

Deep Into Deep

김성찬

Contents

1. Image is Large Input
2. Convolutional Neural Networks
3. What is Convolution?
4. Convolution Layer
5. Receptive Fields
6. Hierarchy of Convolution Layer
7. A Closer Look at Spatial Dimensions
8. The Brain/Neuron View of Conv Layer
9. Implementation as Matrix Multiplication
10. Backpropagation of Convolutional Layer
11. FC Layer vs. Convolutional Layer
12. Pooling Layer

1. Image is Large Input.



Image Classification Task에서
모델이 받을 입력의 크기는 얼마나 될까?

가로 픽셀 수 : 1024

세로 픽셀 수 : 1024

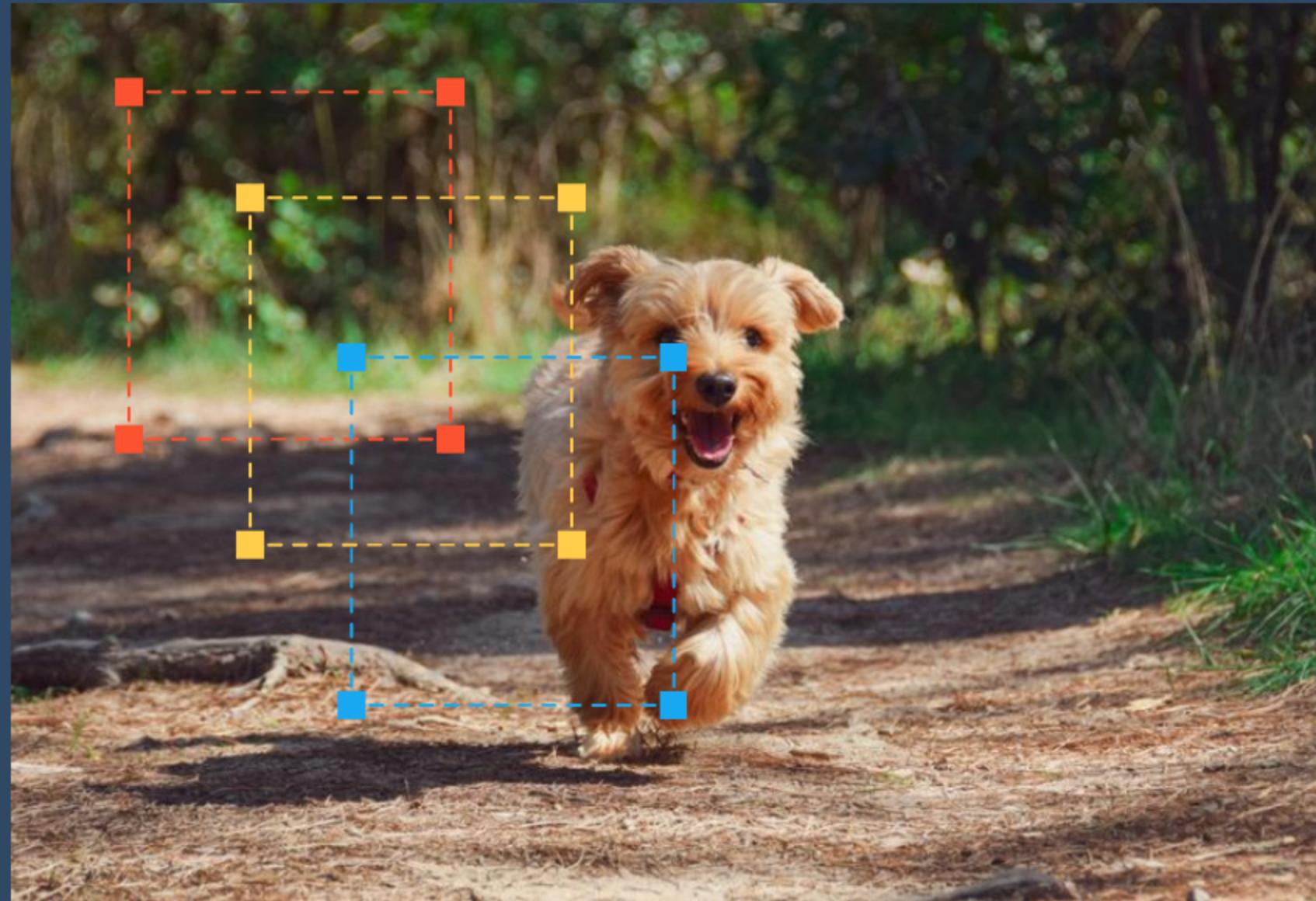
채널 수 : 3

채널이란: 이 이미지에 대한 해석, feature

총 입력의 크기 : $3 * 1024 * 1024 = 3,145,728$

이렇게 큰 입력을 MLP로 잘 처리할 수 있을까?

2. Convolutional Neural Networks



Local Processing of Data

1. Parameter Sharing

같은 필터를 서로 다른 Local에 적용한다.

2. Sparsity of Connections

하나의 이미지를 그대로 처리하지 않고
Local별로 처리한다.

3. What is Convolution?

ConvNet : ConvNet은 여러 Layer들을 모아 놓은 네트워크다. 이때 네트워크를 구성하는 Layer 중에 Convolution Layer가 있다.

ConvNet 기본 구성

1. Convolution Layer :

입력 내 local region들과 연관된 출력을 계산한다.

2. Pooling Layer :

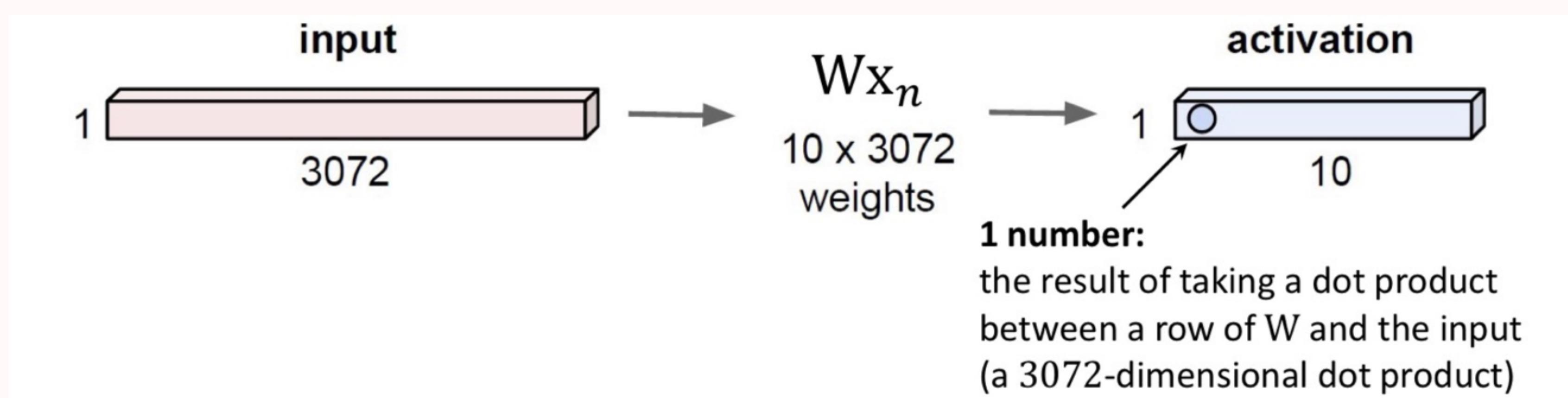
Downsampling을 한다.

3. Fully-Connected Layer :

입력과 출력이 모두 연관된다.

3. What is Convolution?

MLP or FC Layer : $32 \times 32 \times 2$ image \rightarrow stretch to 3072×1



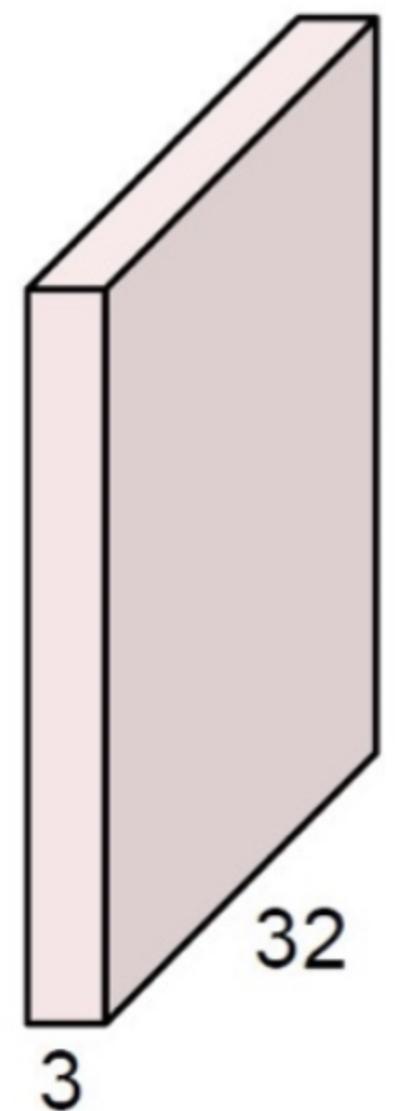
2D 이미지를 1D 벡터로 만드는 것은 2D 이미지의 spatial context를 고려하지 못한다.

작은 사이즈의 이미지를 처리하는 데에도 너무 많은 모델 파라미터가 필요하다.

4. Convolutional Layer

Convolution Layer : $32 \times 32 \times 3$ 이미지의 spatial structure을 유지한다.

