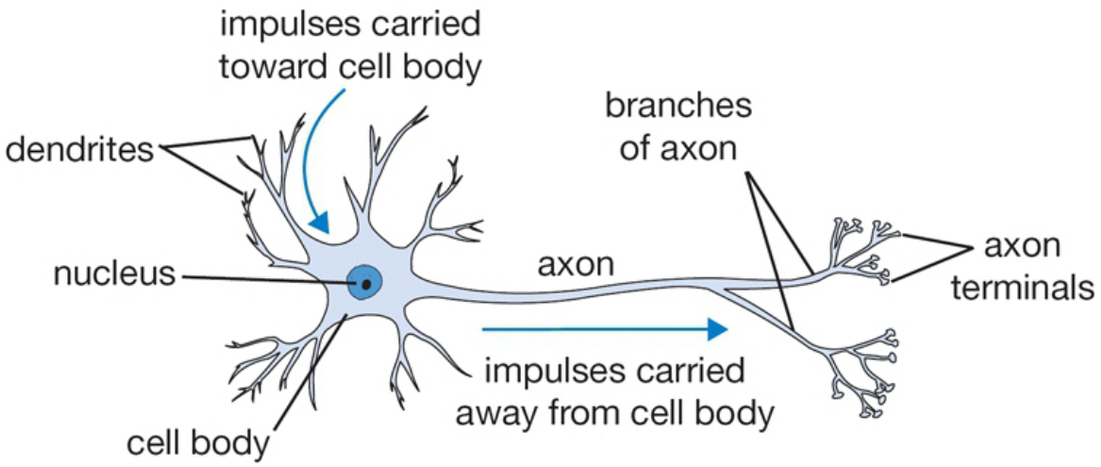
Computer Vision and Cnn

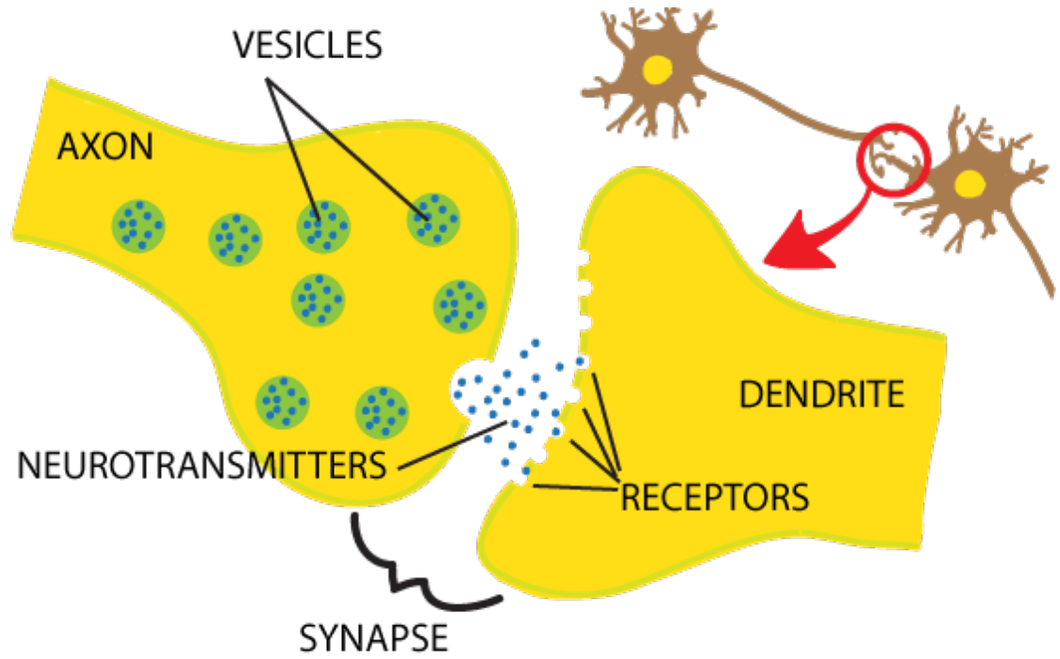
Raghavendra Singh

Ingredients of Neural Networks

* Data
    * Learning Concepts represented by Data
* Problem
    * Converted to a differential loss function
* Neurons
    * Non Linearity
* Architecture
    * How are neurons interconnected

