Convolutional Neural Network

* A bit of history

⇒ The Mark 1 perceptron machine was the first implementation of the perceptron algorithm.

$$f(x) = \begin{cases} 1 & \text{if } w^T x + b > 0 \\ 0 & \text{otherwise} \end{cases}$$

⇒ Update rule:

$$w_i(t+1) = w_i(t) + \alpha(d_j - y_j(t))x_{j,i}$$

Frank Rosenblatt ~ [1957]: Perceptron

⇒ Stack the linear layer to multi-layer perceptron

Widrow & Hoff, ~ [1960]: Adaline/Madaline

⇒ First time back-propagation became popular.

Rumelhart et al. [1986]

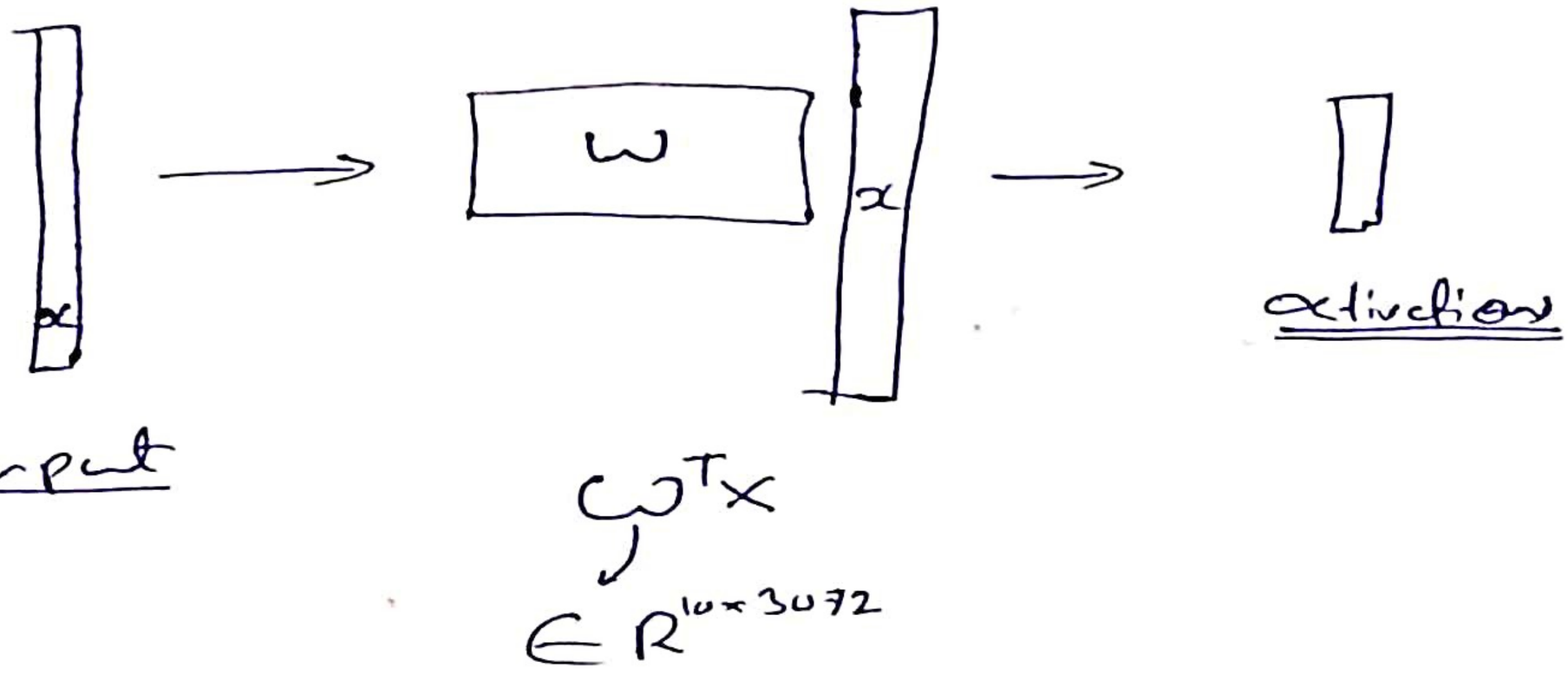
⇒ Image net classification with deep convolutional neural network

Alex Krizhevsky, Ilya Sutskever, Geoffrey E Hinton [2012]

Classification → Detection → Segmentation

* Fully Connected layer

32 × 32 × 3 image ⟶ Stretch to 3072 × 1



Input

$W^T x$

$\in R^{10 \times 3072}$

* Convolution Layer

x 32 × 32 × 3 image ⟶ Preserve spatial Structure

+

W 5 × 5 × 3 filter

⇒ Convolve the filter with the image.

{ Slide over the image spatially, computing dot product }

⇒ Filters always extend the full depth of the input volume.

$W^T x + b$

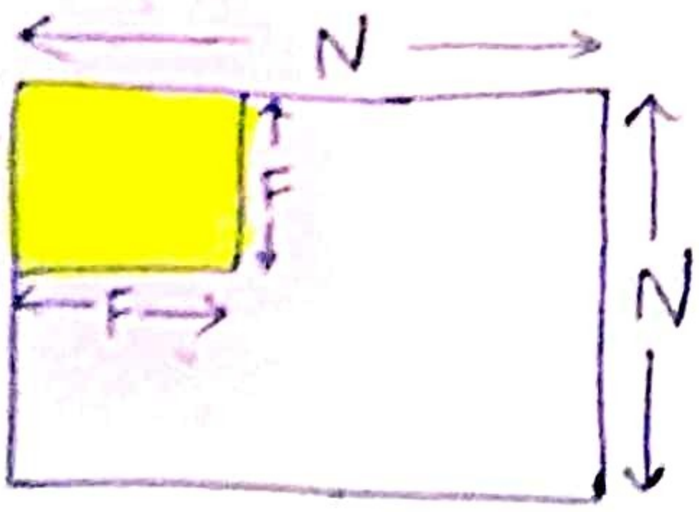
⇒ We can have many filters in each layer.
  ↳ Exaple, if we had 6 5×5 filters, we'll get 6 Separate activation maps.

$3 \times 32 \times 32$ — Convolution Layer → $6 \times (28 \times 28)$

Input Image — Activation Map

⇒ ConvNet is a sequence of Convolution Layers, interspersed with activation functions.
{Exaple ReLU}

Stride
↳ Amount of movement of filter over image
Example: 1, 2 etc.



Output Size: $\frac{N-F}{\text{stride}} + 1$

⇒ In practice it is common to zero pad the border.
  ↳ To maintain input & output size, in Convolution.

F = 3 → Zero pad with 1
F = 5 → Zero pad with 2
F = 7 → Zero pad with 3

⇒ Total number of filters in a layer is generally power of 2
  Eg: 32, 64, 128 etc...

5×5 filter

↳ 5×5 receptive field for each neuron

★ Pooling layer

→ Makes the representation smaller & more manageable.

→ Operates over each activation map independently.

$224 \times 224 \times 64 \xrightarrow{\text{pool}} 112 \times 112 \times 64$

Input → Output

Example: Max Pooling, Avg Pooling

↓

1	1	2	4
5	6	7	8
3	2	1	0
1	2	3	3

→ max pool with 2×2 filter & stride 2 →

6	8
3	4

★ Fully Connected Layer (FC Layer)

→ Contains neurons that connect to the entire input volume, as in ordinary Neural Network.