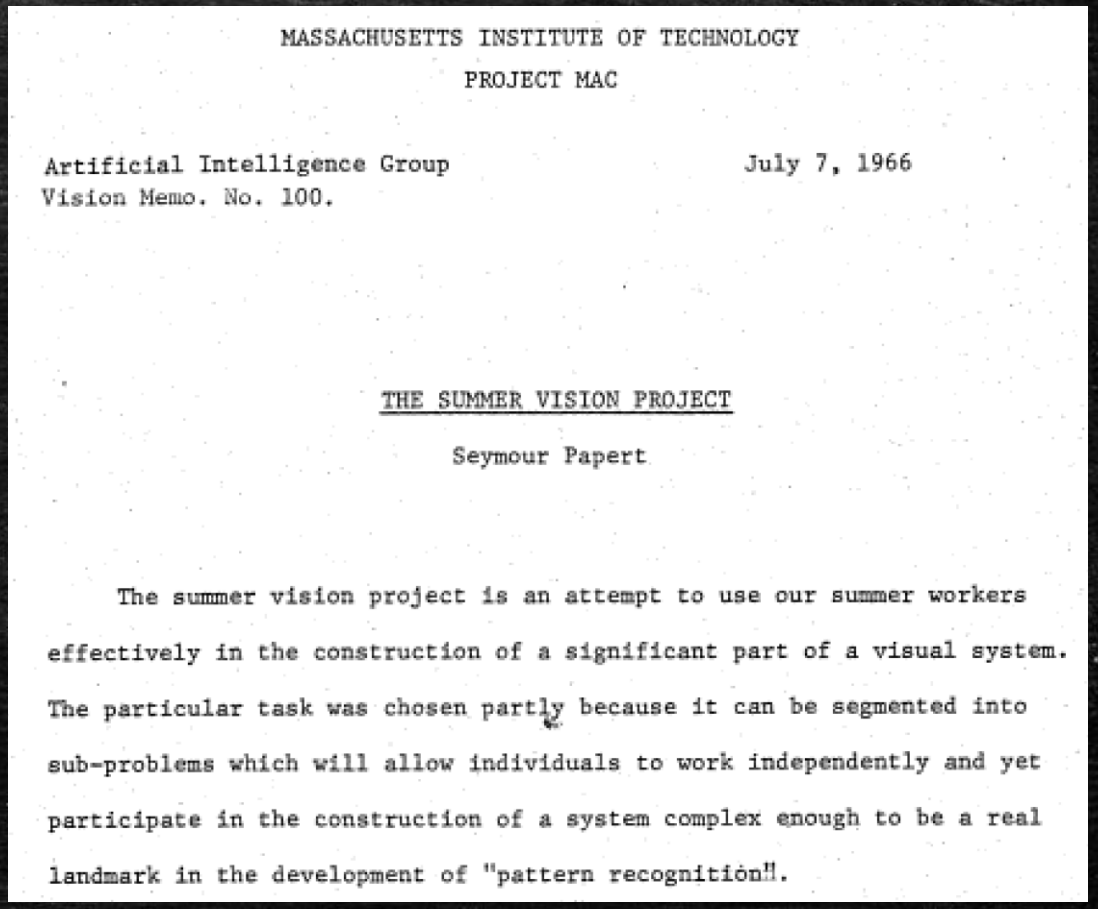
Convolutional Neural Network

Joe Yeh, M.D.

A Brief History on Computer Vision

MIT Summer Vision Project

...in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".



MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

General Goals of MIT Summer Vision Project

Goals - General

The primary goal of the project is to construct a system of programs which will divide a vidisector picture into regions such as

likely objects

likely background areas

chaos.

We shall call this part of its operation FIGURE-GROUND analysis.

It will be impossible to do this without considerable analysis of shape and surface properties, so FIGURE-GROUND analysis is really inseparable in practice from the second goal which is REGION DESCRIPTION.