$32 \times 32 \times 3$ image



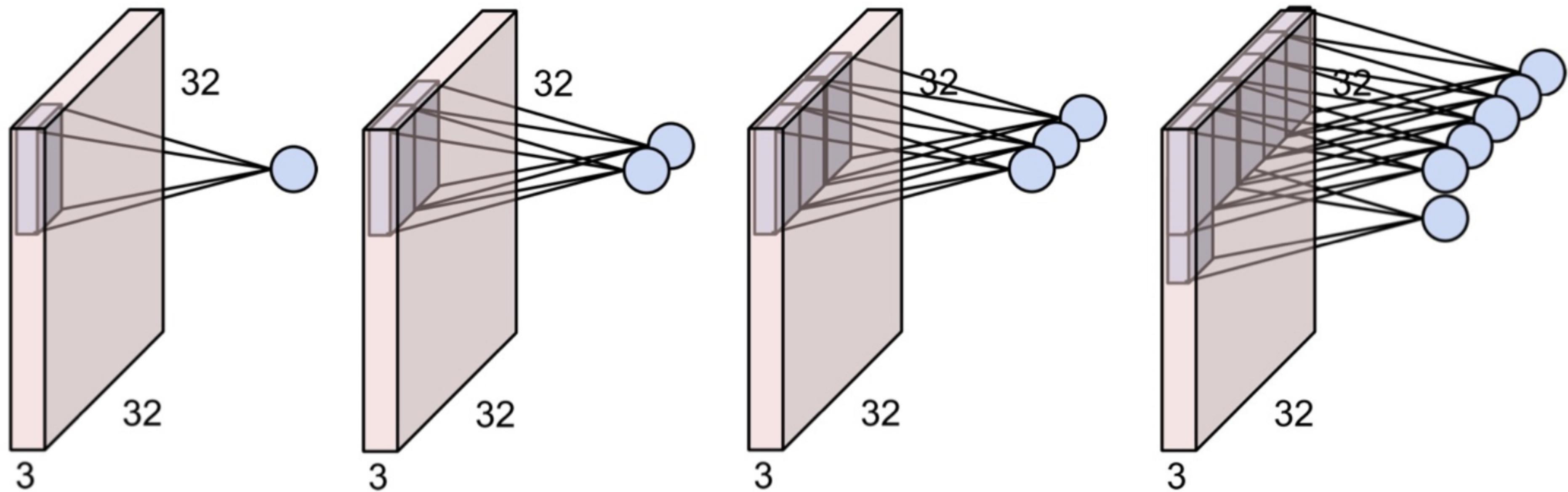
$5 \times 5 \times 3$ filter



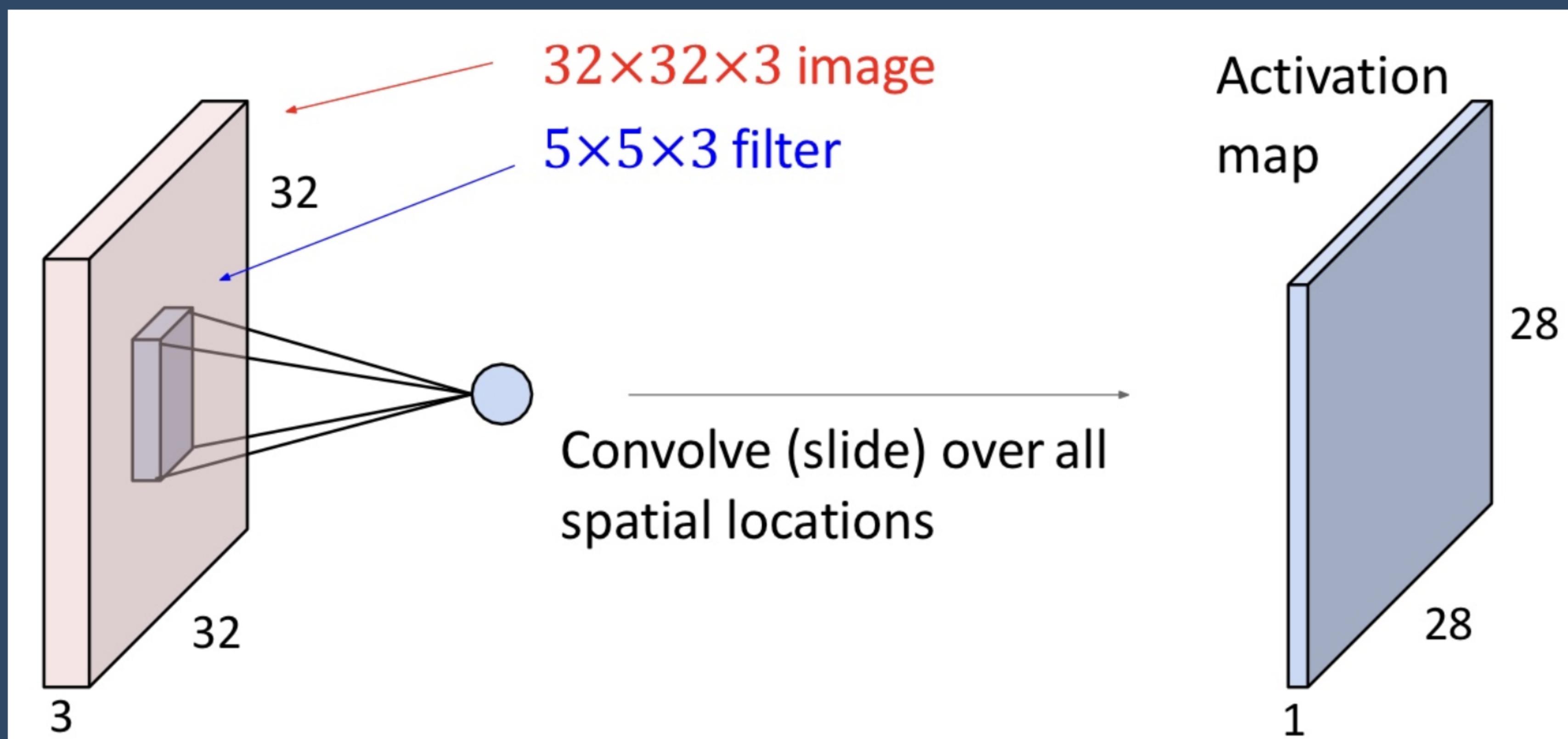
필터와 이미지의 채널 수는 항상 동일해야 한다.

Convolve the filter with the image,
i.e., “slide over the image spatially,
computing dot products”

