CNN

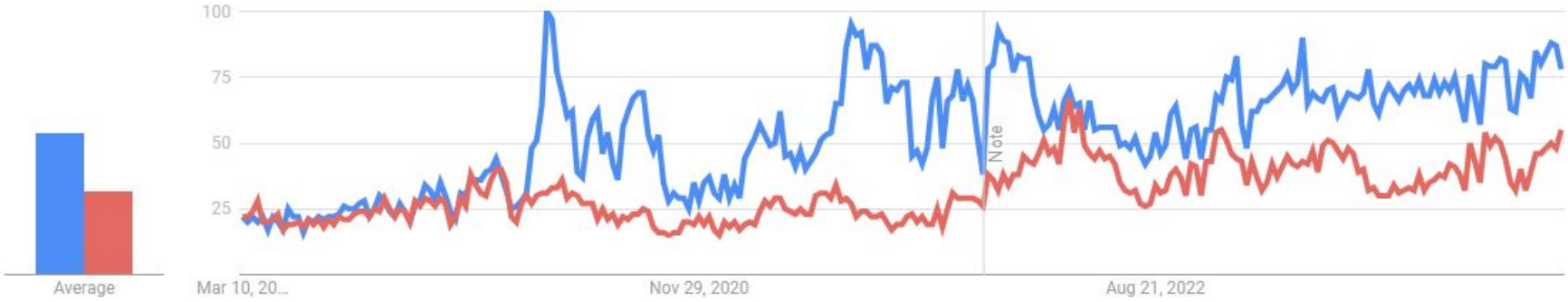
Natural language processing
Topic

Convolutional neural network
Topic

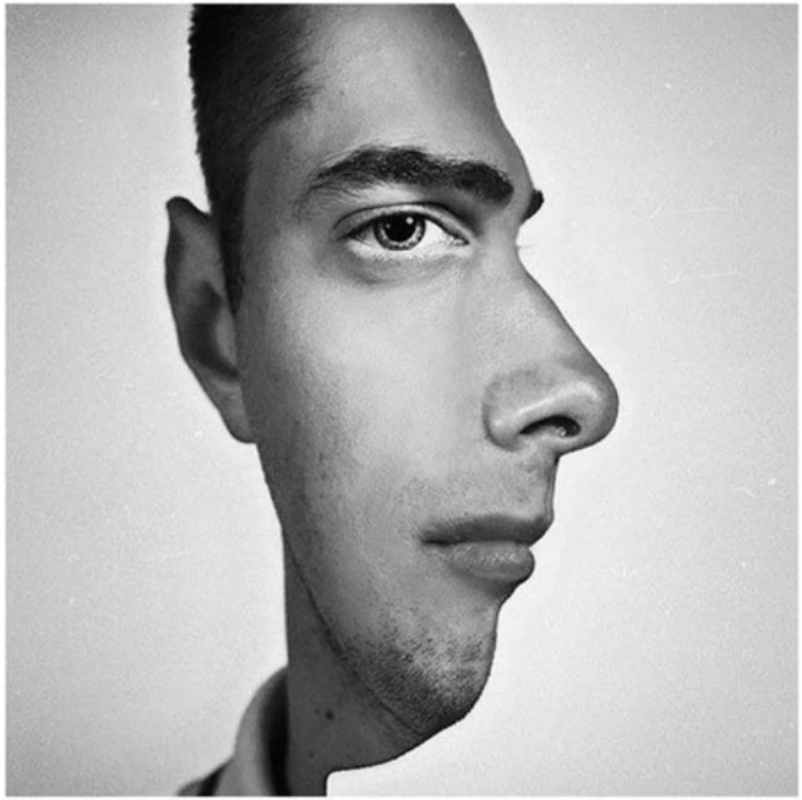
+ Add comparison

India | Past 5 years | All categories | Web Search

