

An Introduction to Convolutional Neural Network

Neural Networks and Learning Systems - 1

Indian Institute of Science

Motivation

Computer Vision: Image Classification



Suppose we had a set of classes : cat, truck, boat.

Motivation

Computer Vision: Image Classification

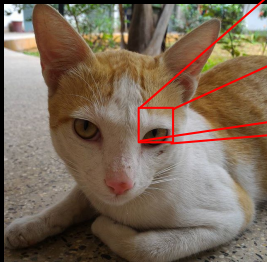


Suppose we had a set of classes : cat, truck, boat.

→ Cat

Problem: The semantic gap

Images are represented as 3D arrays of numbers, with integers between $[0,255]$.



105	111	117	121	119	120	117	116	115	121	125
104	110	114	115	112	118	115	114	111	120	133
108	111	112	114	112	112	114	115	111	117	129
116	115	112	114	116	107	113	118	116	118	130
111	115	116	114	111	106	111	105	110	132	130
106	108	108	107	105	98	105	106	117	140	130
107	106	106	106	106	102	104	106	122	143	125
113	110	109	111	112	112	111	110	127	142	116
113	111	110	111	113	112	110	114	134	143	111
108	107	107	107	107	100	106	117	136	142	107
108	109	109	109	107	101	110	117	132	135	103
113	116	117	115	113	114	119	119	124	124	98

Image classifier



CIFAR-10 dataset

10 classes

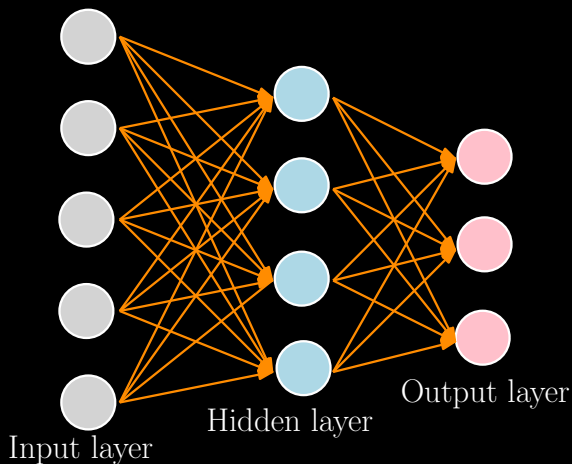
50,000 training images

10,000 testing images

32×32 image size



Multilayer Perceptron

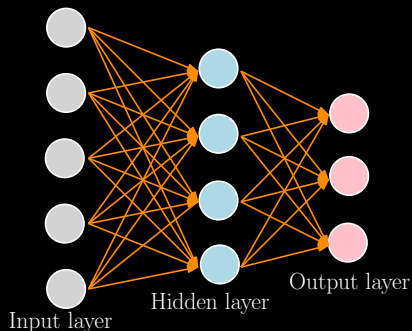


Multilayer Perceptron

Steps in training a neural network:

1. Forward propagation: Feed image and obtain loss function.
2. Back propagation : Calculate the gradients with respect to loss.
3. Update parameters: Update using the gradient.

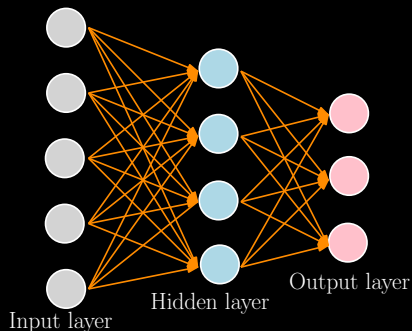
Neural Network



Say 100 hidden neurons.

- ▶ Size of image = $32 \times 32 \times 3 = 3072$.
Number of parameters = 3,07,200.

Neural Network

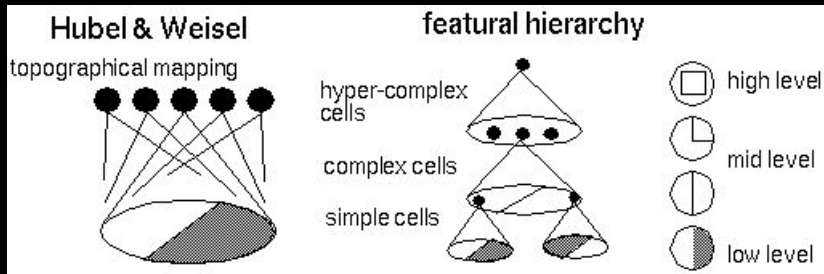


Say 100 hidden neurons.

- ▶ Size of image = $32 \times 32 \times 3 = 3072$.
Number of parameters = 3,07,200.
- ▶ Size of image = $960 \times 720 \times 3 = 20,73,600$.
Number of parameters = 20,73,60,000.

