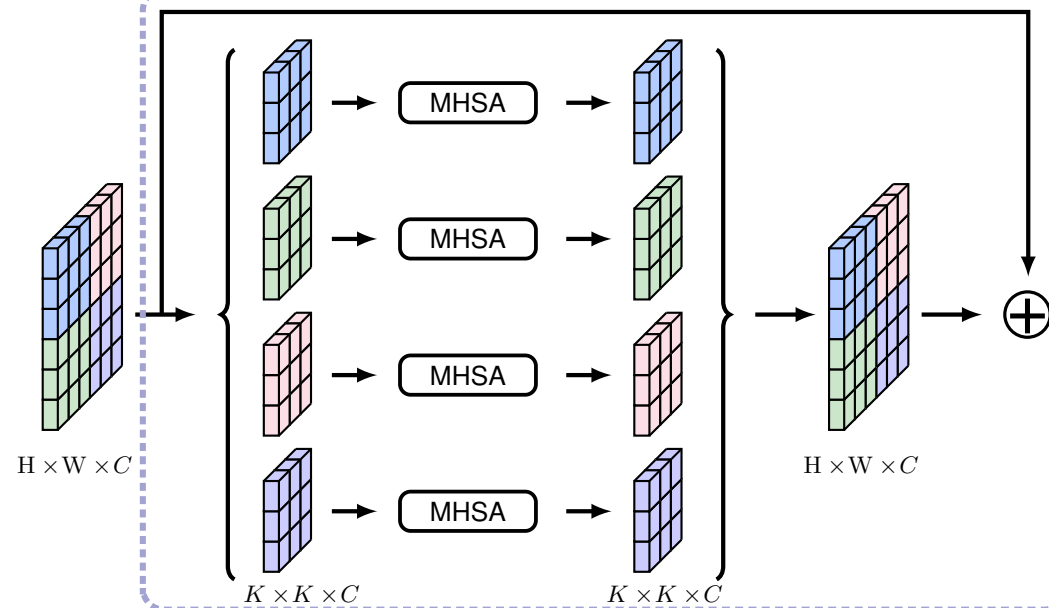


Local-Window Self-Attention

FFN with  $3 \times 3$  Depth-Wise Convolution