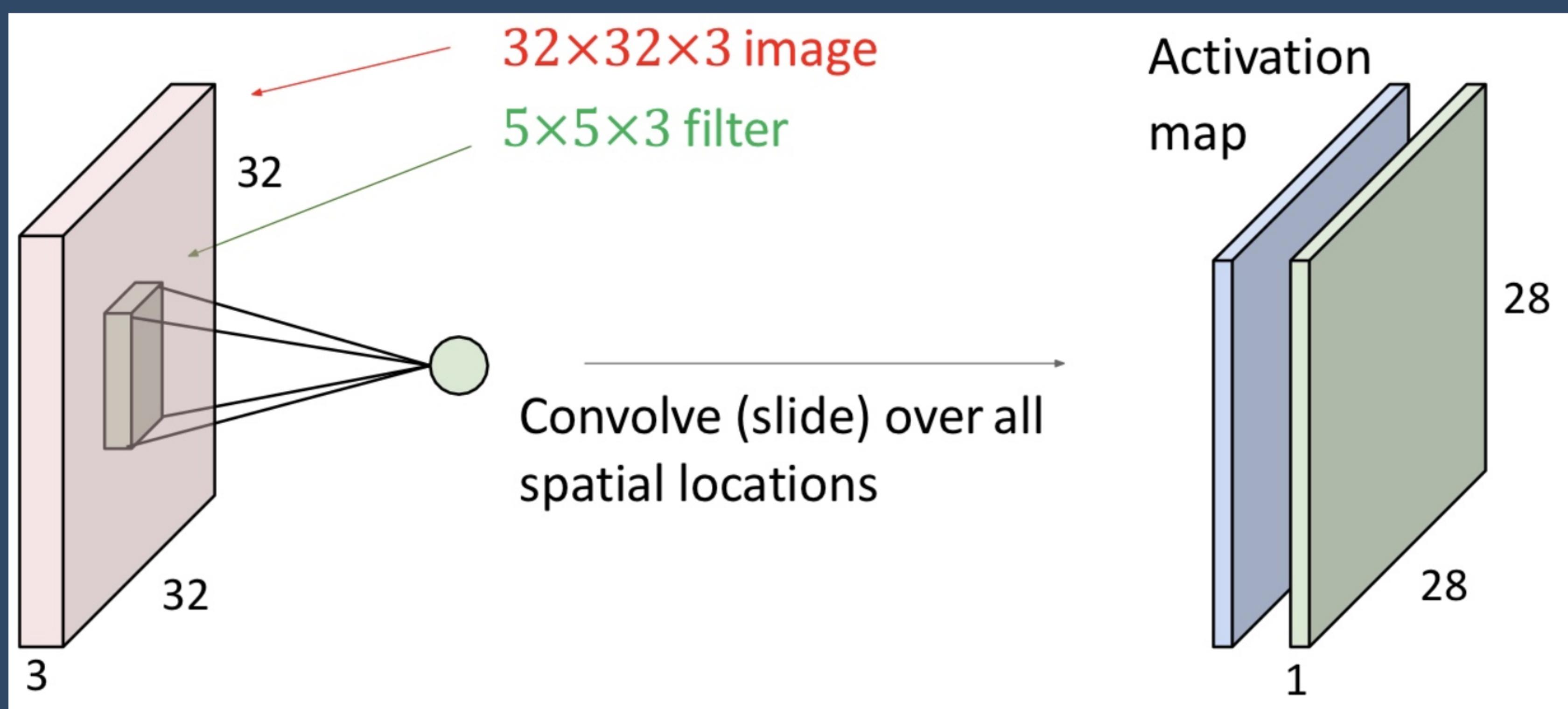
4. Convolutional Layer



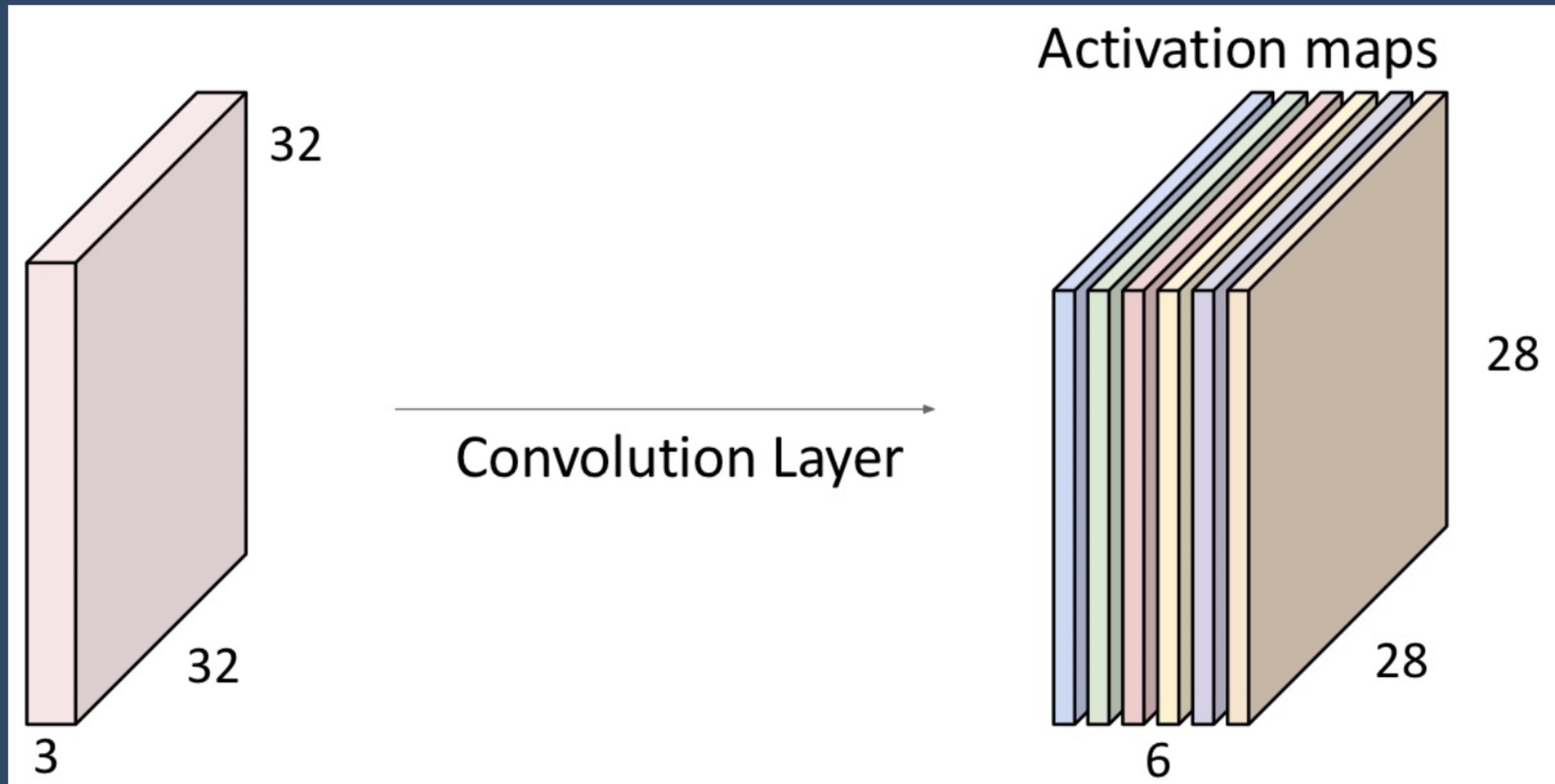
4. Convolutional Layer



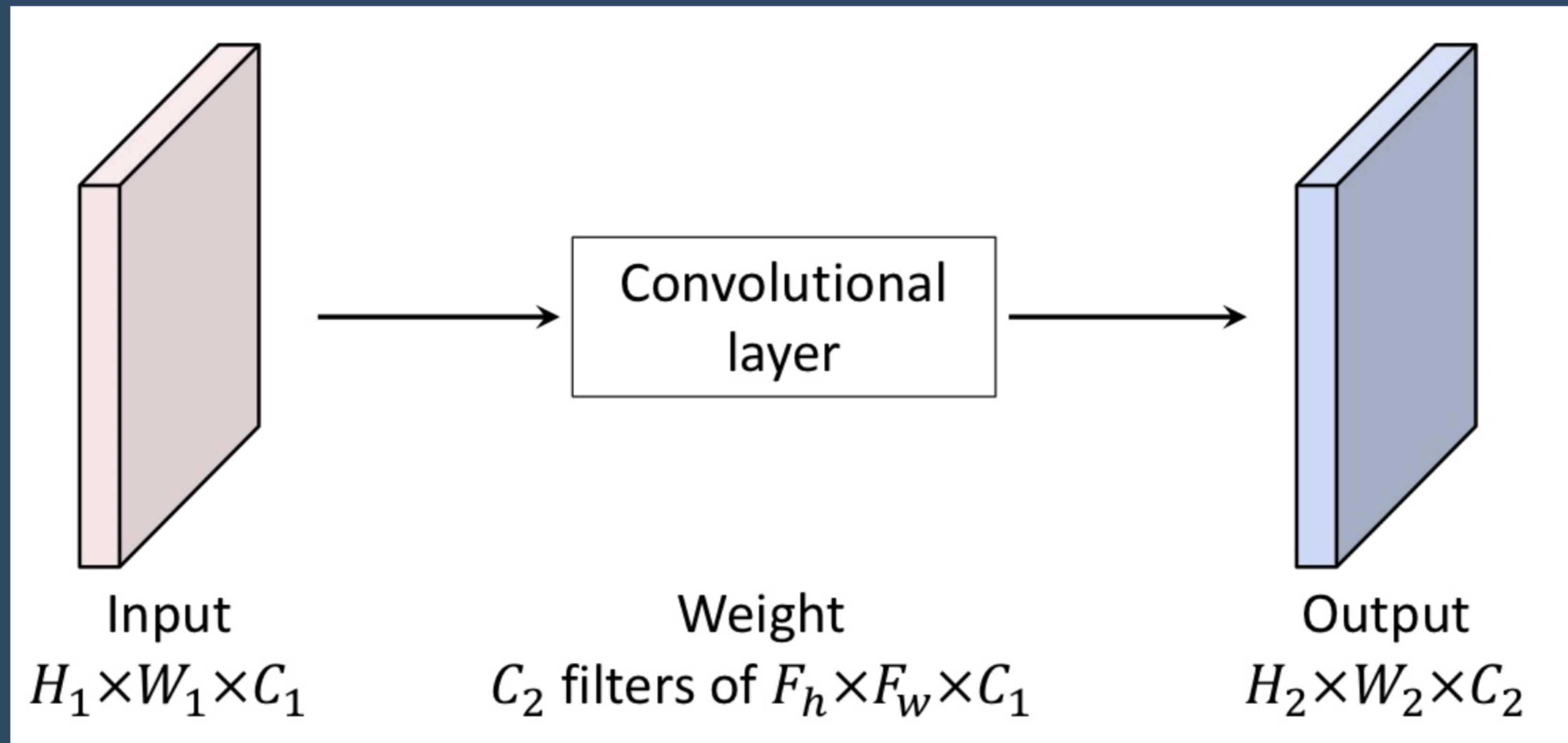
4. Convolutional Layer



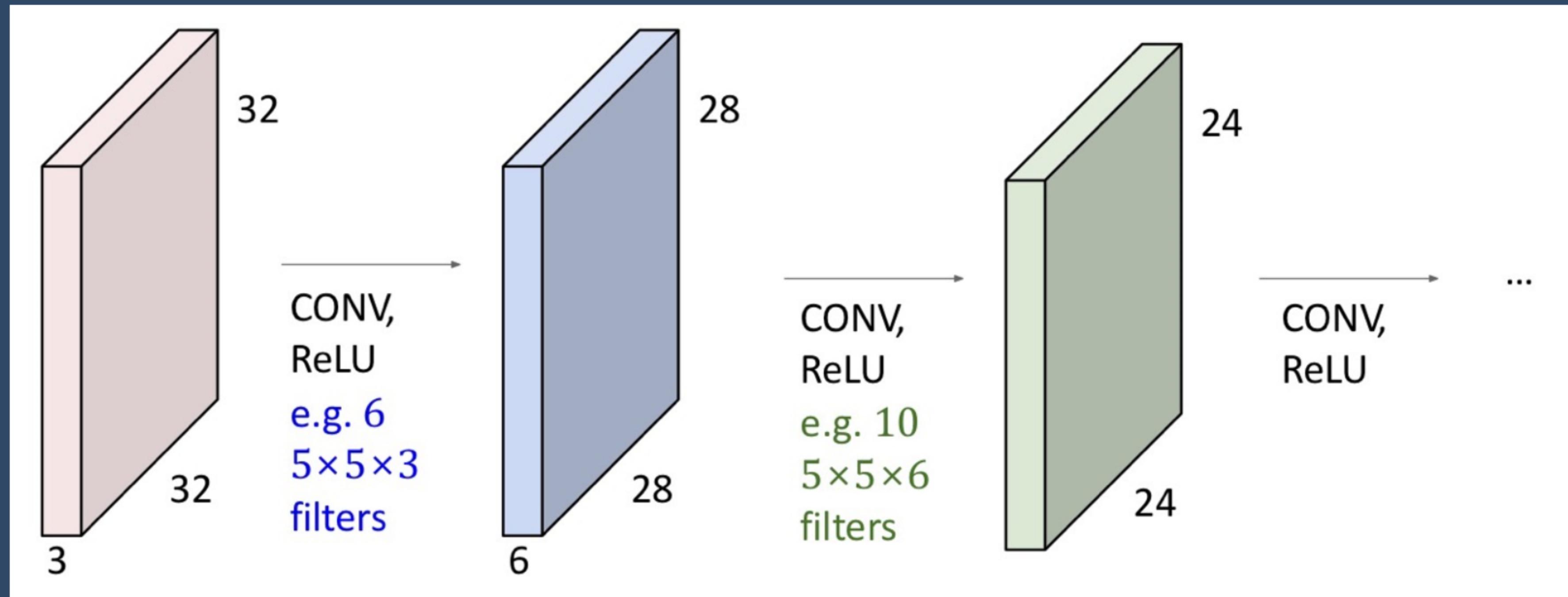
