ConvDownEncoder (Contrastive Learning)

[B, L, 3]

Conv1d(3→256, k=3, s=2)

LayerNorm(256)

GELU

[B, L/2, 256]

Conv1d(256→256, k=3, s=2)

LayerNorm(256)

GELU

[B, L/4, 256]

LayerNorm(256)

Legend

[B,L,C]	Tensor shapes
Op	Operations
Act	Activations
VQ	Special modules
→	Data flow
⇢	Output conn.

Encoder

Path A: Masking

Linear(256→256)

Mask 65%

(span=10, learnable emb)

[B, L/4, 256]

→ U-Net Context Encoder

Path B: Vector Quantization

Gumbel-Softmax VQ

(640 codes, 2 groups)

Linear(256→256)

[B, L/4, 256]

→ Contrastive Targets

Model