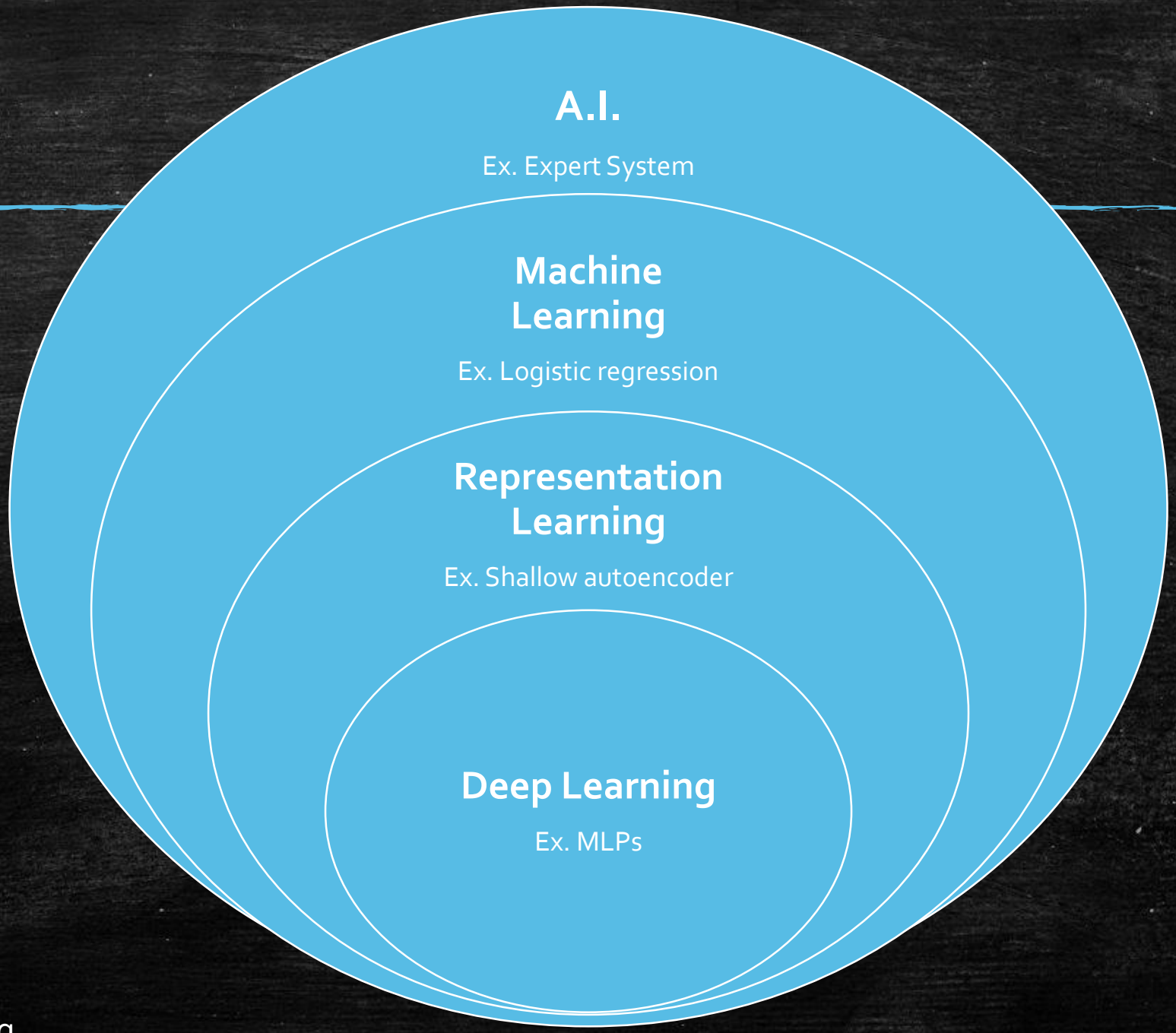
The final goal is OBJECT IDENTIFICATION which will actually name objects by matching them with a vocabulary of known objects.

Artificial Intelligence : The beginning

- Dartmouth Summer Research Project on Artificial Intelligence (1959)
  - Proposed by John McCarthy, Marvin Minsky, Nathaniel Rochester and Claude Shannon
  - to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.

AI@50 From left to right: Trenchard More, John McCarthy, Marvin Minsky, Oliver Selfriege, Ray Solomonoff