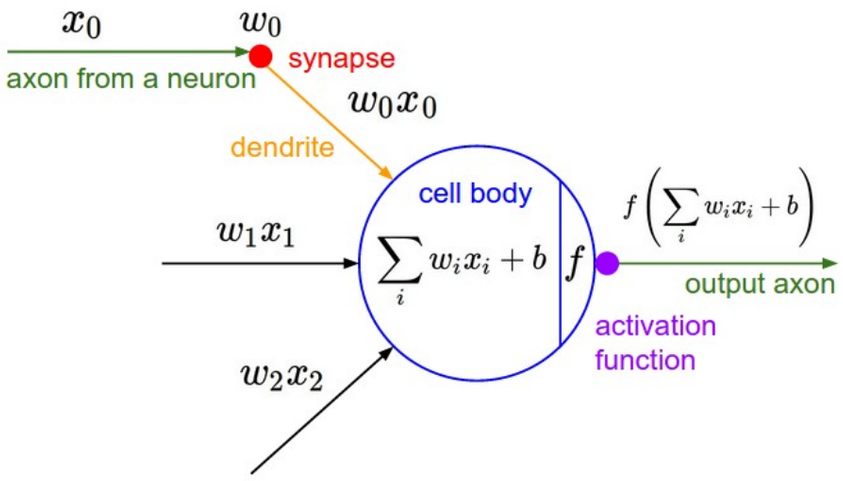
Biological Neurons

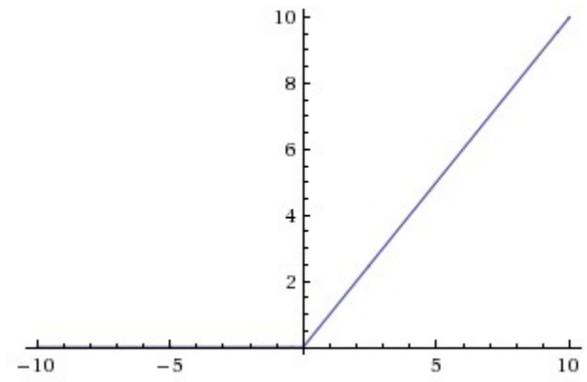




* **Integrate** signals on dendrites
* If result greater than a threshold send out an **impulse** along axon
    * Leak potential if no incoming signal
* Impulse carries **time stamp**
* Neurotransmitters in **synapse modulate** signal strength
    * Neurotransmitter properties change as brain "learns"

