Deep Learning

Vazgen Mikayelyan

YSU, Krisp

October 17, 2019

Outline

What is convolution?

Definition 1

Convolution of the functions $f, g : \mathbb{R} \to \mathbb{R}$ is defined as the integral of the product of the two functions after one is reversed and shifted:

$$(f * g)(t) =: \int_{-\infty}^{+\infty} f(x)\, g(t - x)\, dx.$$

What is convolution?

Definition 1

Convolution of the functions $f, g : \mathbb{R} \to \mathbb{R}$ is defined as the integral of the product of the two functions after one is reversed and shifted:

$$(f * g)(t) =: \int_{-\infty}^{+\infty} f(x)\, g(t - x)\, dx.$$

It easy to see that $f * g = g * f$.

What is convolution?

Definition 1

Convolution of the functions $f, g : \mathbb{R} \to \mathbb{R}$ is defined as the integral of the product of the two functions after one is reversed and shifted:

$$(f * g)(t) =: \int_{-\infty}^{+\infty} f(x)\, g(t - x)\, dx.$$

It easy to see that $f * g = g * f$.

Definition 2

Convolution of the sequences of real numbers $\{f_n\}_{n=-\infty}^{+\infty}, \{g_n\}_{n=-\infty}^{+\infty}$ is the following sequence:

$$z_n =: \sum_{k=-\infty}^{+\infty} f_k g_{n-k}.$$

2D Convolution

Definition 3

Convolution of the functions $f, g : \mathbb{R}^2 \to \mathbb{R}^2$ is the following function:

$$(f * g)(t, \tau) =: \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\, g(t - x, \tau - y)\, dxdy.$$

2D Convolution

Definition 3

Convolution of the functions $f, g : \mathbb{R}^2 \to \mathbb{R}^2$ is the following function:

$$(f * g)(t, \tau) =: \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\, g(t - x, \tau - y)\, dxdy.$$

It easy to see that $f * g = g * f$.

2D Convolution

Definition 3

Convolution of the functions $f, g : \mathbb{R}^2 \to \mathbb{R}^2$ is the following function:

$$(f * g)(t, \tau) =: \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\, g(t - x, \tau - y)\, dxdy.$$

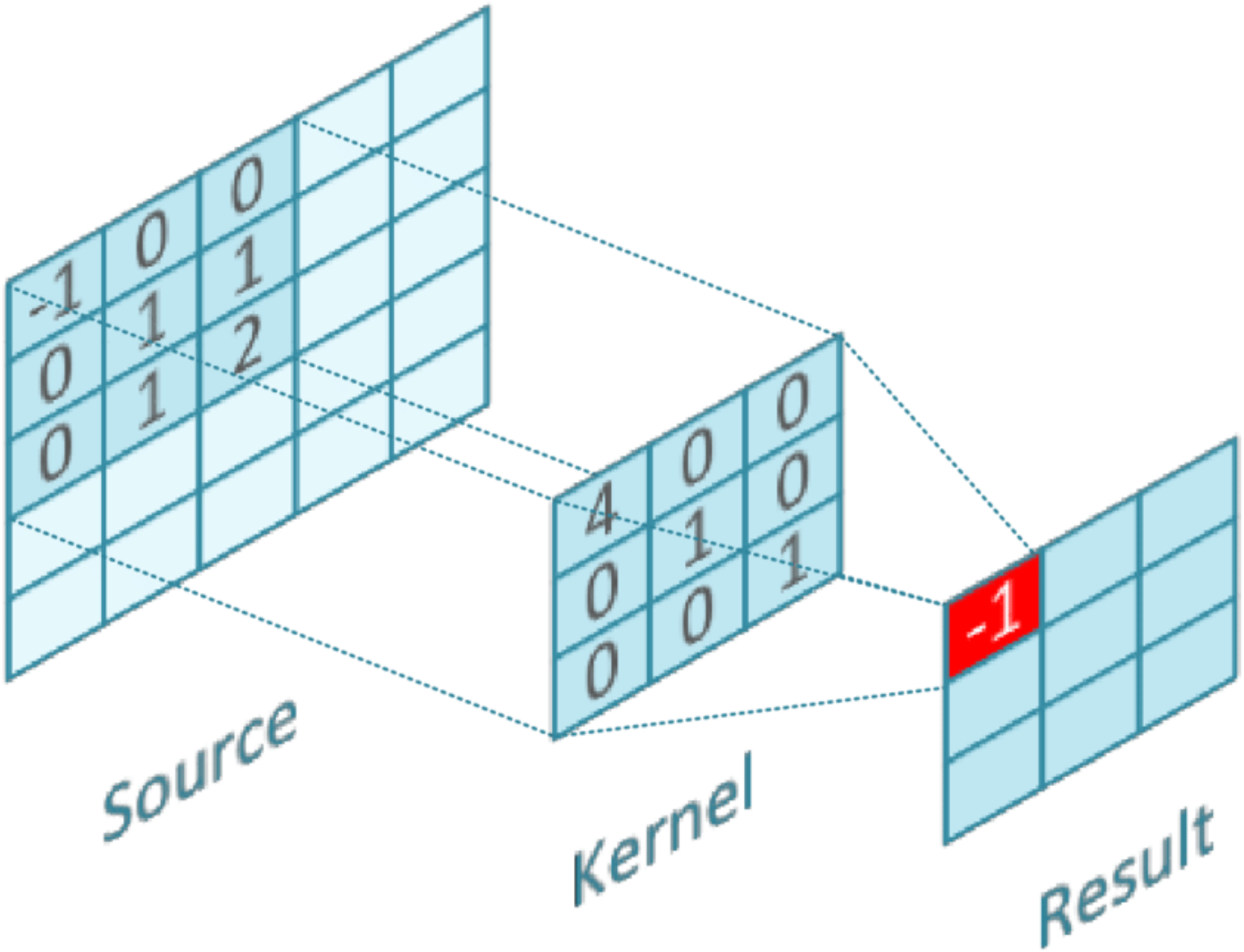
It easy to see that $f * g = g * f$.

Definition 4

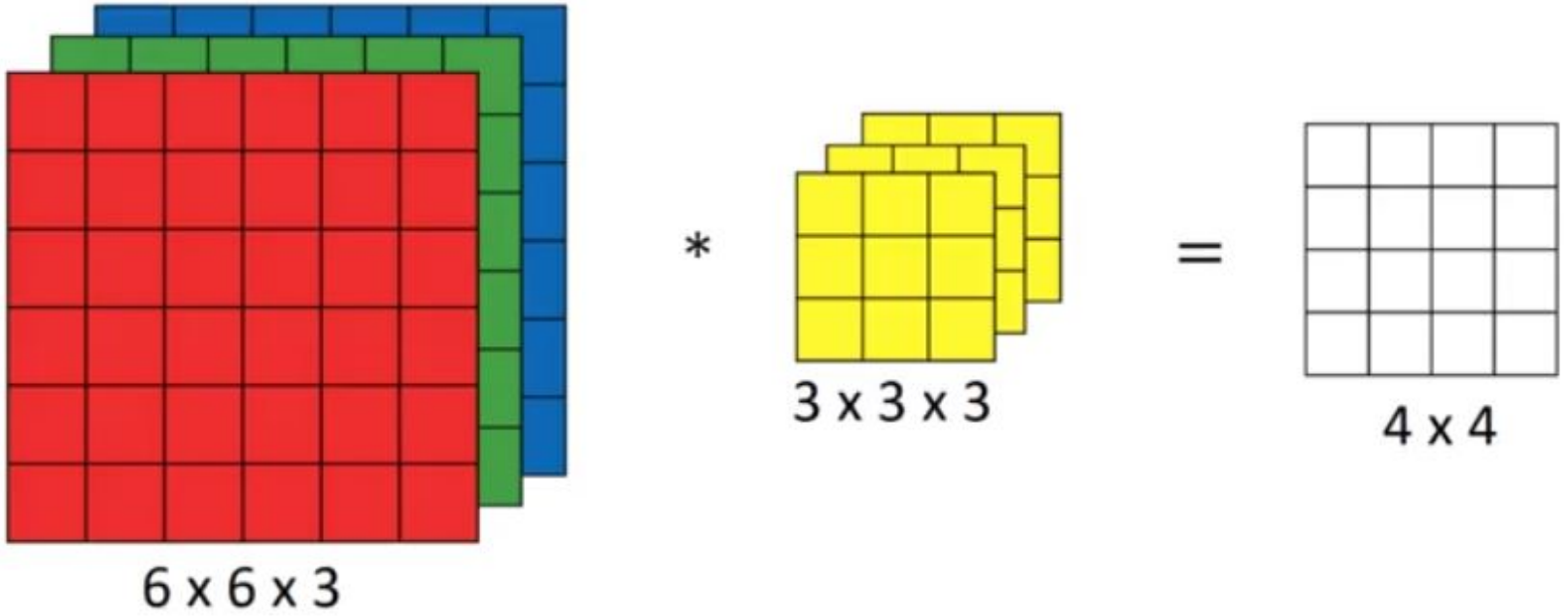
Let $f(x, y)$ is an image and $w(s, t)$ is a kernel where $s \in [a, b]$, $t \in [c, d]$, $x, y, s, t, a, b, c, d \in \mathbb{Z}$. The convolution between kernel w and image f is the following function