Interest over time



Convolutional Neural Networks



Convolutional Neural Networks

Convolutional Neural Networks

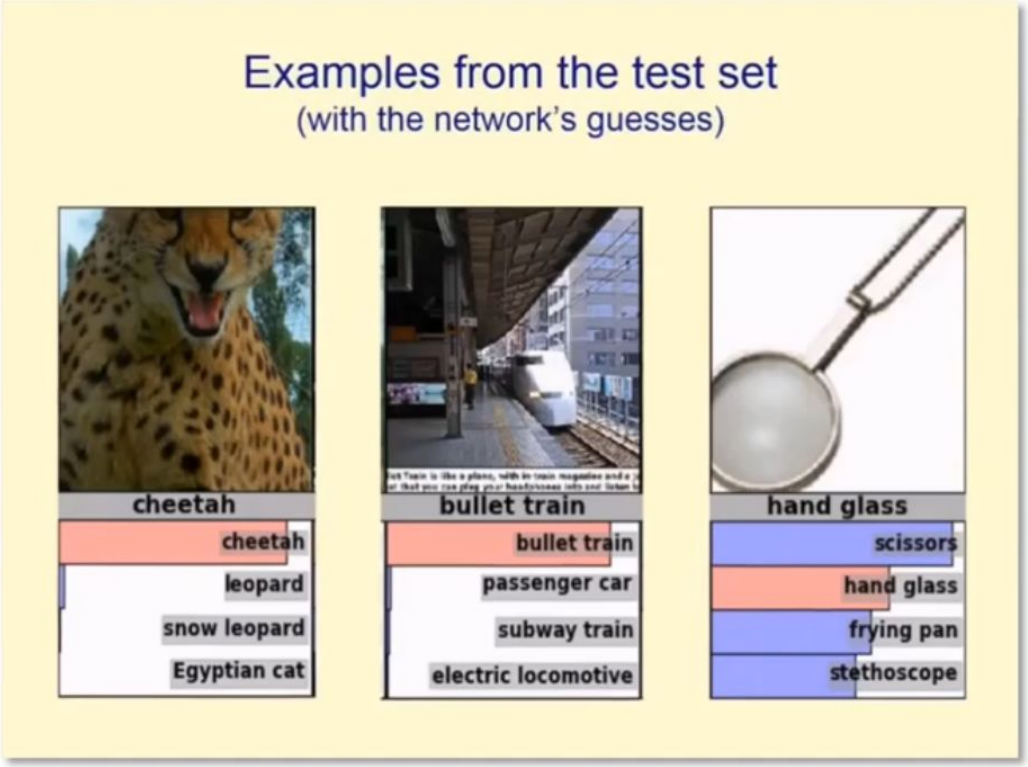
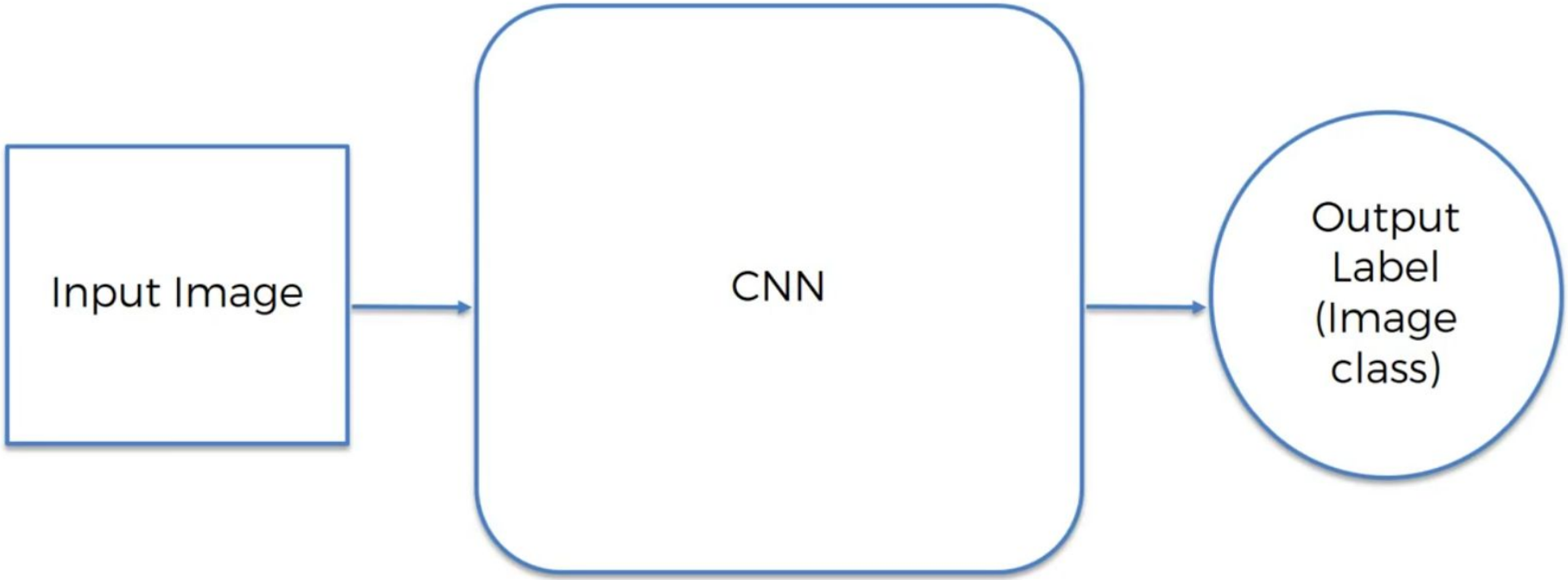