A.I and Deep Learning

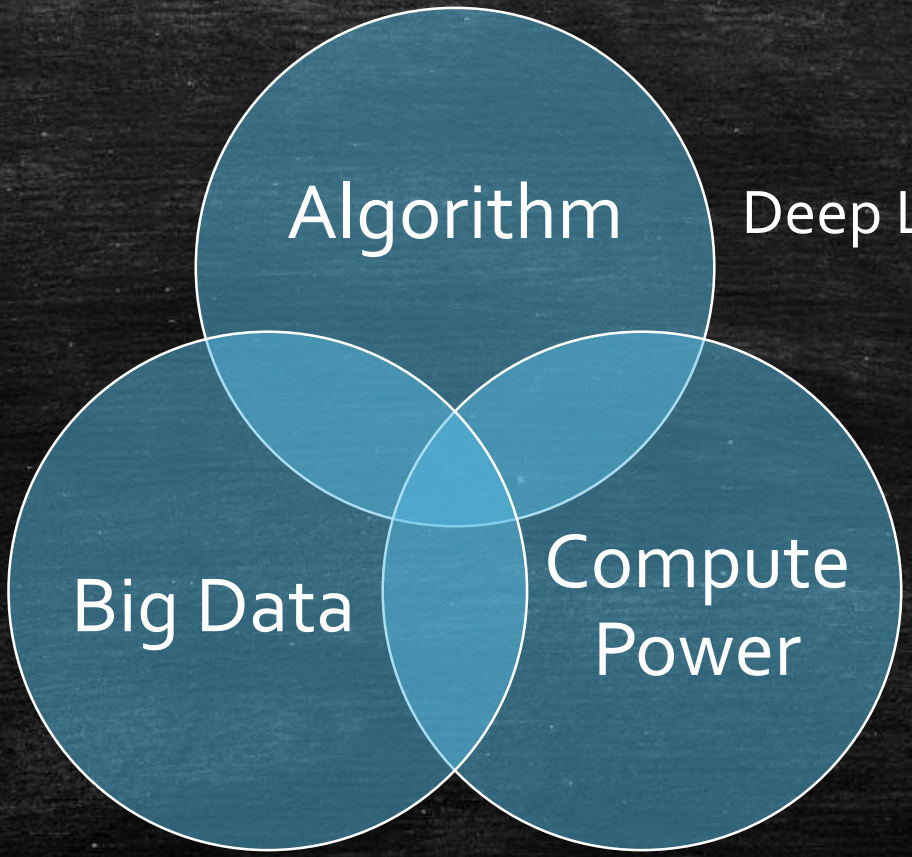


Ian Goodfellow et al. Deep Learning

Three A.I. winters

- Machine translation : 1950-1960s
  - Georgetown Experiment showing Russian to English translation 1954
  - Automated Language Processing Advisory Committee says progress is slow 1966
- Making AI in a controlled environment: 1970s
  - Teaching AI to perform task in micro world
  - Chatbot for talk therapy
  - 1974 UK Lighthill report : …utter failure of AI to achieve its grandiose objectives
- Expert systems : 1980s
  - Symbolic Lisp machines, IBM's Integrated Reasoning Shell
  - Collapse of Symbolics

What's different this time ?



Deep Learning !

Internet,
Digitization

Compute accelerators
GPU, TPU, NPU

What is Convolutional Neural Network ?

