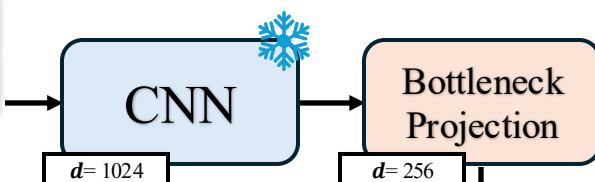




Frames



$d$	Feature dimension
	Frozen model
$\oplus$	Sum
$\odot$	Positional Encoding
TDE	Temporal-Difference Embedding
FFN	Feed Forward Network
MHSA	Multi-Head Self Attention
MSCM	Multi-Scale Convolution Module
DW	Depth-Wise Convolution

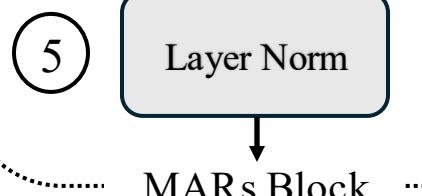
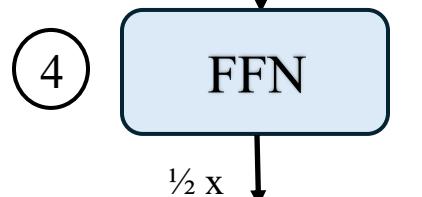
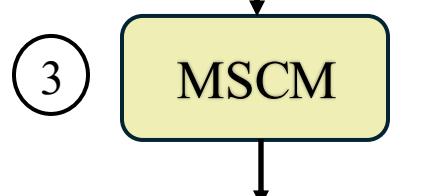
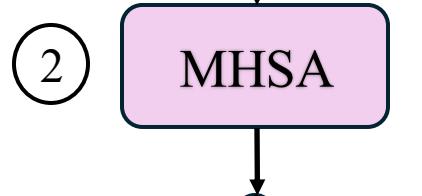
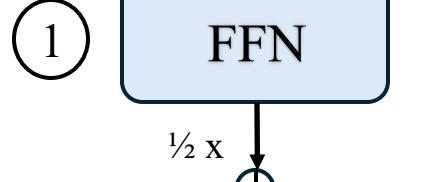
TDE

MARs Block

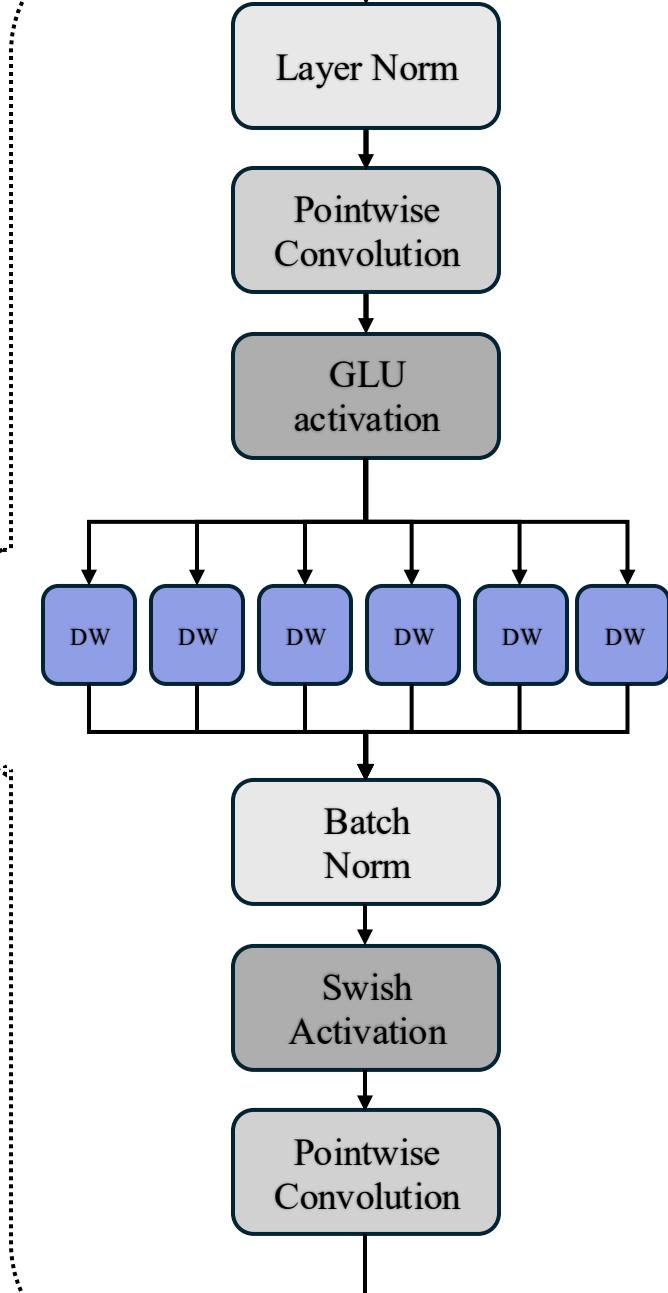
MARs Block

Prediction Head

$d = 1$   
 $\{0.2, 0.3, \dots, 0.7\}$   
 Scores



Architecture of Model



Multi-Scale Convolution