

Layer Name	Input Size	Hyperparameter	Output Size
Up-Sampling	$1 \times 64000$	-	$1 \times 256340$
Encoding Layer 1	$1 \times 256340$	8,4,48	$48 \times 64084$
Encoding Layer 2	$48 \times 64048$	8,4,96	$96 \times 16020$
Encoding Layer 3	$96 \times 16020$	8,4,192	$192 \times 4004$
Encoding Layer 4	$192 \times 4004$	8,4,384	$384 \times 1000$
Encoding Layer 5	$384 \times 1000$	8,4,768	$768 \times 249$
Casual Attention	$768 \times 249$	-	$768 \times 249$
Decoding Layer 1	$1536 \times 249$	8,4, 384	$384 \times 1000$
Decoding Layer 2	$768 \times 1000$	8,4,192	$192 \times 4004$
Decoding Layer 3	$384 \times 4004$	8,4,96	$96 \times 16020$
Decoding Layer 4	$192 \times 16020$	8,4,48	$48 \times 64048$
Decoding Layer 5	$96 \times 64048$	8,4,1	$1 \times 256340$
Down-Sampling	$1 \times 256340$	-	$1 \times 64000$