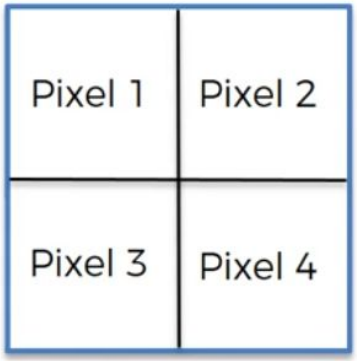
Image Source: a talk by Geoffrey Hinton

Convolutional Neural Networks

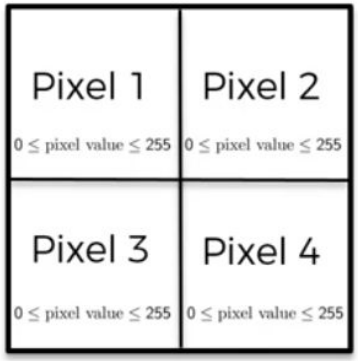


Convolutional Neural Networks

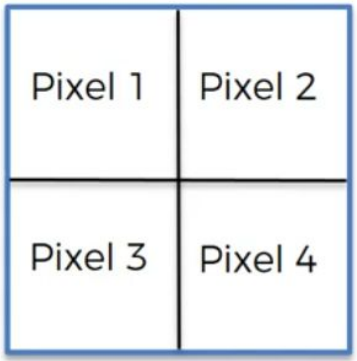
B / W Image 2x2px



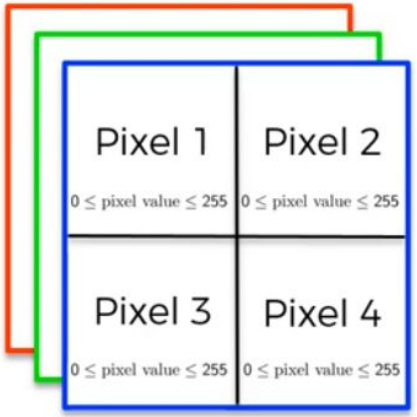
2d array



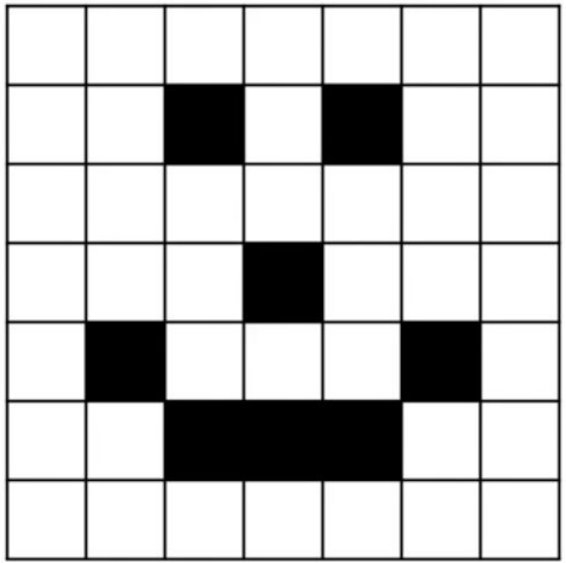
Colored Image 2x2px



3d array



Convolutional Neural Networks



0	0	0	0	0	0	0
0	1	0	0	0	1	0
0	0	0	0	0	0	0
0	0	0	1	0	0	0
0	1	0	0	0	1	0
0	0	1	1	1	0	0
0	0	0	0	0	0	0

http://yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf

Convolutional Neural Networks

STEP 1: Convolution

STEP 2: Max Pooling

STEP 3: Flattening

STEP 4: Full Connection

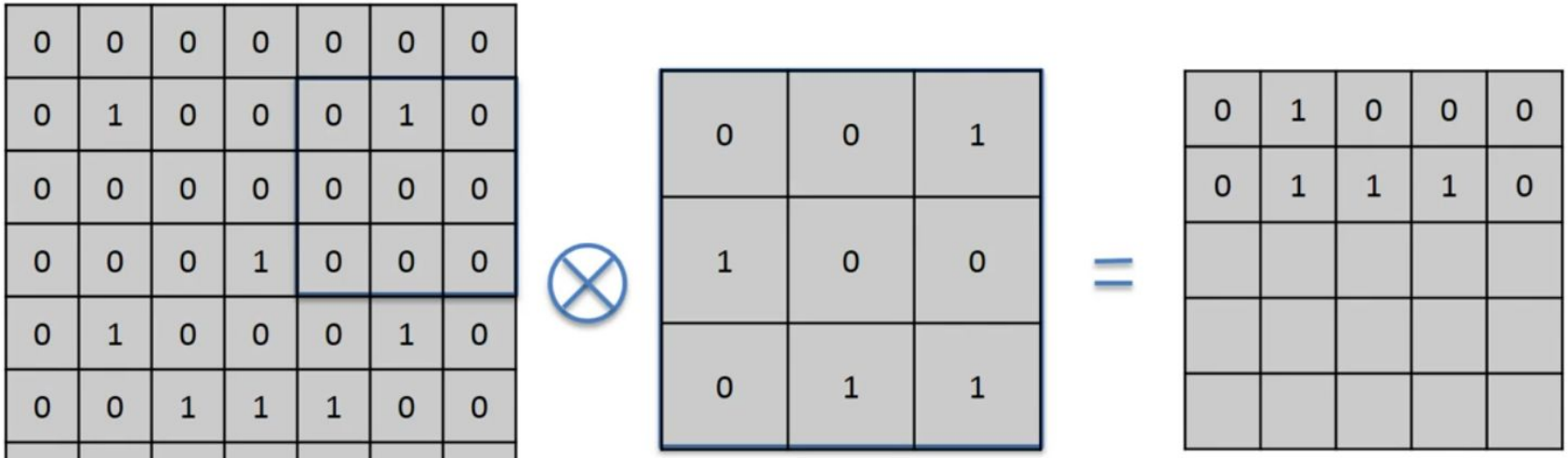
Introduction to Convolutional Neural Networks

