

	input size	hyperparameters	output size
format	(channels ×sampling points)	(kernel size, stride, channels)	(channels× sampling points)
up-sampling	1×64000	-	1×256340
layer 1 encoding layer	1×256340	8,4,48	48×64084
layer 2 encoding layer	48×64048	8,4,96	96×16020
layer 3 encoding layer	96×16020	8,4,192	192×4004
layer 4 encoding layer	192×4004	8,4,384	384×1000
layer 5 encoding layer	384×1000	8,4,768	768×249
Transformer module	768×249	dual-path Transformer × 4	768×249
layer 5 decoding layer	1536×249	8,4, 384	384×1000
layer 4 decoding layer	768×1000	8,4,192	192×4004
layer 3 decoding layer	384×4004	8,4,96	96×16020
layer 2 decoding layer	192×16020	8,4,48	48×64048
layer 1 decoding layer	96×64048	8,4,1	1×256340
down-sampling	1×256340	-	1×64000