



(a) MDT

(b) MDM

(c) RSAS

RS-SA

Radial Stripe Self-attention

FFN

Feed Forward Network

LN

Layer Normalization

$\oplus$  Addition

ARPE

Angular Relative Position Encoding

DCConv

Deformable Convolution

Linear

Linear Projection

$\odot$  Concatenation