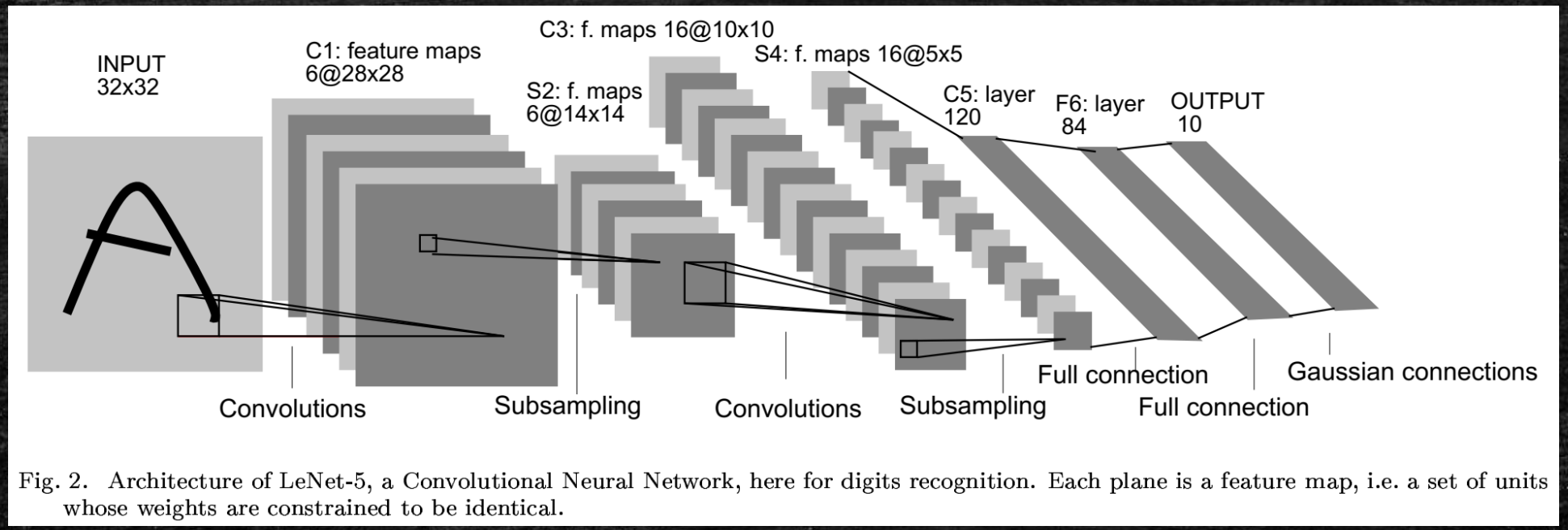


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

Landmark CNN Architectures

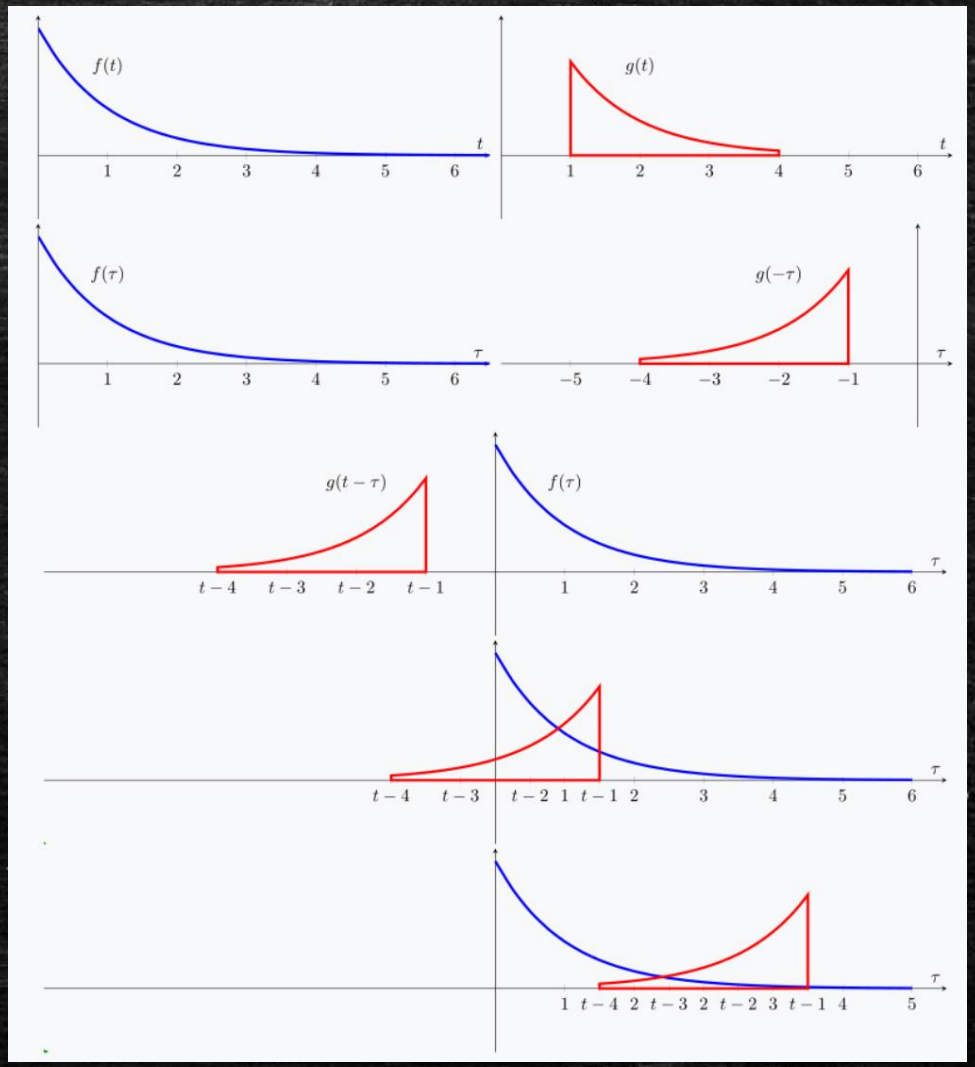
- LeNet (1998, U. Montreal, LeCun, Bengio)
- AlexNet (2012, U. Toronto, Krizhevsky, Hinton)
- VGG (2014, Oxford)
- Inception (2014, Google)
- ResNet (2015, Microsoft)
- DenseNet(2017, Facebook)

Computer Vision Tasks

- Image classification (categorical output)
  - Example : chest x-ray → diagnosis of pneumonia
- Image regression (continuous real number output)
  - Example : CT image→ bone age
- Object detection
  - Example : Lung CT image → 3-D bounding box enclosing tumor
- Image segmentation
  - Example : Brain MR image → contour of tumor

What is Convolution?

