Deep Learning Basics

Lecture 4: Convolutional Neural Networks

최성준 (고려대학교 인공지능학과)

Convolution

- Continuous convolution

$$(f * g)(t) = \int f(\tau)g(t-\tau)d\tau = \int f(t-\tau)g(t)d\tau$$

- Discrete convolution

$$(f * g)(t) = \sum_{i=-\infty}^{\infty} f(i)g(t-i) = \sum_{i=-\infty}^{\infty} f(t-i)g(i)$$

- 2D image convolution

$$(I * K)(i,j) = \sum_m \sum_n I(m,n)K(i-m, j-n) = \sum_m \sum_n I(i-m, i-n)K(m,n)$$

Convolution

K (3x3 filter)

K_{11}	K_{12}	K_{13}
K_{21}	K_{22}	K_{23}
K_{31}	K_{32}	K_{33}

$*$

I (7x7 image)

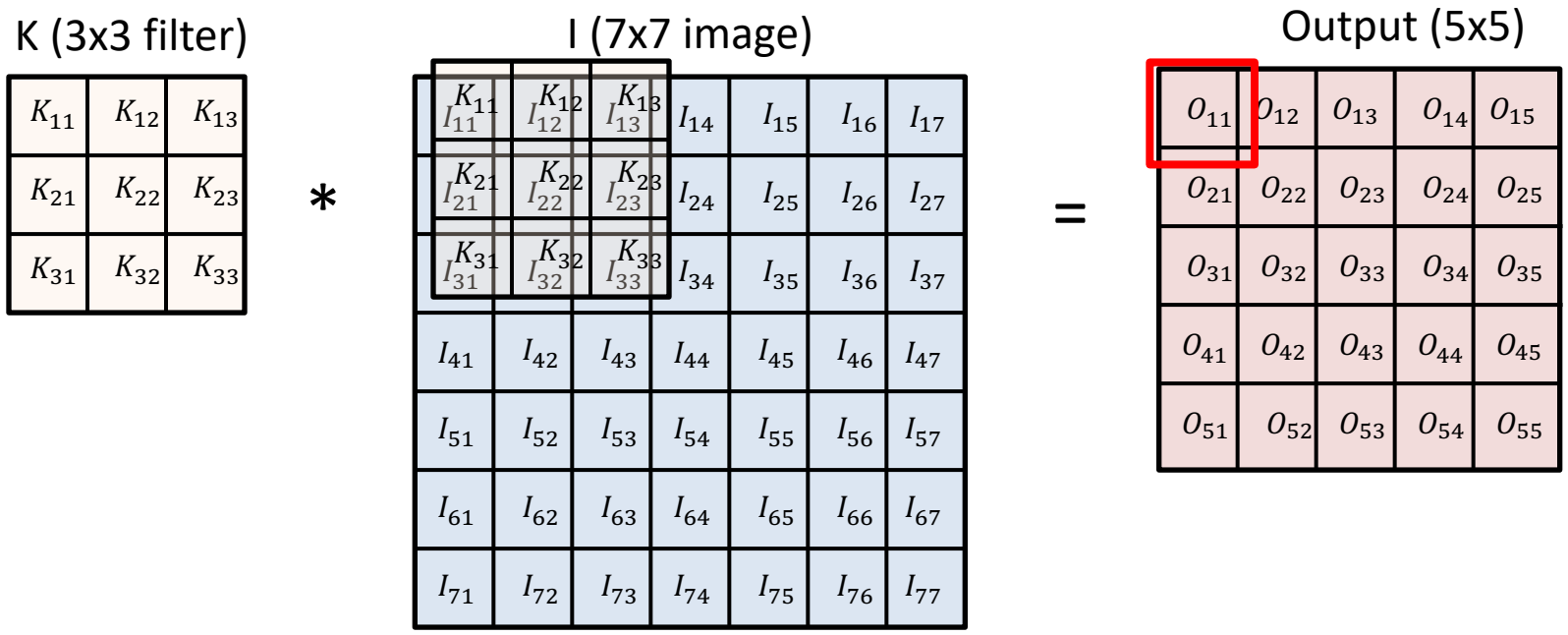
I_{11}	I_{12}	I_{13}	I_{14}	I_{15}	I_{16}	I_{17}
I_{21}	I_{22}	I_{23}	I_{24}	I_{25}	I_{26}	I_{27}
I_{31}	I_{32}	I_{33}	I_{34}	I_{35}	I_{36}	I_{37}
I_{41}	I_{42}	I_{43}	I_{44}	I_{45}	I_{46}	I_{47}
I_{51}	I_{52}	I_{53}	I_{54}	I_{55}	I_{56}	I_{57}
I_{61}	I_{62}	I_{63}	I_{64}	I_{65}	I_{66}	I_{67}
I_{71}	I_{72}	I_{73}	I_{74}	I_{75}	I_{76}	I_{77}

$=$

Output (5x5)

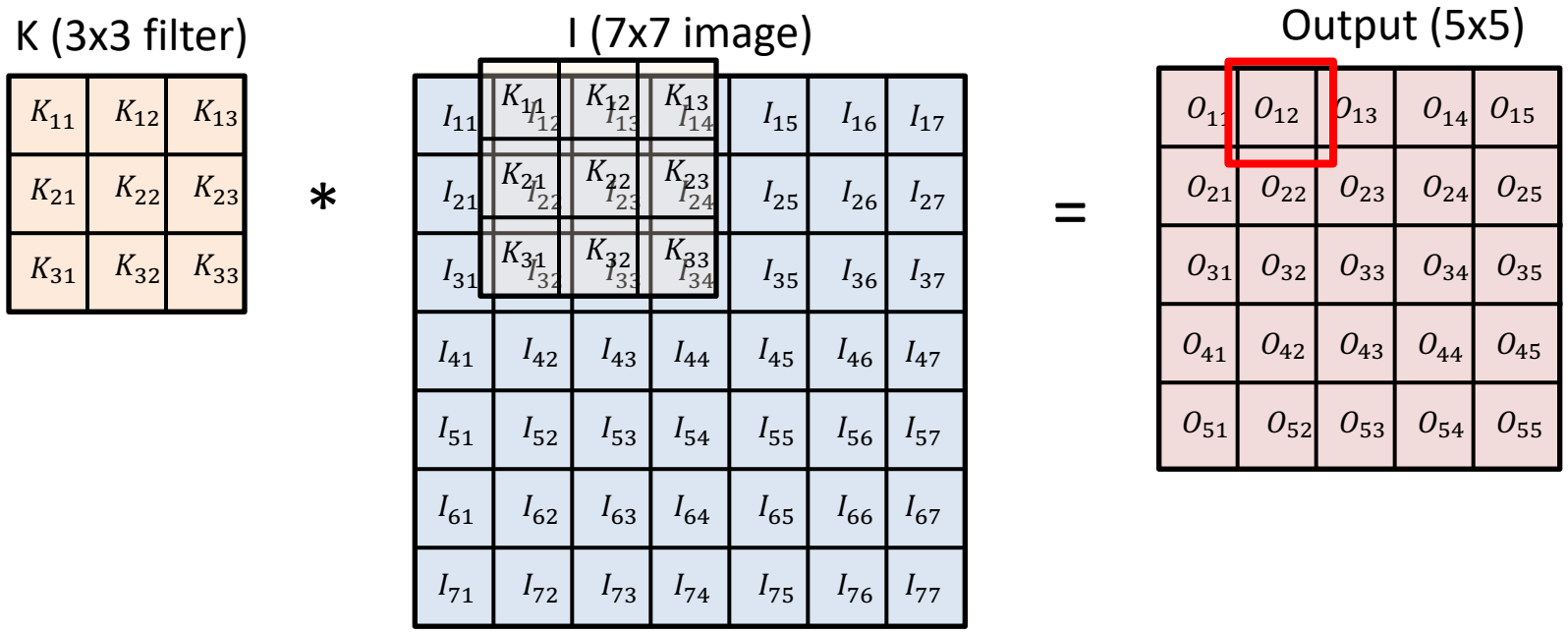
O_{11}	O_{12}	O_{13}	O_{14}	O_{15}
O_{21}	O_{22}	O_{23}	O_{24}	O_{25}
O_{31}	O_{32}	O_{33}	O_{34}	O_{35}
O_{41}	O_{42}	O_{43}	O_{44}	O_{45}
O_{51}	O_{52}	O_{53}	O_{54}	O_{55}

Convolution



$$O_{11} = I_{11}K_{11}+I_{12}K_{12}+I_{13}K_{13}+I_{21}K_{21}+I_{22}K_{22}+I_{23}K_{23}+I_{31}K_{31}+I_{32}K_{32}+I_{33}K_{33}+bias$$

