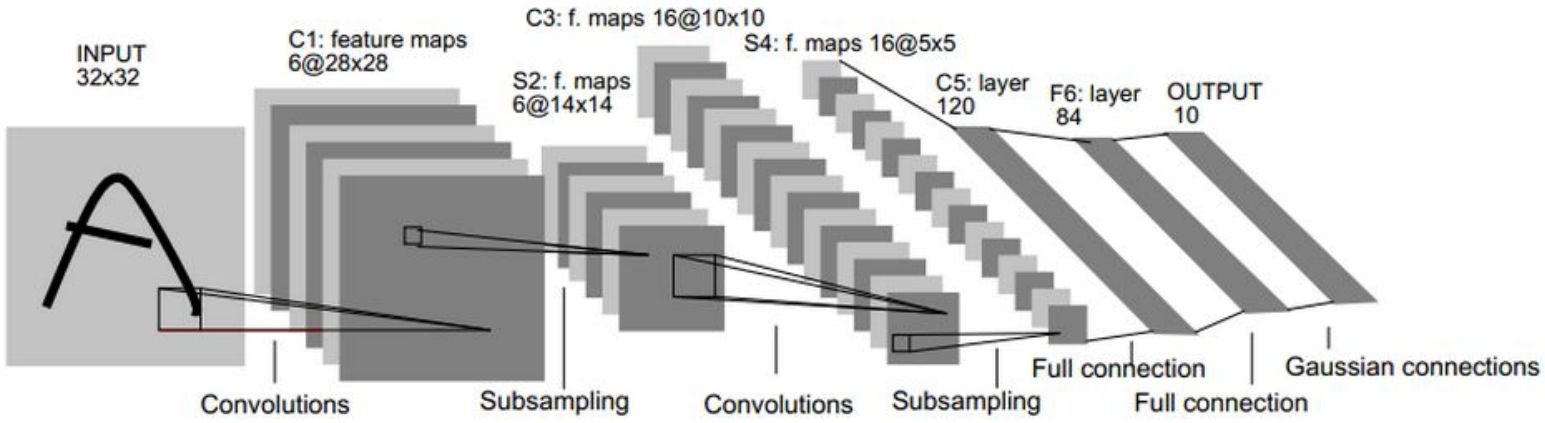
# COSI 165B
# Deep Learning

Chuxu Zhang
Computer Science Department
Brandeis University

3/3/2021

# Last Lecture

- What is CNN

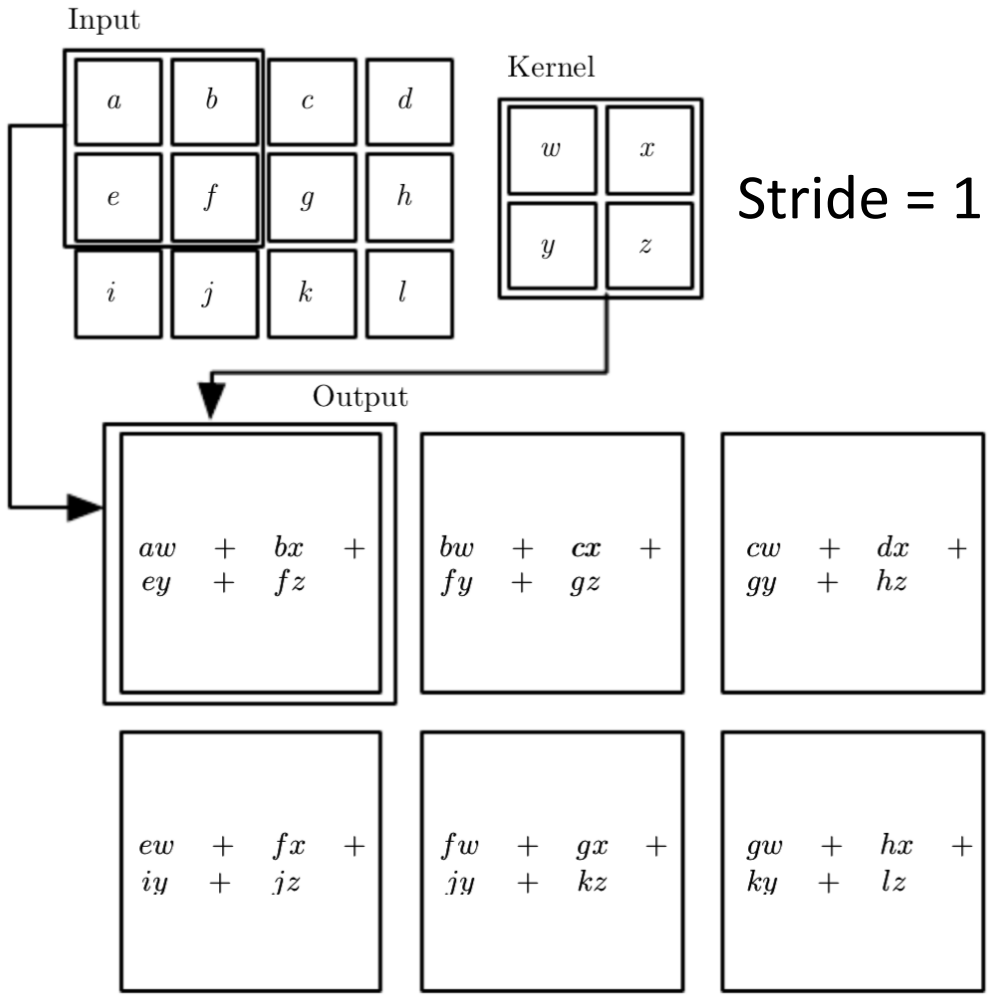



Convolutional Layer: convolution operation

Pooling Layer: max/mean pooling

Fully Connected Layer: feed-forward neural network

## ❑ Convolution: example




Stride = 1

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

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

# Last Lecture

## ❑ Motivation/Strengths

### Sparse Interactions

MLP: fully connected, every output unit interacts with every input unit. $O(m \times n)$

Convolution: making the kernel smaller than the input, sparse connection. $O(k \times n)$