4. Convolutional Layer



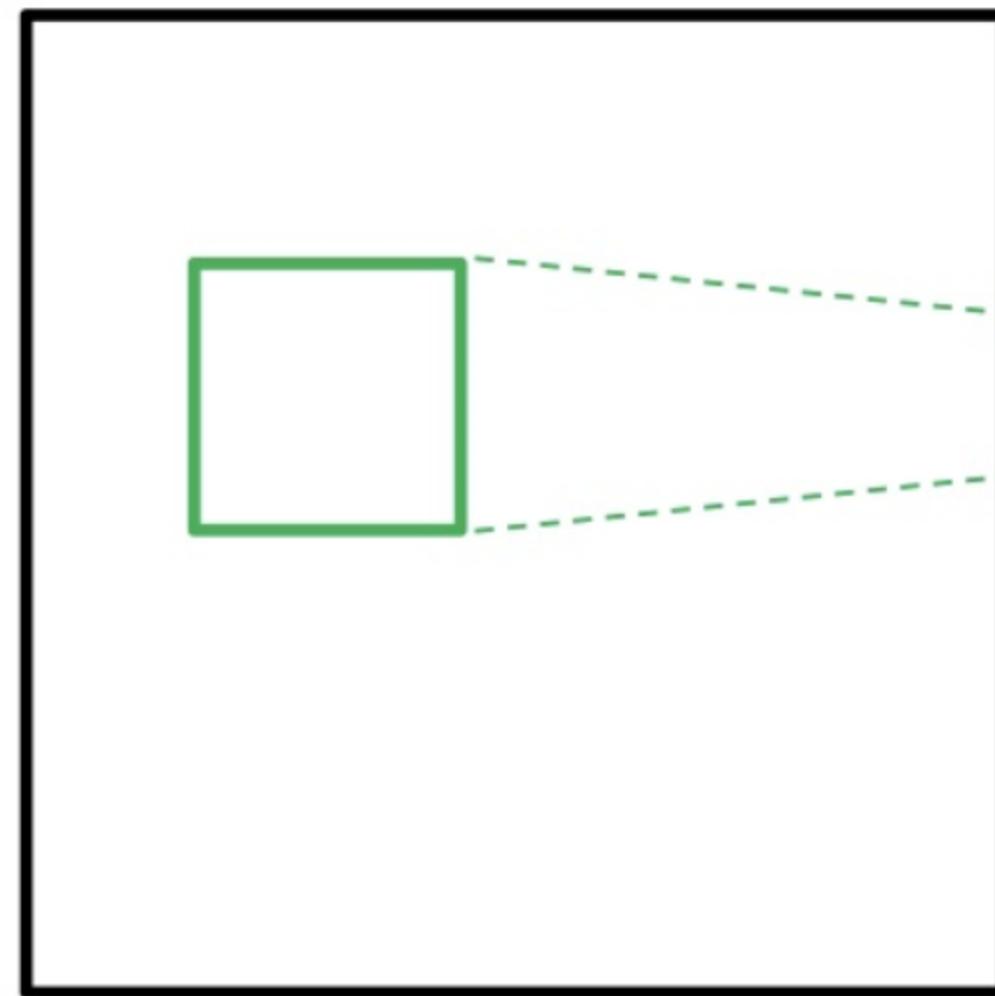
4. Convolutional Layer



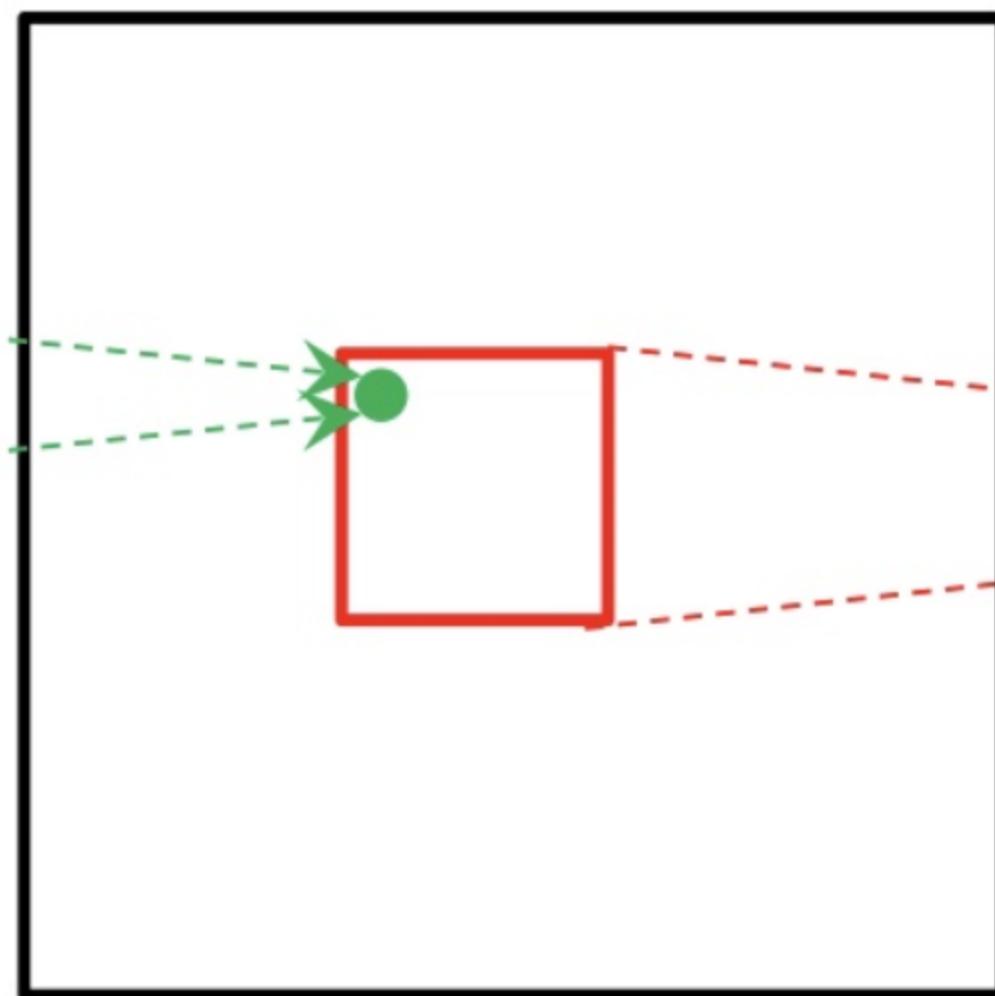
4. Convolutional Layer



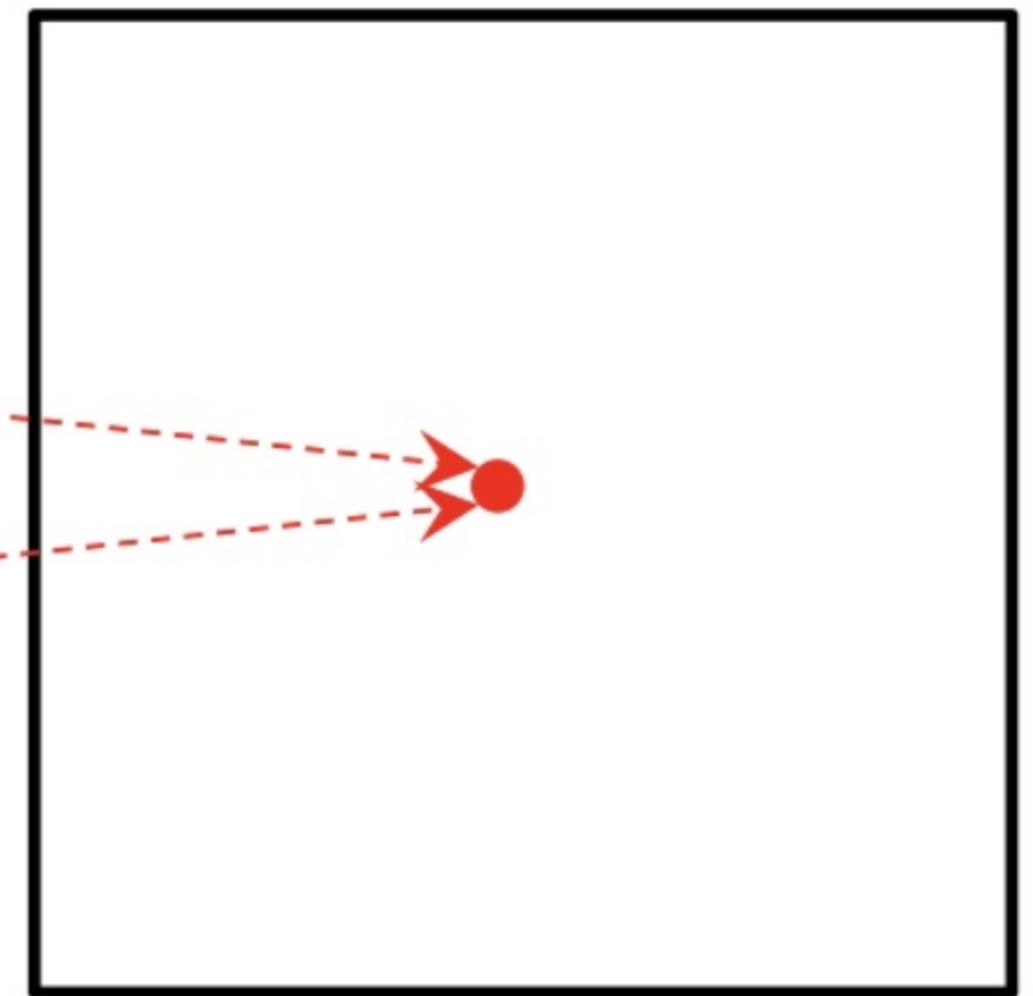