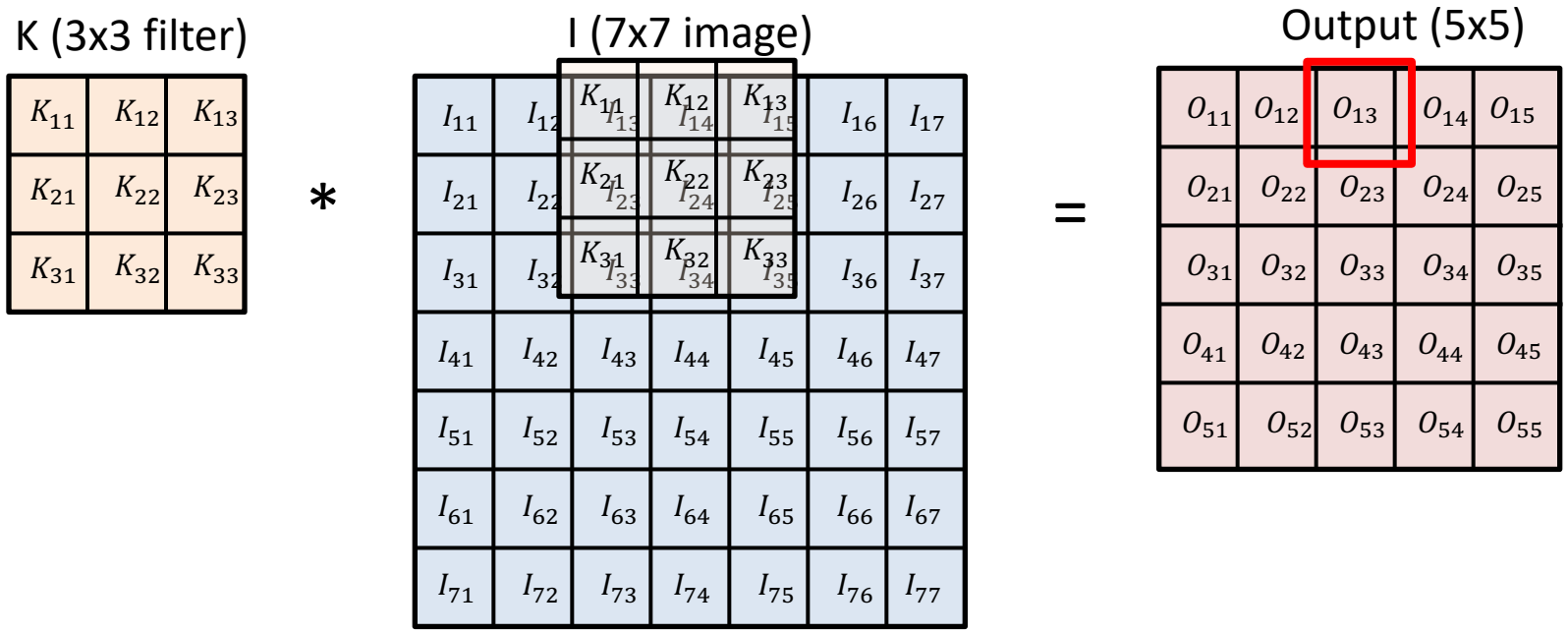
Convolution



$$O_{11} = I_{11}K_{11}+I_{12}K_{12}+I_{13}K_{13}+I_{21}K_{21}+I_{22}K_{22}+I_{23}K_{23}+I_{31}K_{31}+I_{32}K_{32}+I_{33}K_{33}+bias$$

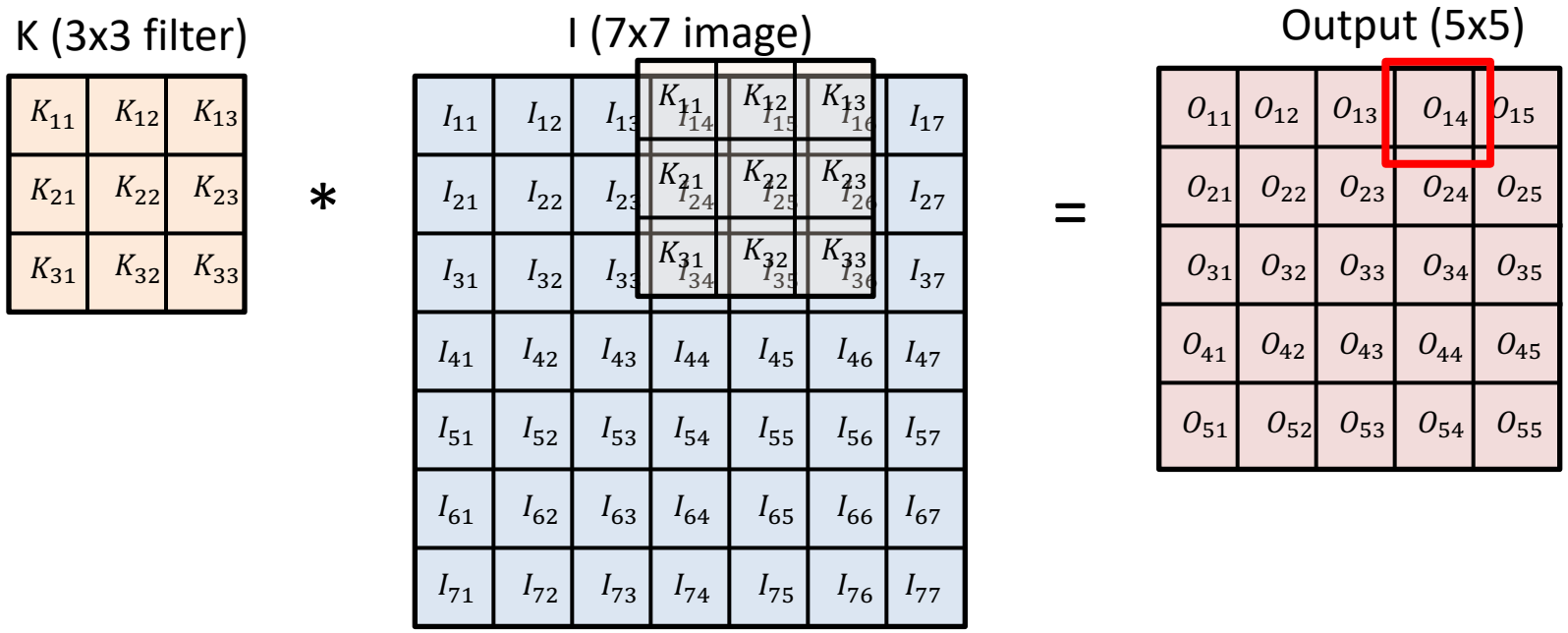
$$O_{12} = I_{12}K_{11}+I_{13}K_{12}+I_{14}K_{13}+I_{22}K_{21}+I_{23}K_{22}+I_{24}K_{23}+I_{32}K_{31}+I_{33}K_{32}+I_{34}K_{33}+bias$$

Convolution



$$O_{11} = I_{11}K_{11}+I_{12}K_{12}+I_{13}K_{13}+I_{21}K_{21}+I_{22}K_{22}+I_{23}K_{23}+I_{31}K_{31}+I_{32}K_{32}+I_{33}K_{33}+bias$$
$$O_{12} = I_{12}K_{11}+I_{13}K_{12}+I_{14}K_{13}+I_{22}K_{21}+I_{23}K_{22}+I_{24}K_{23}+I_{32}K_{31}+I_{33}K_{32}+I_{34}K_{33}+bias$$
$$O_{13} = I_{13}K_{11}+I_{14}K_{12}+I_{15}K_{13}+I_{23}K_{21}+I_{24}K_{22}+I_{25}K_{23}+I_{33}K_{31}+I_{34}K_{32}+I_{35}K_{33}+bias$$

