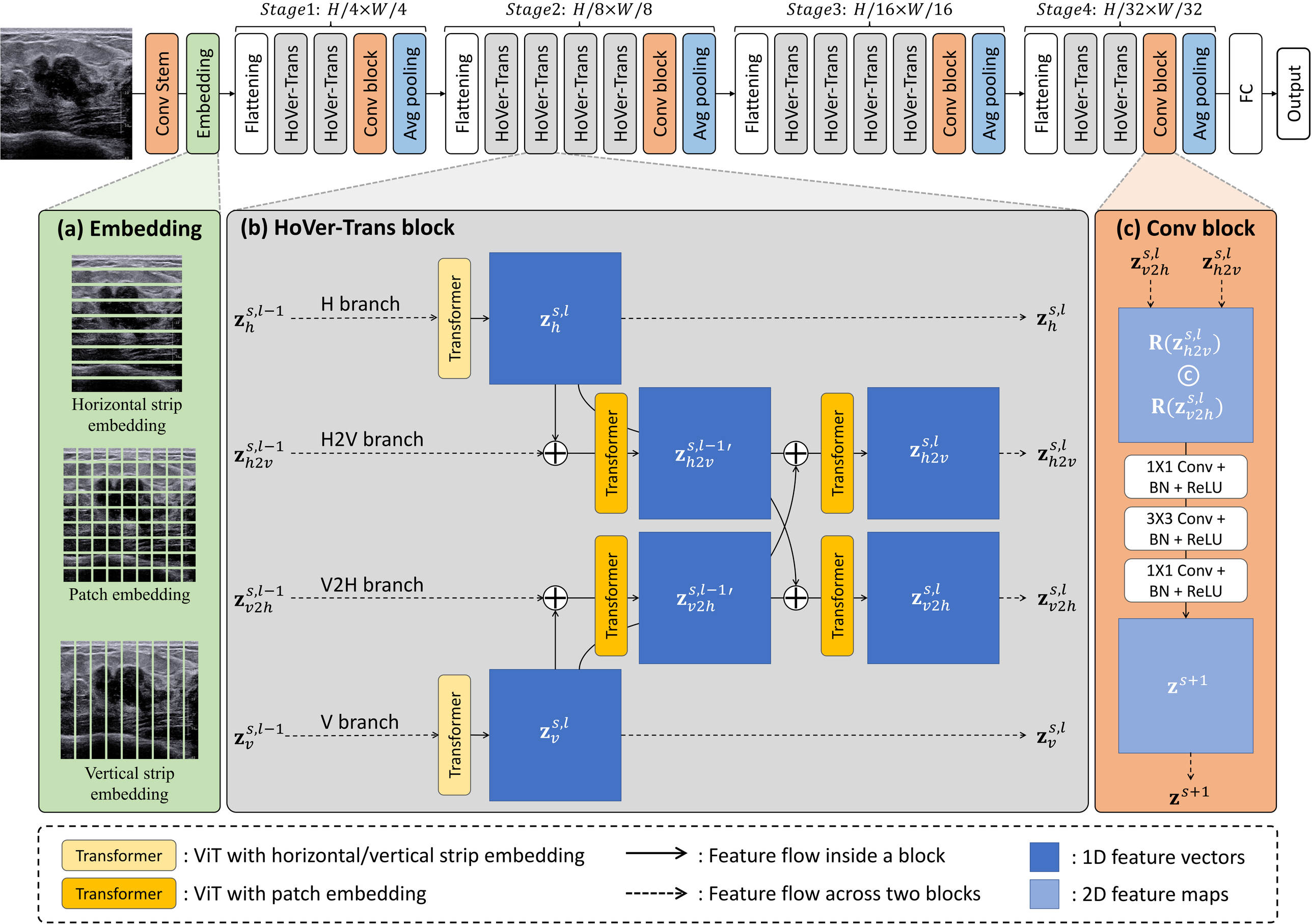
Stage1: H/4×W/4
Stage2: H/8×W/8
Stage3: H/16×W/16
Stage4: H/32×W/32
Conv Stem
Embedding
Flattening
HoVer-Trans
Conv block
Avg pooling
FC
Output
(a) Embedding
Horizontal strip embedding
Patch embedding
Vertical strip embedding
(b) HoVer-Trans block
H branch
H2V branch
V2H branch
V branch
Transformer
(c) Conv block
1X1 Conv + BN + ReLU
3X3 Conv + BN + ReLU
Transformer : ViT with horizontal/vertical strip embedding
Transformer : ViT with patch embedding
: Feature flow inside a block
: Feature flow across two blocks
: 1D feature vectors
: 2D feature maps