$$(w * f)(x, y) = \sum_{s=a}^{b} \sum_{t=c}^{d} w(s, t)\, f(x - s, y - t)$$

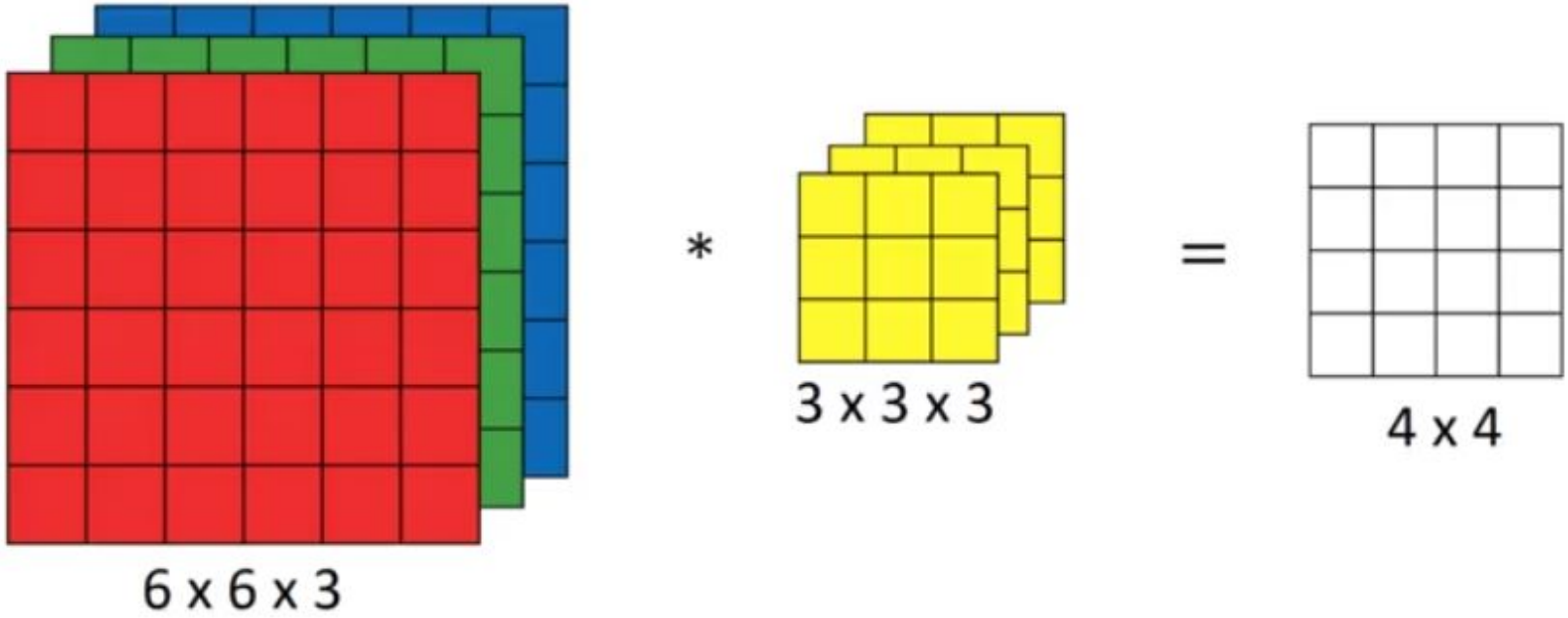
2D Convolution



Convolution on RGB Images

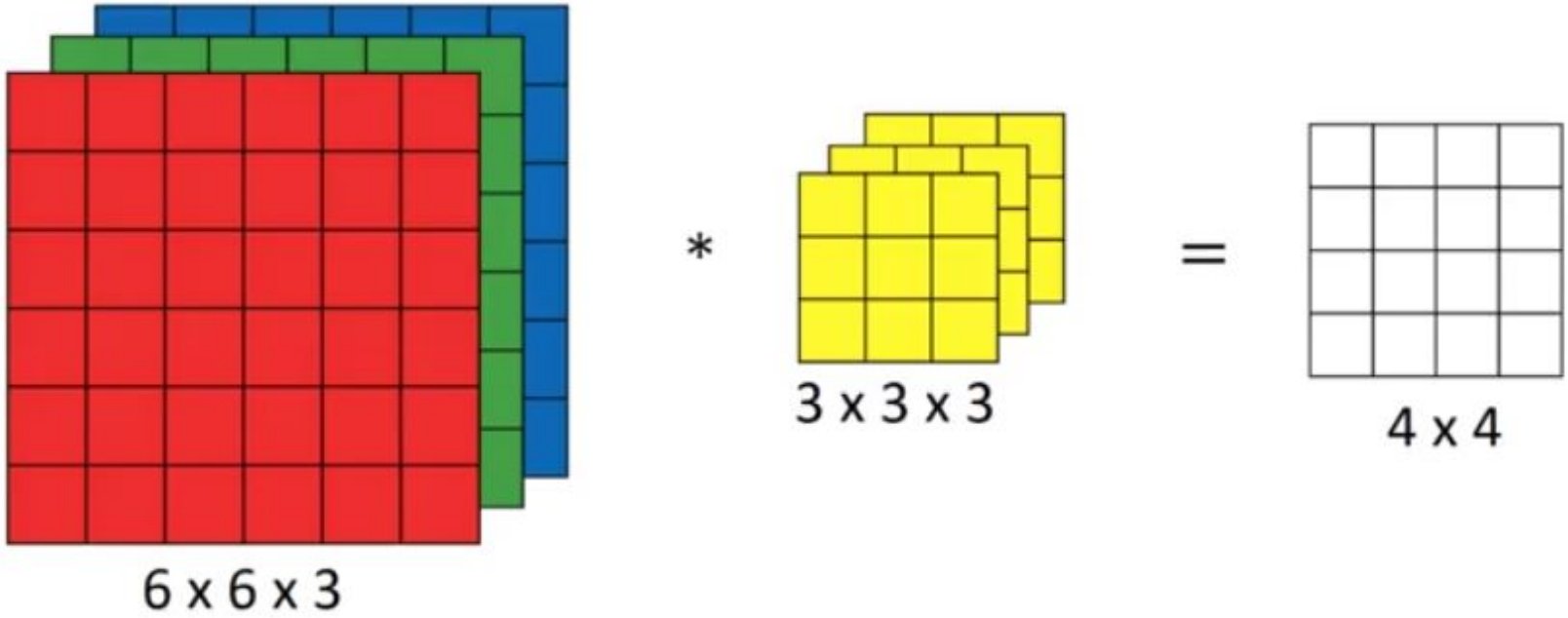


Convolution on RGB Images



- How many dimensions has convolution kernel in general?

Convolution on RGB Images



- How many dimensions has convolution kernel in general?
- What does 1×1 convolution do?

Valid and Same Convolution

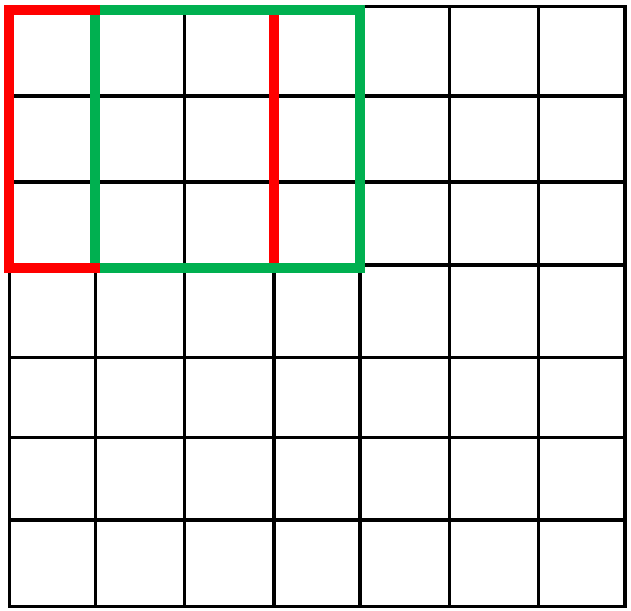
- Padding = Same: means the input image ought to have zero padding so that the output in convolution doesn't differ in size as input.

Valid and Same Convolution

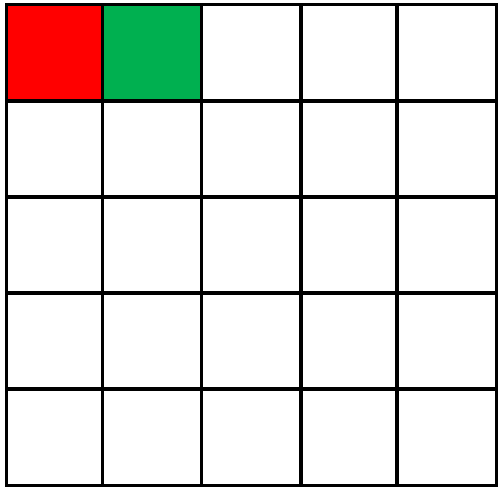
- Padding = Same: means the input image ought to have zero padding so that the output in convolution doesn't differ in size as input.
- Padding = Valid: means we don't add the zero pixel padding around the input matrix, and its like saying, we are ready to loose some information.

