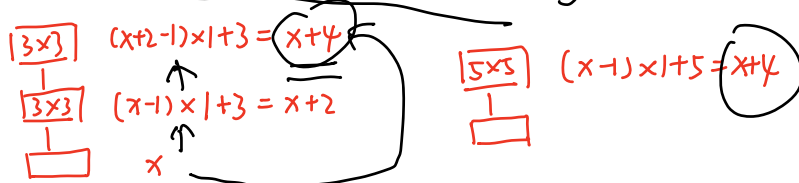


为什么要选择  $3 \times 3$  stride=1 的卷积?

① 2个  $3 \times 3$  卷积作用和 1个  $5 \times 5$  卷积 (stride=1) 作用一样 ✓

3个  $3 \times 3$  卷积作用和 1个  $7 \times 7$  卷积 (stride=1) 作用一样 ✓

proof 1. 感受野  $F(i) = [F(i+1) - 1] \times \text{stride} + \text{Kernel Size}$   
假设最后一层的感受野大小为  $x$



proof 2. 特征图 输入为  $224 \times 224$

$$\begin{aligned} & \boxed{3 \times 3} \left\lfloor \frac{224-3}{1} \right\rfloor + 1 = 222 \\ & \boxed{3 \times 3} \left\lfloor \frac{222-3}{1} \right\rfloor + 1 = 220 \end{aligned}$$

$$\boxed{5 \times 5} \left\lfloor \frac{224-5}{1} \right\rfloor + 1 = 220$$

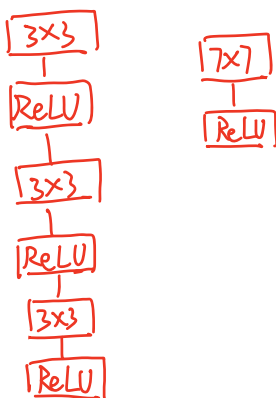
② 参数  $\text{Channels}_{in} = \text{Channels}_{out} = C$

$$2 \text{ 个 } 3 \times 3 \text{ 卷积: } C \times 3 \times 3 \times C + C \times 3 \times 3 \times C = 18C^2$$

$$1 \text{ 个 } 5 \times 5 \text{ 卷积: } C \times 5 \times 5 \times C = 25C^2$$

③ 非线性

relu



3个  $3 \times 3$  的卷积比 1个  $7 \times 7$  的卷积有更多的非线性!

(决策函数更具辨别力)

④  $3 \times 3$  是最小的能够捕获像素八邻域信息的尺寸