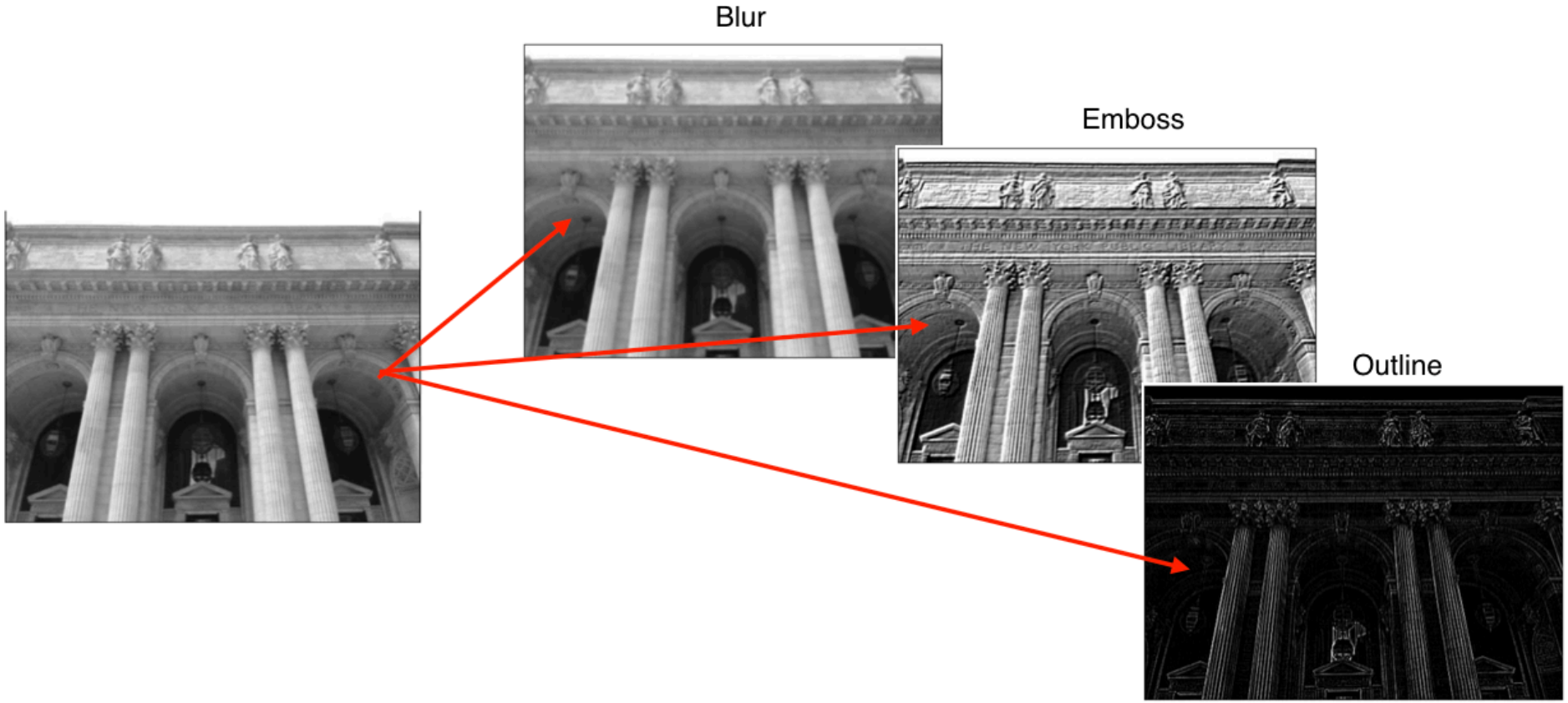
Convolution



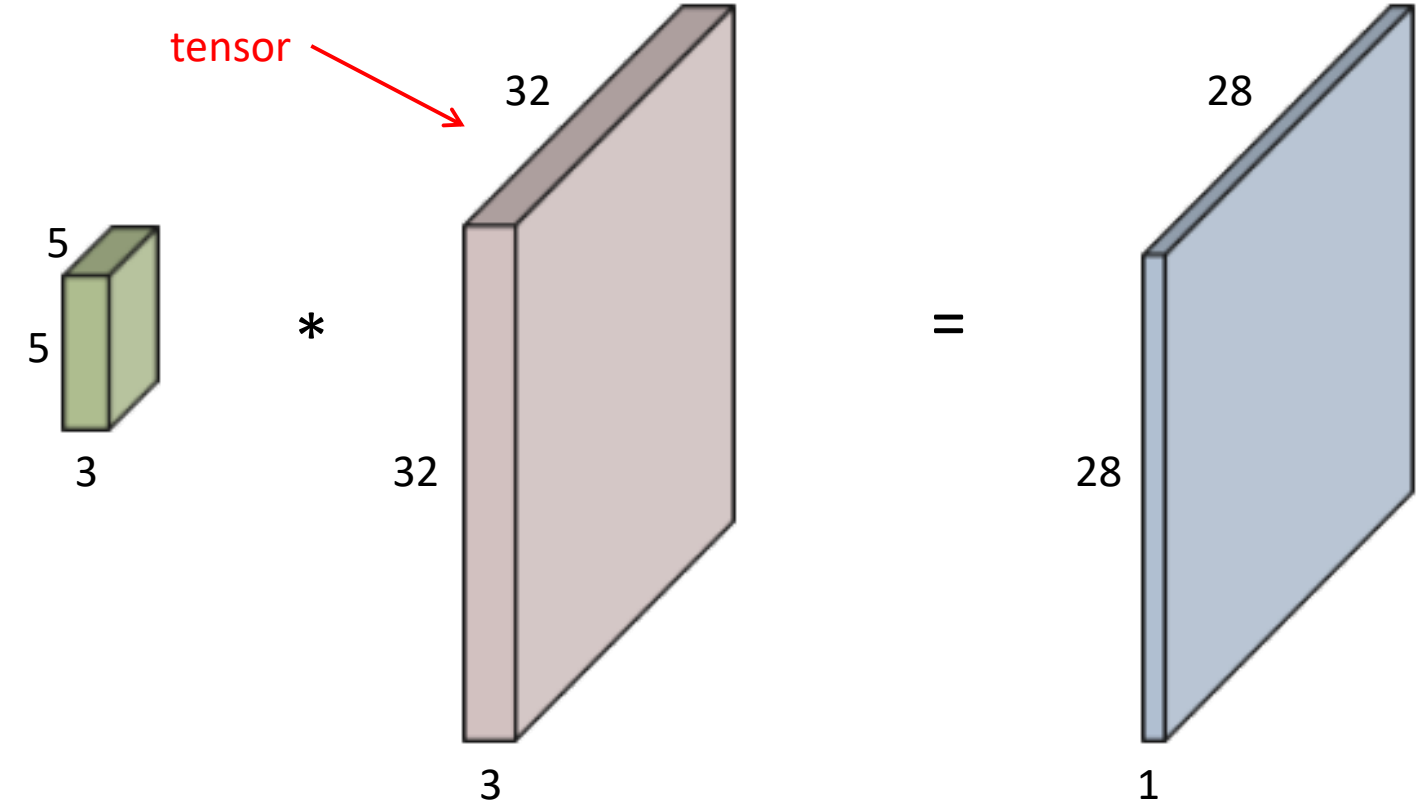
$$O_{11} = I_{11}K_{11}+I_{12}K_{12}+I_{13}K_{13}+I_{21}K_{21}+I_{22}K_{22}+I_{23}K_{23}+I_{31}K_{31}+I_{32}K_{32}+I_{33}K_{33}+bias$$
$$O_{12} = I_{12}K_{11}+I_{13}K_{12}+I_{14}K_{13}+I_{22}K_{21}+I_{23}K_{22}+I_{24}K_{23}+I_{32}K_{31}+I_{33}K_{32}+I_{34}K_{33}+bias$$
$$O_{13} = I_{13}K_{11}+I_{14}K_{12}+I_{15}K_{13}+I_{23}K_{21}+I_{24}K_{22}+I_{25}K_{23}+I_{33}K_{31}+I_{34}K_{32}+I_{35}K_{33}+bias$$
$$O_{14} = I_{14}K_{11}+I_{15}K_{12}+I_{16}K_{13}+I_{24}K_{21}+I_{25}K_{22}+I_{26}K_{23}+I_{34}K_{31}+I_{35}K_{32}+I_{36}K_{33}+bias$$