Strided Convolution

7 x 7 Input Volume

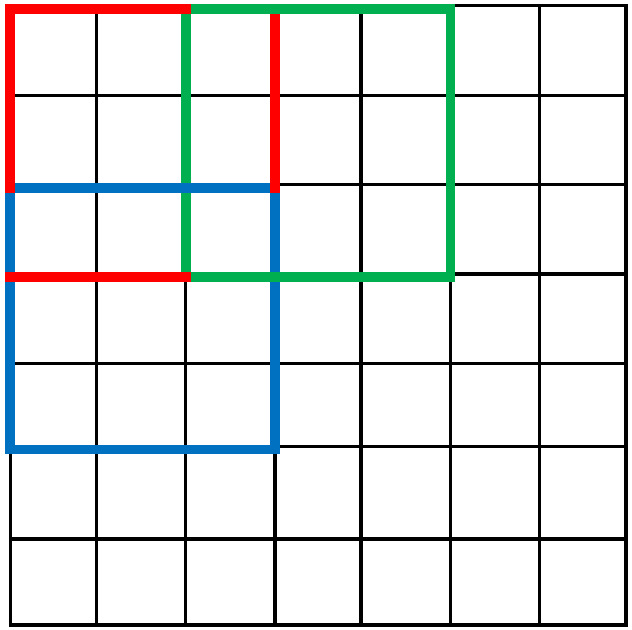


5 x 5 Output Volume

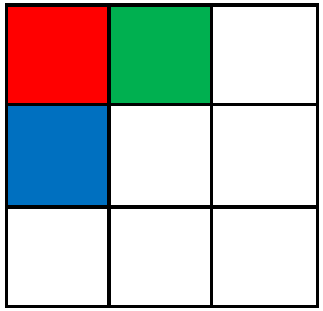


Strided Convolution

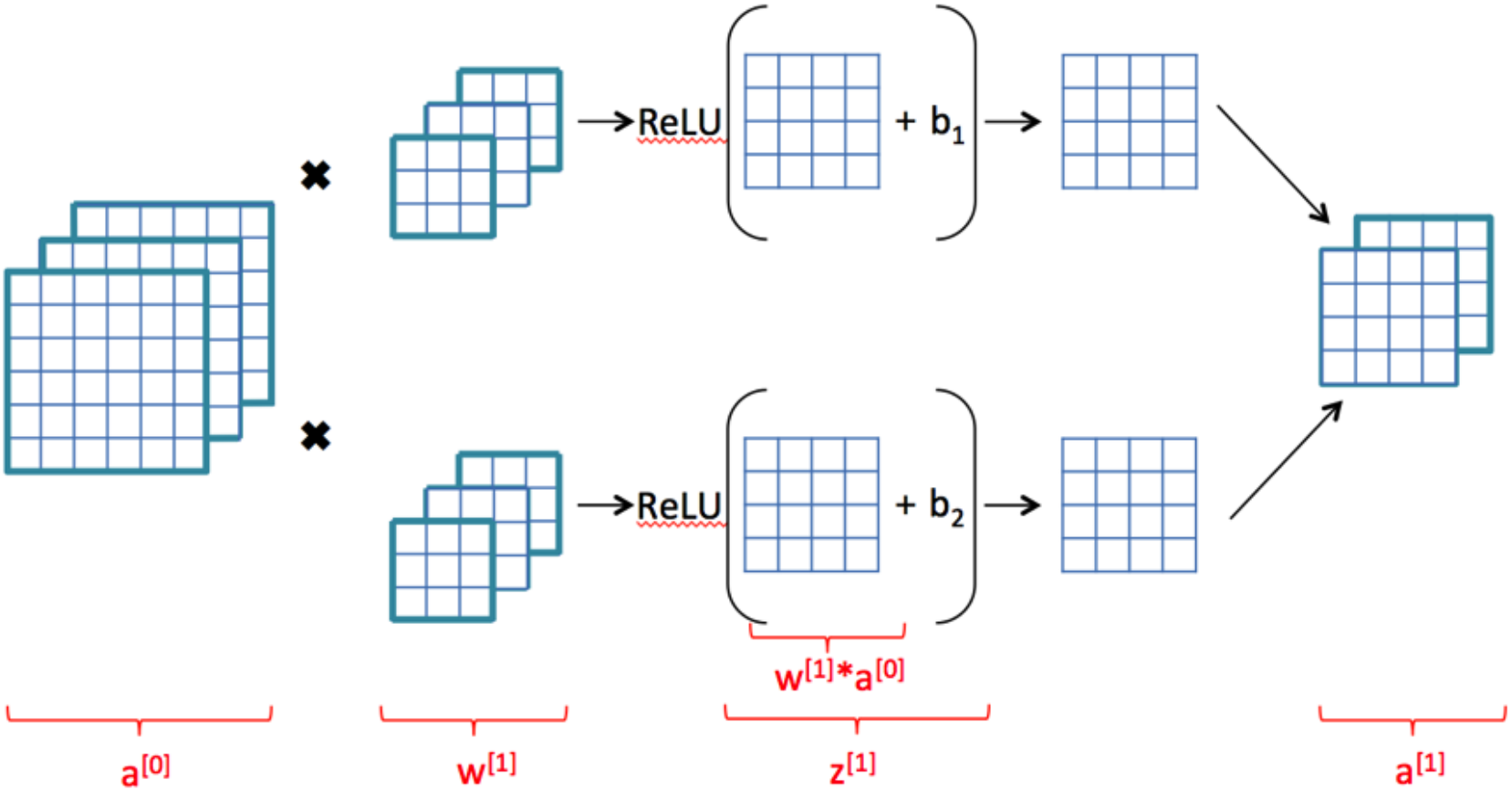
7 x 7 Input Volume



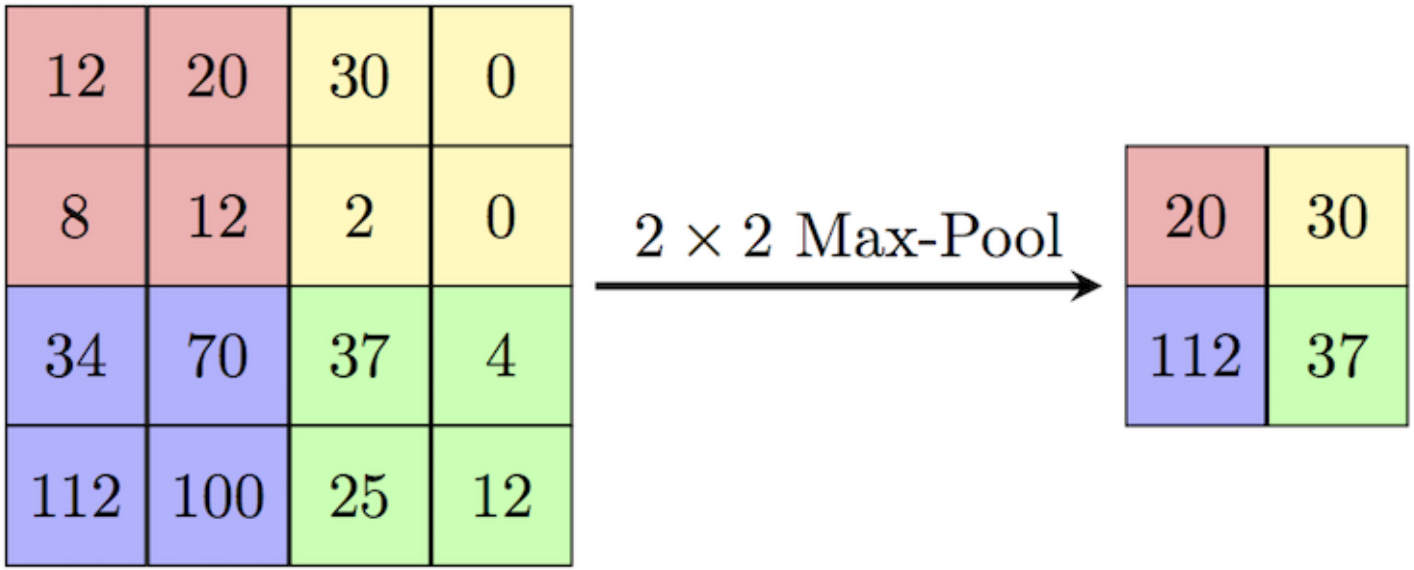
3 x 3 Output Volume



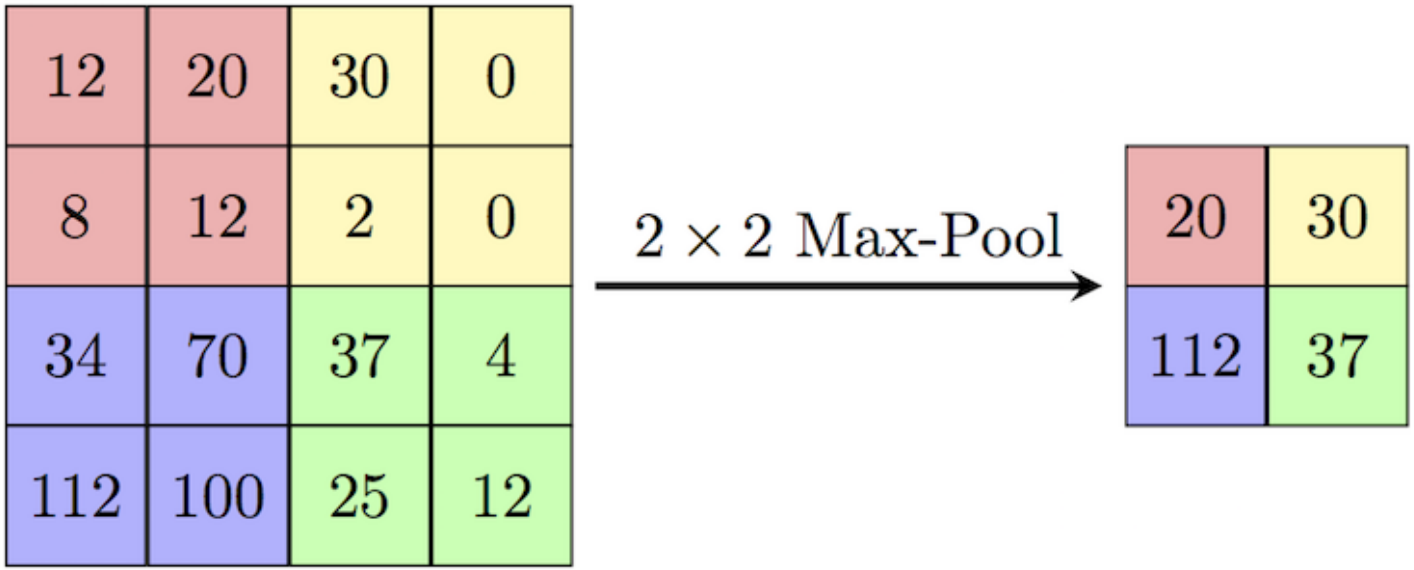
One Layer of CNN



Pooling Layers

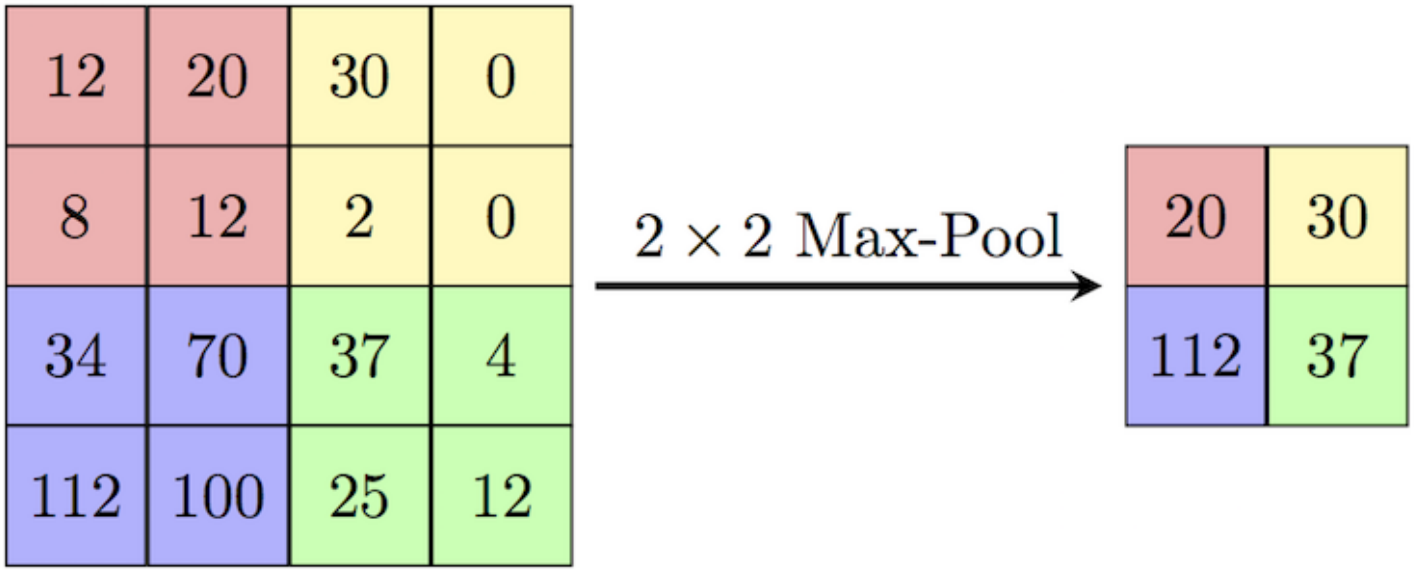


Pooling Layers



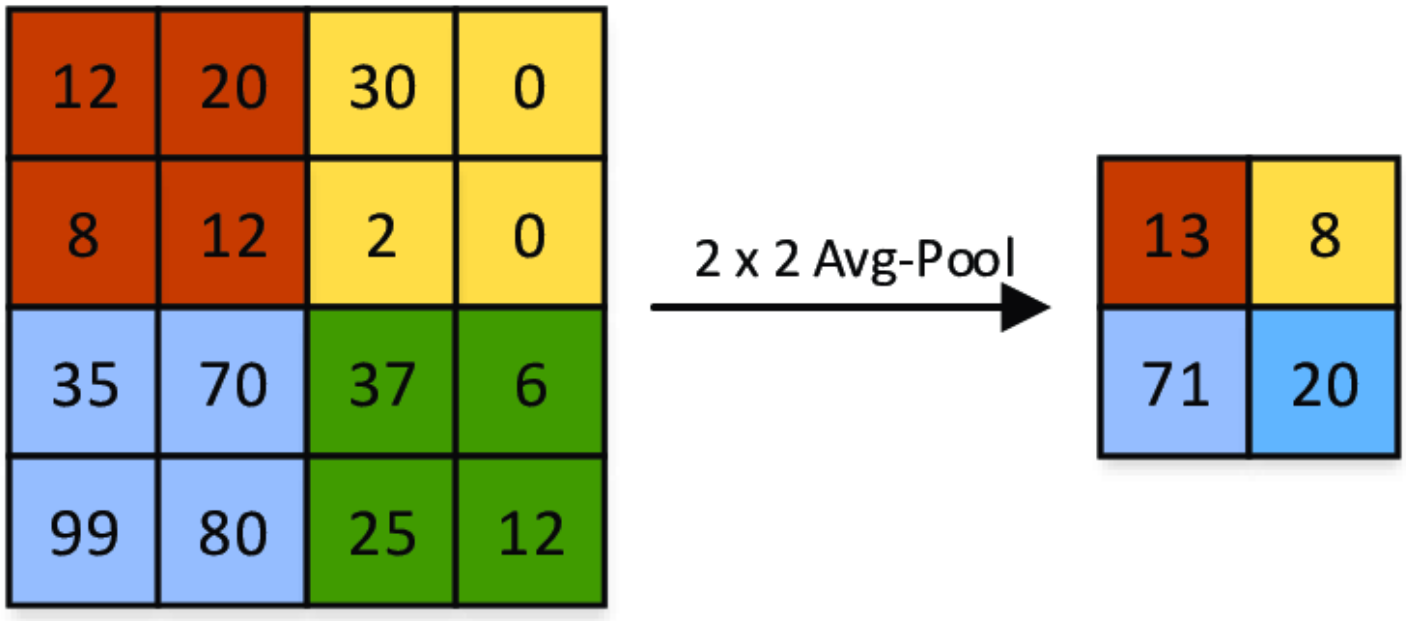
- Number of channels is the same after pooling layer.