Artificial Neurons

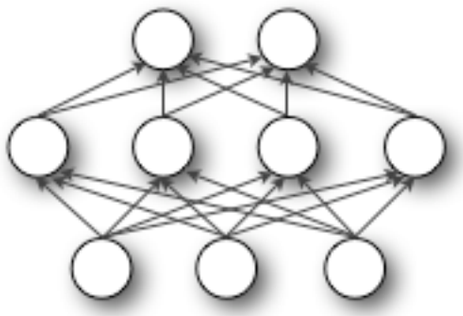


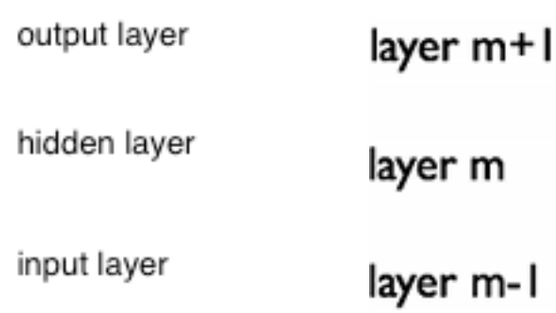


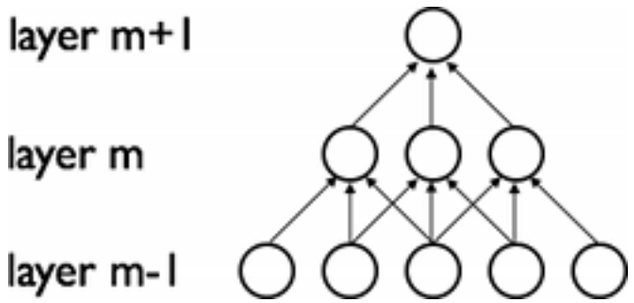
ReLU

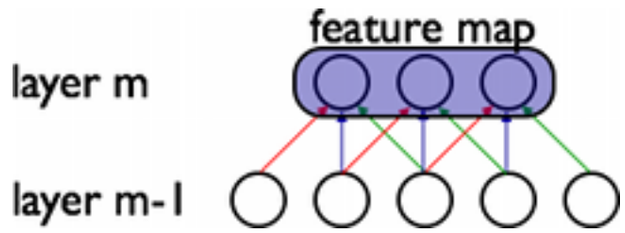
- **Sum weighted** inputs
- Activation function is a **non-linear** function, e.g. ReLu
- **No time** stamp
- Synapse — weights that **multiply** input
  - Weights and biases are "learned" from data

Feedforward Architectures



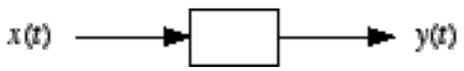






* MLP
    * All neurons in this layer connected to all neurons in the next layer
* Sparse MLP
    * Some neurons in this layer connected to a neuron in next layer
        * Some usually defined by concept of locality
* Sparse + Shared MLP == CNN
    * Connections have shared weights

Linear Time Invariant Systems



$$\delta(t) = \begin{cases} 1 & t = 0 \\ 0 & t \neq 0 \end{cases}$$

$$x(t) \equiv \sum_i \delta(t - \tau_i)\, x(\tau_i)$$

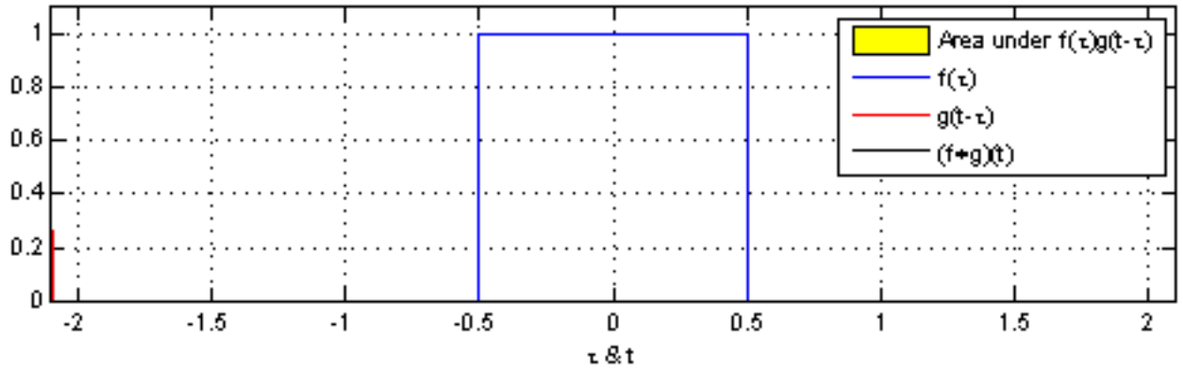
linearity -principle of supposition

h(t)
Impulse response

$$\begin{aligned} \delta(t) &\rightarrow h(t) \\ \delta(t-\tau_i) &\rightarrow h(t-\tau_i) \\ x(\tau_i)\,\delta(t-\tau_i) &\rightarrow x(\tau_i)\,h(t-\tau_i) \\ \sum_i x(\tau_i)\,\delta(t-\tau_i) &\rightarrow \sum_i x(\tau_i)\,h(t-\tau_i) \end{aligned}$$