### Parameter Sharing

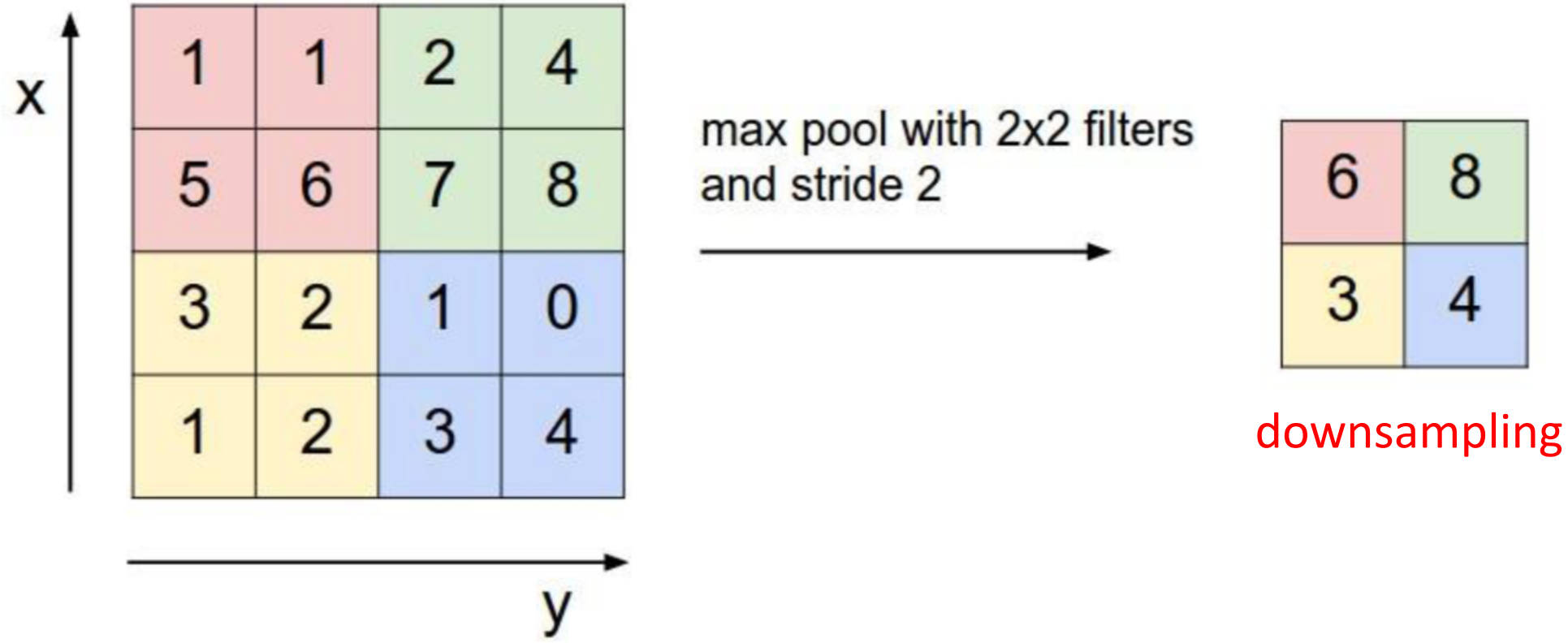
MLP: each element of the weight matrix is used exactly once when computing the output of a layer.

Convolution: each member of the kernel is used at every position of the input, learn only one set of parameters for every position.

### Equivariant Representations

If the input changes, the output changes in the same way: shift-convolution, convolution-shift.