Convolution

- 2D convolution in action



RGB Image Convolution

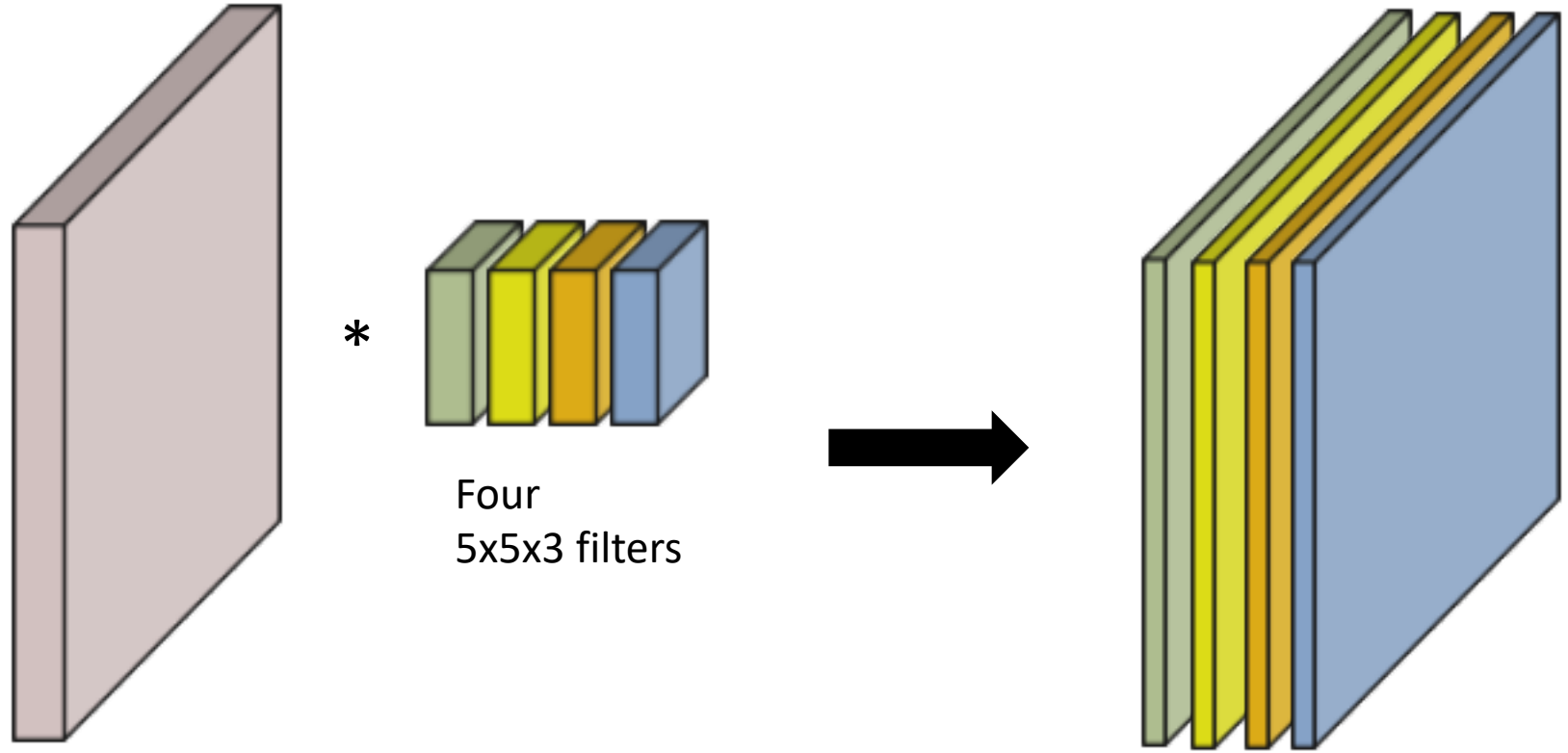


5x5x3 filter

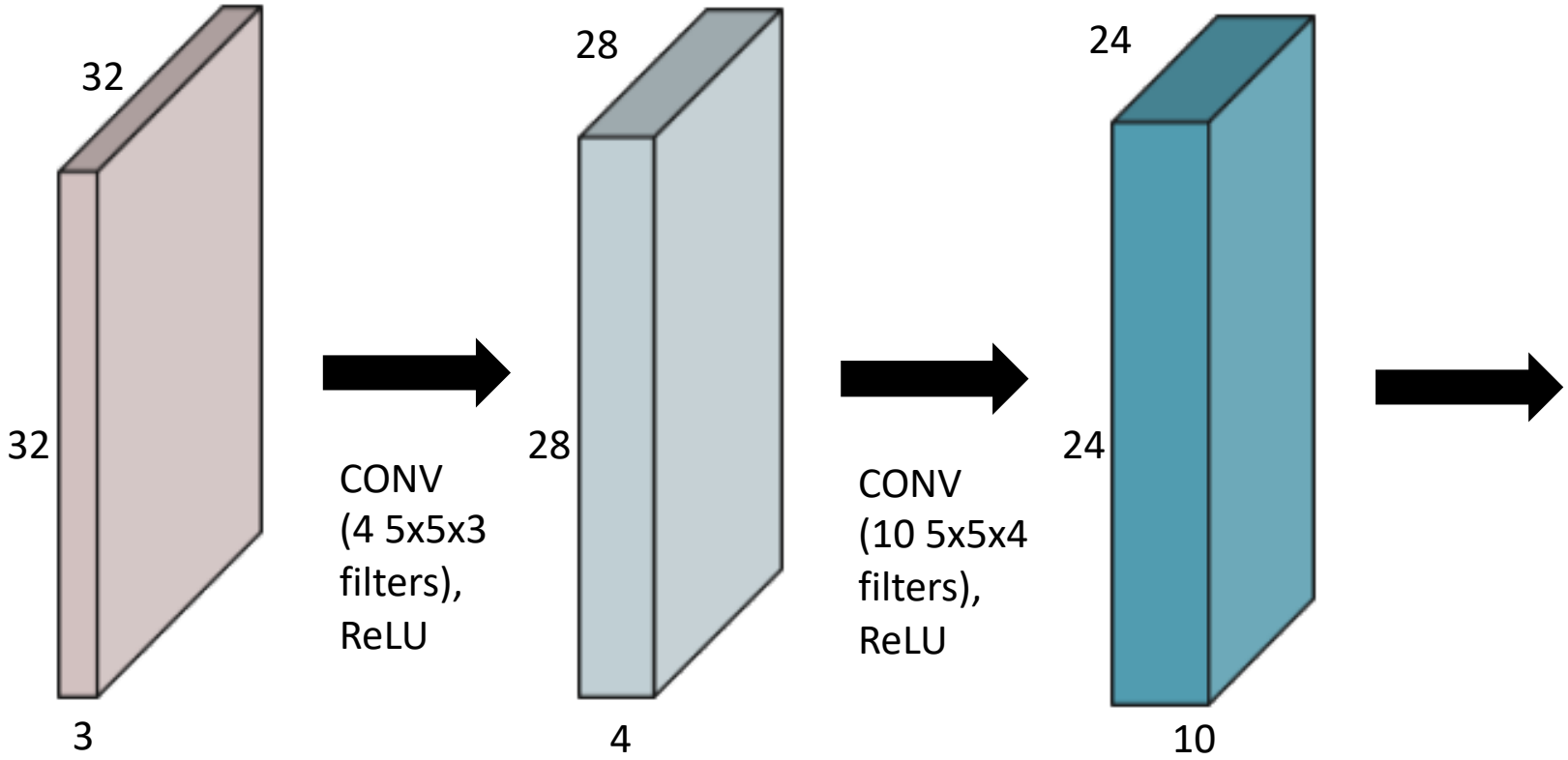
32x32x3 Image

28x28x1 feature

RGB Image Convolution

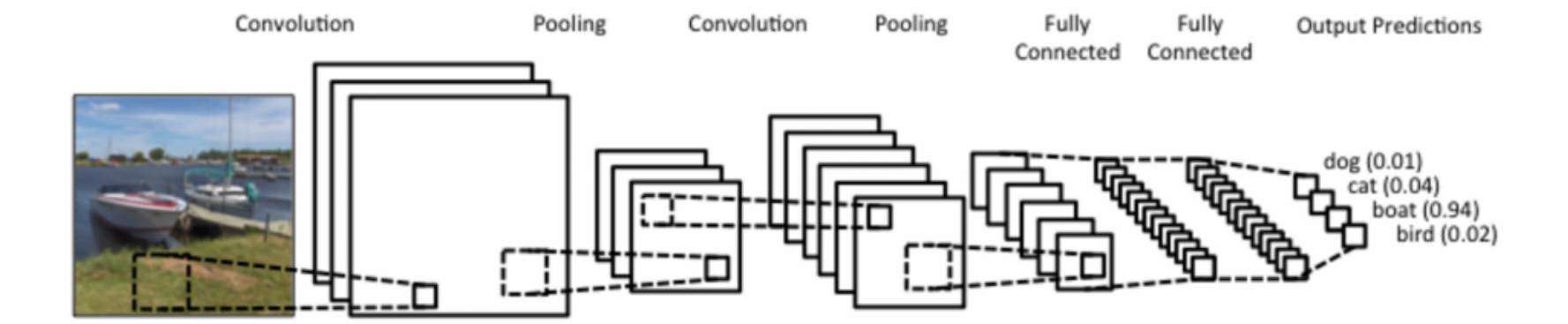


Stack of Convolutions



Convolutional Neural Networks

- CNN consists of convolution layer, pooling layer, and fully connected layer.
  - Convolution and pooling layers: feature extraction
  - Fully connected layer: decision making (e.g., classification)

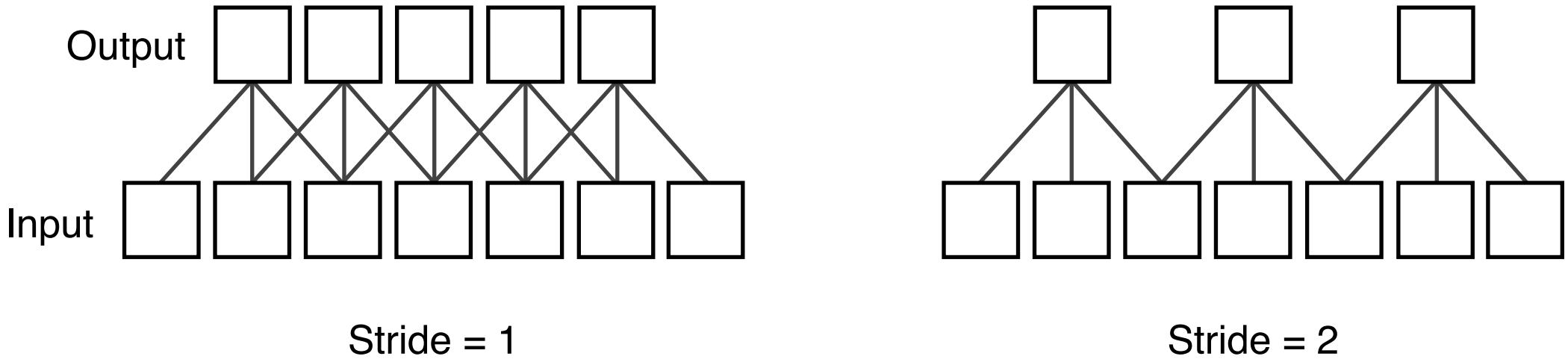


Convolution Arithmetic (of GoogLeNet)

Stride Channel #param

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

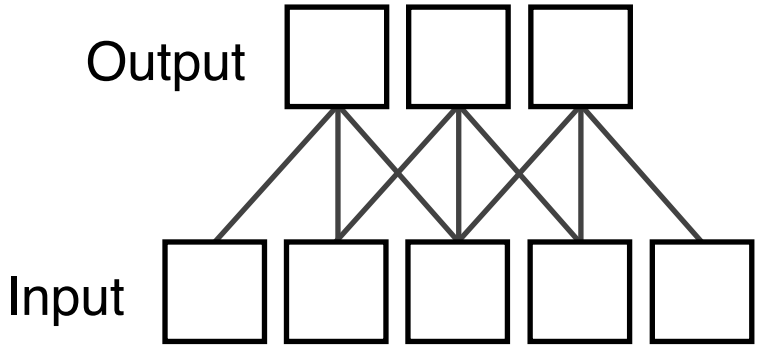
Stride



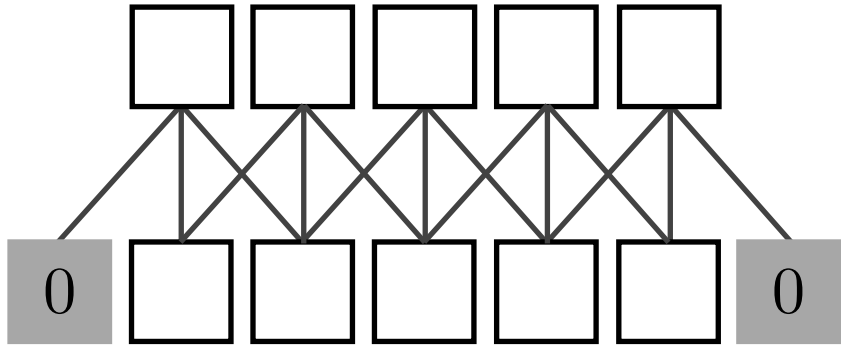
Stride = 1

Stride = 2

Padding

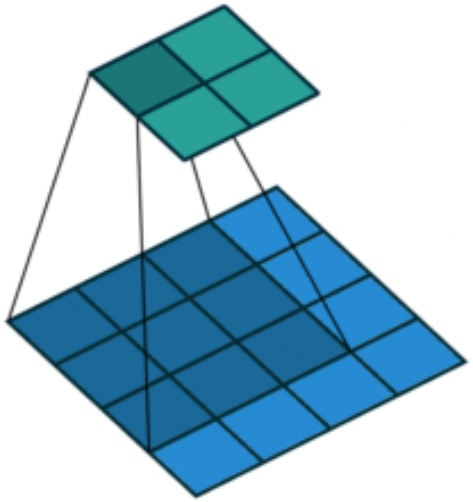


No padding (stride=1)