Pooling Layers



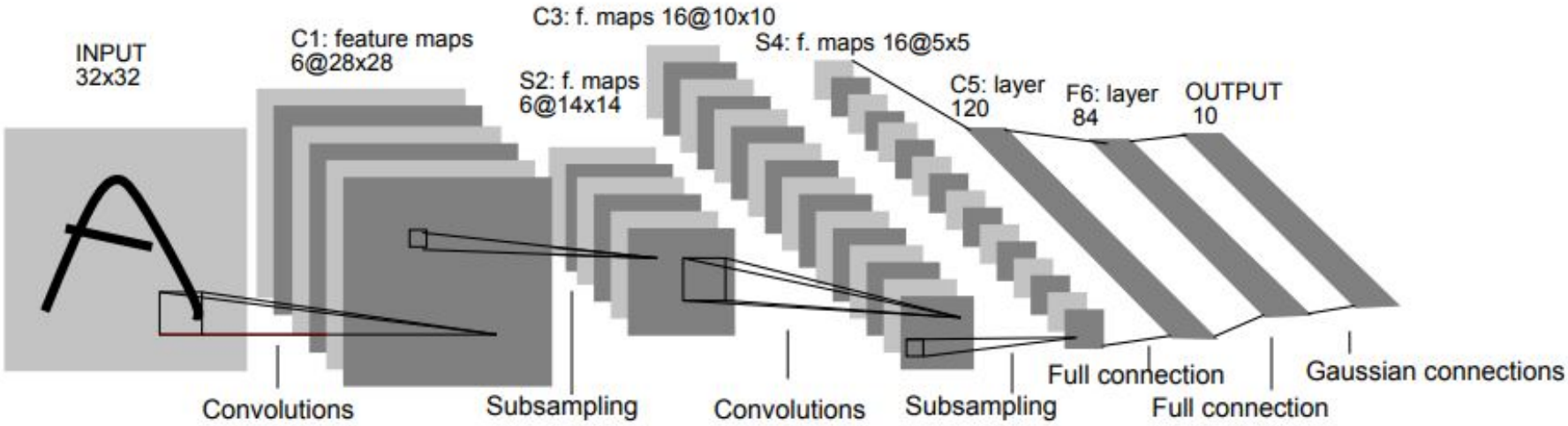
- Number of channels is the same after pooling layer.
- There are not trainable parameters in this layer.

Pooling Layers

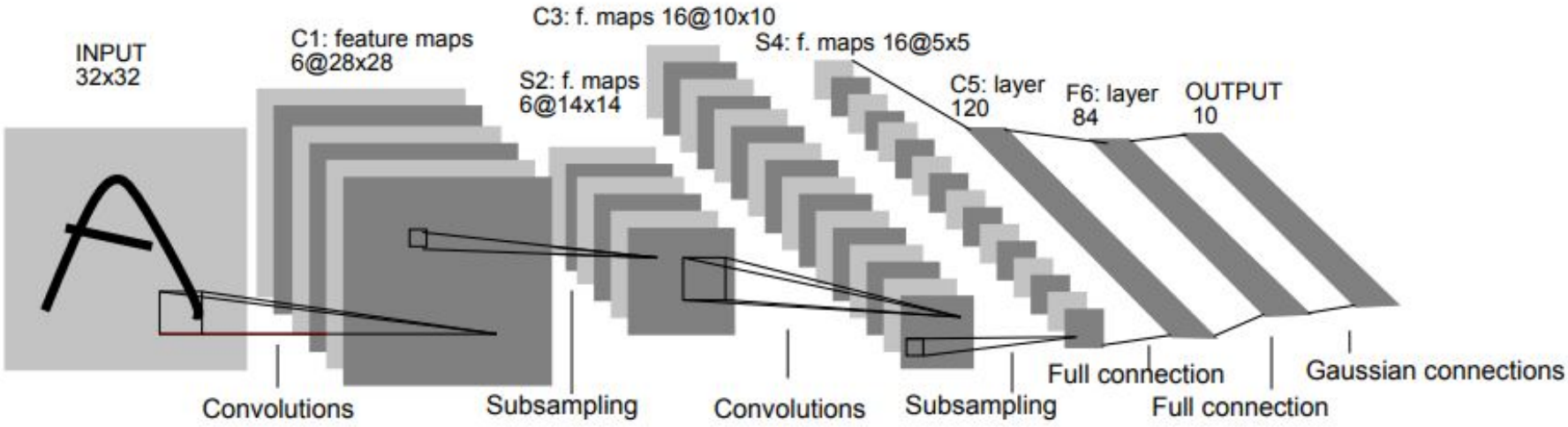


Outline

LeNet-5 (1998)

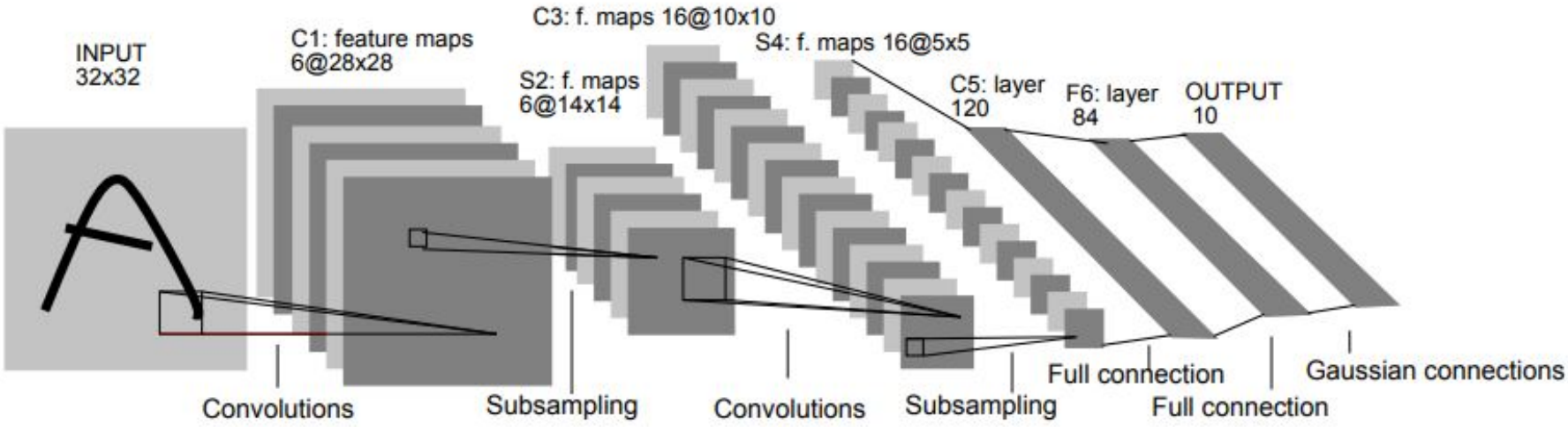

INPUT
32x32
C1: feature maps
6@28x28
S2: f. maps
6@14x14
C3: f. maps 16@10x10
S4: f. maps 16@5x5
C5: layer
120
F6: layer
84
OUTPUT
10
Convolutions
Subsampling
Convolutions
Subsampling
Full connection
Full connection
Gaussian connections

LeNet-5 (1998)



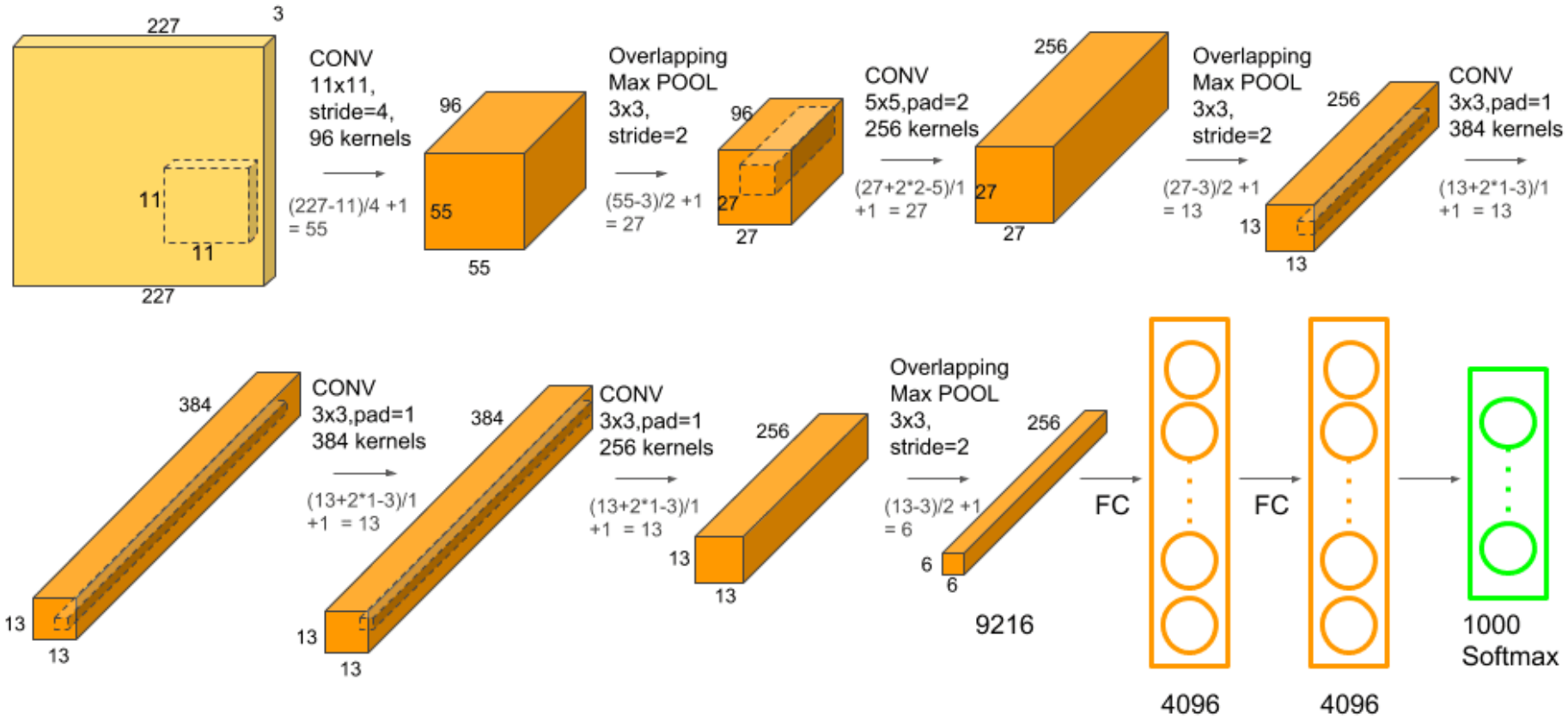
- Activation functions are sigmoids and hyperbolic tangents.

LeNet-5 (1998)



- Activation functions are sigmoids and hyperbolic tangents.
- LeNet-5 has approximately 60k parameters.

AlexNet (2012)



AlexNet (2012)

- The ReLU non-linearity is applied to the output of every convolutional and fully-connected layer.

AlexNet (2012)

- The ReLU non-linearity is applied to the output of every convolutional and fully-connected layer.
- Local response normalization?

