ConvDown Encoder

[B, L, 3]

| *Input*

Conv, Stride=2

GELU

[B, L/2, 3]

|

Conv, Stride=2

GELU

[B, L/4, H]

ConvTranspose

GELU

[B, L/4, H]

VQ Path →   *← Masking Path*

Linear

[B, L/4, H]

Encoder

Model

Decoder