- Pooling: 2D example



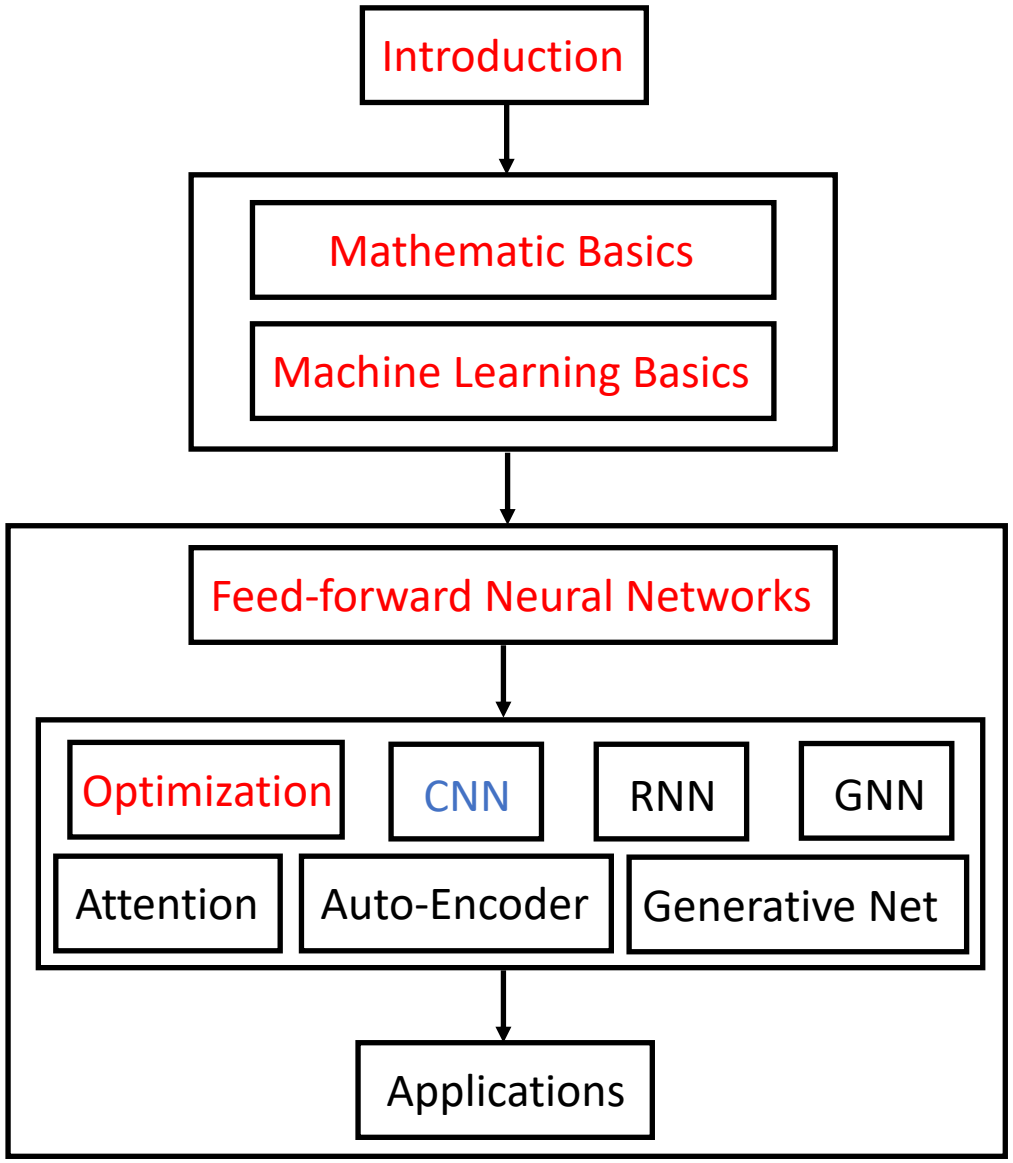
## ❑ Motivation/Strengths

Use a pooling function to modify the output of the layer.

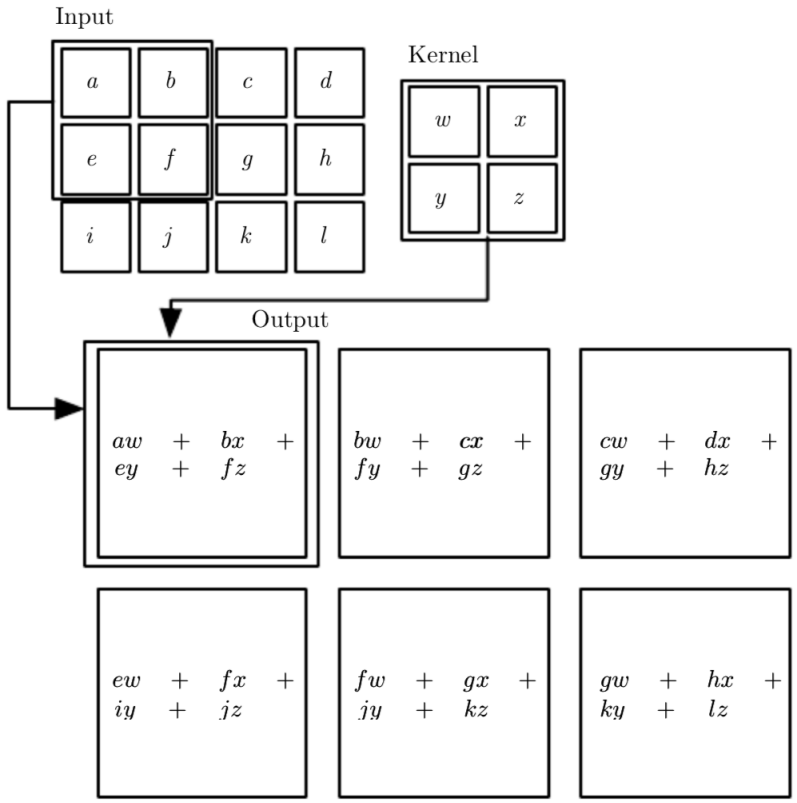
A pooling function replaces the output of the net at a certain location with a summary statistic of the nearby outputs.
e.g., max pooling: reports the maximum output within a rectangular neighborhood

In all cases, pooling helps to make the representation become approximately invariant to small translations of the input. Invariance to translation means that if we translate the input by a small amount, the values of most of the pooled outputs do not change.