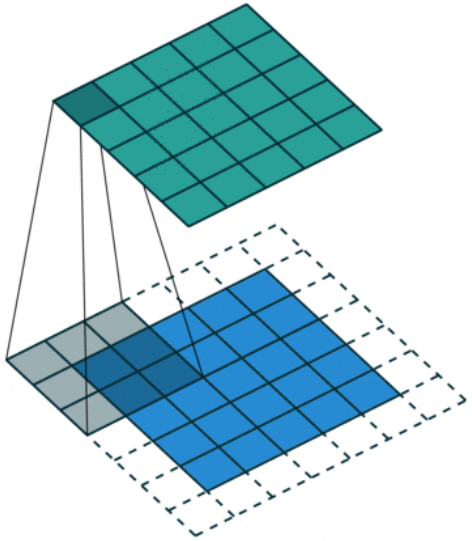


Zero padding (stride=1)

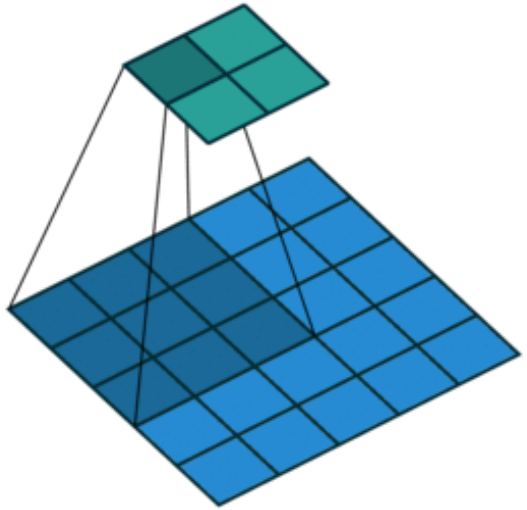
Stride? Padding?



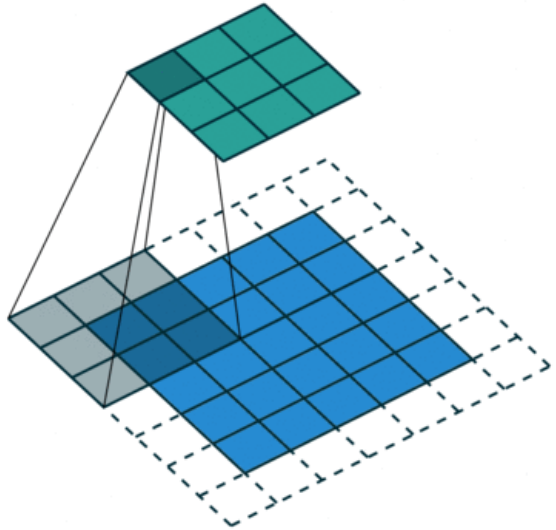
Stride (1)
Padding (0)



Stride (1)
Padding (1)



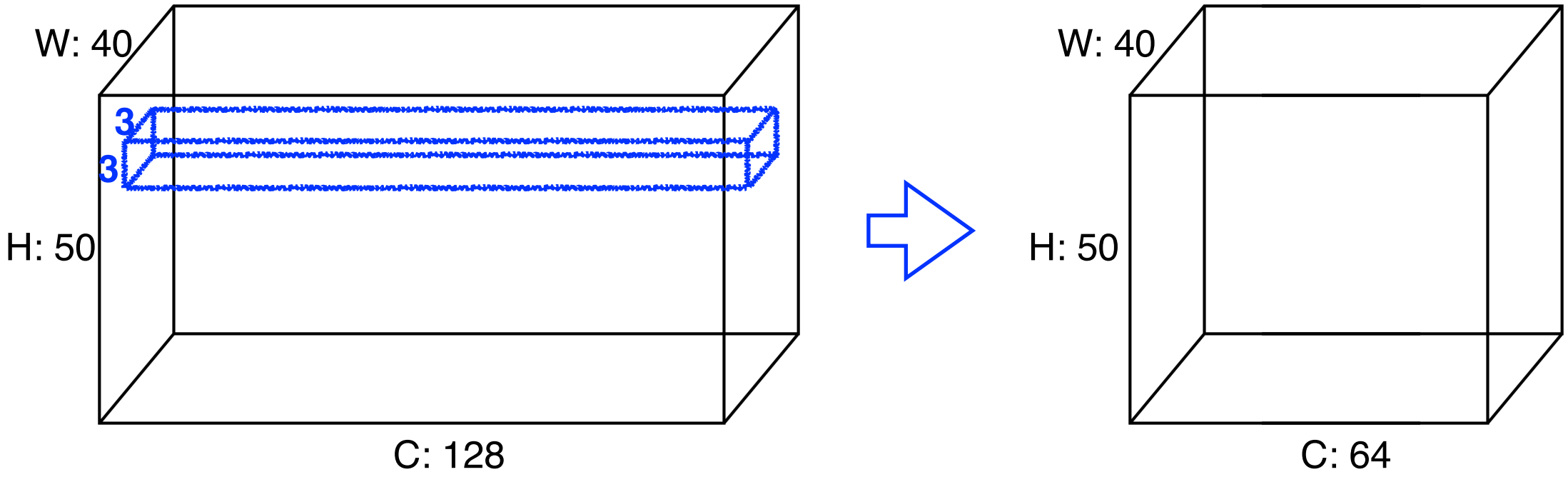
Stride (2)
Padding (0)



Stride (2)
Padding (1)

Convolution Arithmetic

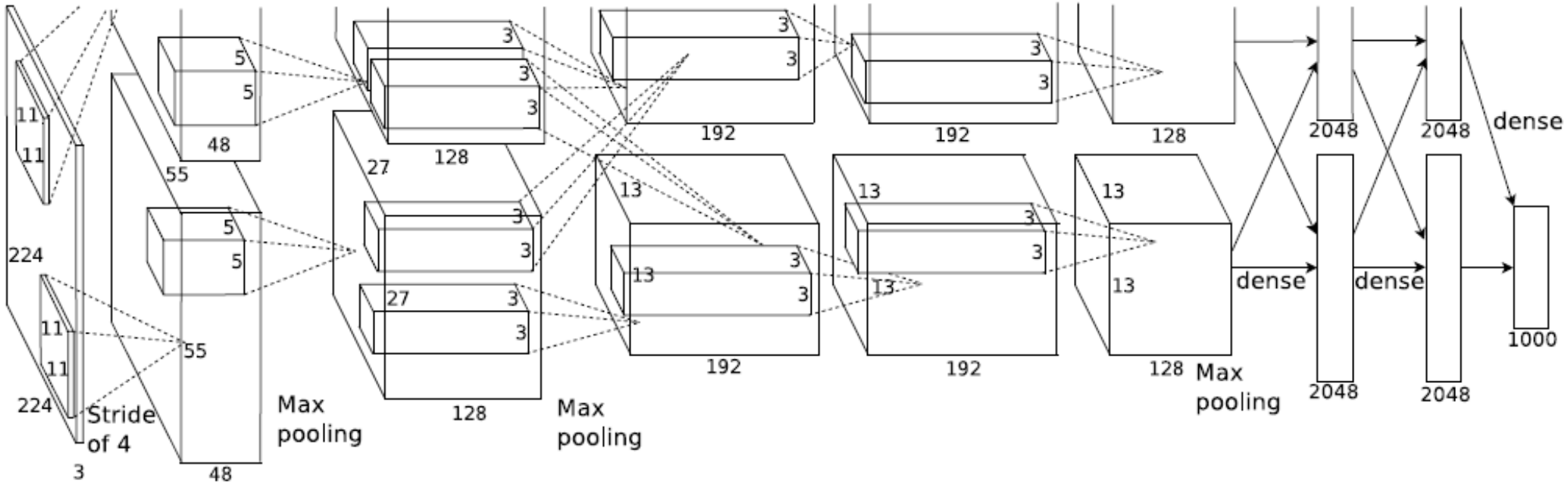
- Padding (1), Stride (1), 3×3 Kernel



What is the **number of parameters** of this model?