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau$$



1D Convolution : step by step

```
Result
Padded original signal  0 0 0 1 0 0 0 0
Kernel                  1 2 3
```

```
Result                  0
Padded original signal  0 0 0 1 0 0 0 0
Kernel                    1 2 3
```

```
Result                  0 0
Padded original signal  0 0 0 1 0 0 0 0
Kernel                    1 2 3
```

```
Result                  0 0 3
Padded original signal  0 0 0 1 0 0 0 0
Kernel                      1 2 3
```

```
Result                  0 0 3 2
Padded original signal  0 0 0 1 0 0 0 0
Kernel                        1 2 3
```

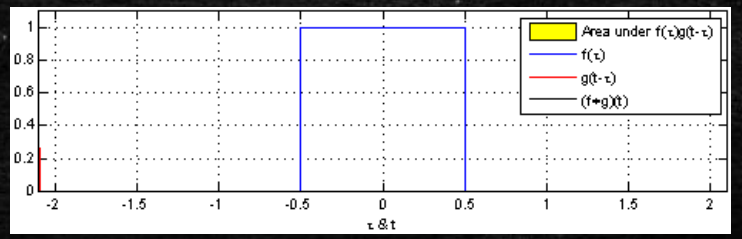
```
Result                  0 0 3 2 1
Padded original signal  0 0 0 1 0 0 0 0
Kernel                          1 2 3
```

```
Result                  0 0 3 2 1 0
Padded original signal  0 0 0 1 0 0 0 0
Kernel                            1 2 3
```

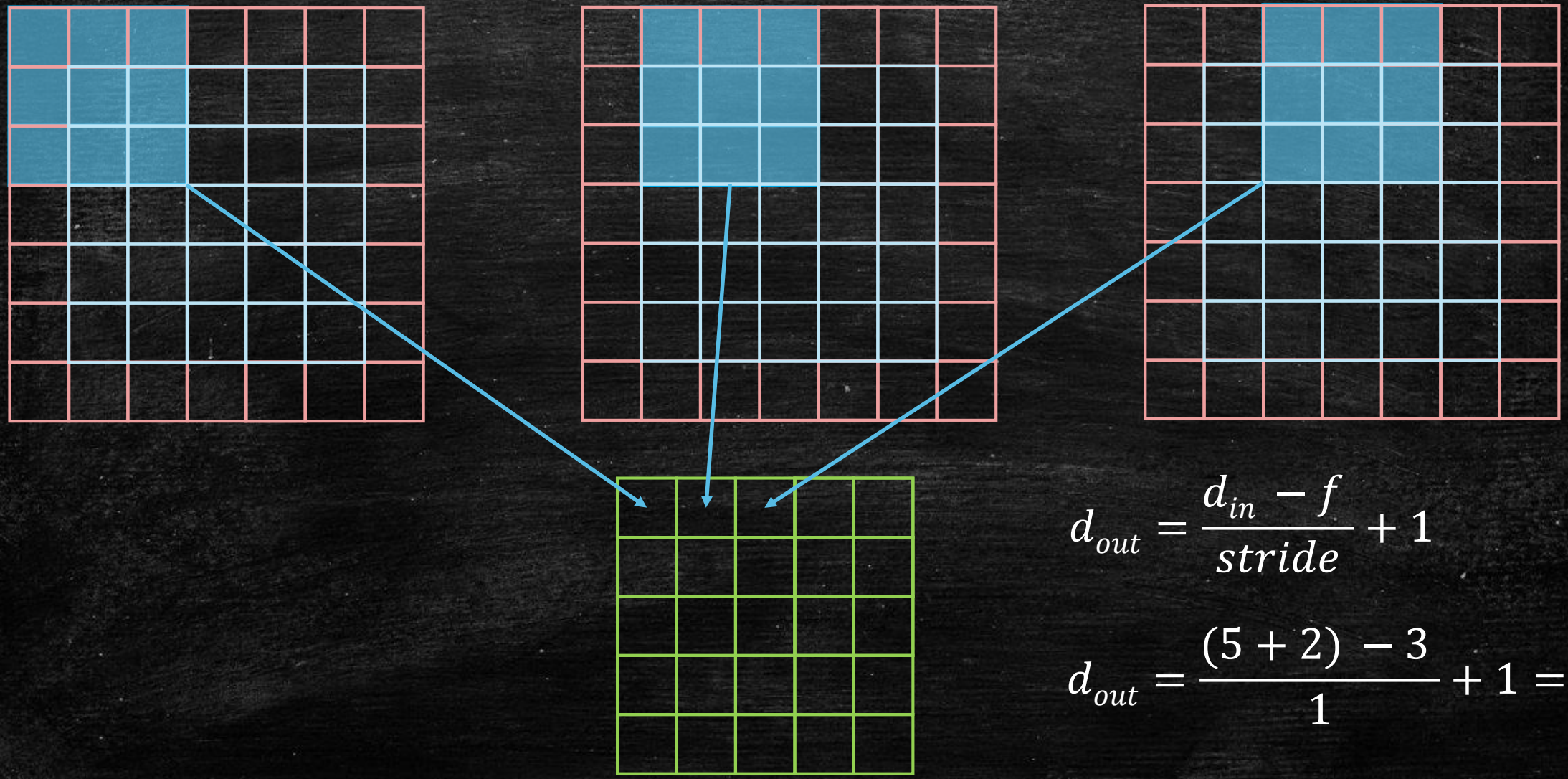
```
Result                  0 0 3 2 1 0 0
Padded original signal  0 0 0 1 0 0 0 0
Kernel                              1 2 3
```