# Structure of This Course

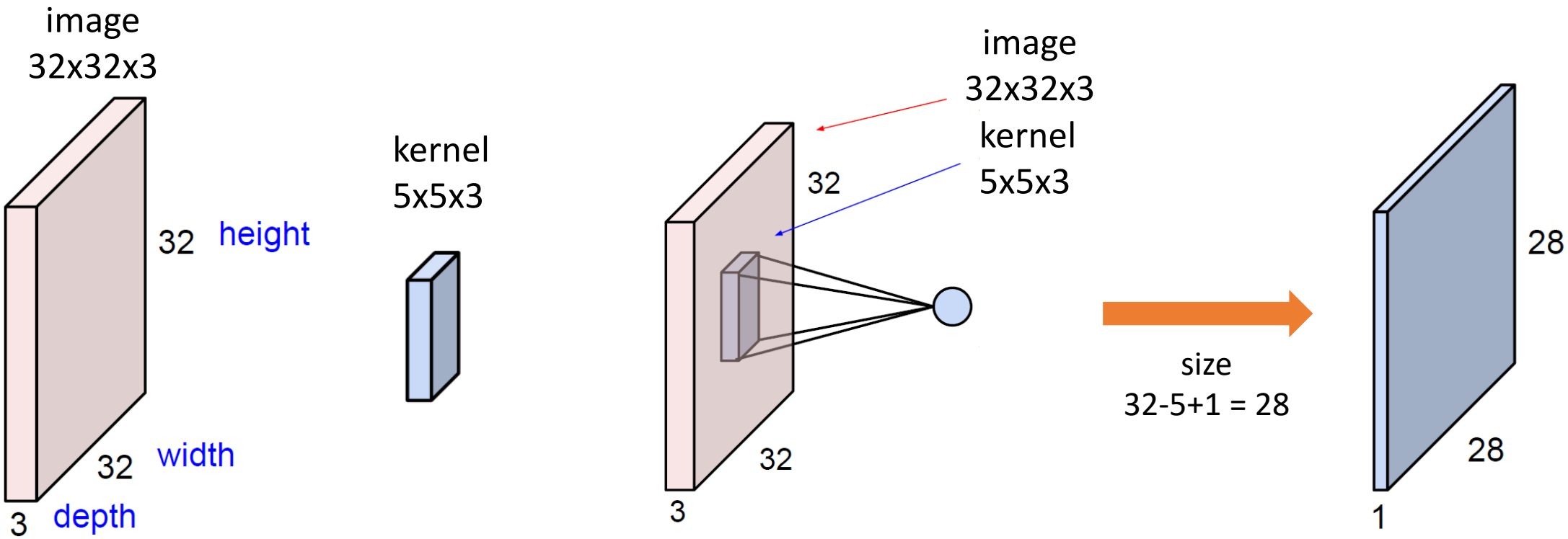


## ❑ Convolution: more example

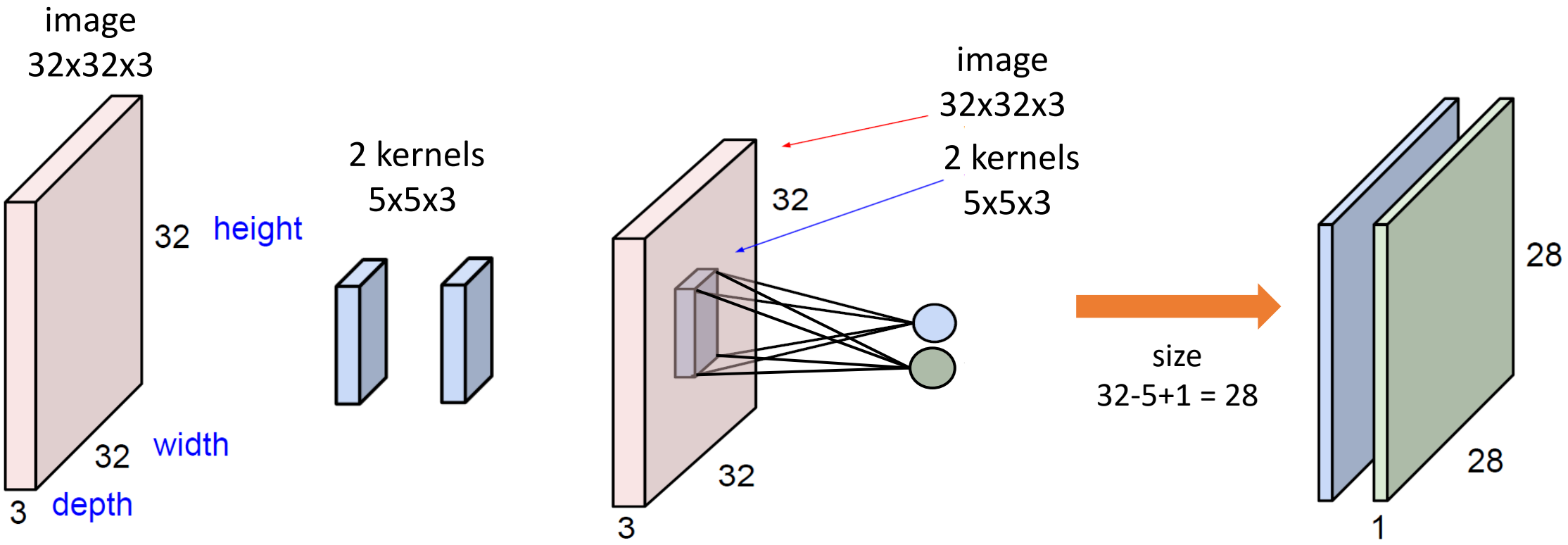



Kernel number = 1
What's about more

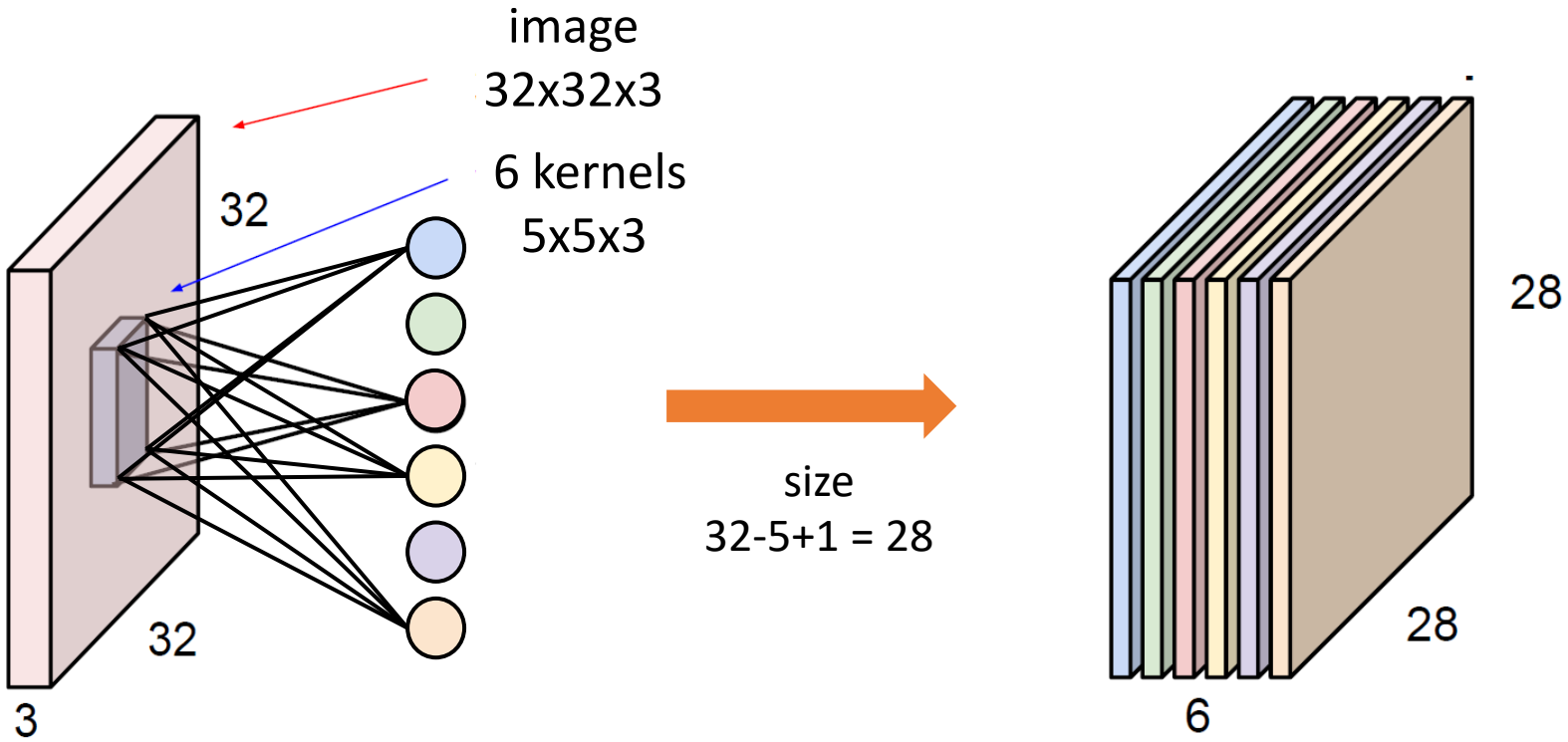
## ❑ Convolution: more example



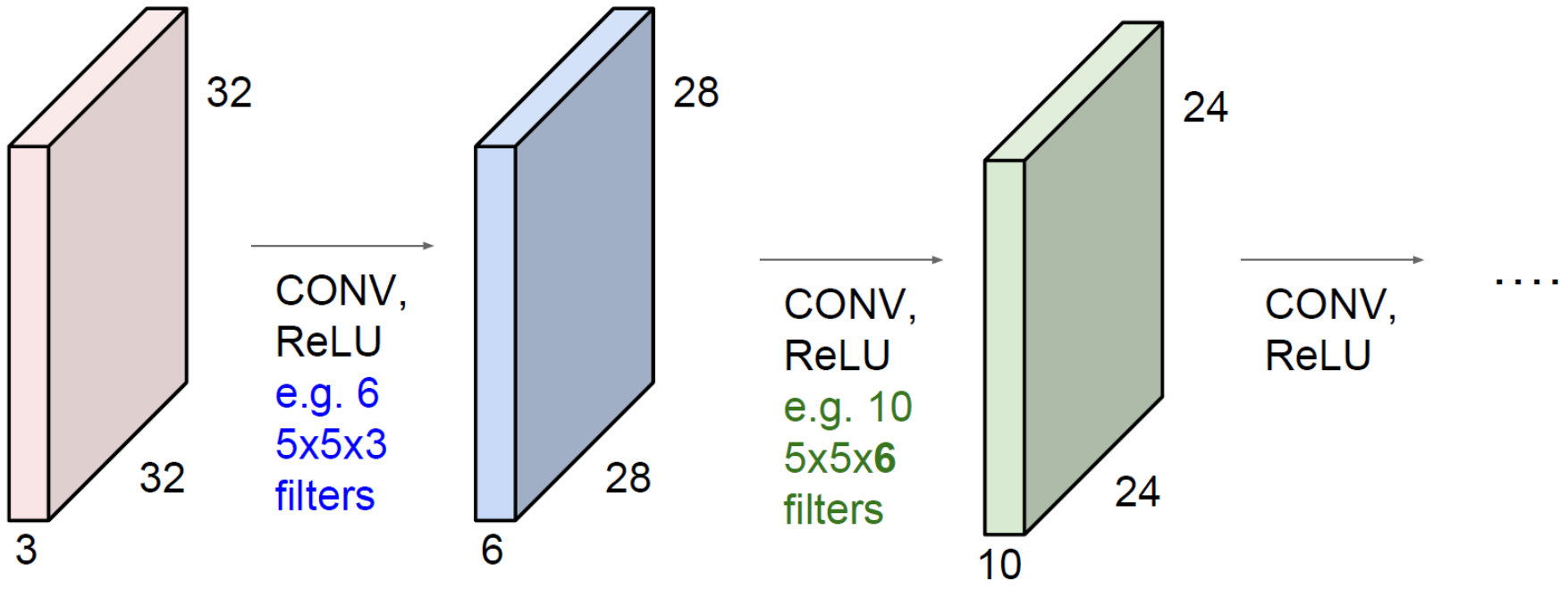
## ❑ Convolution: more example



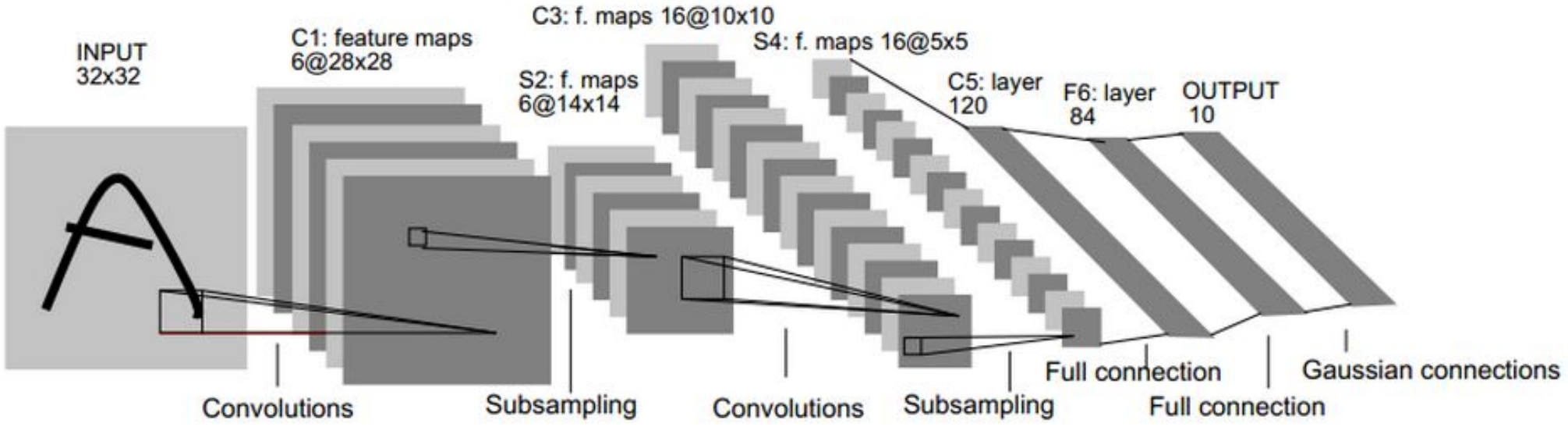
## ❑ Convolution: more example



## ❑ Convolution: more example



## ❑ A Full Model




6 kernels
5x5
s=1

2x2
s=2

16 kernels
6x5x5
s=1

2x2
s=2

Output Size = (N-F)/S + 1

# Convolutional Neural Networks

## ❑ Data

| | Single channel | Multi-channel |
|---|---|---|
| 1-D | Audio waveform: The axis we convolve over corresponds to time. We discretize time and measure the amplitude of the waveform once per time step. | Skeleton animation data: Animations of 3-D computer-rendered characters are generated by altering the pose of a "skeleton" over time. At each point in time, the pose of the character is described by a specification of the angles of each of the joints in the character's skeleton. Each channel in the data we feed to the convolutional model represents the angle about one axis of one joint. |