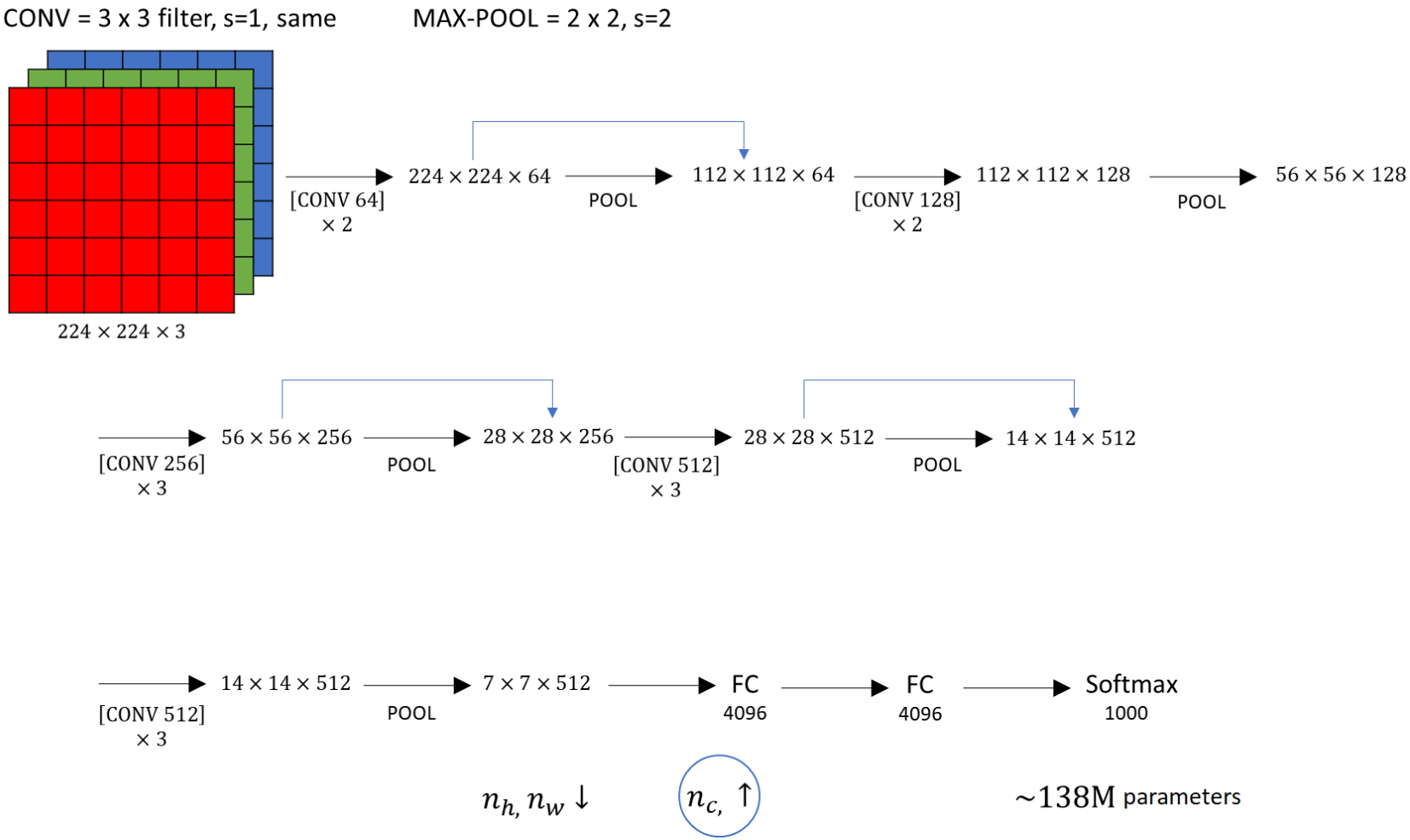
AlexNet (2012)

- The ReLU non-linearity is applied to the output of every convolutional and fully-connected layer.
- Local response normalization?
- AlexNet has approximately 60M parameters.

AlexNet (2012)

- The ReLU non-linearity is applied to the output of every convolutional and fully-connected layer.
- Local response normalization?
- AlexNet has approximately 60M parameters.
- Accuracies on ImageNet: Top1=63.3%, Top5=84.6%.

VGG-16 (2014)


CONV = 3 x 3 filter, s=1, same
MAX-POOL = 2 x 2, s=2
224 × 224 × 3
[CONV 64] × 2
224 × 224 × 64
POOL
112 × 112 × 64
[CONV 128] × 2
112 × 112 × 128
POOL
56 × 56 × 128
[CONV 256] × 3
56 × 56 × 256
POOL
28 × 28 × 256
[CONV 512] × 3
28 × 28 × 512
POOL
14 × 14 × 512
[CONV 512] × 3
14 × 14 × 512
POOL
7 × 7 × 512
FC 4096
FC 4096
Softmax 1000
$n_h, n_w \downarrow$
$n_c, \uparrow$
~138M parameters

VGG-16 (2014)

- Fixed kernel size.

VGG-16 (2014)

- Fixed kernel size.
- The ReLU non-linearity is applied to the output of every convolutional and fully-connected layer.

VGG-16 (2014)

- Fixed kernel size.
- The ReLU non-linearity is applied to the output of every convolutional and fully-connected layer.
- VGG-16 has approximately 138M parameters.

VGG-16 (2014)

- Fixed kernel size.
- The ReLU non-linearity is applied to the output of every convolutional and fully-connected layer.
- VGG-16 has approximately 138M parameters.
- Accuracies on ImageNet: Top1=74.4%, Top5=91.9%.

VGG-19 (2014)

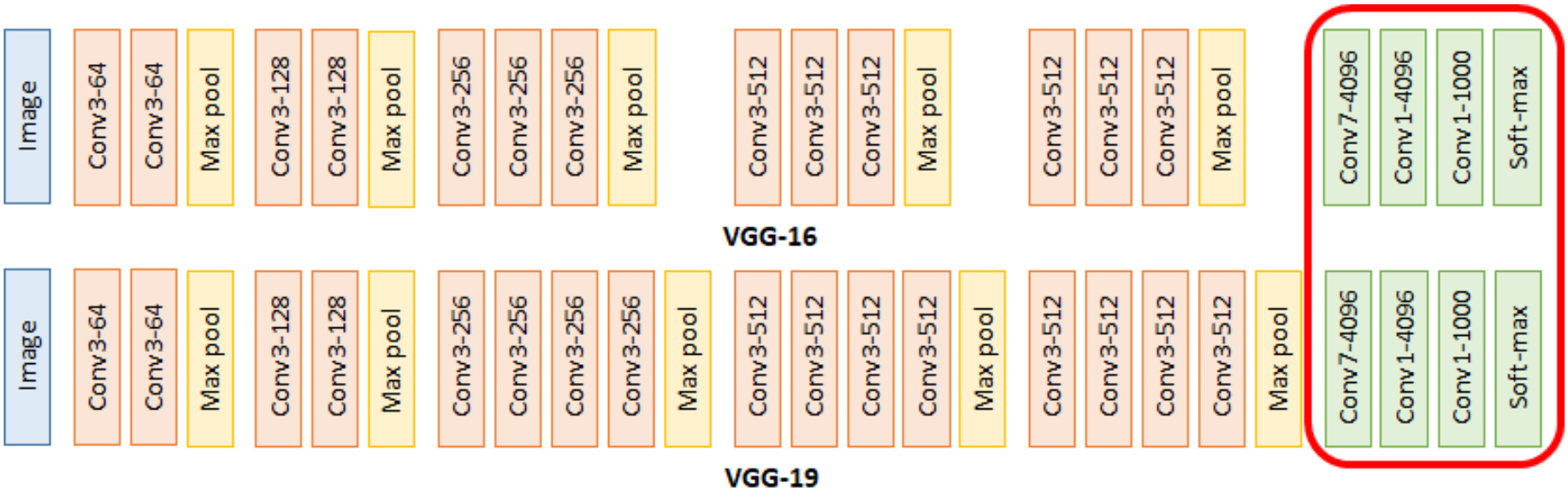


Image
Conv3-64
Conv3-64
Max pool
Conv3-128
Conv3-128
Max pool
Conv3-256
Conv3-256
Conv3-256
Max pool
Conv3-512
Conv3-512
Conv3-512
Max pool
Conv3-512
Conv3-512
Conv3-512
Max pool
Conv7-4096
Conv1-4096
Conv1-1000
Soft-max
VGG-16
Image
Conv3-64
Conv3-64
Max pool
Conv3-128
Conv3-128
Max pool
Conv3-256
Conv3-256
Conv3-256
Conv3-256
Max pool
Conv3-512
Conv3-512
Conv3-512
Conv3-512
Max pool
Conv3-512
Conv3-512
Conv3-512
Conv3-512
Max pool
Conv7-4096
Conv1-4096
Conv1-1000
Soft-max
VGG-19

VGG-19 (2014)

- The ReLU non-linearity is applied to the output of every convolutional and fully-connected layer.

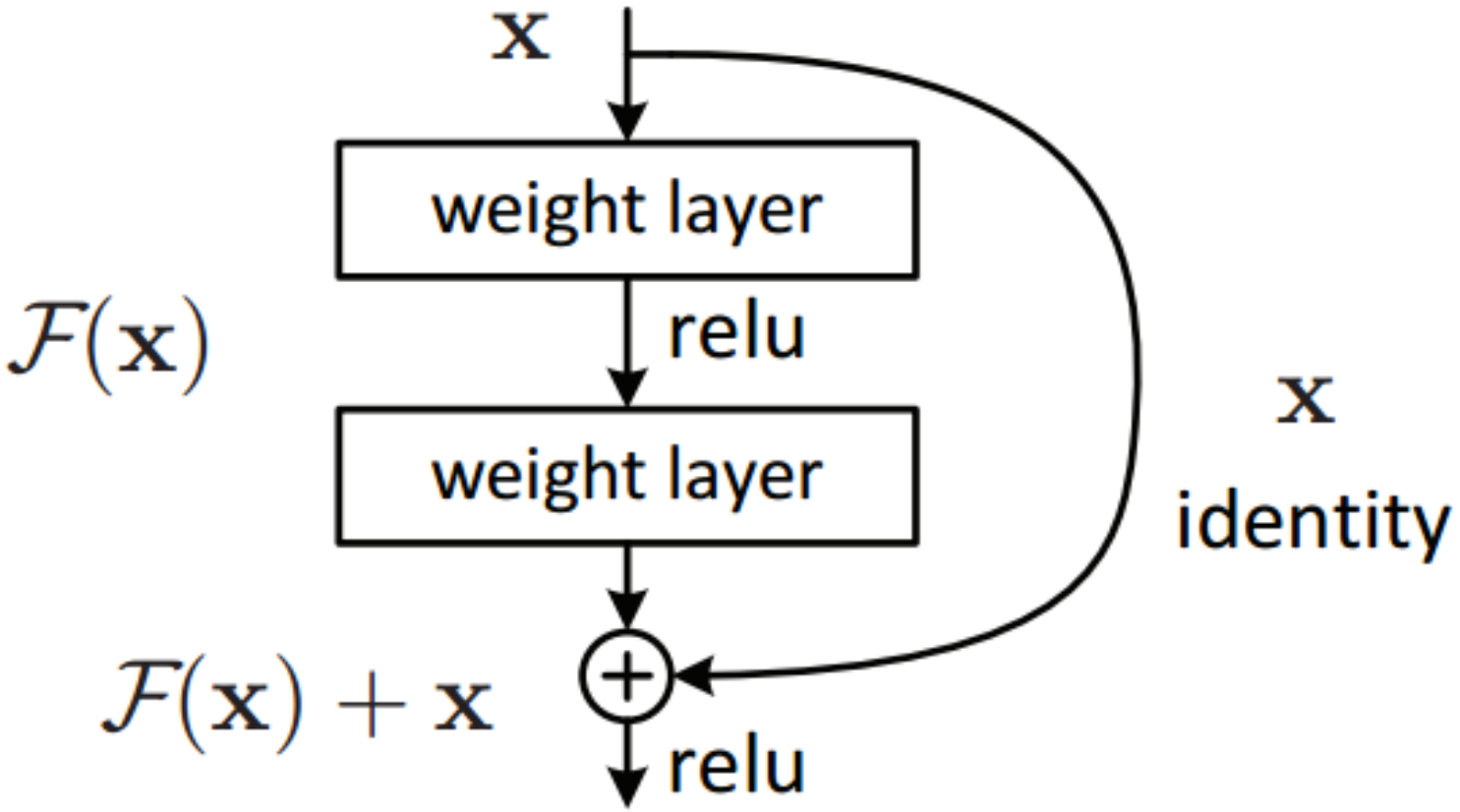
VGG-19 (2014)

- The ReLU non-linearity is applied to the output of every convolutional and fully-connected layer.
- VGG-19 has approximately 144M parameters.