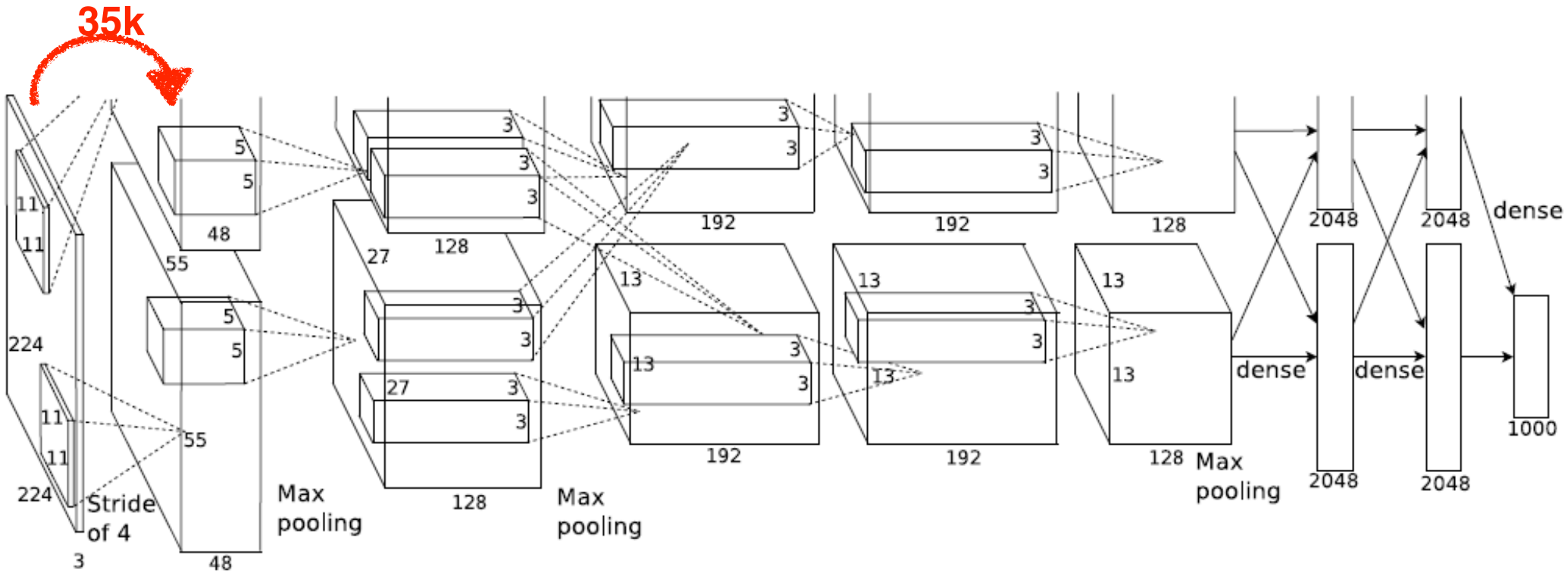
The answer is $3 \times 3 \times 128 \times 64 = 73{,}728$

Exercise



What is the **number of parameters** of this model?

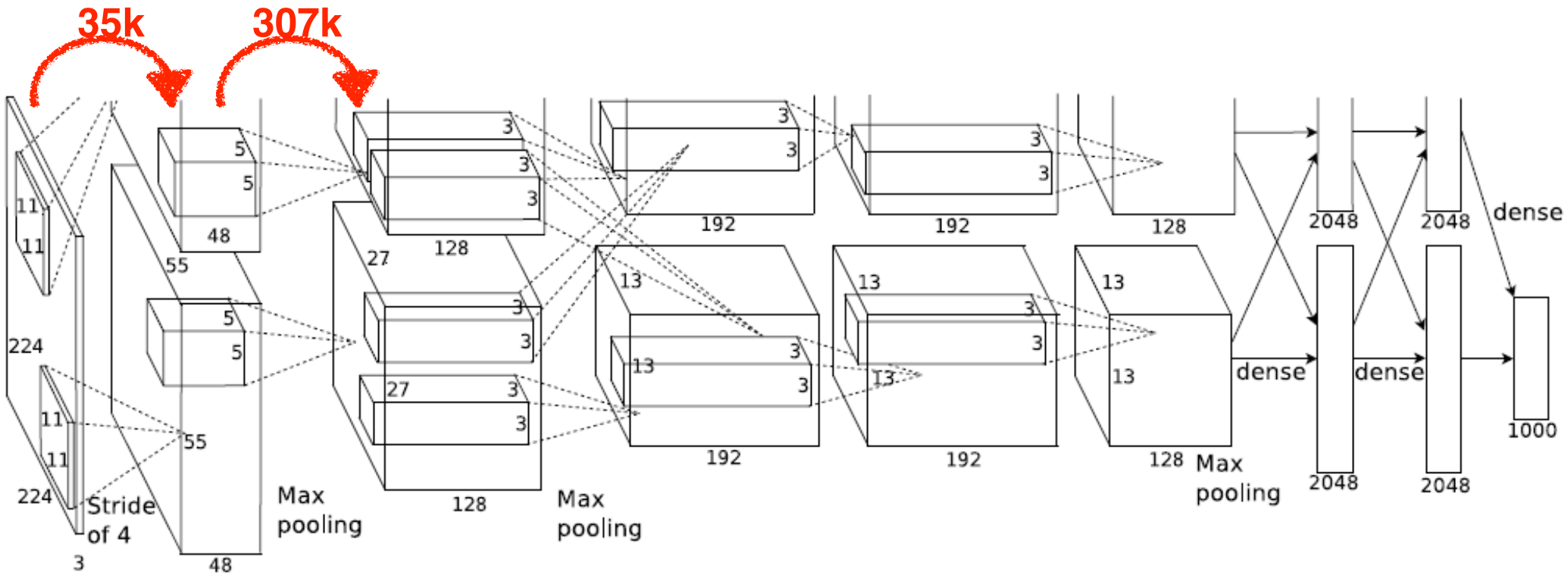
Exercise



$$11 \times 11 \times 3 \times 48 * 2 \approx 35k$$

What is the **number of parameters** of this model?

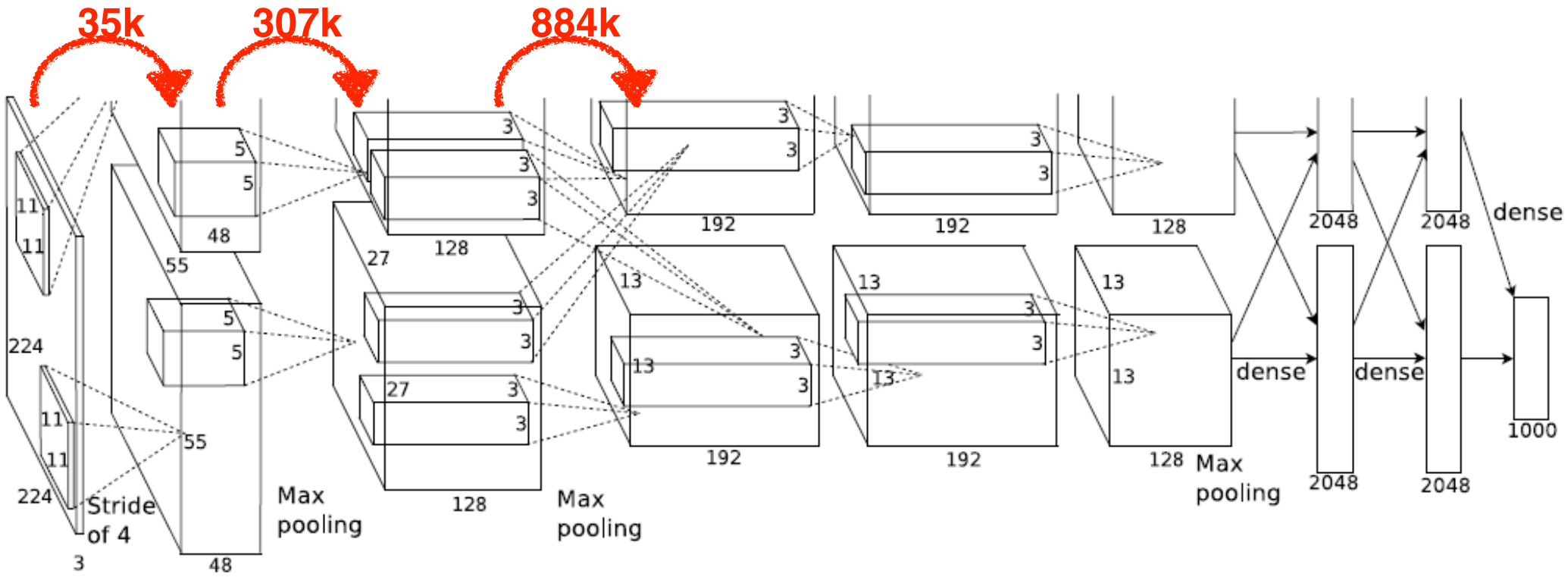
Exercise



$$5 \times 5 \times 48 \times 128 * 2 \approx 307k$$

What is the **number of parameters** of this model?

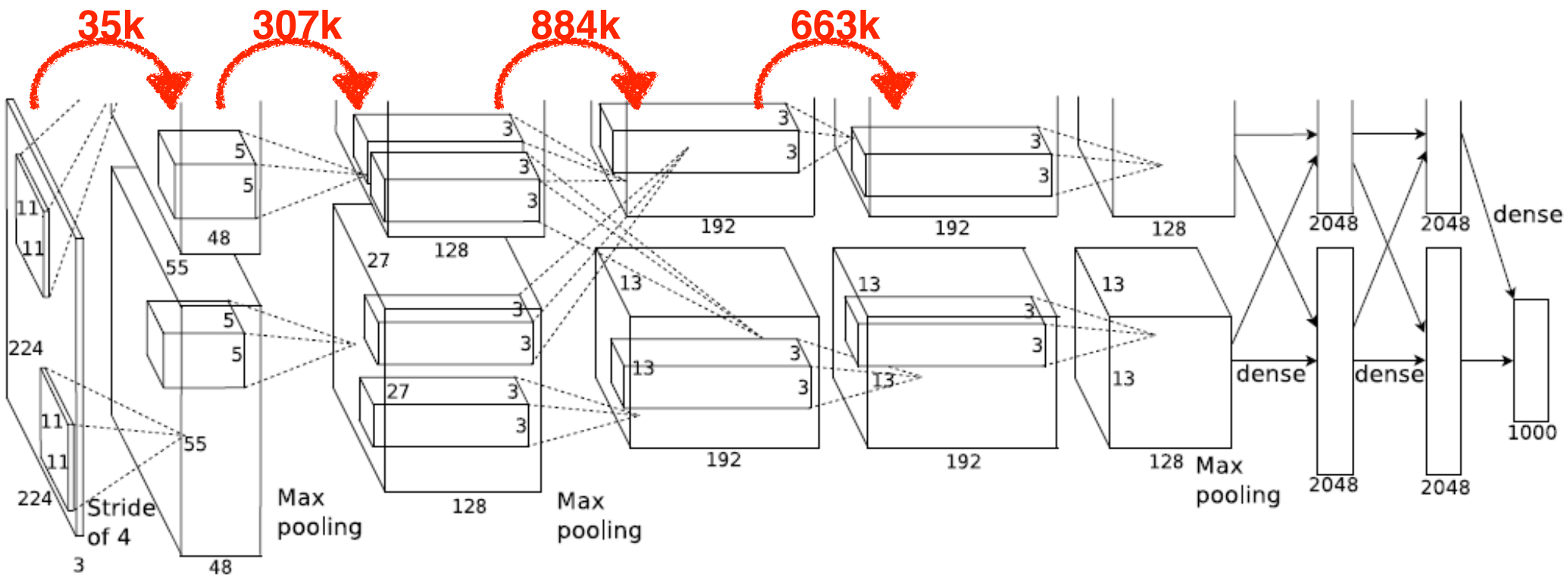
Exercise



$$3 \times 3 \times 128 * 2 \times 192 * 2 \approx 884k$$

What is the **number of parameters** of this model?