‘Same’ padding result:
0 3 2 1 0

‘Valid’ padding result:
3 2 1



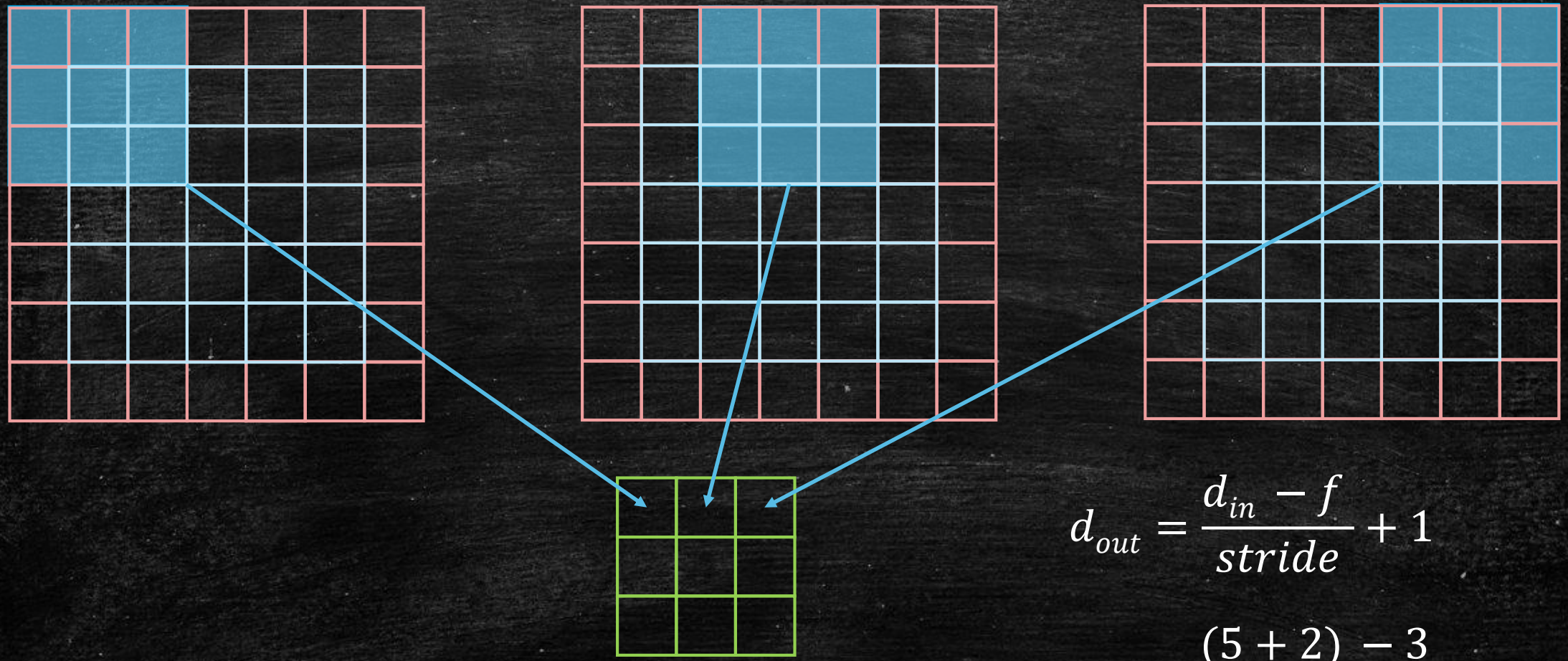
2D Convolution : Same padding, stride = 1