Visual Cortex

Hubel and Weisel: Feature identification using low, mid and high level features.¹



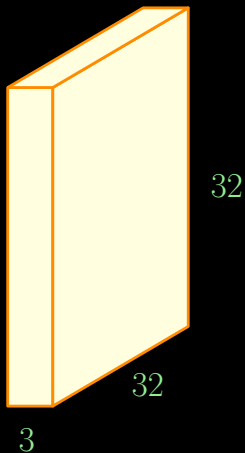
¹Hubel, David H., and Torsten N. Wiesel. "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex." *The Journal of physiology* 160.1 (1962): 106-154.

Convolution Neural Network

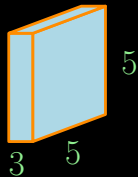
- ▶ Identifying features : **Convolution**
- ▶ Bringing invariance : **Pooling**

Convolution Neural Network

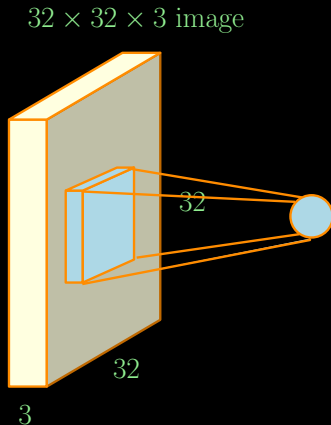
$32 \times 32 \times 3$ image



$5 \times 5 \times 3$ filter

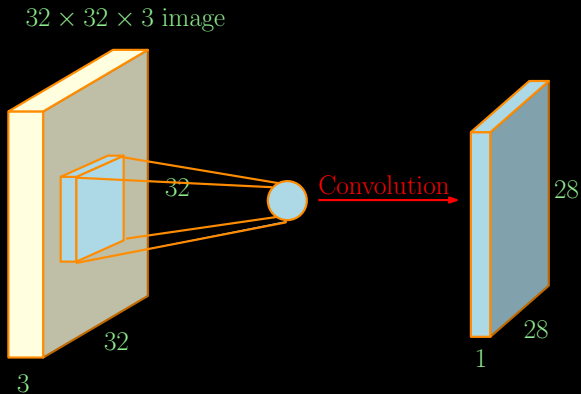


Convolution Neural Network



A single value is obtained by taking the dot product of filter and a $5 \times 5 \times 3$ portion of image.

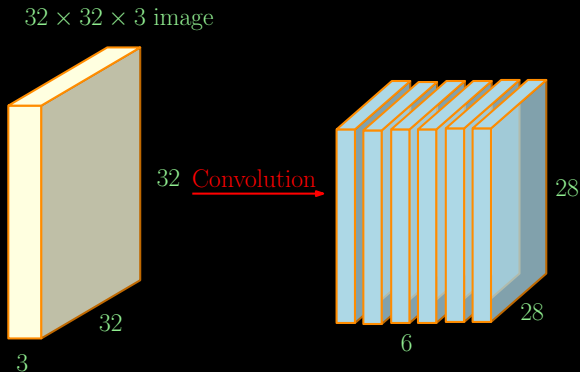
Convolution Neural Network



Convoluting the filter over the image, we obtain a map of size $28 \times 28 \times 1$.

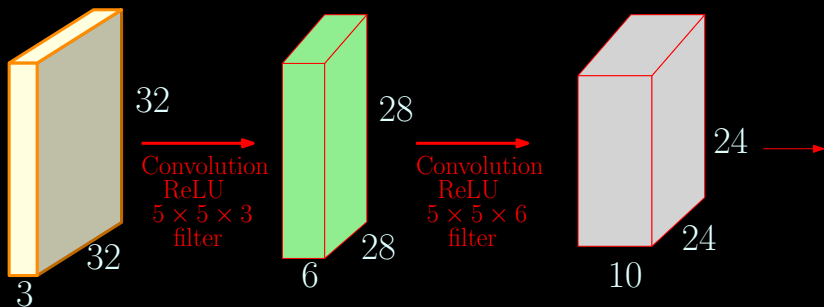
Convolution Neural Network

If we have six 5×5 filters, we will get 6 maps.



This forms the new input of size $28 \times 28 \times 6$.

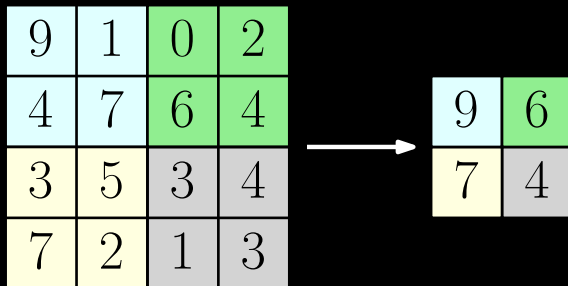
Convolution Neural Network



Depth of filters increases as the layers increase.

Pooling

MAX POOLING - Downsamples the input by 2 along the width and height by performing pooling over 2×2 with a stride of 2. Discards 75% of the activations.



Convolution Neural Network