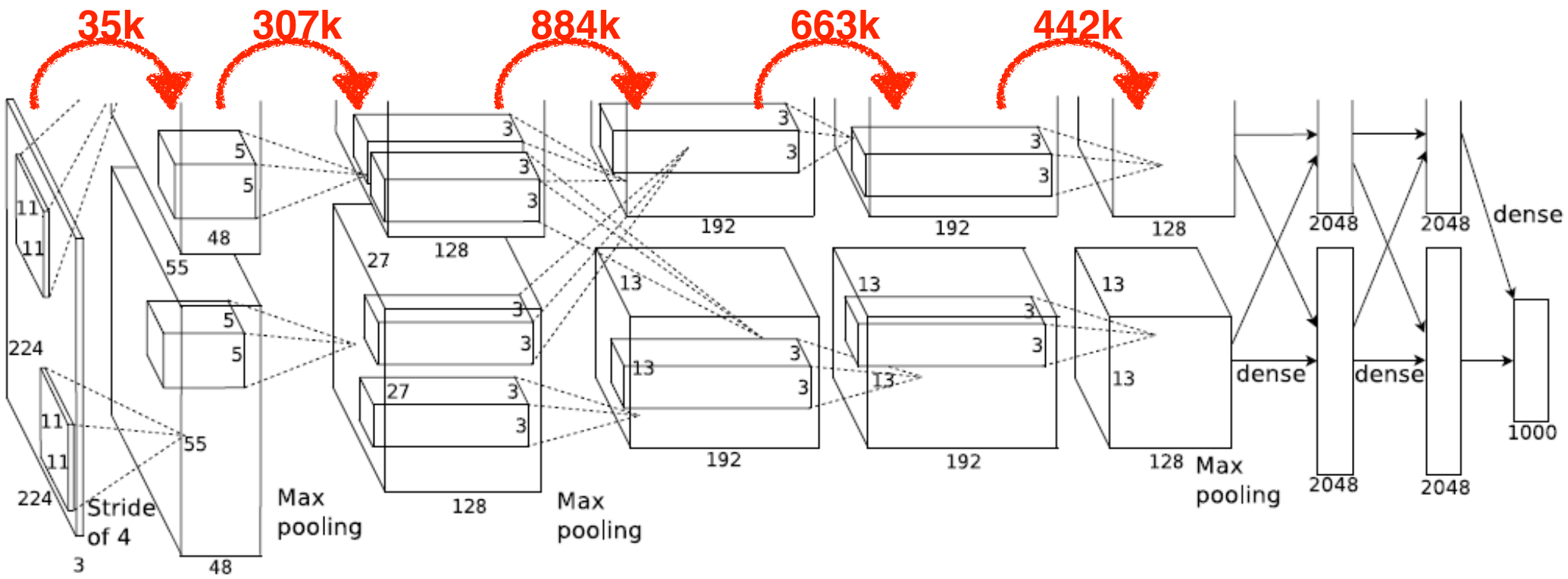
Exercise



$$3 \times 3 \times 192 \times 192 * 2 \approx 663k$$

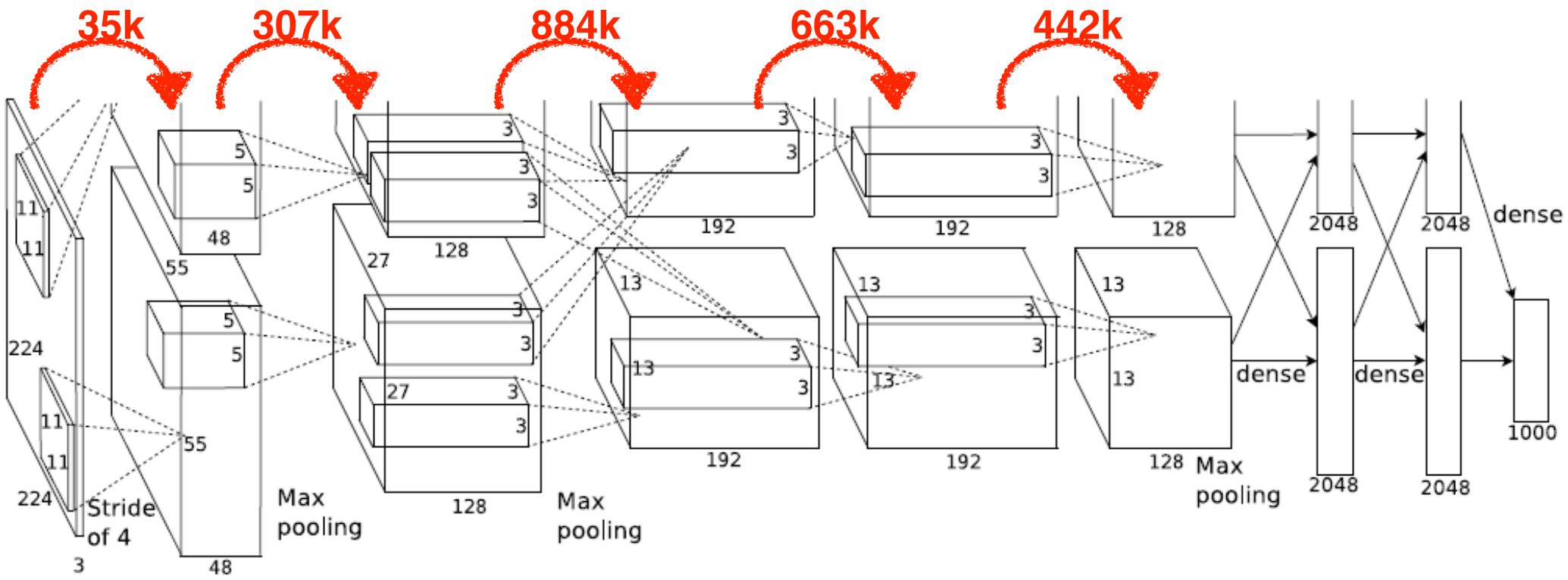
What is the **number of parameters** of this model?

Exercise



$$3 \times 3 \times 192 \times 128 * 2 \approx 442k$$

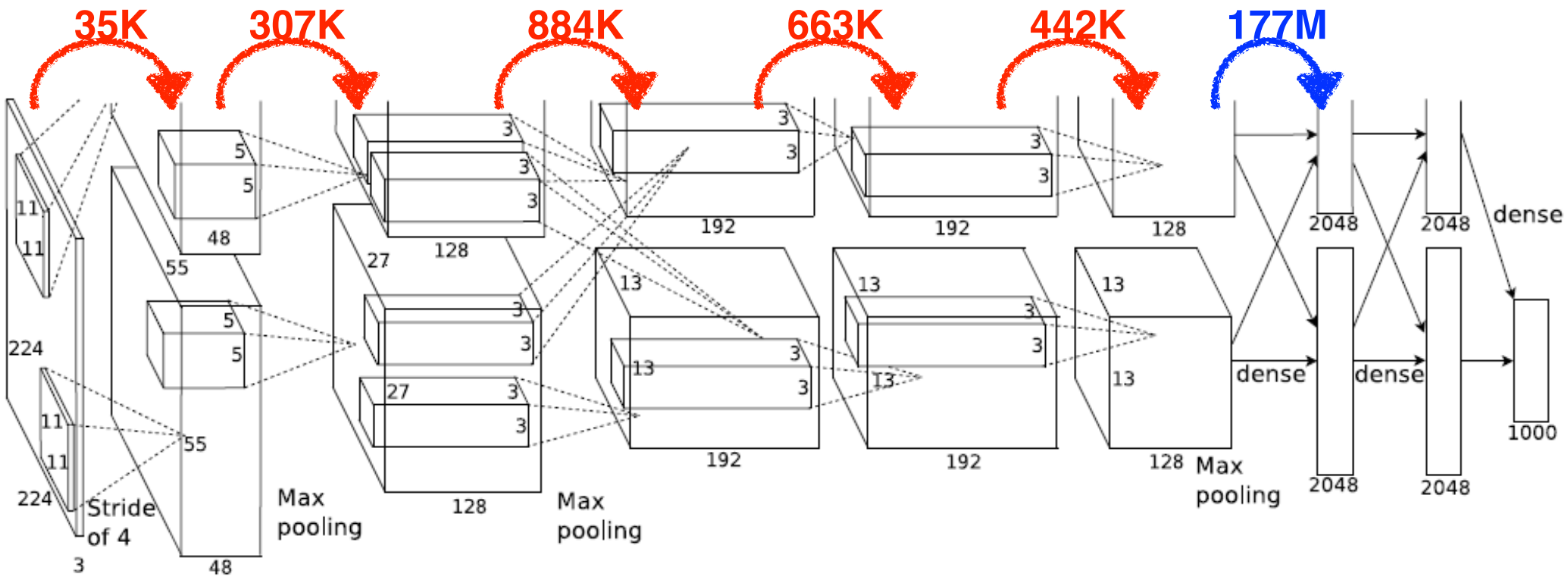
Exercise



$$3 \times 3 \times 192 \times 128 * 2 \approx 442k$$

What is the **number of parameters** of this model?