$$d_{out} = \frac{d_{in} - f}{stride} + 1$$

$$d_{out} = \frac{(5 + 2) - 3}{1} + 1 = 5$$

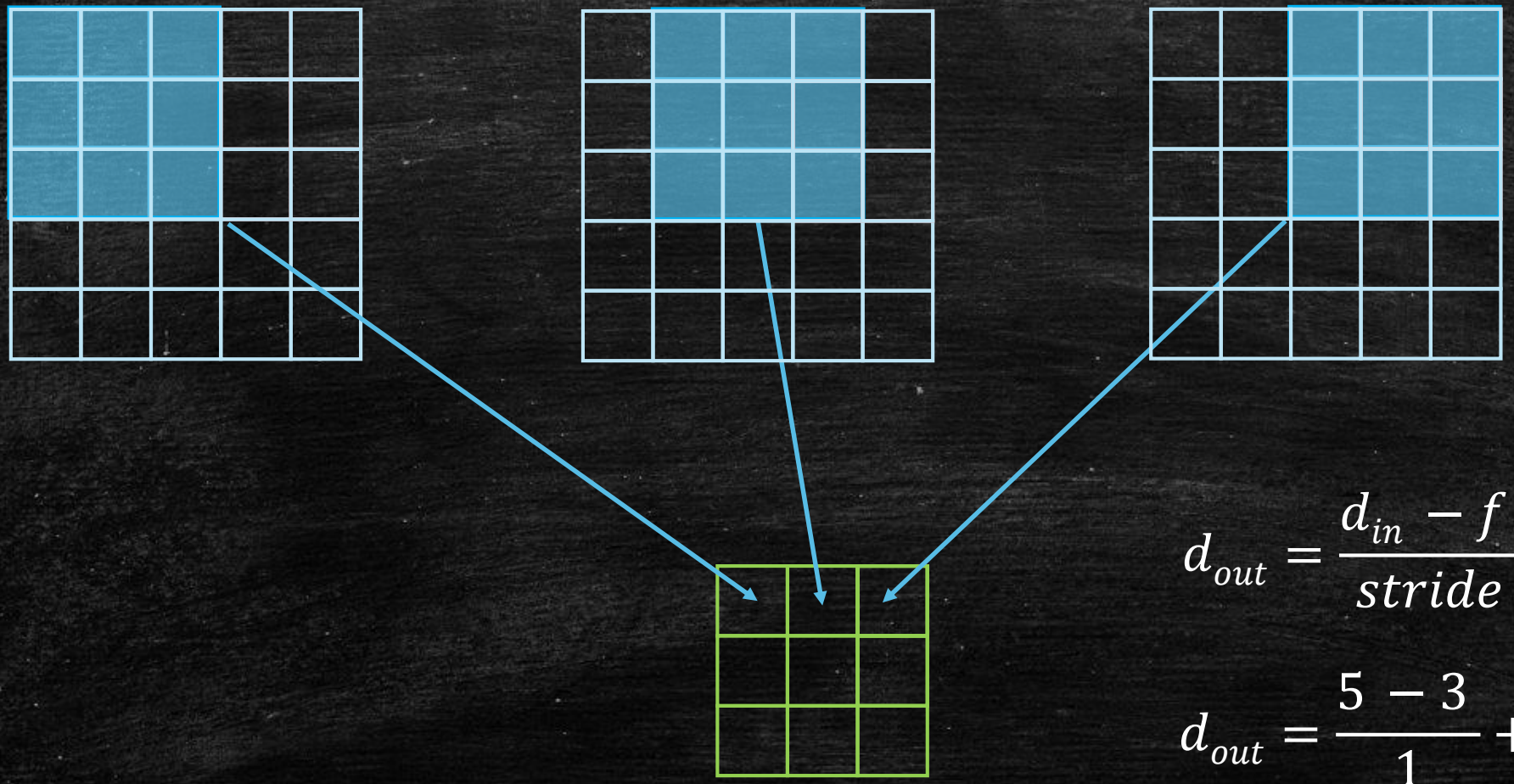
2D Convolution : Same padding, stride = 2



$$d_{out} = \frac{d_{in} - f}{stride} + 1$$

$$d_{out} = \frac{(5 + 2) - 3}{2} + 1 = 3$$

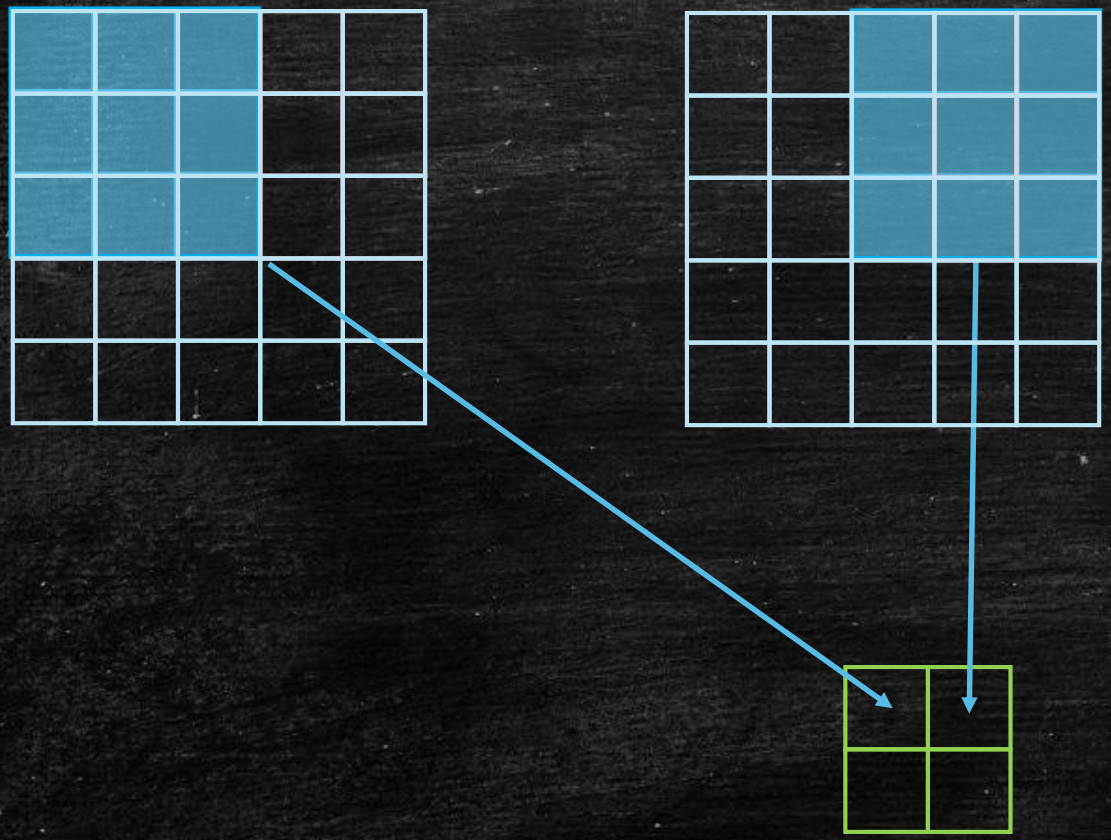
2D Convolution : Valid padding, stride =1



