CNN ENCODER

CNN

Max-Pooling

+

Layer-1 Layer-3
Layer-2 Layer-4

TRANSFORMER ENCODER

Patch Embedding

feat4

Transformer Layer [3,6,9-Dual Attention Block]

⟨n=12⟩

Transformer Layer [3,6,9-Dual Attention Block]

feat3

feat2

Multi-Scale Skip Fusion

feat4

feat3

feat2

feat1

DECODER

(64, H, W)

(128, H/2, W/2)

(256, H/4, W/4)

(512, H/8, W/8)