https://cs.nju.edu.cn/wujx/

https://cs.nju.edu.cn/wujx/paper/CNN.pdf

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

Step 1 - Convolution

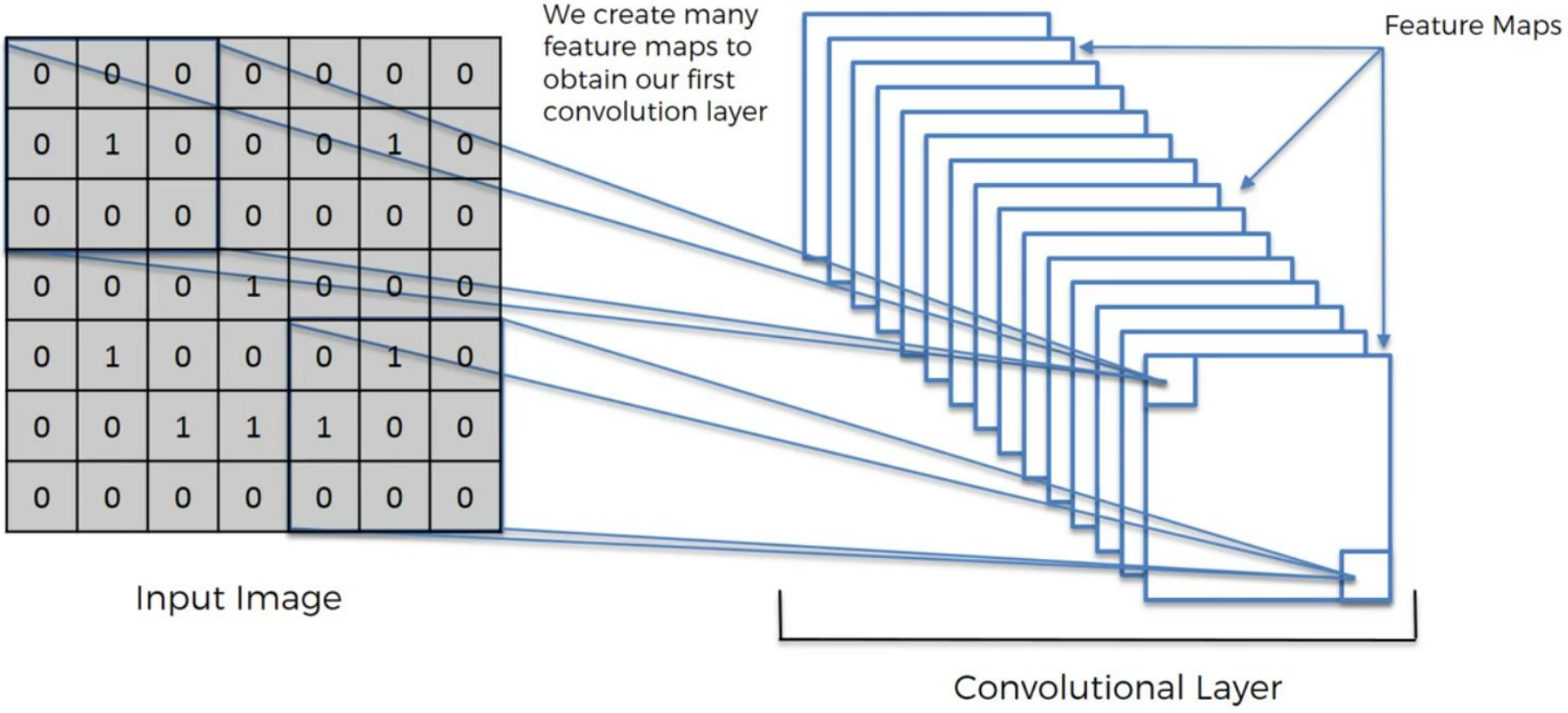


Input Image

Feature Detector

Feature Map

Step 1 - Convolution



Filters - How they work

https://docs.gimp.org/2.10/en/

Step 1 - Convolution

Sharpen:

0	0	0	0	0
0	0	-1	0	0
0	-1	5	-1	0
0	0	-1	0	0
0	0	0	0	0

Step 1 - Convolution

Blur:

0	0	0	0	0
0	1	1	1	0
0	1	1	1	0
0	1	1	1	0
0	0	0	0	0

Step 1 - Convolution

Edge Enhance:

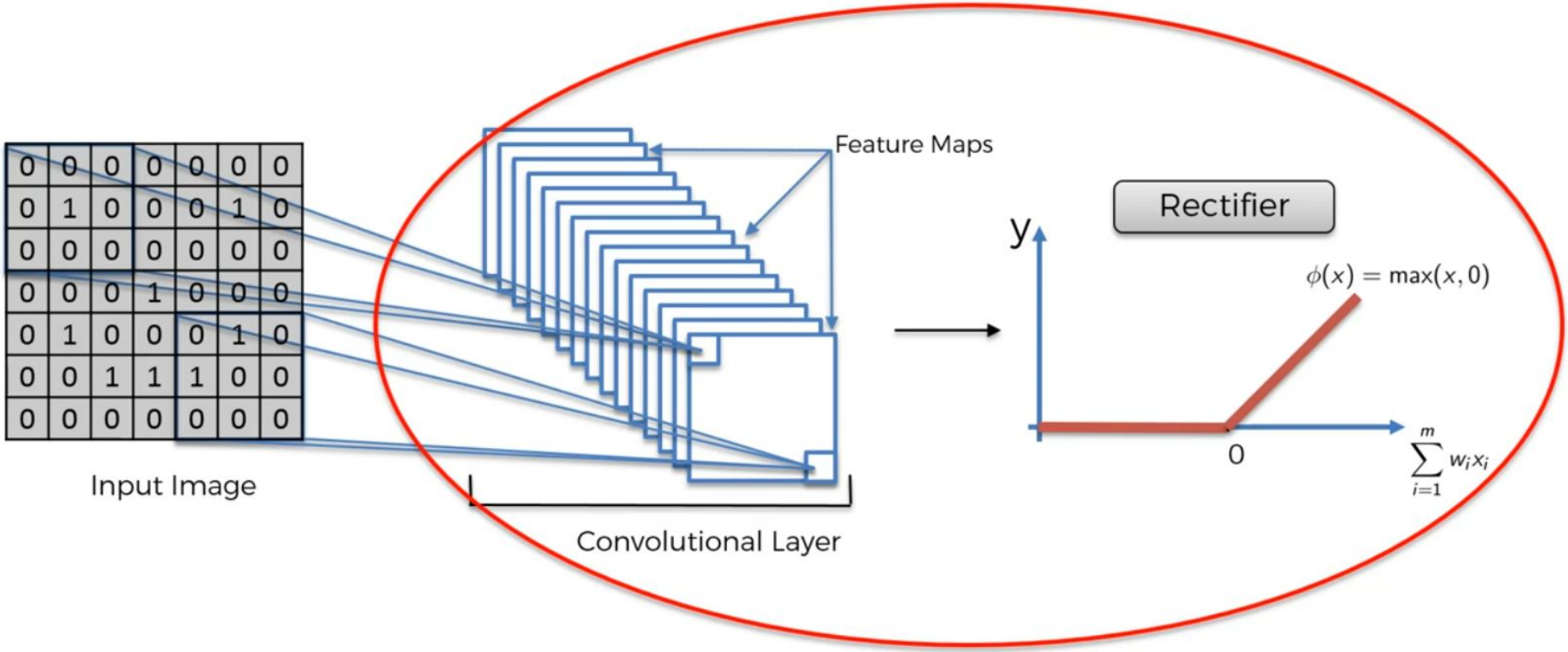
	0	0	0	
	-1	1	0	
	0	0	0	

Step 1 - Convolution

Edge Detect:

