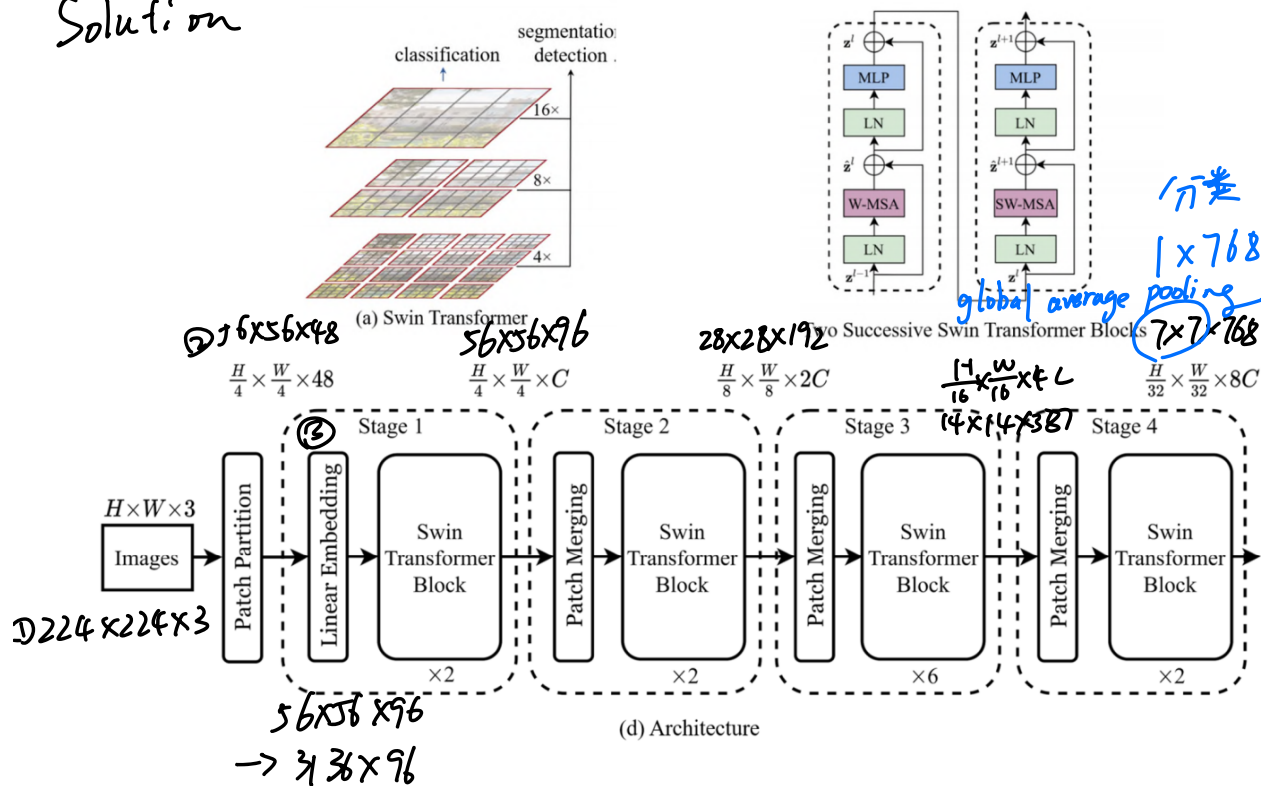
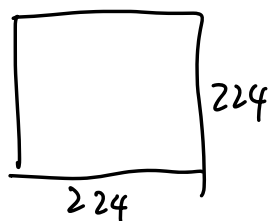


