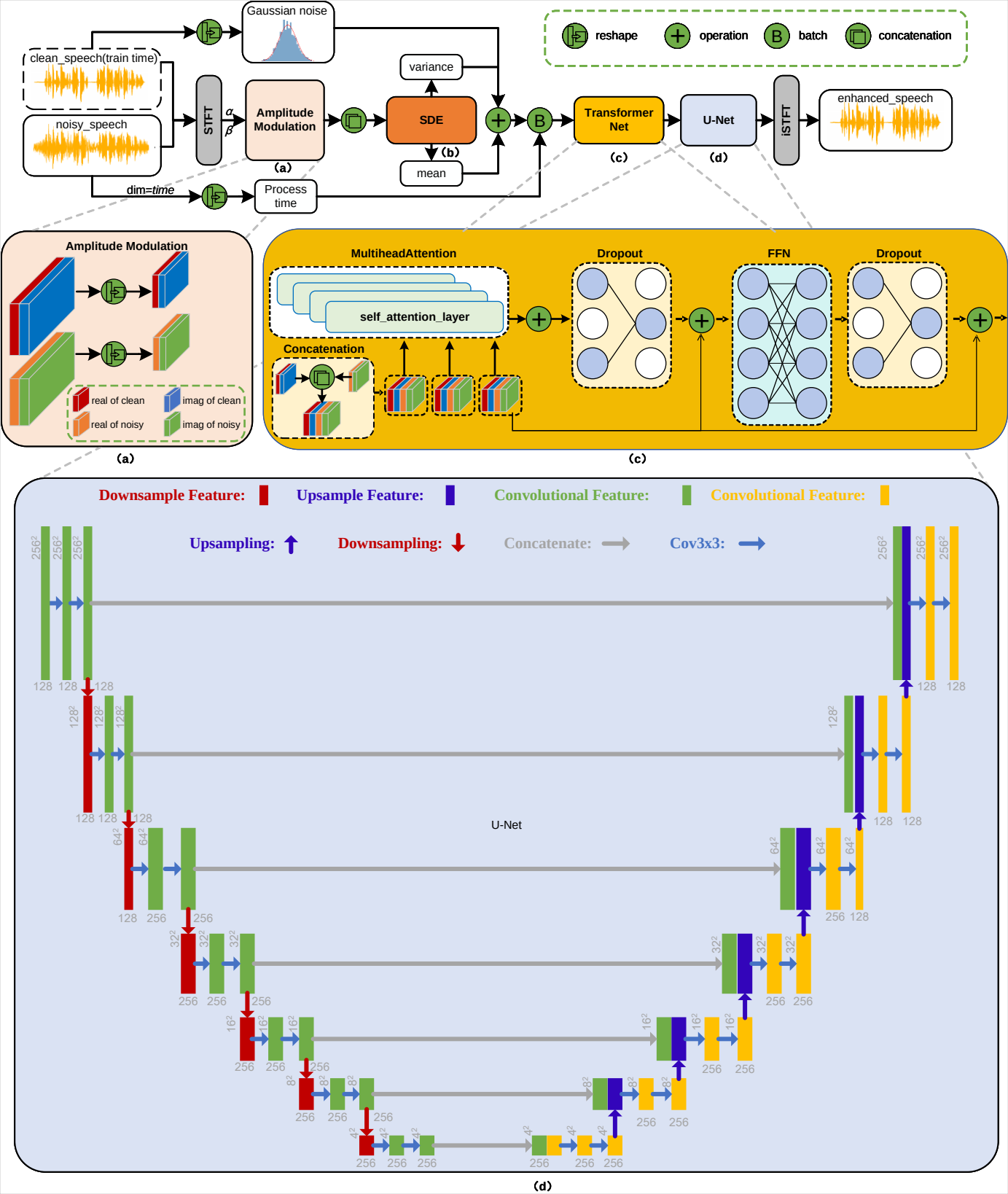
Gaussian noise
clean_speech(train time)
noisy_speech
STFT
α
β
Amplitude Modulation
(a)
variance
SDE
(b)
mean
B
Transformer Net
(c)
U-Net
(d)
iSTFT
enhanced_speech
dim=time
Process time
reshape
operation
batch
concatenation
Amplitude Modulation
real of clean
imag of clean
real of noisy
imag of noisy
(a)
MultiheadAttention
self_attention_layer
Concatenation
Dropout
FFN
Dropout
(c)
Downsample Feature:
Upsample Feature:
Convolutional Feature:
Convolutional Feature:
Upsampling:
Downsampling:
Concatenate:
Cov3x3:
U-Net
(d)