	0	1	0	
	1	-4	1	
	0	1	0	

Step 1(B) – ReLU Layer



Step 1(B) – ReLU Layer

Step 1(B) – ReLU Layer

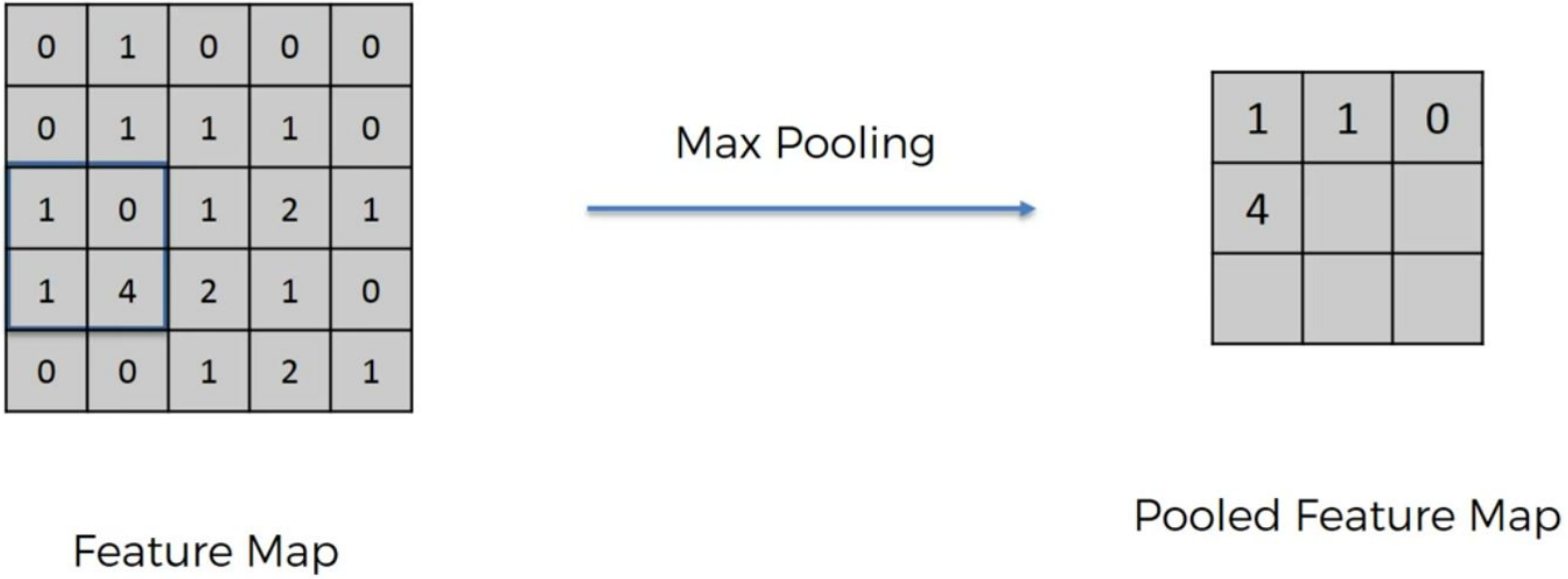
Step 1(B) – ReLU Layer

https://arxiv.org/pdf/1609.04112.pdf

Step 2 - Max Pooling

Step 2 - Max Pooling

Step 2 - Max Pooling



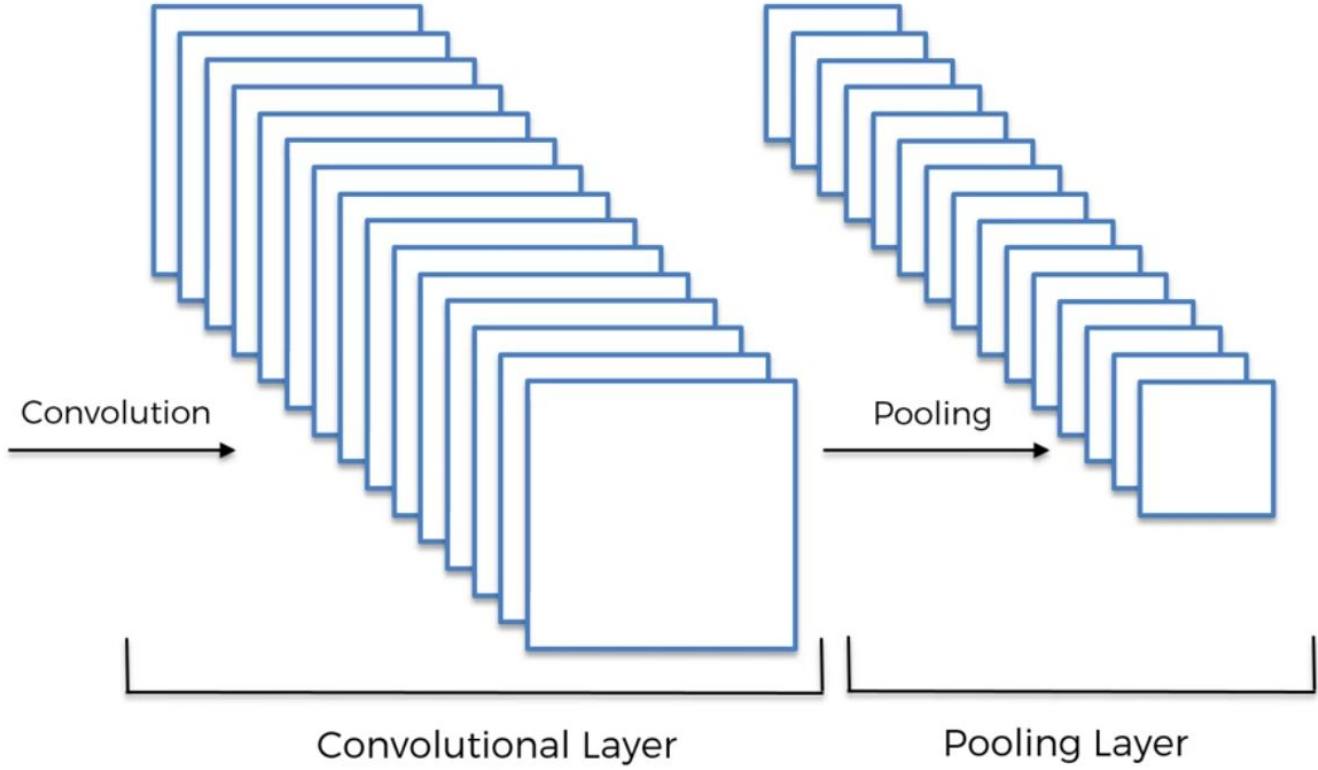
Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition

https://cs.nju.edu.cn/wujx/paper/CNN.pdf

Step 2 - Max Pooling

0	0	0	0	0	0	0
0	1	0	0	0	1	0
0	0	0	0	0	0	0
0	0	0	1	0	0	0
0	1	0	0	0	1	0
0	0	1	1	1	0	0
0	0	0	0	0	0	0