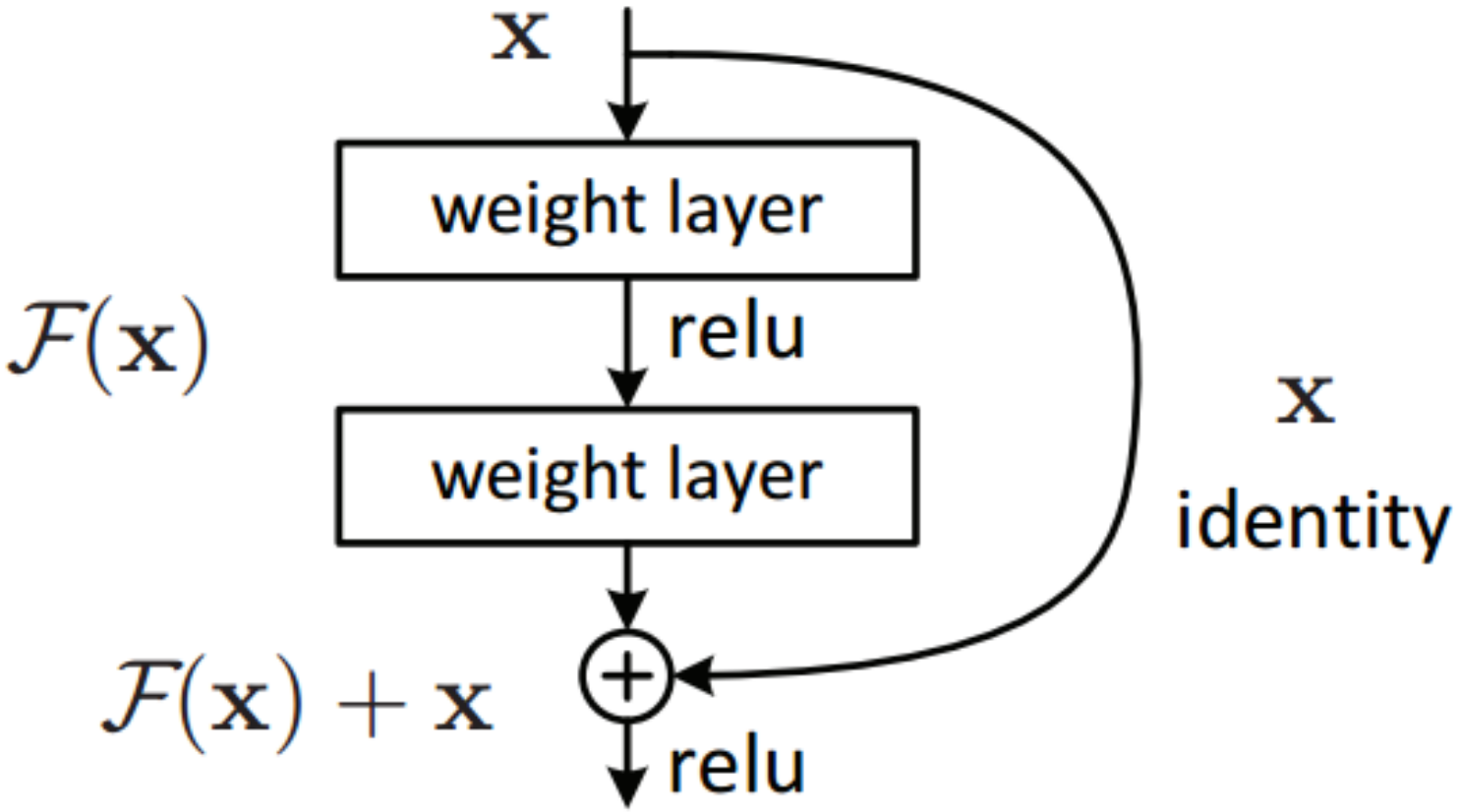
VGG-19 (2014)

- The ReLU non-linearity is applied to the output of every convolutional and fully-connected layer.
- VGG-19 has approximately 144M parameters.
- Accuracies on ImageNet: Top1=74.5%, Top5=92%.

Resnet-50 (2015)

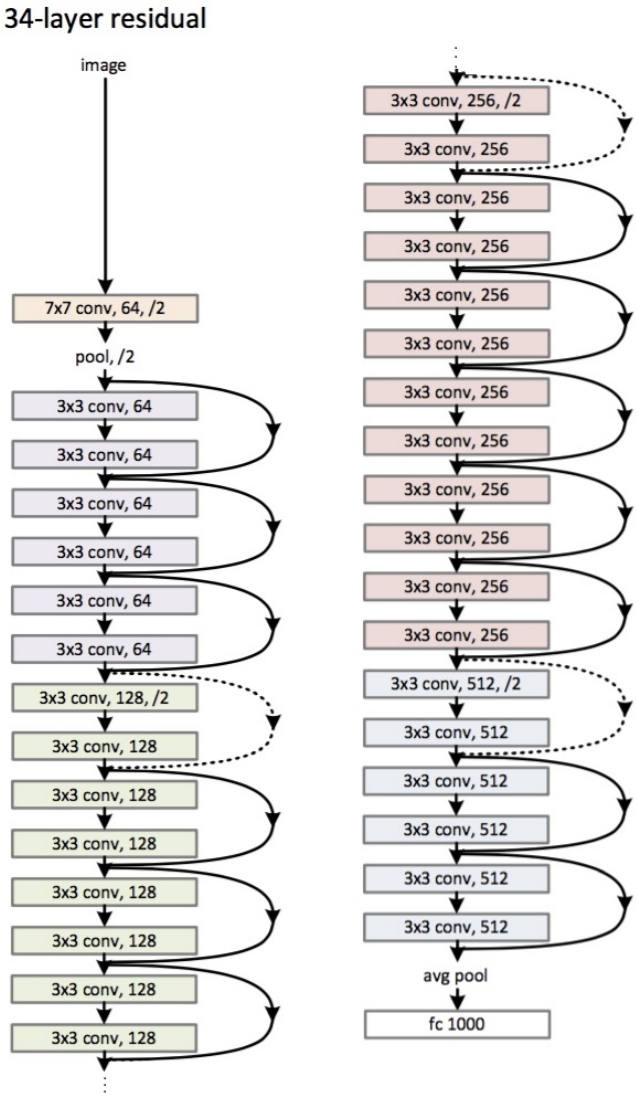


Resnet-50 (2015)



Identity function is easy to learn for residual block.

Resnet-50 (2015)



Resnet-50 (2015)

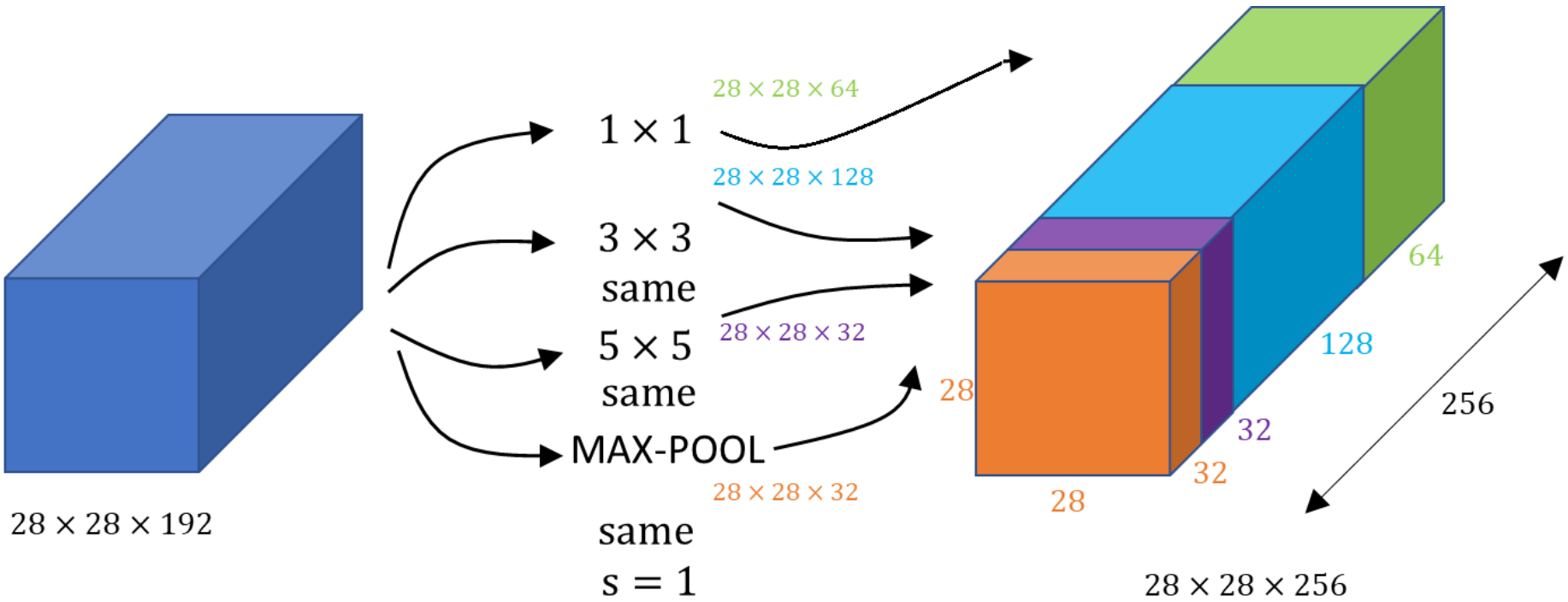
- Resnet-50 has approximately 25.6M parameters.

Resnet-50 (2015)

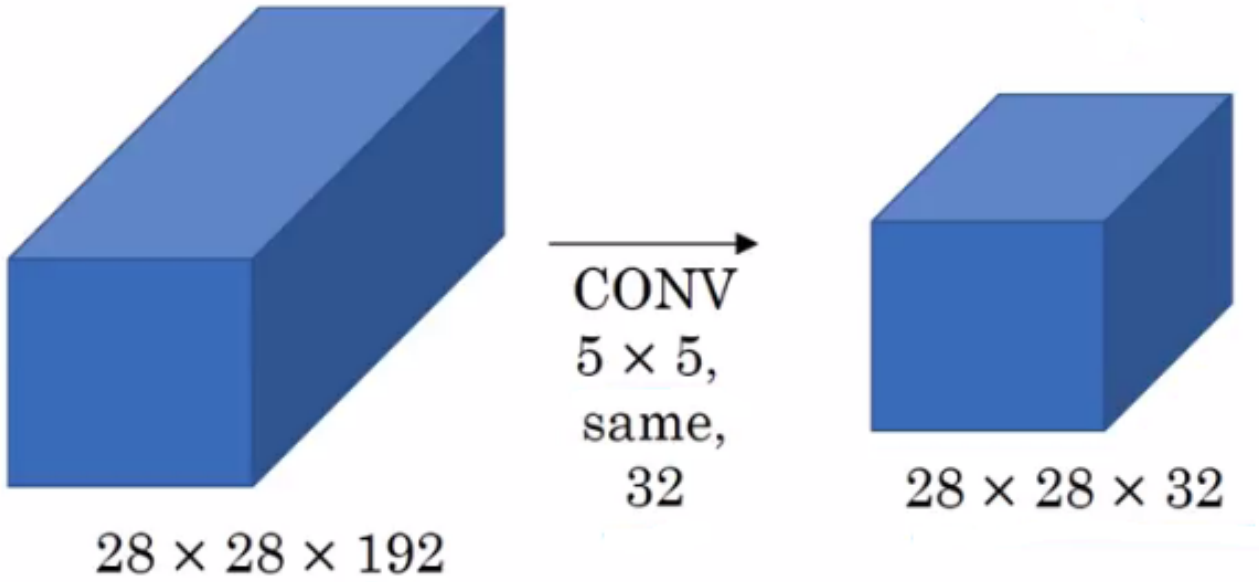
- Resnet-50 has approximately 25.6M parameters.
- Accuracies on ImageNet: Top1=77.15%, Top5=93.29%.