$$y(t) \equiv \sum_i x(\tau_i)\, h(t - \tau_i)$$

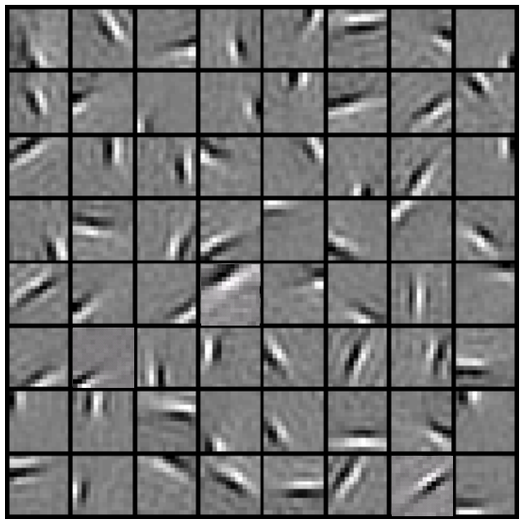
Convolution



Images == LSI Signals

* Images often modelled as linear shift invariant signals

Natural Images



Learned Basis

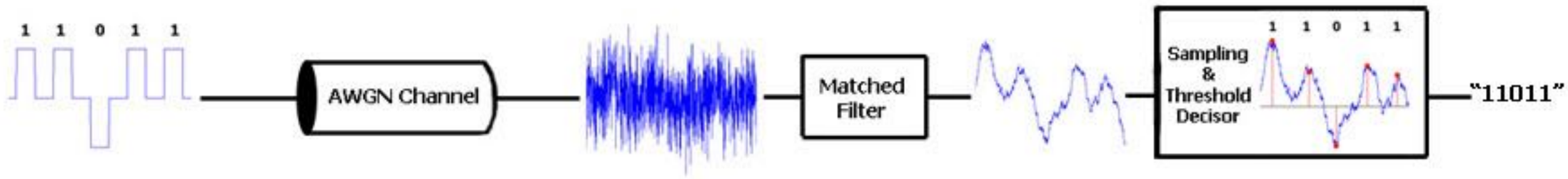
Decomposition

Image Filtering