5. Receptive Fields



$5 \times 5 \times 1$ filter



$5 \times 5 \times 1$ filter



5. Receptive Fields : Hierarchy of Convolution Layers

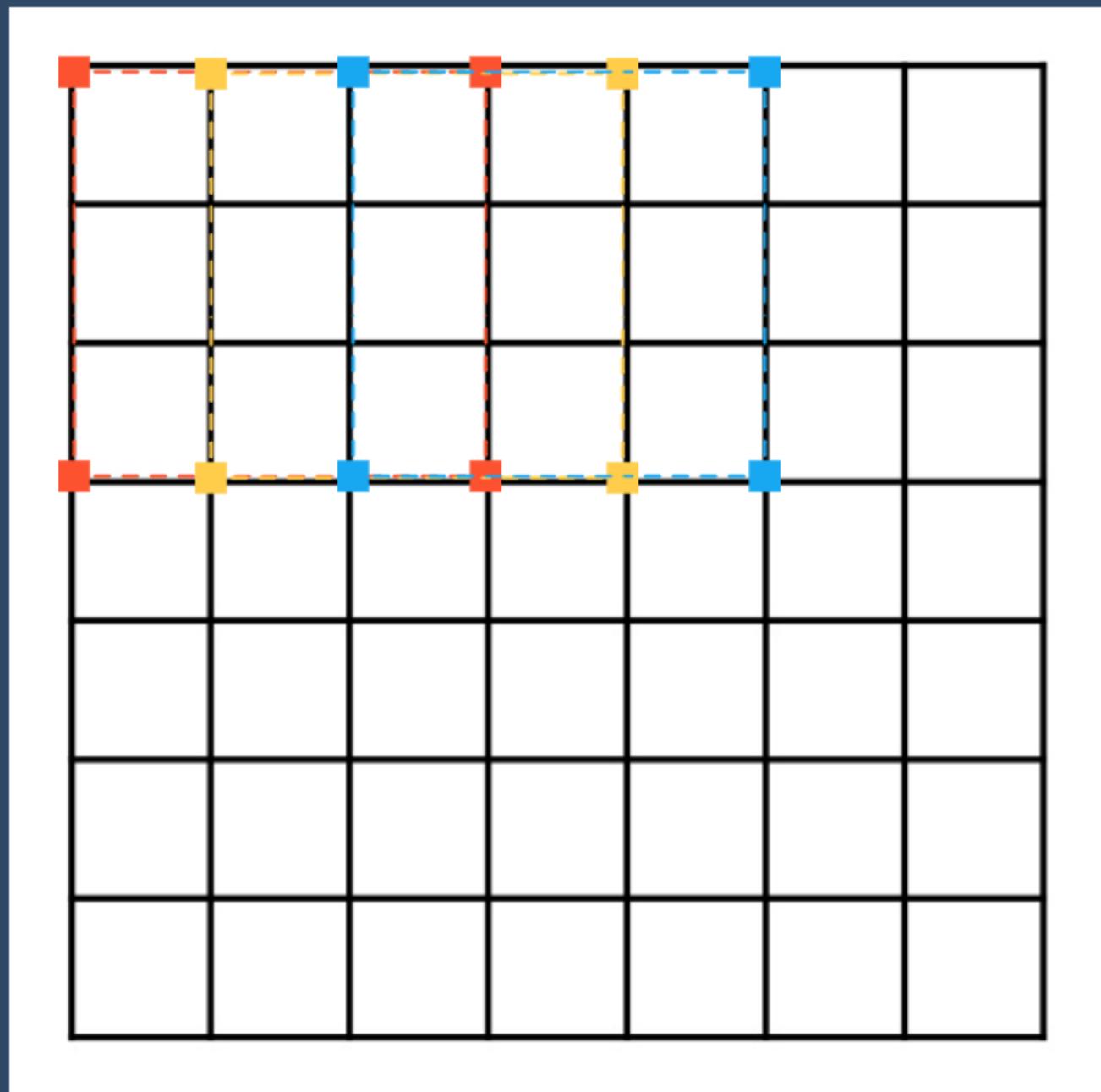
Visualizing and Understanding Convolutional Networks