$$d_{out} = \frac{d_{in} - f}{stride} + 1$$

$$d_{out} = \frac{5 - 3}{1} + 1 = 3$$

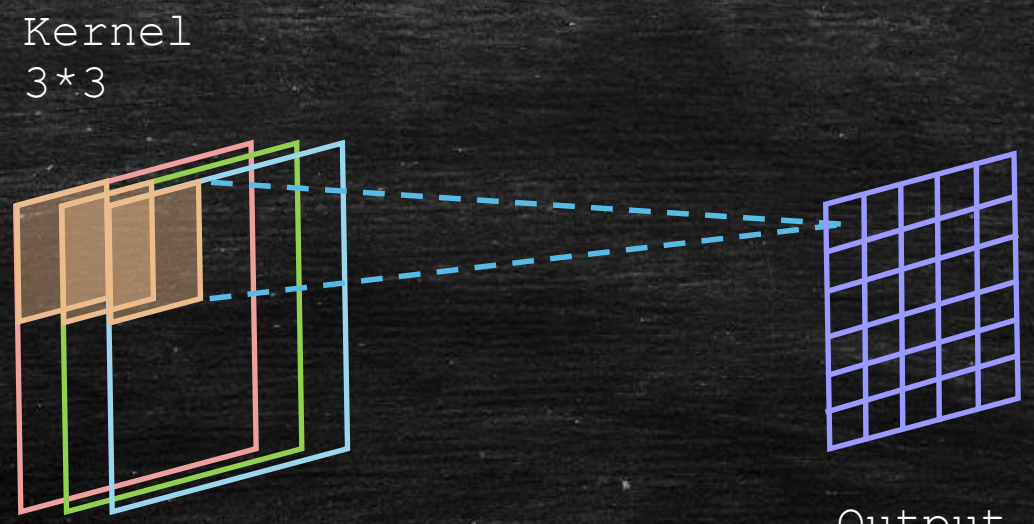
2D Convolution : Valid padding, stride = 2



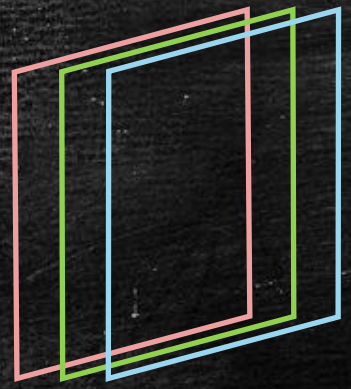
$$d_{out} = \frac{d_{in} - f}{stride} + 1$$

$$d_{out} = \frac{5 - 3}{2} + 1 = 2$$

2D Convolution on a 3D Volume

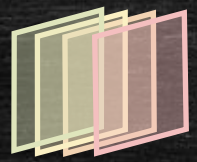


2D Convolution on a 3D Volume



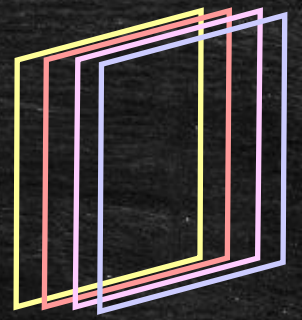
Input
7*7*3

Convolved by



Kernel
3*3*4

=



Feature Maps
5*5*4