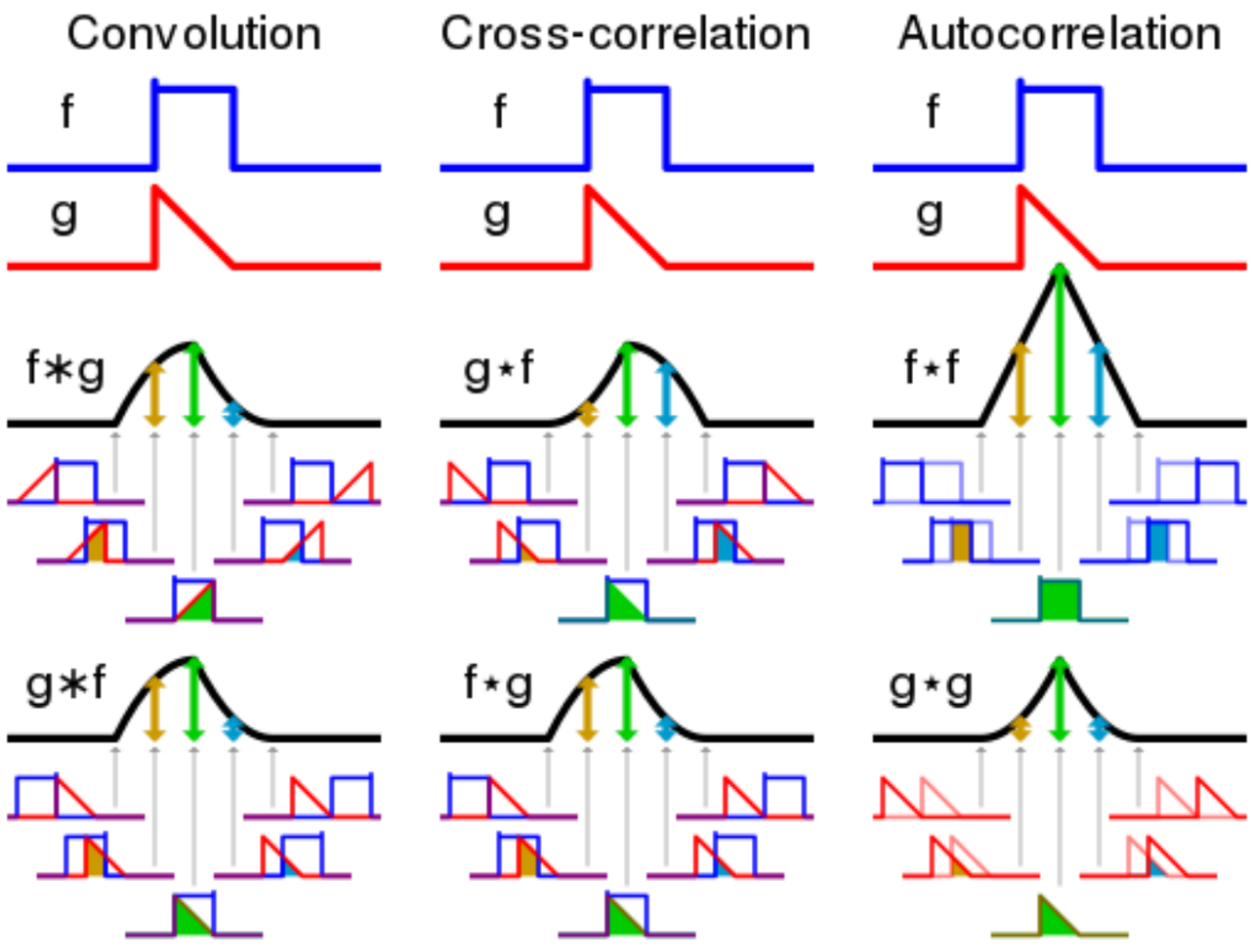
- To "filter" an image, or "transform" an image often involves convolution operation

Matching Filter

* Matched filter: To maximise SNR at receiver

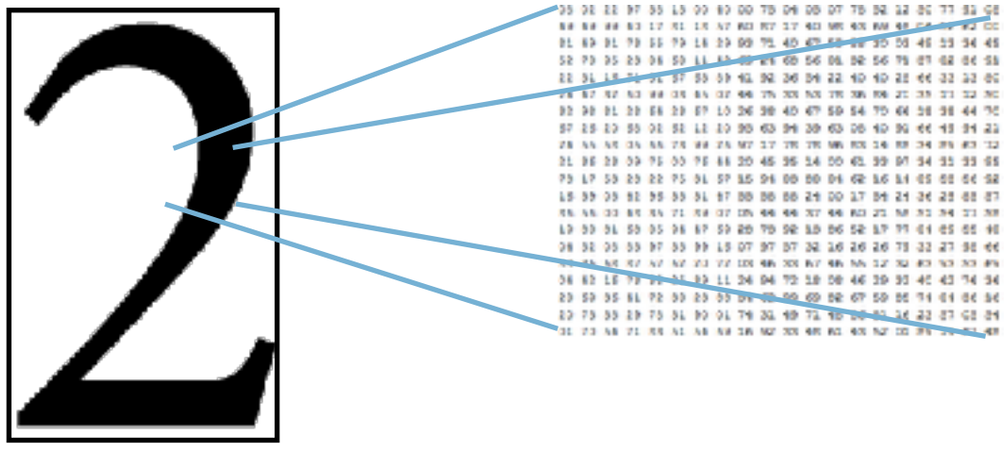


Convolution and Correlation



Computer "Vision"

