

Local-Window Self-Attention

FFN with 3×3 Depth-Wise Convolution