<https://arxiv.org/abs/1311.2901>

6. A Closer Look at Spatial Dimensions

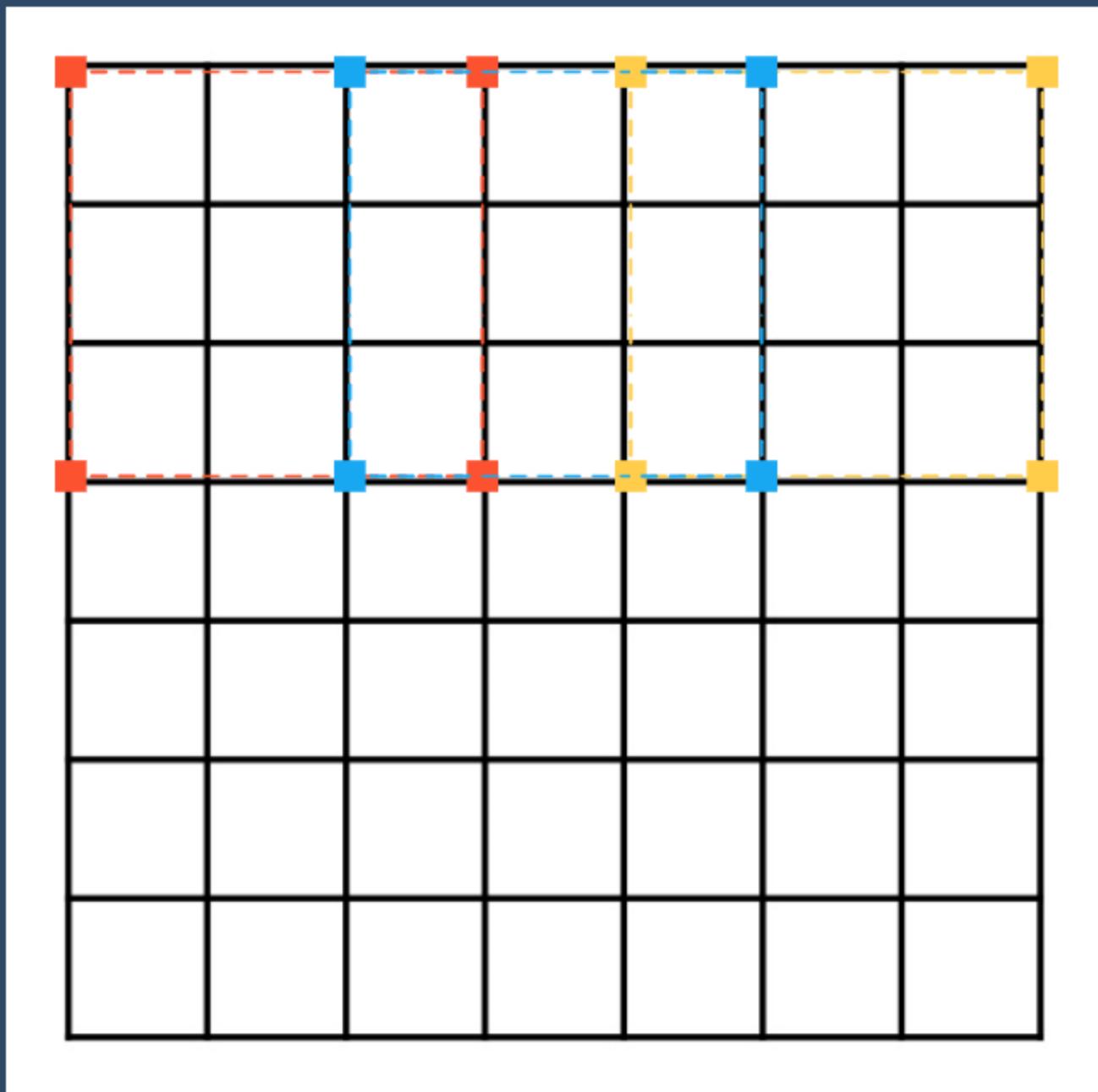
7x7 input, 3x3 filter, with stride 1

stride : 몇 칸을 건너뛸건지



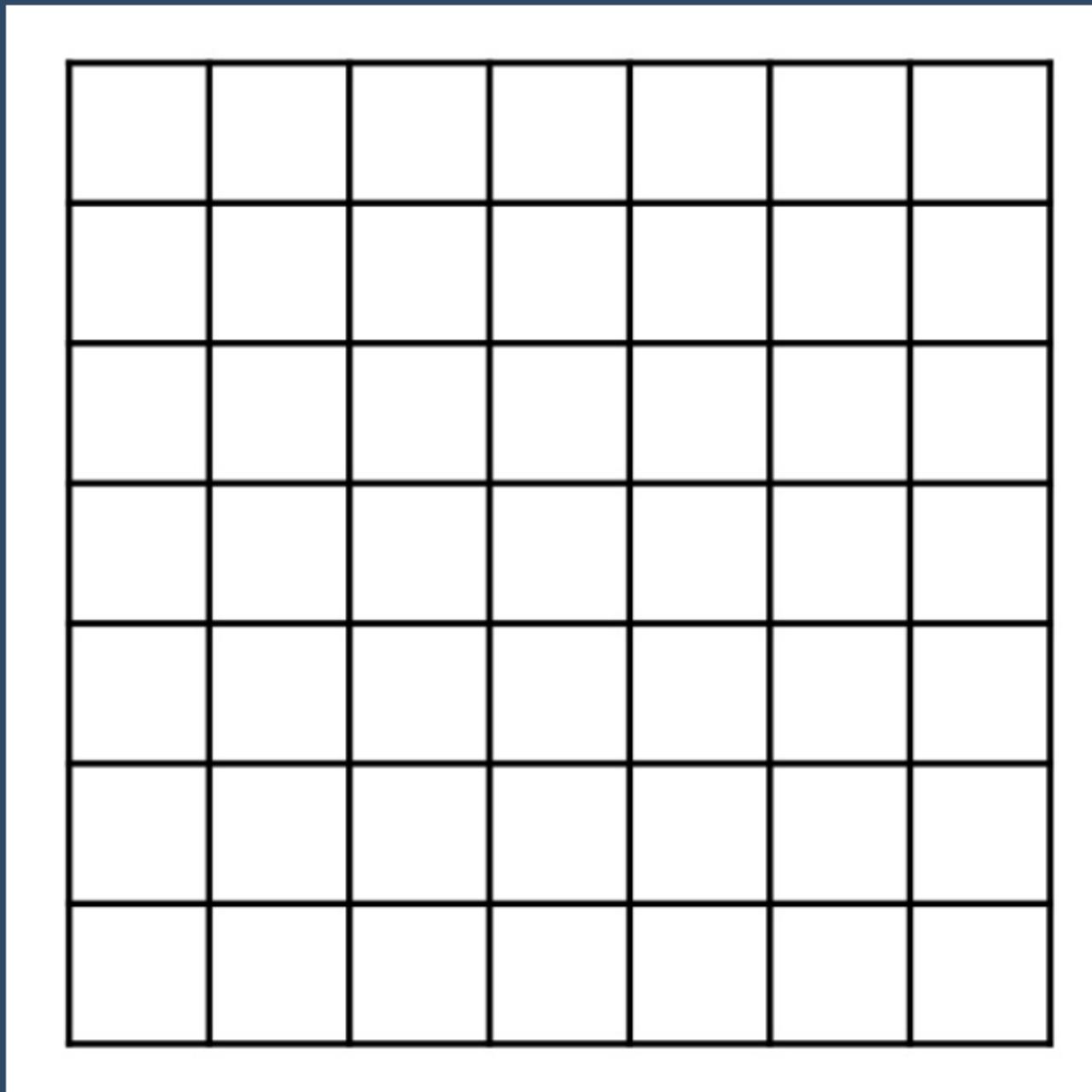
6. A Closer Look at Spatial Dimensions

7x7 input, 3x3 filter, with stride 2



6. A Closer Look at Spatial Dimensions

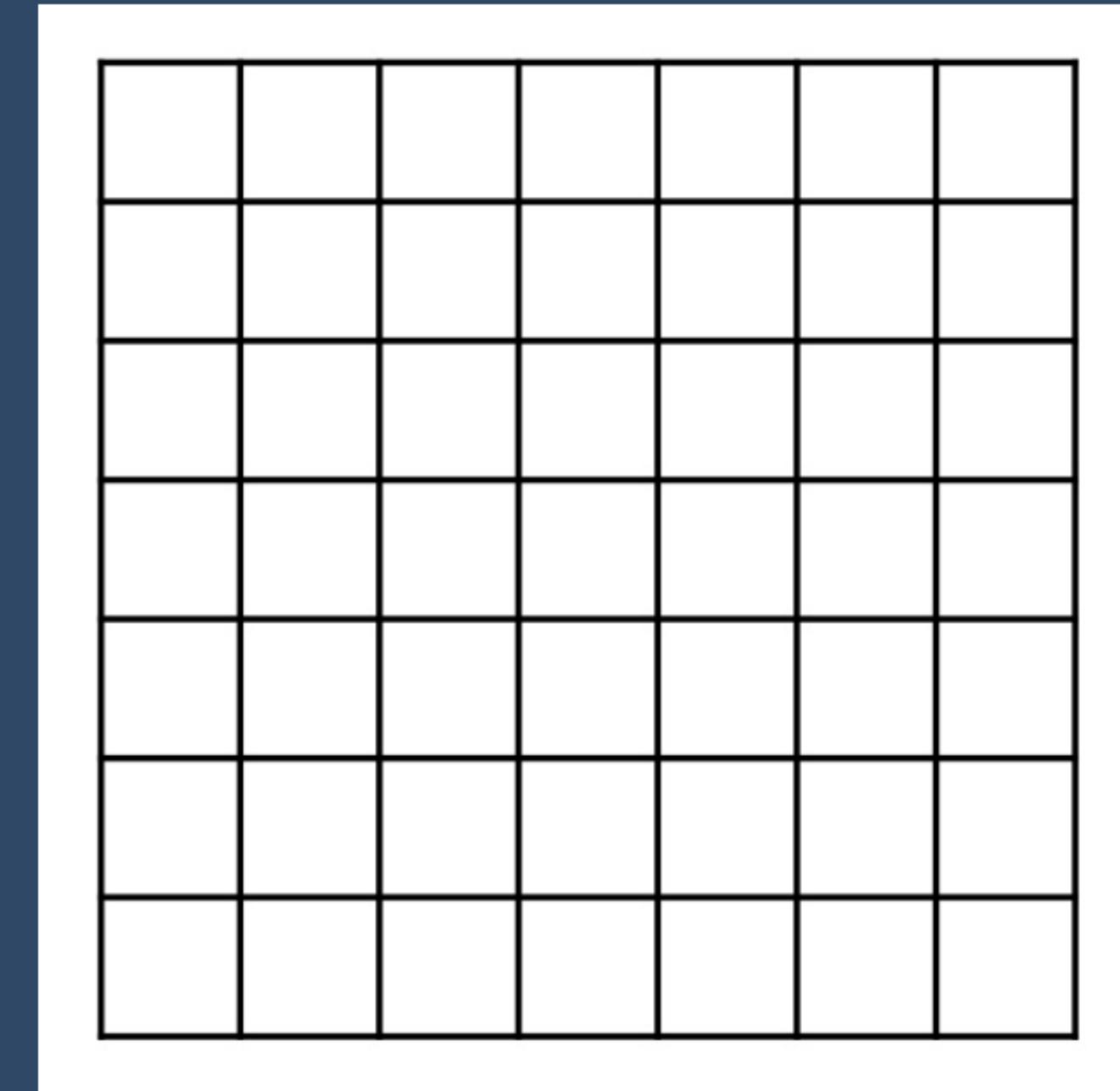
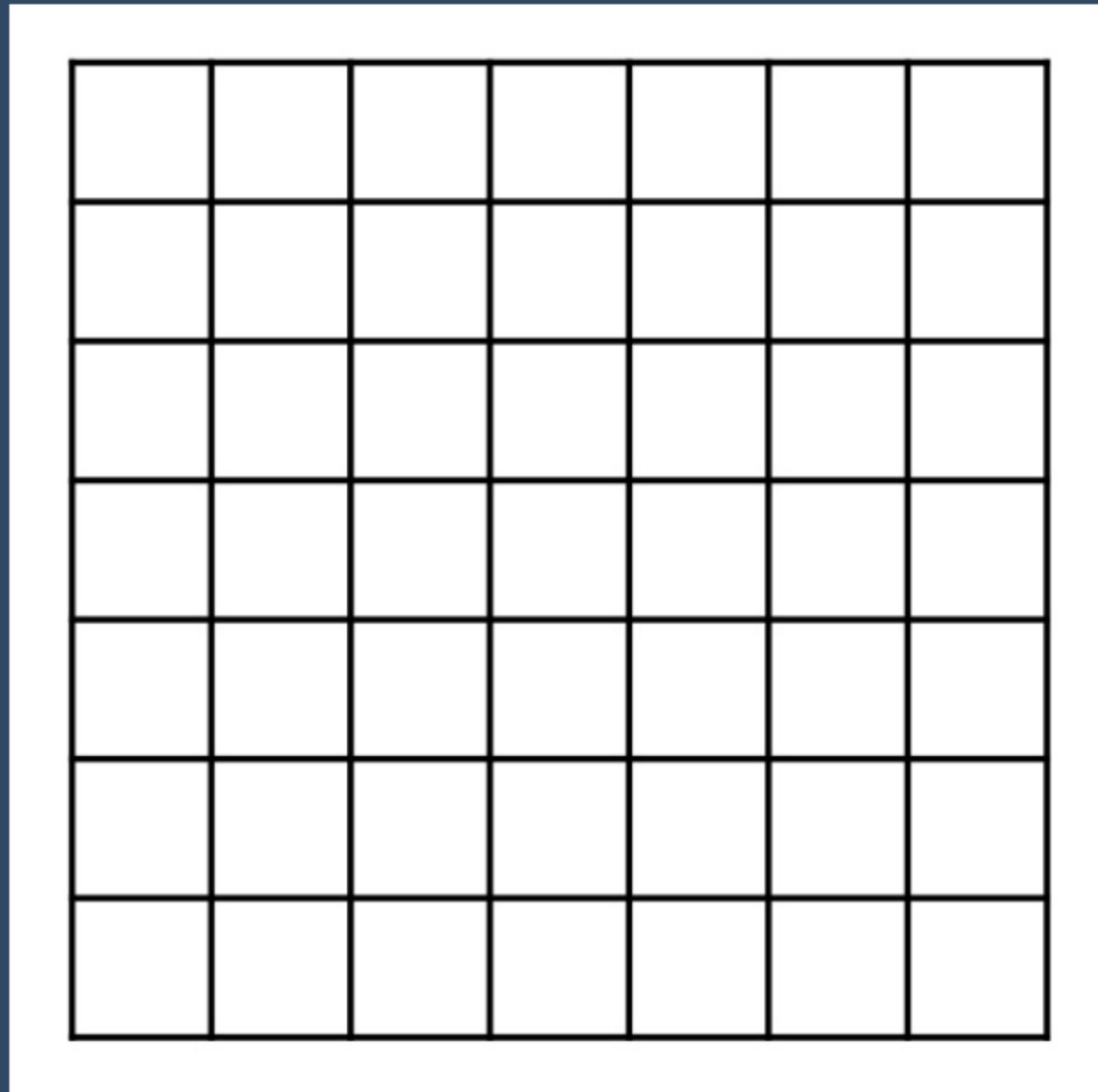
$N \times N$ input, $F \times F$ filter, with stride S