| 2-D | Audio data that has been preprocessed with a Fourier transform: We can transform the audio waveform into a 2D tensor with different rows corresponding to different frequencies and different columns corresponding to different points in time. Using convolution in the time makes the model equivariant to shifts in time. Using convolution across the frequency axis makes the model equivariant to frequency, so that the same melody played in a different octave produces the same representation but at a different height in the network's output. | Color image data: One channel contains the red pixels, one the green pixels, and one the blue pixels. The convolution kernel moves over both the horizontal and vertical axes of the image, conferring translation equivariance in both directions. |
| 3-D | Volumetric data: A common source of this kind of data is medical imaging technology, such as CT scans. | Color video data: One axis corresponds to time, one to the height of the video frame, and one to the width of the video frame. |

## Neuroscientific Basis

The history of convolutional networks begins with neuroscientific experiments.

Neurophysiologists David Hubel and Torsten Wiesel collaborated for several years to determine many of the most basic facts about how the mammalian vision system works.

Their accomplishments were eventually recognized with a Nobel prize. Their work helped to characterize many aspects of brain function.

In this simplified view, we focus on a part of the brain called V1, also known as the primary visual cortex. V1 is the first area of the brain that begins to perform significantly advanced processing of visual input.

## ❑ Neuroscientific Basis

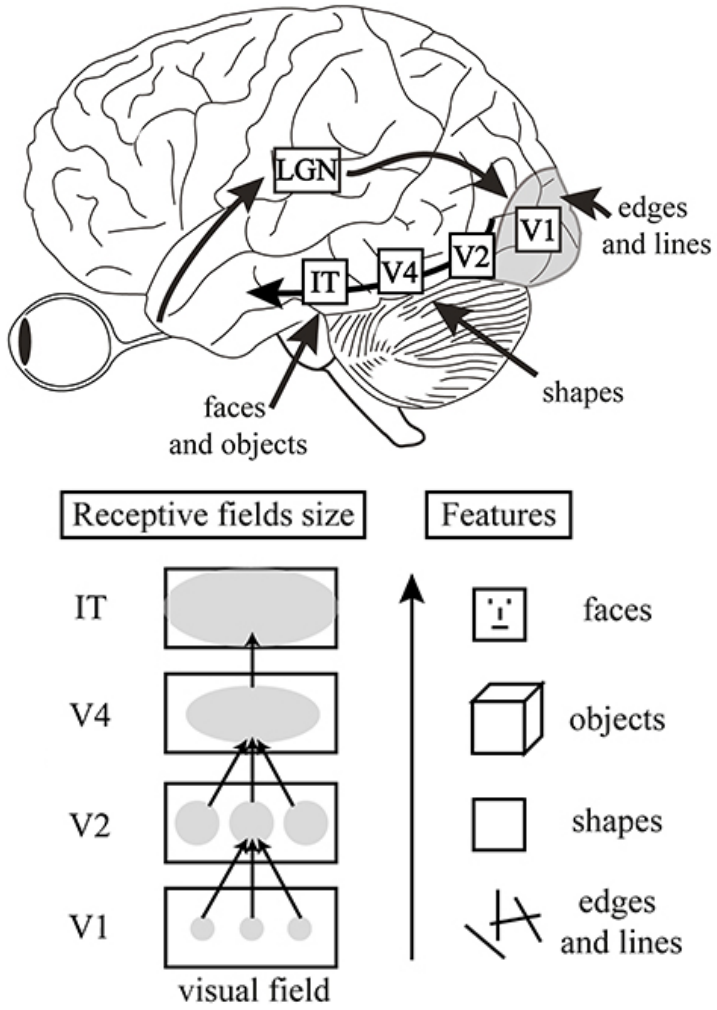
A convolutional network layer is designed to capture three properties of V1:

V1 is arranged in a spatial map. It actually has a two-dimensional structure mirroring the structure of the image in the retina. Convolutional networks capture this property by having their features defined in terms of two dimensional maps.

V1 contains many simple cells. A simple cell's activity can to some extent be characterized by a linear function of the image in a small, spatially localized receptive field. The detector units of a convolutional network are designed to emulate these properties of simple cells.

V1 also contains many complex cells. These cells respond to features that are similar to those detected by simple cells, but complex cells are invariant to small shifts in the position of the feature. This inspires the pooling units of convolutional networks.