GoogLeNet/Inception v1 (2014)

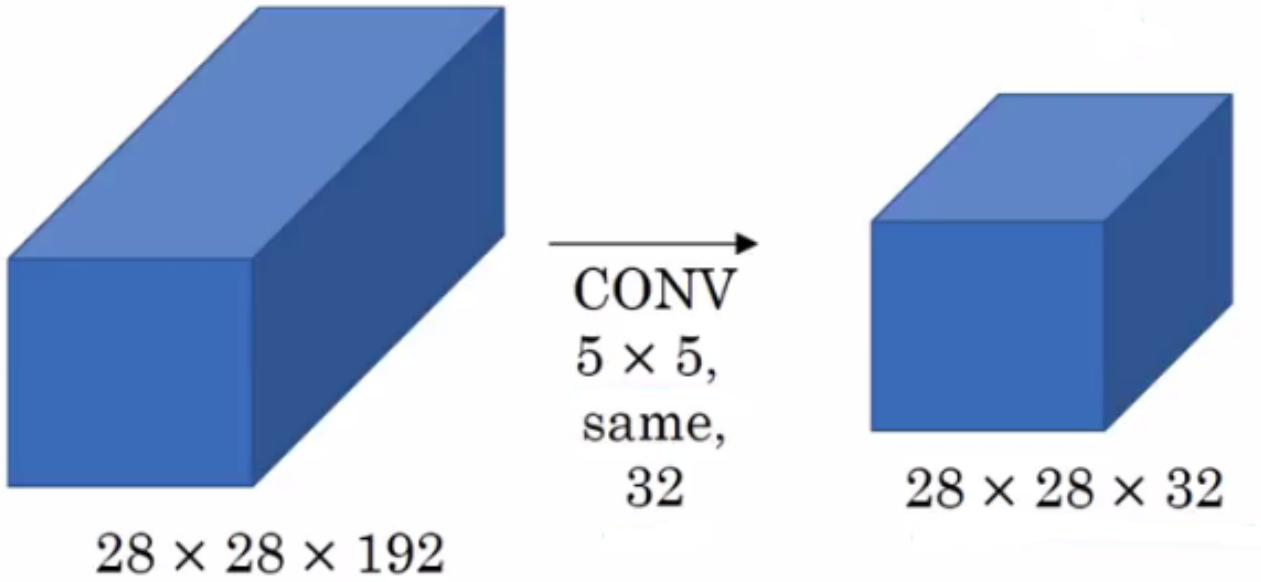
GoogLeNet/Inception v1 (2014)



GoogLeNet/Inception v1 (2014)

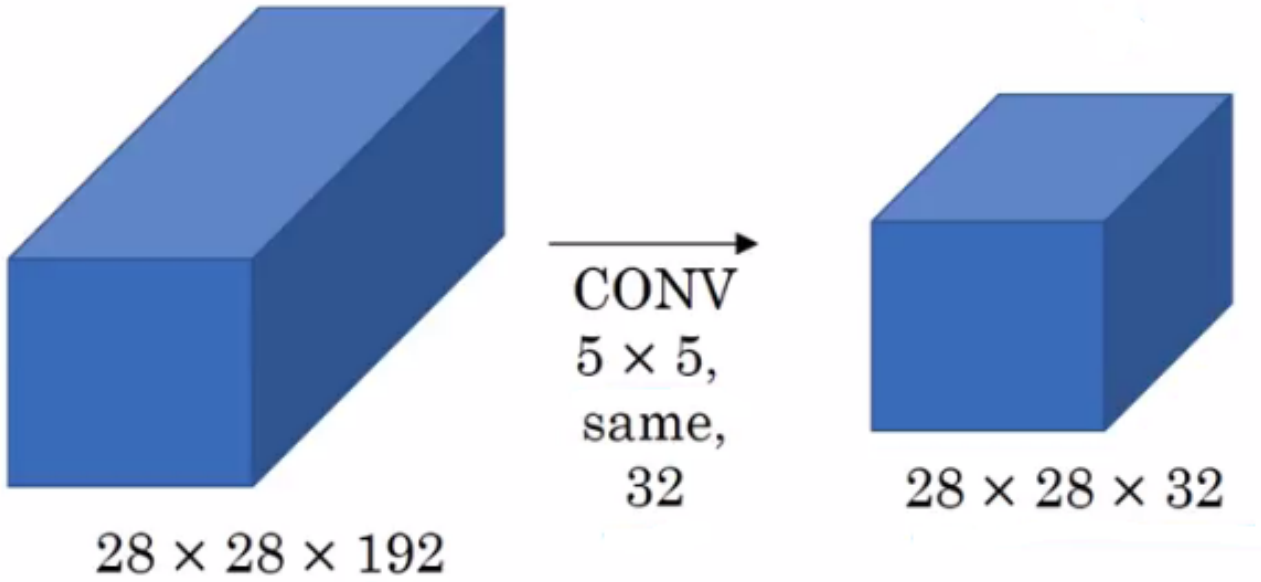


GoogLeNet/Inception v1 (2014)



How much is the number of multiplications?

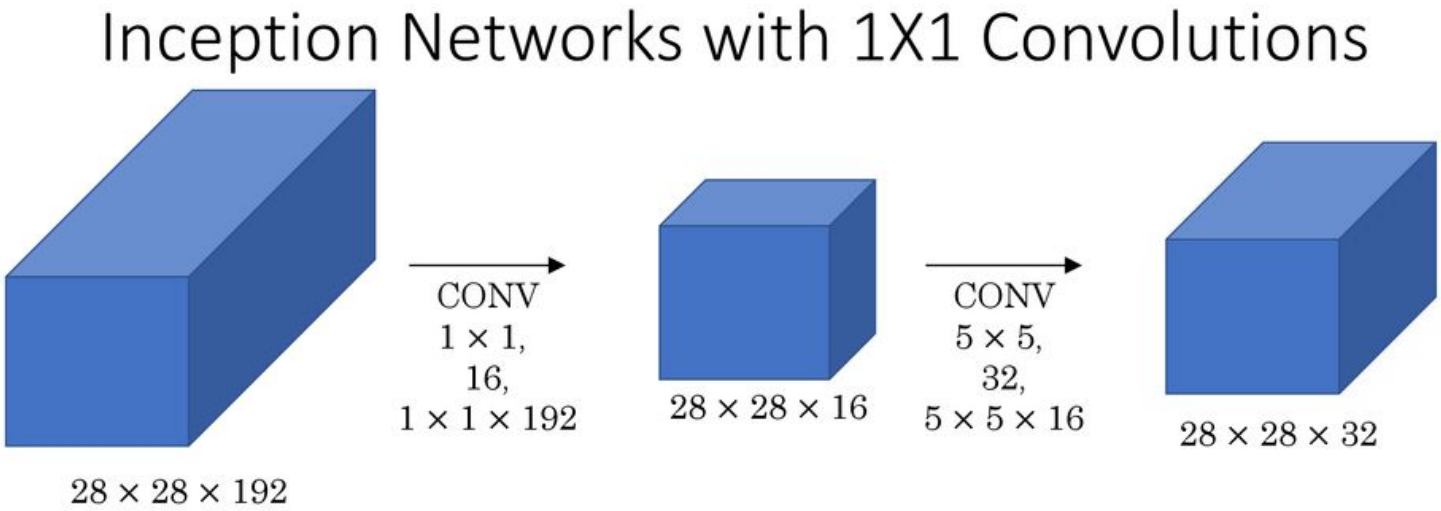
GoogLeNet/Inception v1 (2014)



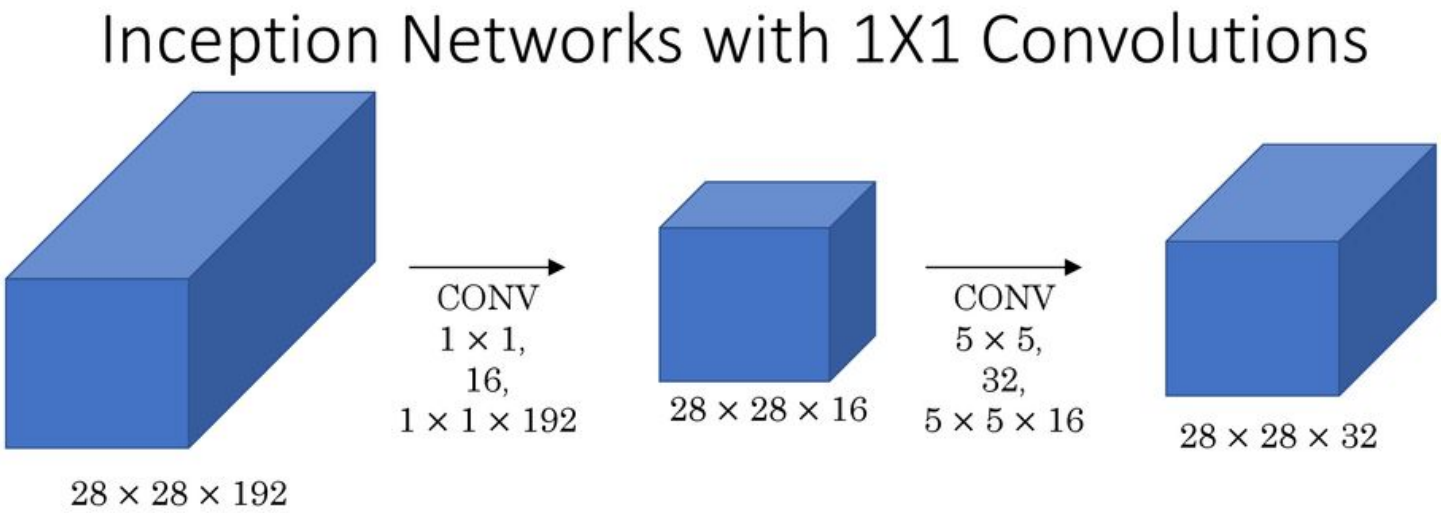
How much is the number of multiplications?

$$28 \cdot 28 \cdot 32 \cdot 5 \cdot 5 \cdot 192 \approx 120M$$

GoogLeNet/Inception v1 (2014)

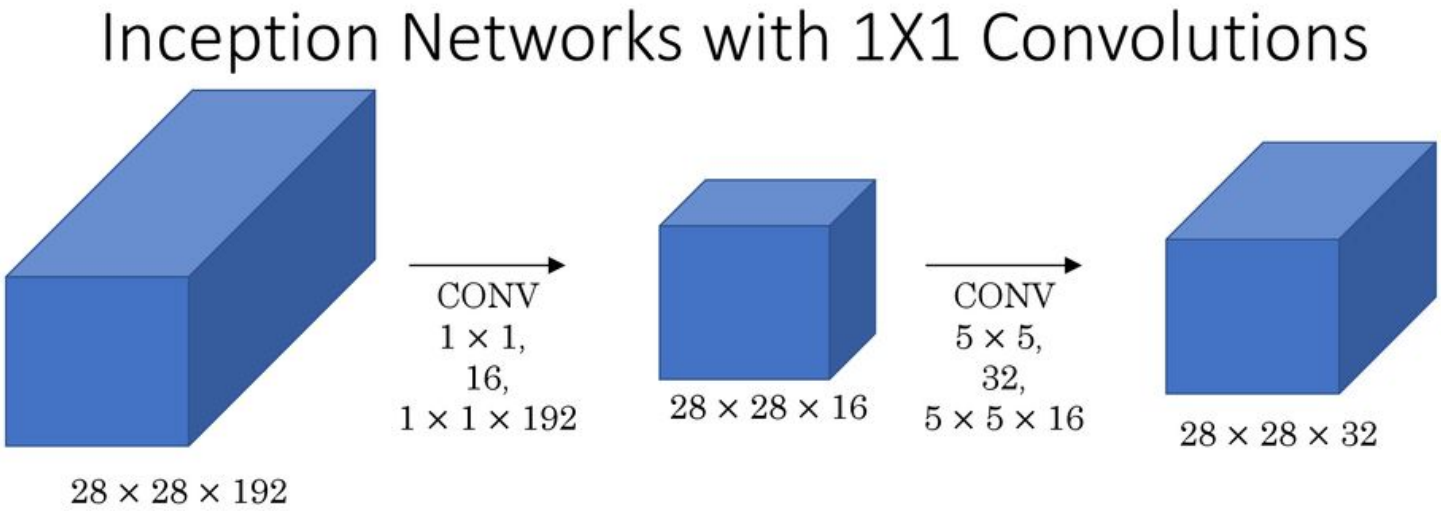


GoogLeNet/Inception v1 (2014)



How much is the number of multiplications?