$$(N - F) / S + 1$$

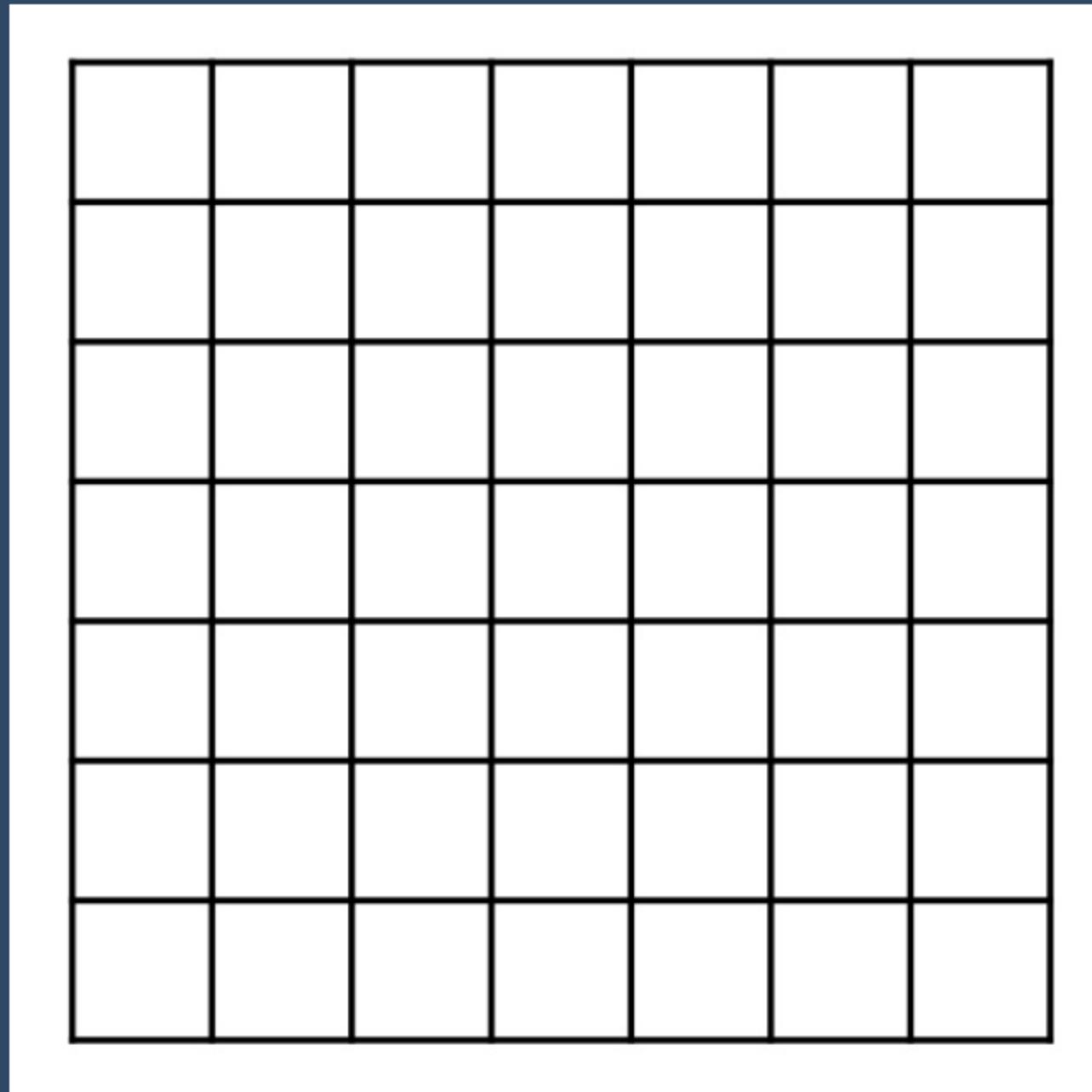
6. A Closer Look at Spatial Dimensions

$N \times N$ input, 3×3 filter, with stride 1, with 1 pixel padding



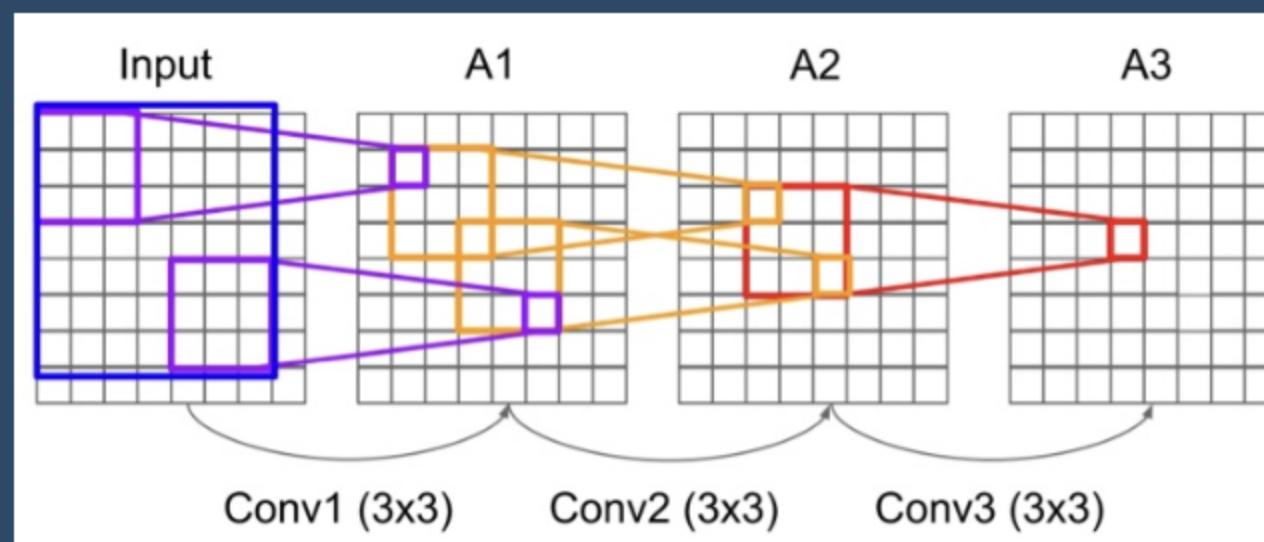
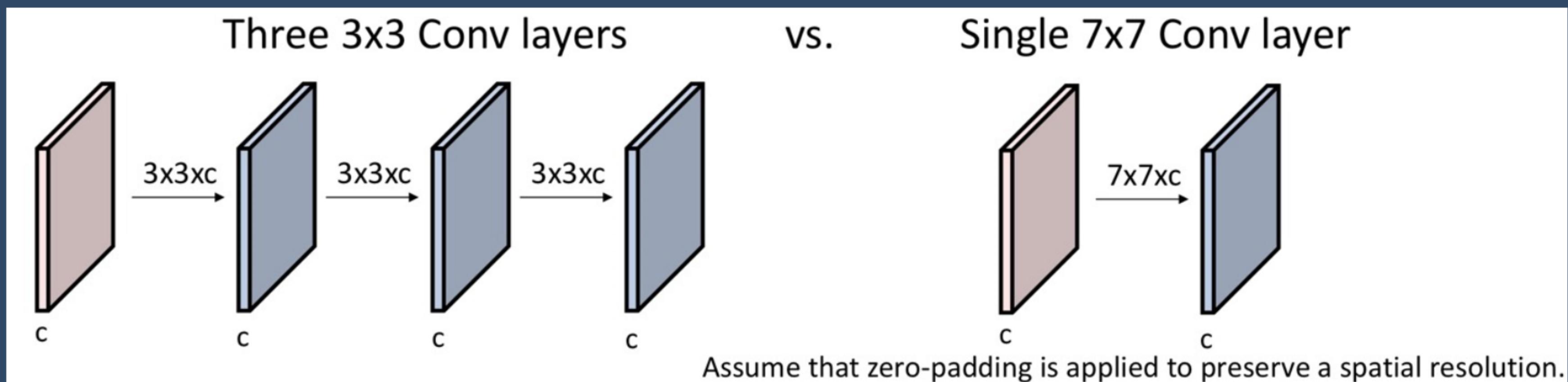
6. A Closer Look at Spatial Dimensions