Input Image



https://www.cs.ryerson.ca/~aharley/vis/conv/flat.html

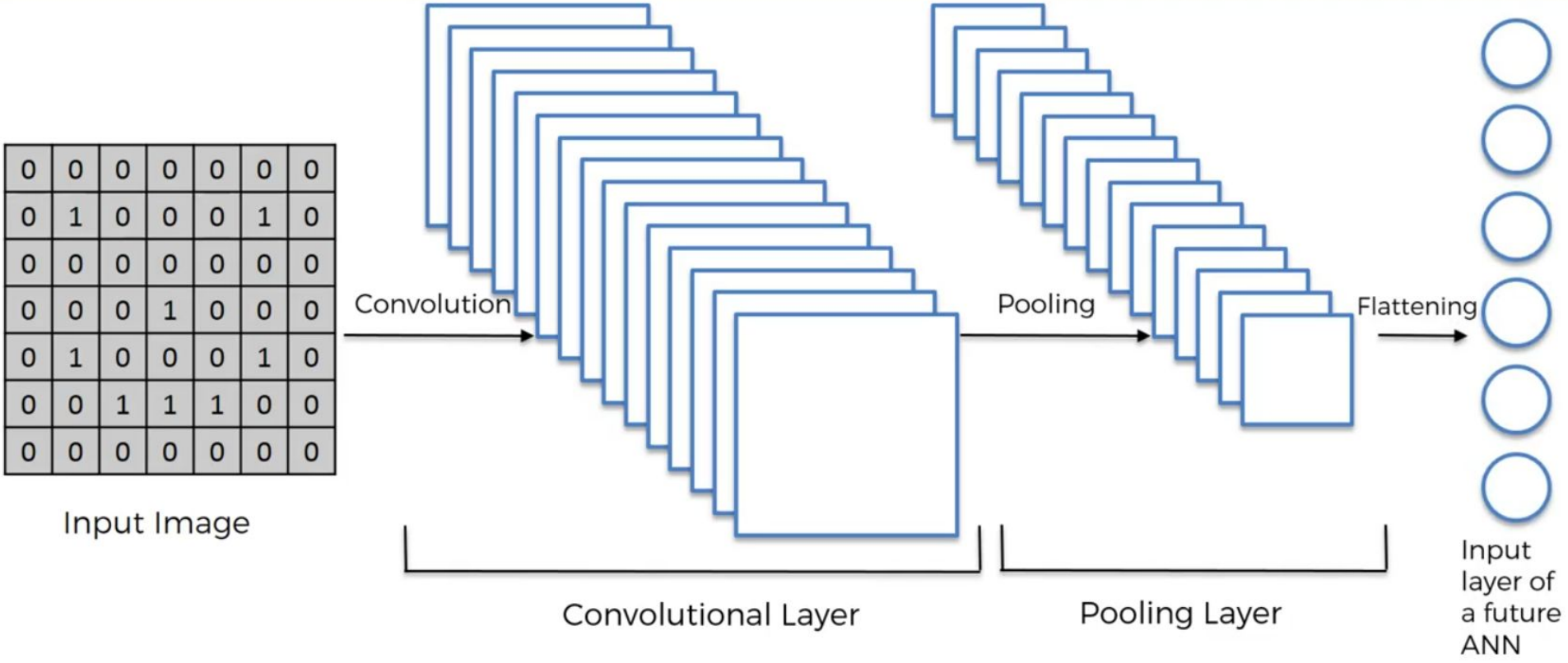
Step 3 - Flattening

1	1	0
4	2	1
0	2	1

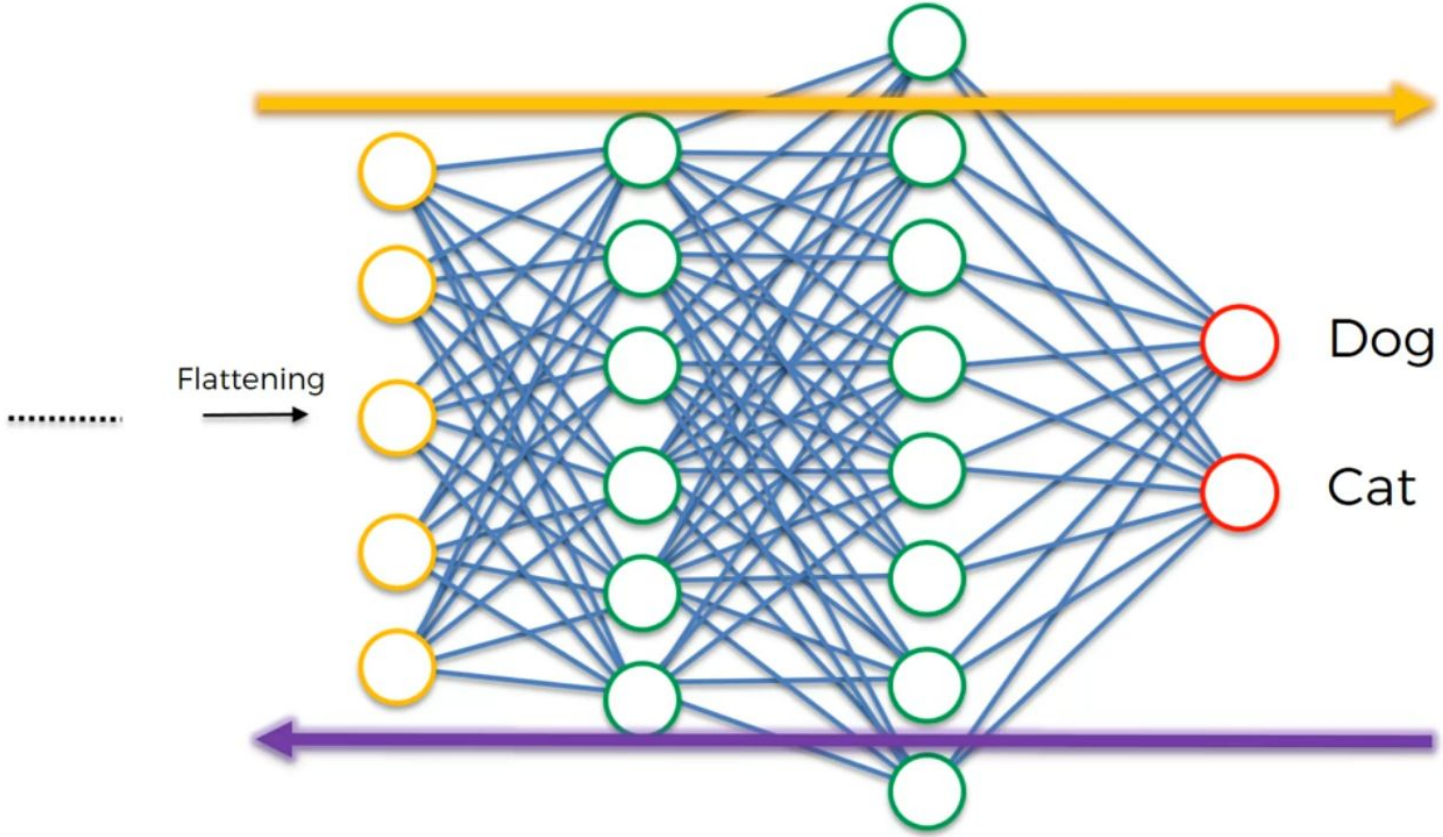
Pooled Feature Map

1
1
0
4
2
1
0
2
1

Step 3 - Flattening



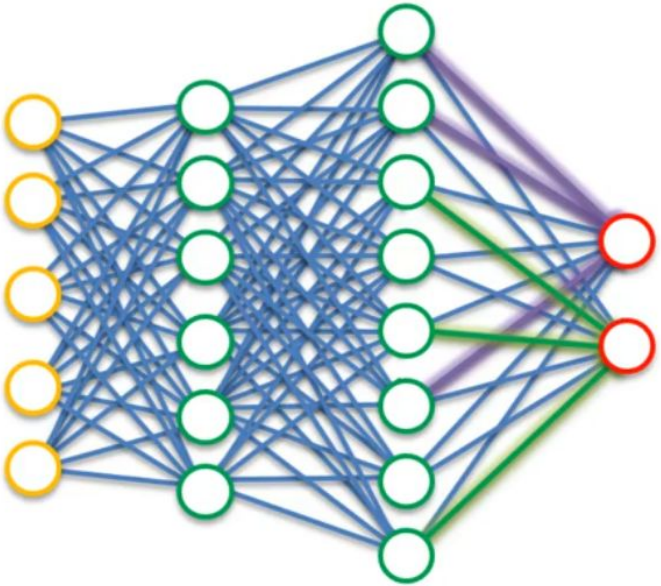
Step 4 - Full Connection



Softmax & Cross-Entropy

$$f_j(z) = \frac{e^{z_j}}{\sum_k e^{z_k}}$$

Flattening



Dog → z_1 → 0.95

Cat → z_2 → 0.05

Categorizing Listing Photos at Airbnb

https://medium.com/airbnb-engineering/categorizing-listing-photos-at-airbnb-f9483f3ab7e3

Airbnb is a marketplace featuring millions of homes

Factors for decision-making during a guest's search journey

- location
- Price
- listing photos

Problem Statement

had no way to help guests find the most informative images,

No way to ensure the information conveyed in the photos was accurate

No way to advise hosts about how to improve the appeal of their images in a scalable way

How categorizing listing photos is useful to Airbnb

same room type can be grouped together

categorization makes it much easier to validate the number of certain rooms

check whether the basic room information is correct

On the guest side

it facilitates re-ranking and re-layout of photos based on distinct room types so that the ones people are most interested in will be surfaced first

On the host side

it helps automatically review listings to ensure they abide by marketplace’s high standards

Model used - ResNet50

Why ResNet50

Due to its good balance between model performance and computation time

To make it compatible with use case,

- added two extra fully connected layers
- a Softmax activation in the end

Re-training ResNet50

- Keep the base ResNet50 model fixed
- only re-train the added two layers using using massive data

two major challenges in Training ResNet50

- Even though there were lots of listing photos uploaded by hosts, it didn’t have accurate room-type labels associated with them
- Re-training a DNN like ResNet50 is highly non-trivial — There were more than 25 million parameters to train and this required substantial GPU support

For the first challenge

- asked vendors to label relatively small number of photos, usually in thousands or tens of thousands
- leveraged image captions created by hosts for room-type information and extracted labels out of it