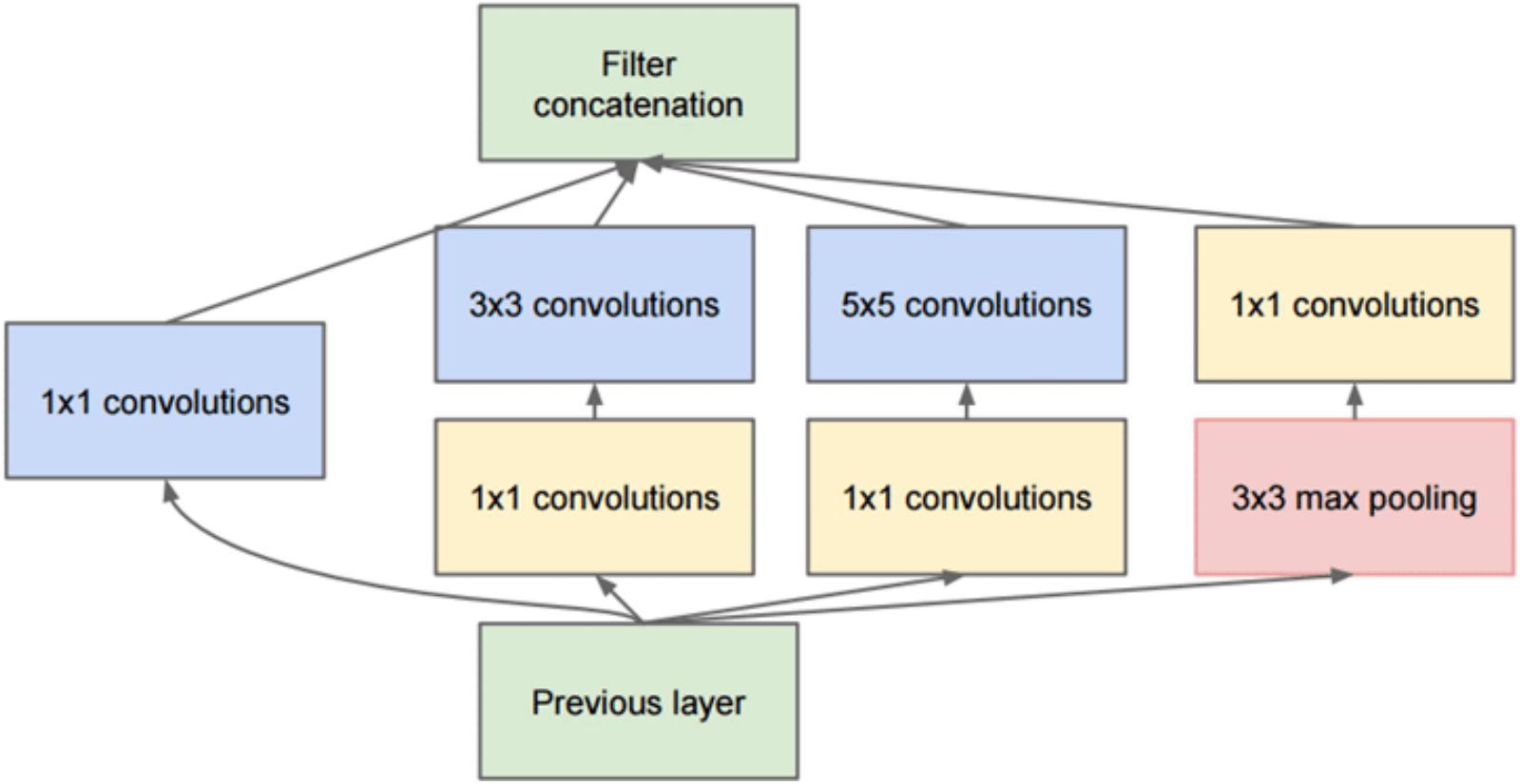
GoogLeNet/Inception v1 (2014)



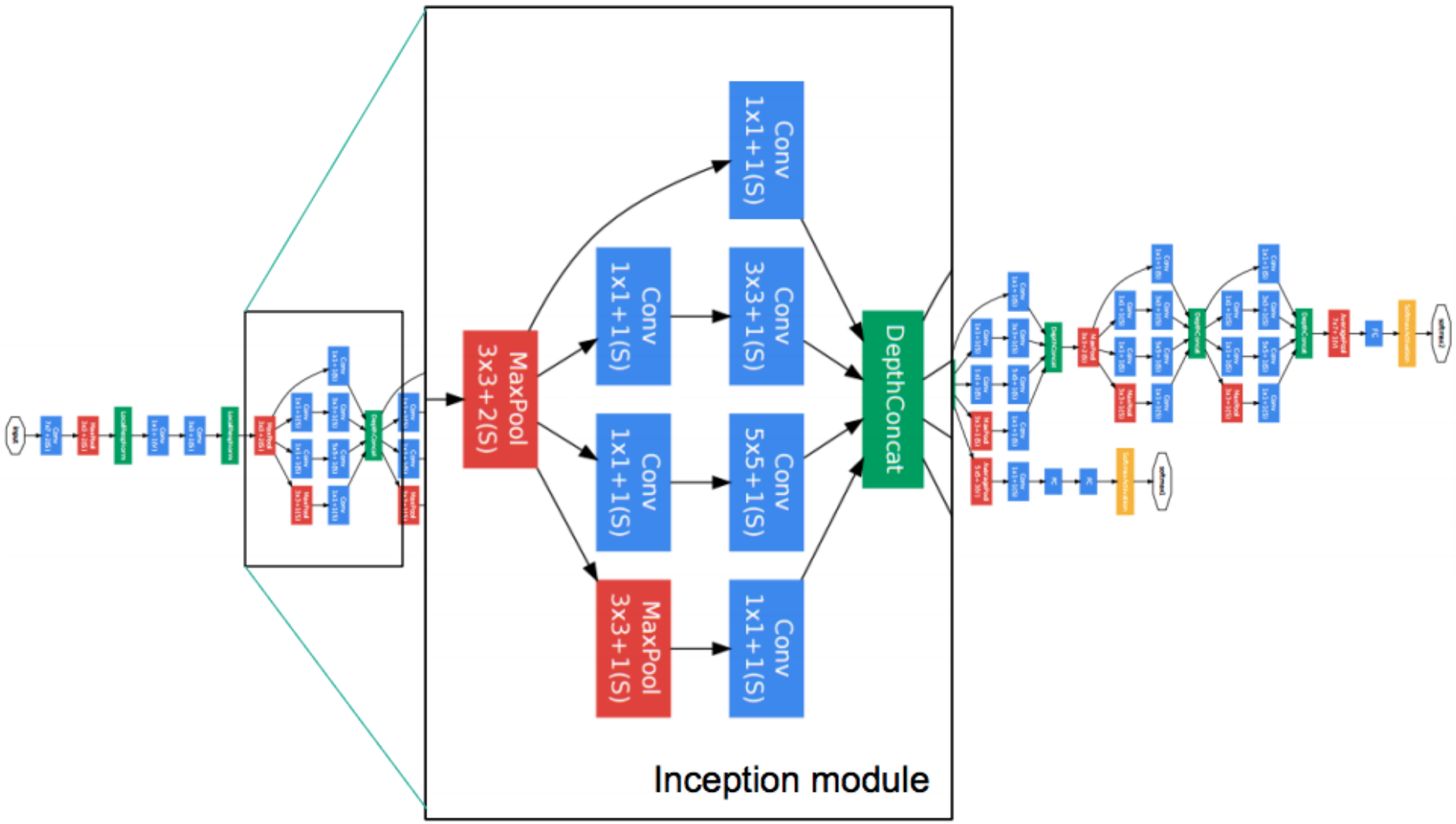
How much is the number of multiplications?

$$28 \cdot 28 \cdot 16 \cdot 1 \cdot 1 \cdot 192 + 28 \cdot 28 \cdot 32 \cdot 5 \cdot 5 \cdot 16 \approx 12.4M$$

GoogLeNet/Inception v1 (2014)



GoogLeNet/Inception v1 (2014)



Inceptions

- GoogLeNet/Inception v1 (2014) has approximately 5M parameters.

Inceptions

- GoogLeNet/Inception v1 (2014) has approximately 5M parameters.
- Accuracies on ImageNet: Top1=69.8%, Top5=89.9%.

Inceptions

- GoogLeNet/Inception v1 (2014) has approximately 5M parameters.
- Accuracies on ImageNet: Top1=69.8%, Top5=89.9%.
- Inception v2 (2015) has approximately 11.2M parameters.

Inceptions

- GoogLeNet/Inception v1 (2014) has approximately 5M parameters.
- Accuracies on ImageNet: Top1=69.8%, Top5=89.9%.
- Inception v2 (2015) has approximately 11.2M parameters.
- Accuracies on ImageNet: Top1=74.8%, Top5=92.2%.

Inceptions

- GoogLeNet/Inception v1 (2014) has approximately 5M parameters.
- Accuracies on ImageNet: Top1=69.8%, Top5=89.9%.
- Inception v2 (2015) has approximately 11.2M parameters.
- Accuracies on ImageNet: Top1=74.8%, Top5=92.2%.
- Inception v3 (2015) has approximately 23.8M parameters.

Inceptions

- GoogLeNet/Inception v1 (2014) has approximately 5M parameters.
- Accuracies on ImageNet: Top1=69.8%, Top5=89.9%.
- Inception v2 (2015) has approximately 11.2M parameters.
- Accuracies on ImageNet: Top1=74.8%, Top5=92.2%.
- Inception v3 (2015) has approximately 23.8M parameters.
- Accuracies on ImageNet: Top1=78.8%, Top5=94.4%.

Inceptions

- GoogLeNet/Inception v1 (2014) has approximately 5M parameters.
- Accuracies on ImageNet: Top1=69.8%, Top5=89.9%.
- Inception v2 (2015) has approximately 11.2M parameters.
- Accuracies on ImageNet: Top1=74.8%, Top5=92.2%.
- Inception v3 (2015) has approximately 23.8M parameters.
- Accuracies on ImageNet: Top1=78.8%, Top5=94.4%.
- Inception v4/Inception-ResNet (2016) has approximately 55.8M parameters.

Inceptions

- GoogLeNet/Inception v1 (2014) has approximately 5M parameters.
- Accuracies on ImageNet: Top1=69.8%, Top5=89.9%.
- Inception v2 (2015) has approximately 11.2M parameters.
- Accuracies on ImageNet: Top1=74.8%, Top5=92.2%.
- Inception v3 (2015) has approximately 23.8M parameters.
- Accuracies on ImageNet: Top1=78.8%, Top5=94.4%.
- Inception v4/Inception-ResNet (2016) has approximately 55.8M parameters.
- Accuracies on ImageNet: Top1=80.1%, Top5=95.1%.

Inceptions

- GoogLeNet/Inception v1 (2014) has approximately 5M parameters.
- Accuracies on ImageNet: Top1=69.8%, Top5=89.9%.
- Inception v2 (2015) has approximately 11.2M parameters.
- Accuracies on ImageNet: Top1=74.8%, Top5=92.2%.
- Inception v3 (2015) has approximately 23.8M parameters.
- Accuracies on ImageNet: Top1=78.8%, Top5=94.4%.
- Inception v4/Inception-ResNet (2016) has approximately 55.8M parameters.
- Accuracies on ImageNet: Top1=80.1%, Top5=95.1%.
- The best result on ImageNet: Top1=86.4%, Top5=98%

Inceptions

- GoogLeNet/Inception v1 (2014) has approximately 5M parameters.
- Accuracies on ImageNet: Top1=69.8%, Top5=89.9%.
- Inception v2 (2015) has approximately 11.2M parameters.
- Accuracies on ImageNet: Top1=74.8%, Top5=92.2%.
- Inception v3 (2015) has approximately 23.8M parameters.
- Accuracies on ImageNet: Top1=78.8%, Top5=94.4%.
- Inception v4/Inception-ResNet (2016) has approximately 55.8M parameters.
- Accuracies on ImageNet: Top1=80.1%, Top5=95.1%.
- The best result on ImageNet: Top1=86.4%, Top5=98% with 829M parameters.