$N \times N$ input, $F \times F$ filter, with stride S , with P pixel padding



$$(N - F + 2P) / S + 1$$

6. A Closer Look at Spatial Dimensions

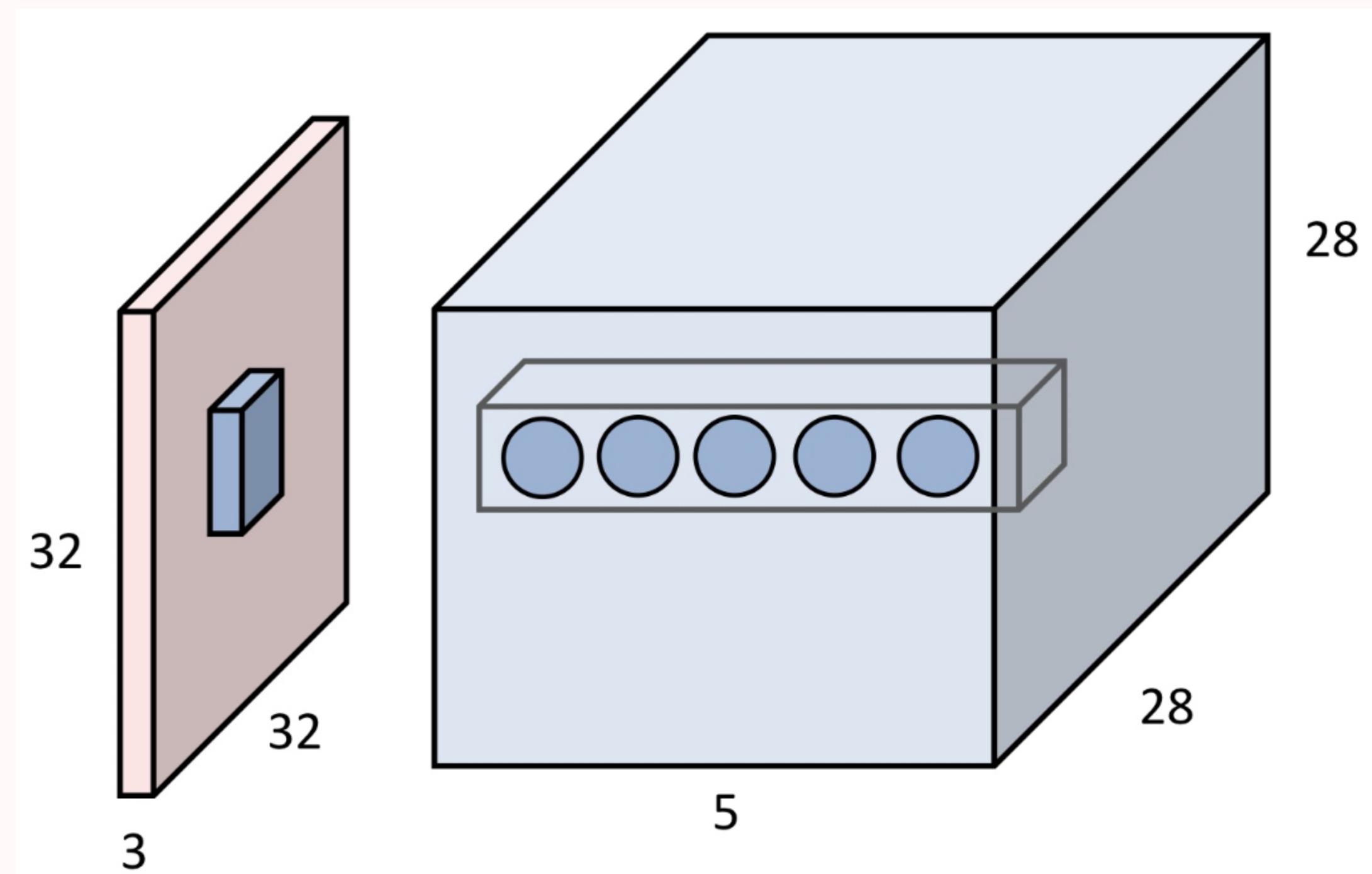


Receptive fields are equal.

of Conv parameters

$$3 \times C \times (3 \times 3 \times C) = 27 C^2 < C \times (7 \times 7 \times C) = 49 C^2$$

7. The Brain/Neuron View of Conv Layer



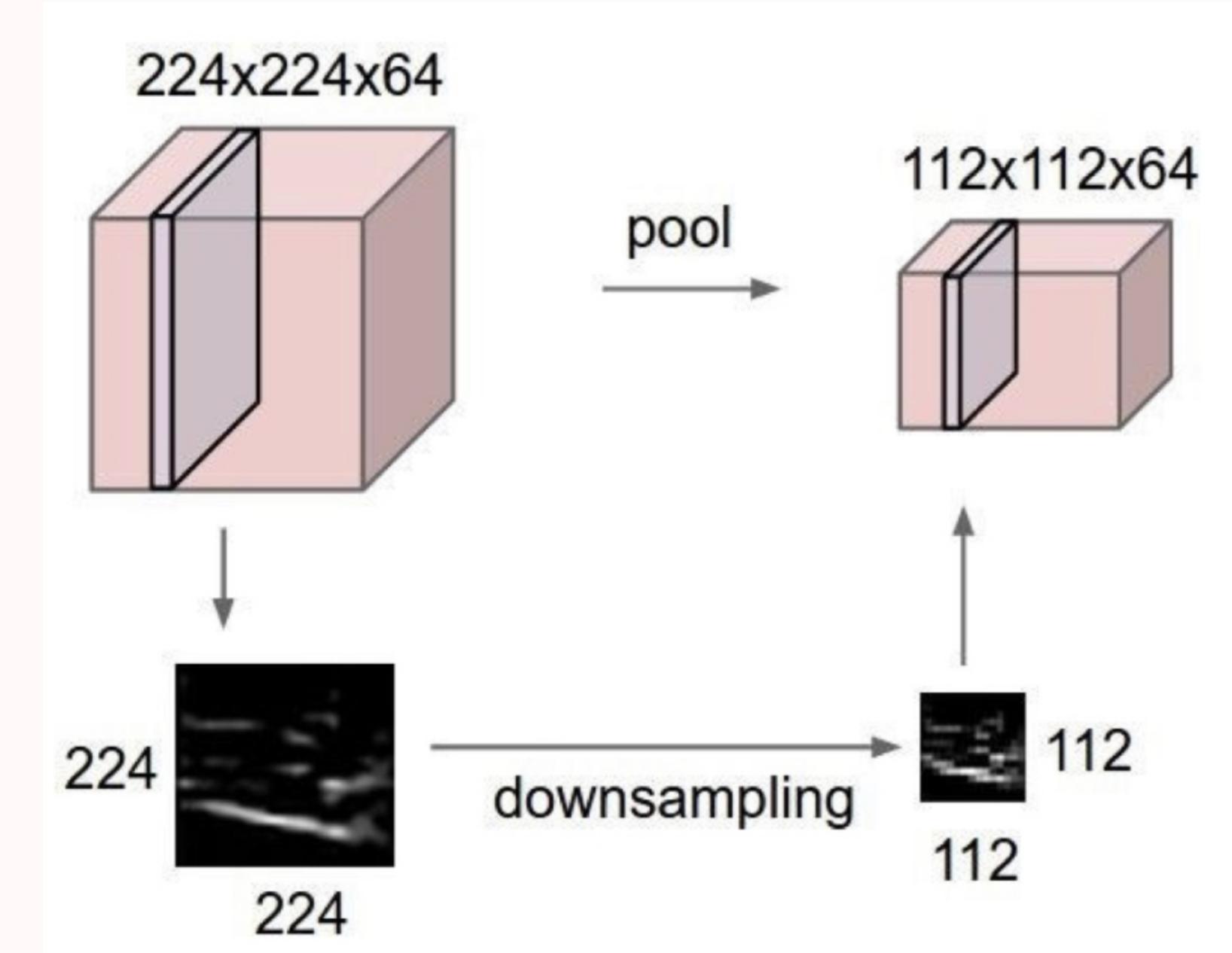
채널이 가지는 의미는 무엇일까?

채널은 특정 입력에 대한
여러가지 해석이라고
생각할 수 있다.

11. Pooling Layer

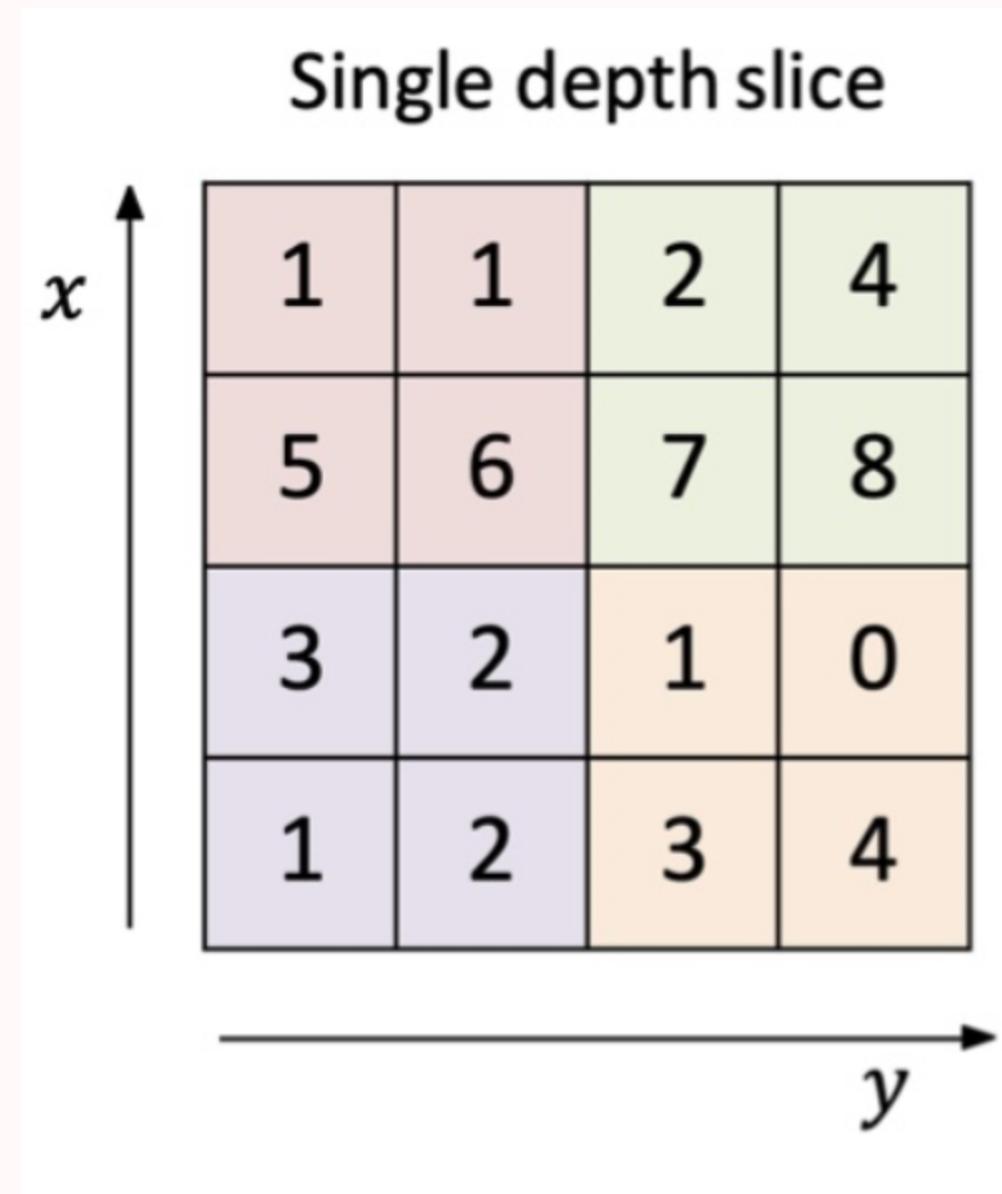
Makes the representations smaller and more manageable.

Operates over each activation map independently.



11. Pooling Layer

Max Pooling



max pool with 2×2 filters and stride 2

6	8
3	4

How to compute
Gradient of Max Pooling?

Thank you for listening

Deep Into Deep