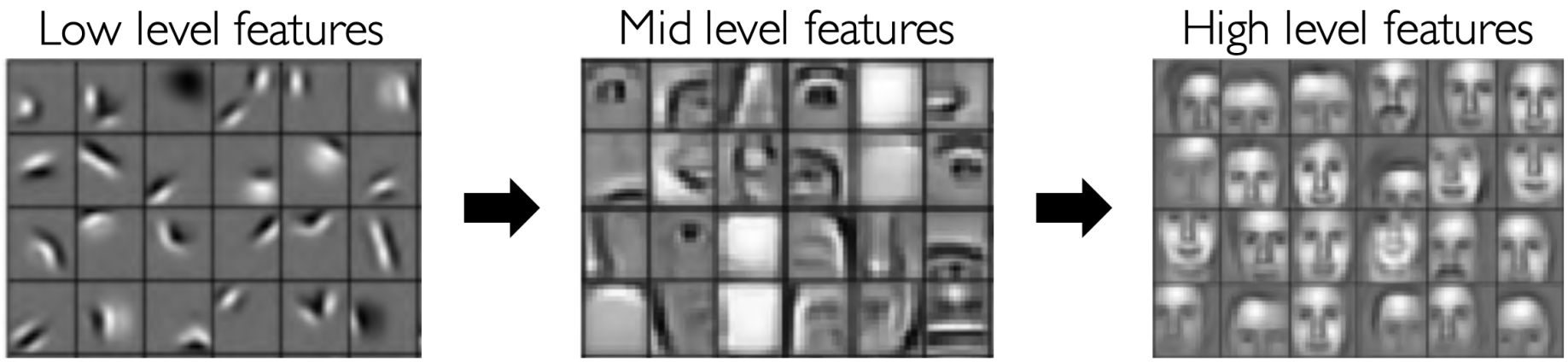
## ❑ A Full Model




LGN: receive visual information from outer environment and pass it V1

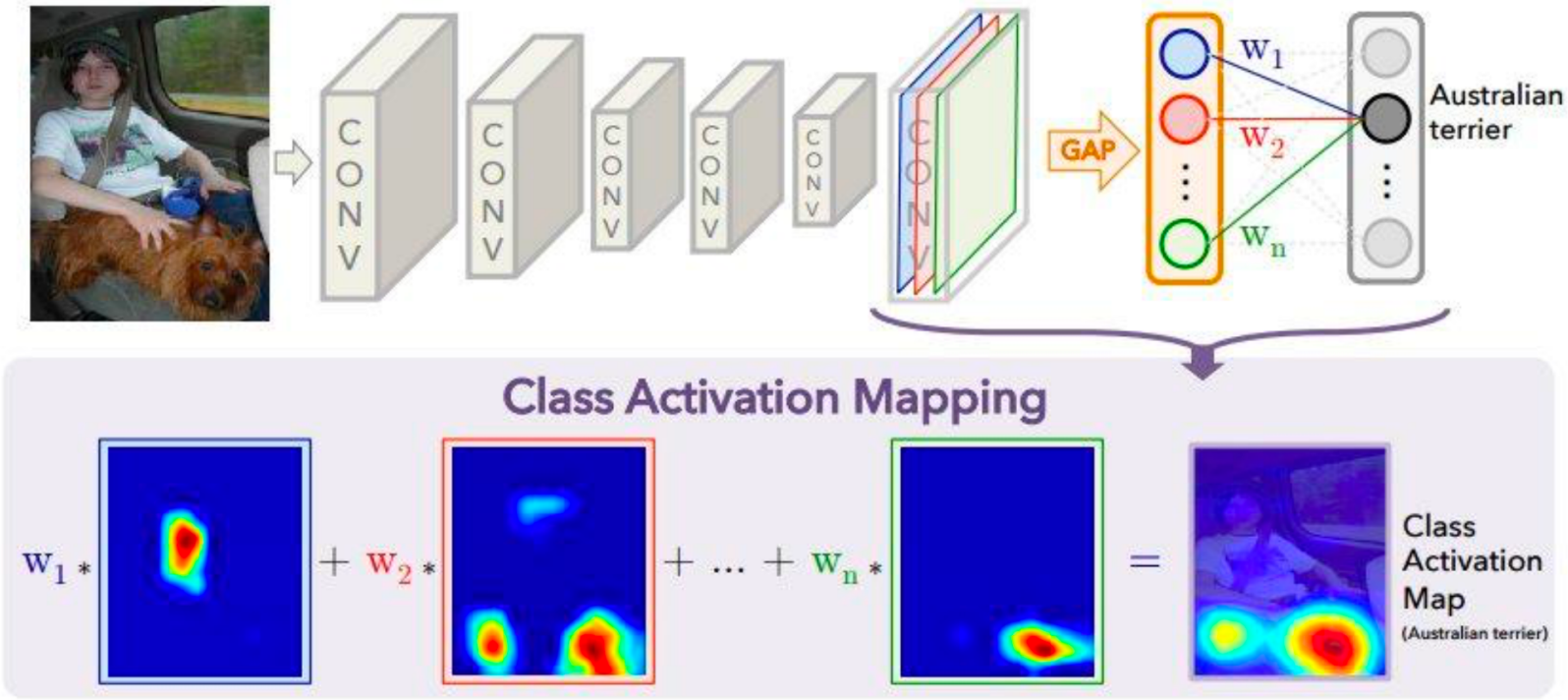
V1, V2, V4, IT
different visual cortex
larger and larger view
more and more complex

## ❑ Feature Visualization




**Face Recognition**

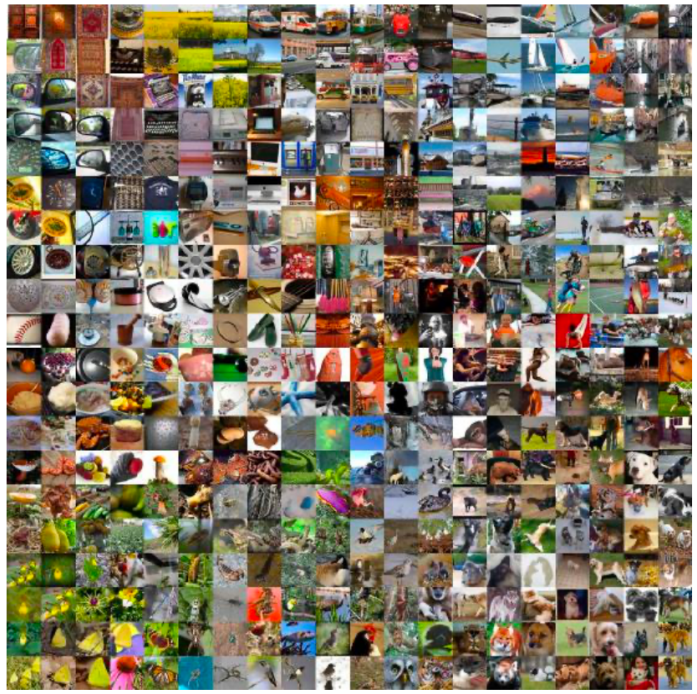
## ❑ Feature Visualization




A **class activation map (CAM)** for a given class highlights the image regions used by the CNN to identify that class.

# Convolutional Neural Networks

## ❑ Computer Vision



ImageNet:
22K categories. 14M images.