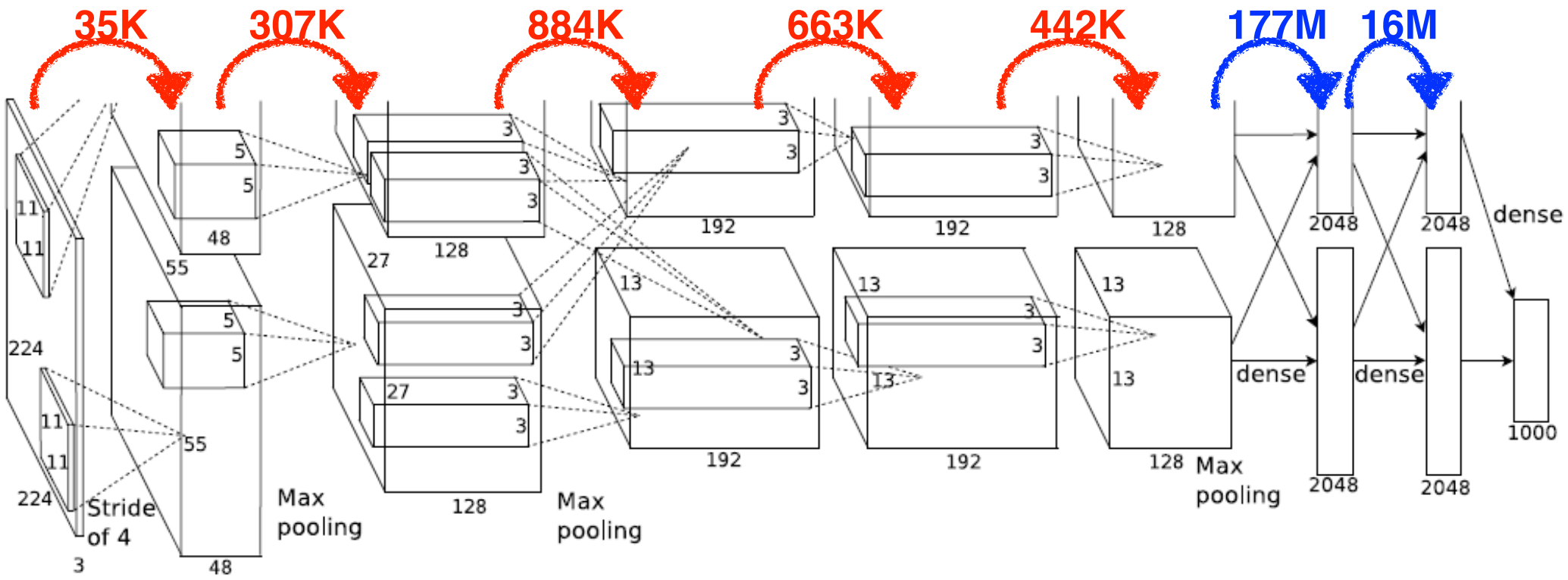
Exercise



$$13 * 13 * 128 * 2 \times 2048 * 2 \approx 177M$$

What is the **number of parameters** of this model?

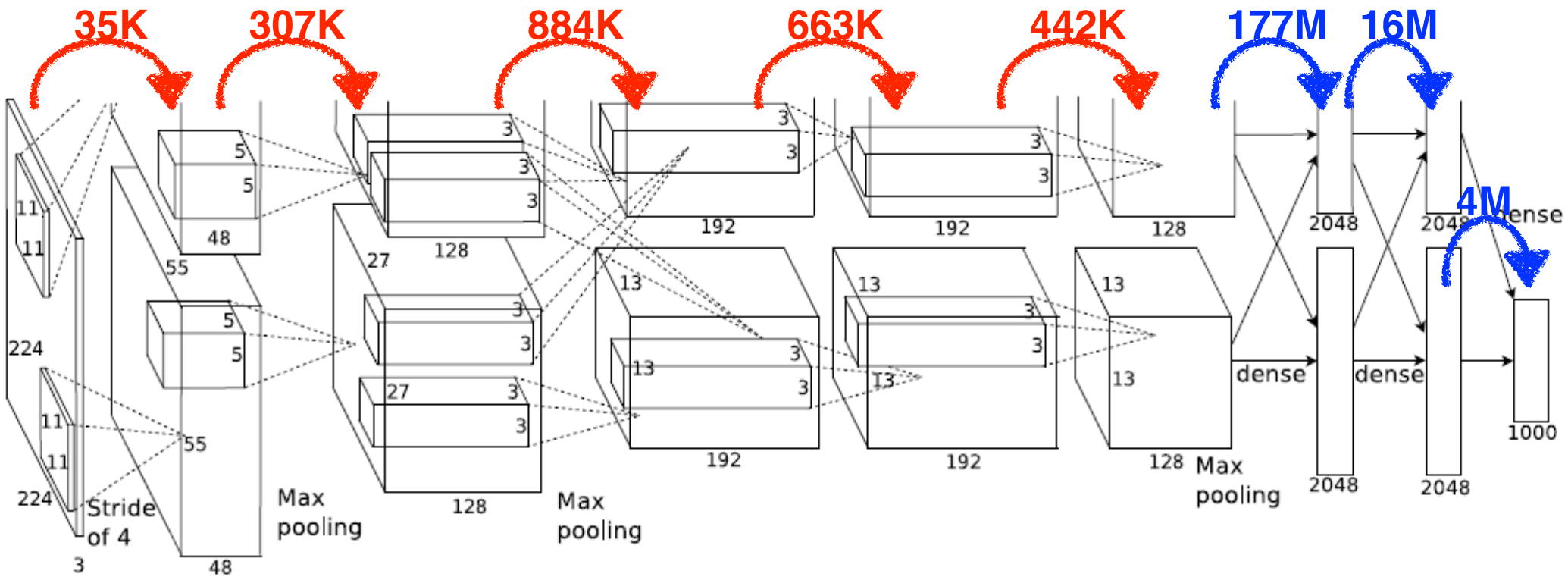
Exercise



$$2048 * 2 \times 2048 * 2 \approx 16M$$

What is the **number of parameters** of this model?

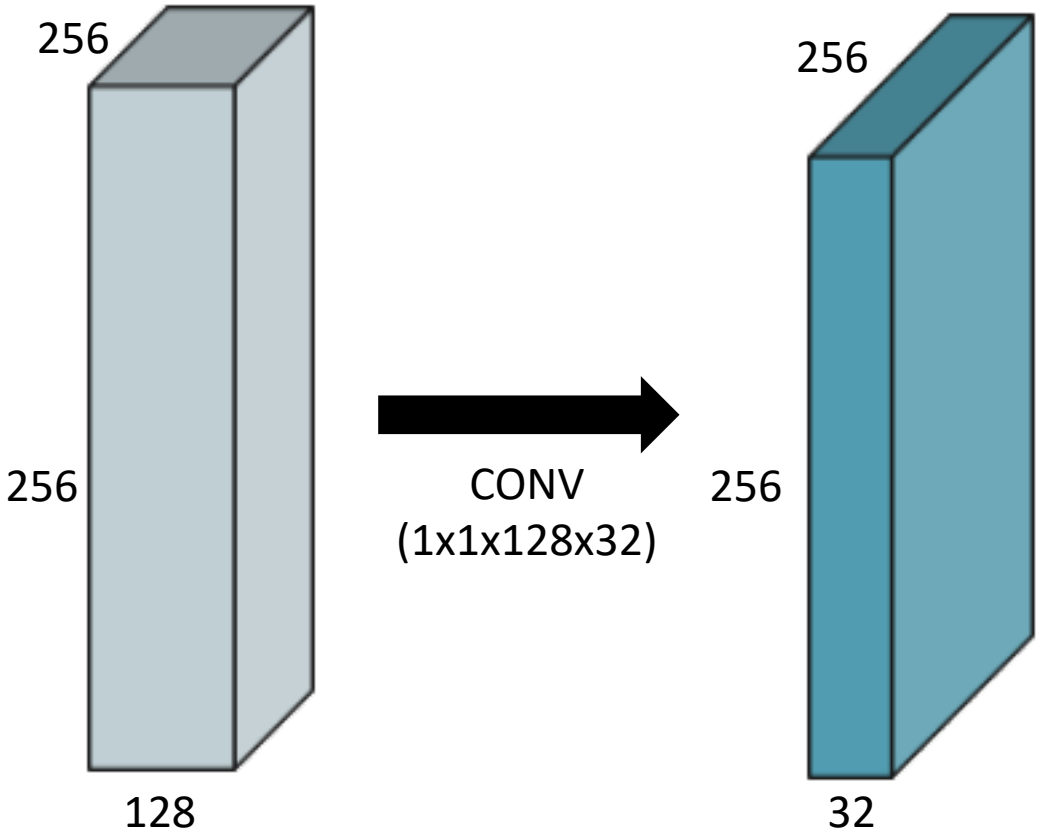
Exercise



$$2048 * 2 \times 1000 \approx 4M$$

What is the **number of parameters** of this model?

1x1 Convolution



- Why?
  - Dimension reduction
  - To reduce the number of parameters while increasing the depth
  - e.g., bottleneck architecture

Thank you for listening