# Example 5.1.3.2 Swin Transform Q under stand

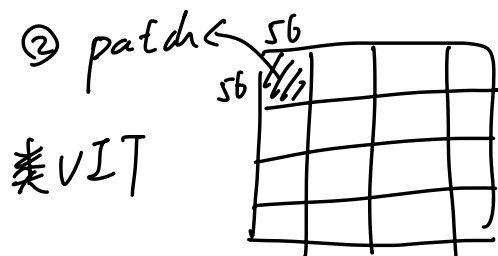
Solution



① 输入图



$H = 224$   
 $W = 224$   
RGB 3



$\frac{H}{4} : 56 \text{ pixel}$

$\frac{W}{4} : 56 \text{ pixel}$

$\frac{H}{4} \times \frac{W}{4} \times 48$   
1 patch pixel

$48 = 4 \times 4 \times 3$

4 patch 4 patch  $\rightarrow$  RGB

③  $56 \times 56 \times 96 \rightarrow 3136 \times 96$

$C=96$ , transform 要求 96 超参数

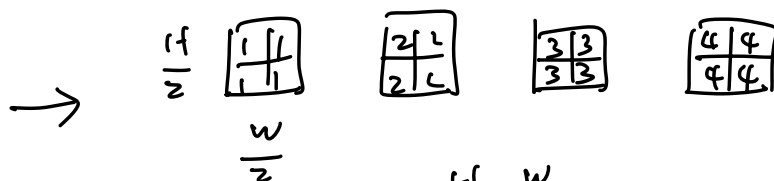
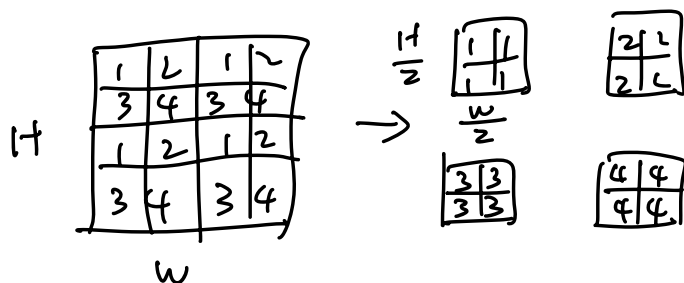
$56 \times 56 = 3136$