MNIST: handwritten digits

Large Data Collection


places: natural scenes

**Airplane**

**Automobile**

**Bird**

**Cat**

**Deer**

**Dog**

**Frog**

**Horse**

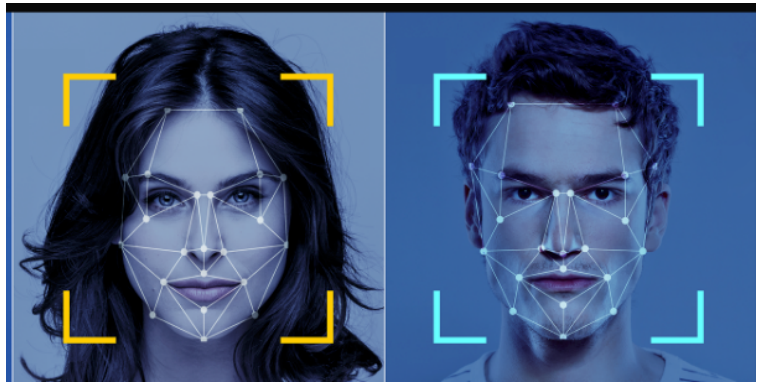
**Ship**

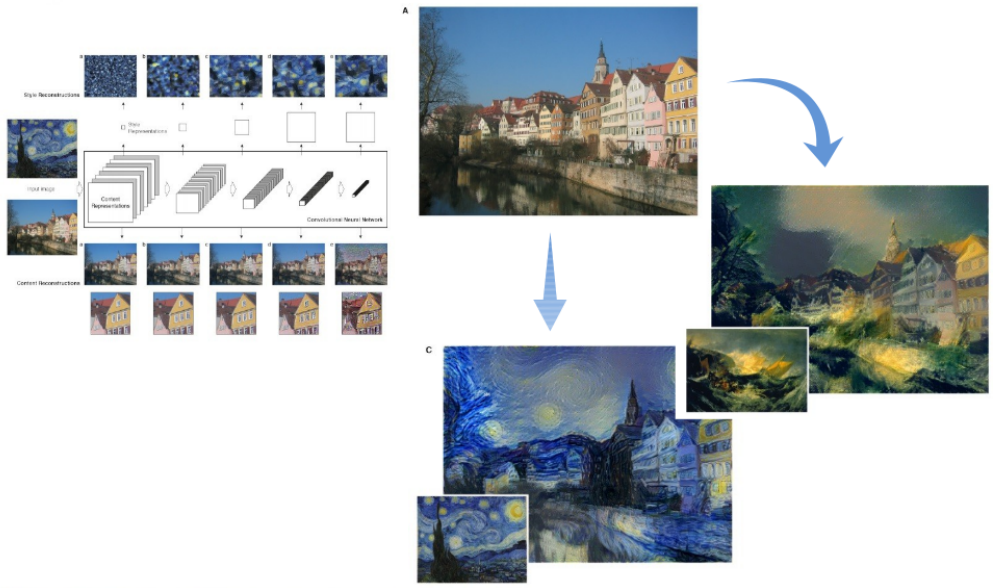
**Truck**

CIFAR-10

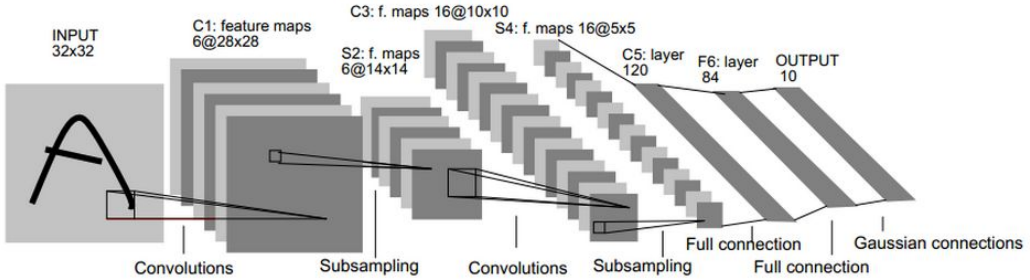
# Convolutional Neural Networks

## ❑ Computer Vision

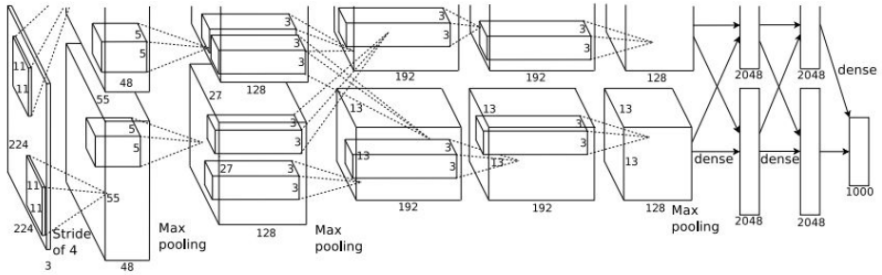


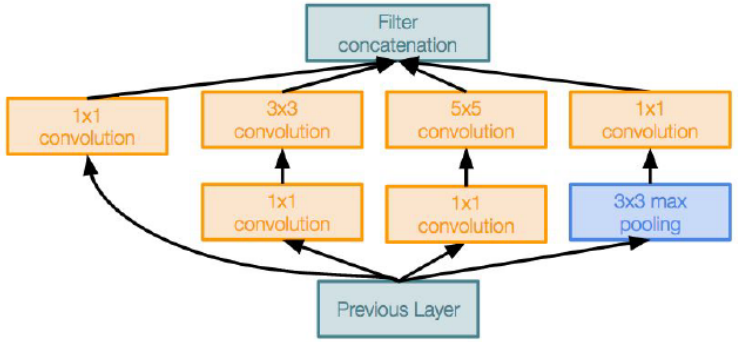
## ❑ Computer Vision




LeNet-5

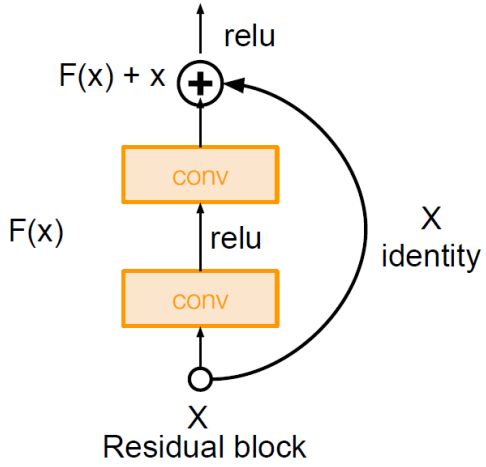



AlexNet




Inception module

GoogLeNet




ResNet

# Q & A