④ transformer

W-MSA 窗口注意力  $\xrightarrow{\text{shift}}$  SW-MSA  
移动窗口注意力

⑤  $H \times W \times C$



$1 \times 1$  卷积  
减少一半通道



$1 \times 1 \times C_{in} \times C_{out}$   
4 2

$\frac{H}{2} \times \frac{W}{2} \times 4C$

$W_1 = [w_1^{(1)} w_2^{(1)} w_3^{(1)} w_4^{(1)}]$

$W_2 = [w_1^{(2)} w_2^{(2)} w_3^{(2)} w_4^{(2)}]$

$\frac{H}{2} \times \frac{W}{2} \times 2C$

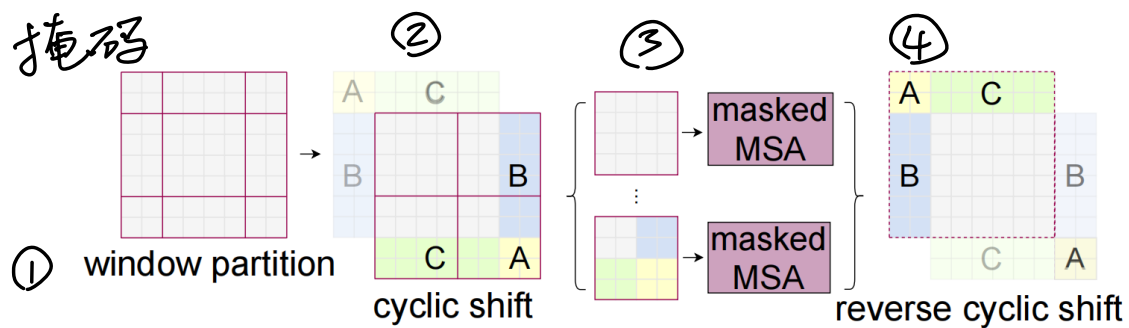
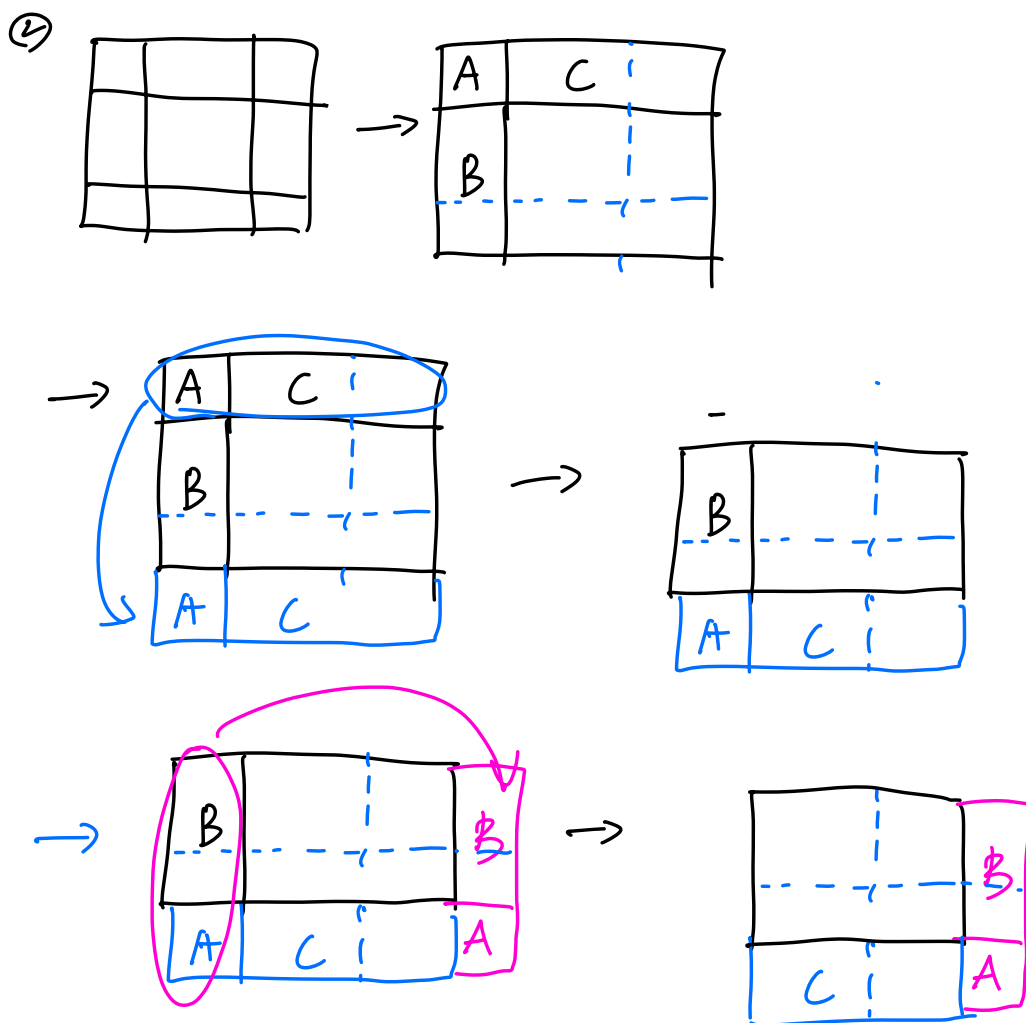


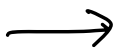
Figure 4. Illustration of an efficient batch computation approach for self-attention in shifted window partitioning.

①目标第9个窗口自注意力,但只算中次



All in All

1	2	3
4	5	6
7	8	9



1	2	3	
4	5	6	4
7	8	9	7
	2	3	1

	5	6	4
	8	9	7
	2	3	1



5	6	4
地 8	9	7
天 2	3	1

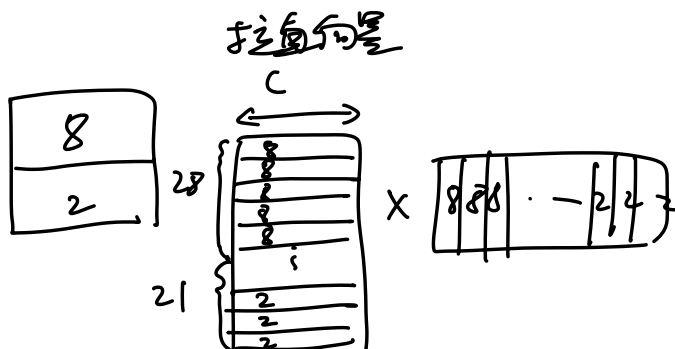
保持4次  
自注意力  
计算

③ 掩码 solve →  
Example 14 x 14 size

问题: 天地不应算自注意力

	5	6	4
4 {	地 8	9	7
3 {	天 2	3	1

patch size 7x7 向下取整



6	4
---	---

→

0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0

9	7
3	1

→

0	0	0	0	-100
0	0	0	0	-100
0	0	0	0	-100
0	0	0	0	-100

④ 归并

mask

88	88	82	82
88	88	82	82
28	28	22	22
28	28	22	22

0	-100
-100	0