Image



Computer sees a matrix of numbers

Classifying an image is to label the image

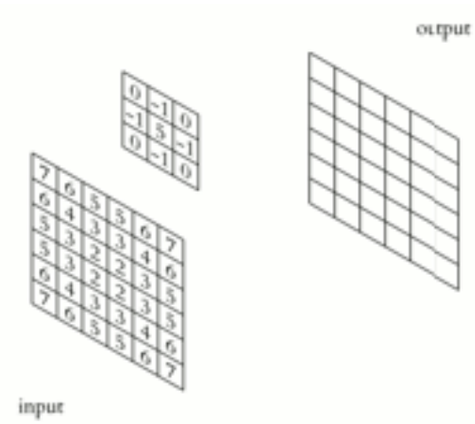
2

Understanding an image is to describe the image

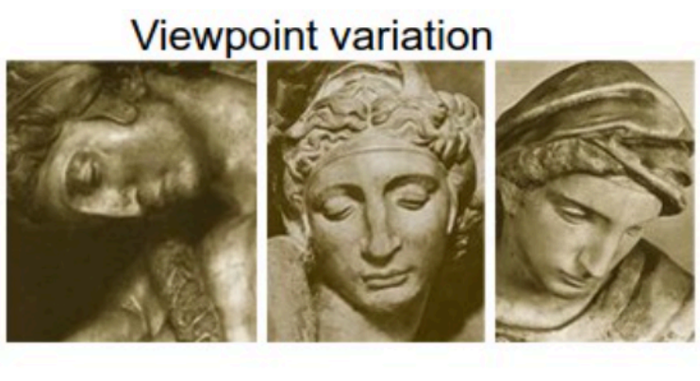
numeral 2

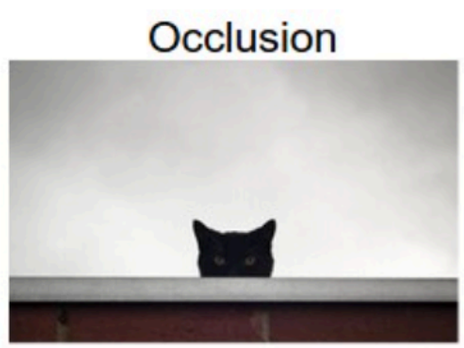
in black on white background

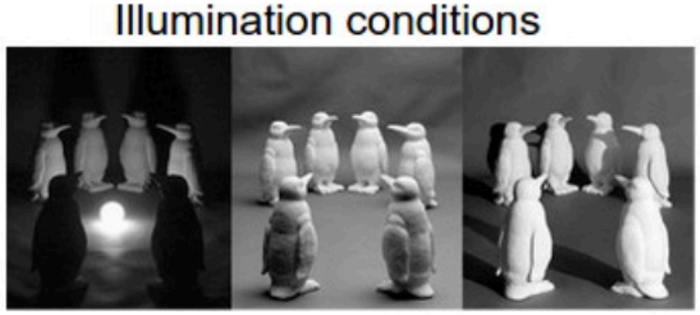
Processing an image is to mathematically operate on it

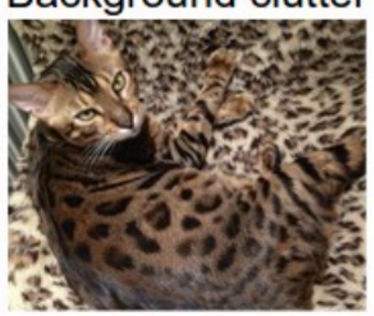


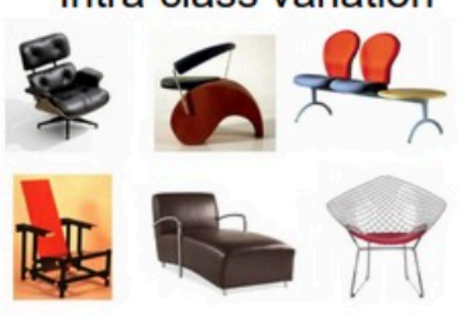
Why is it hard?



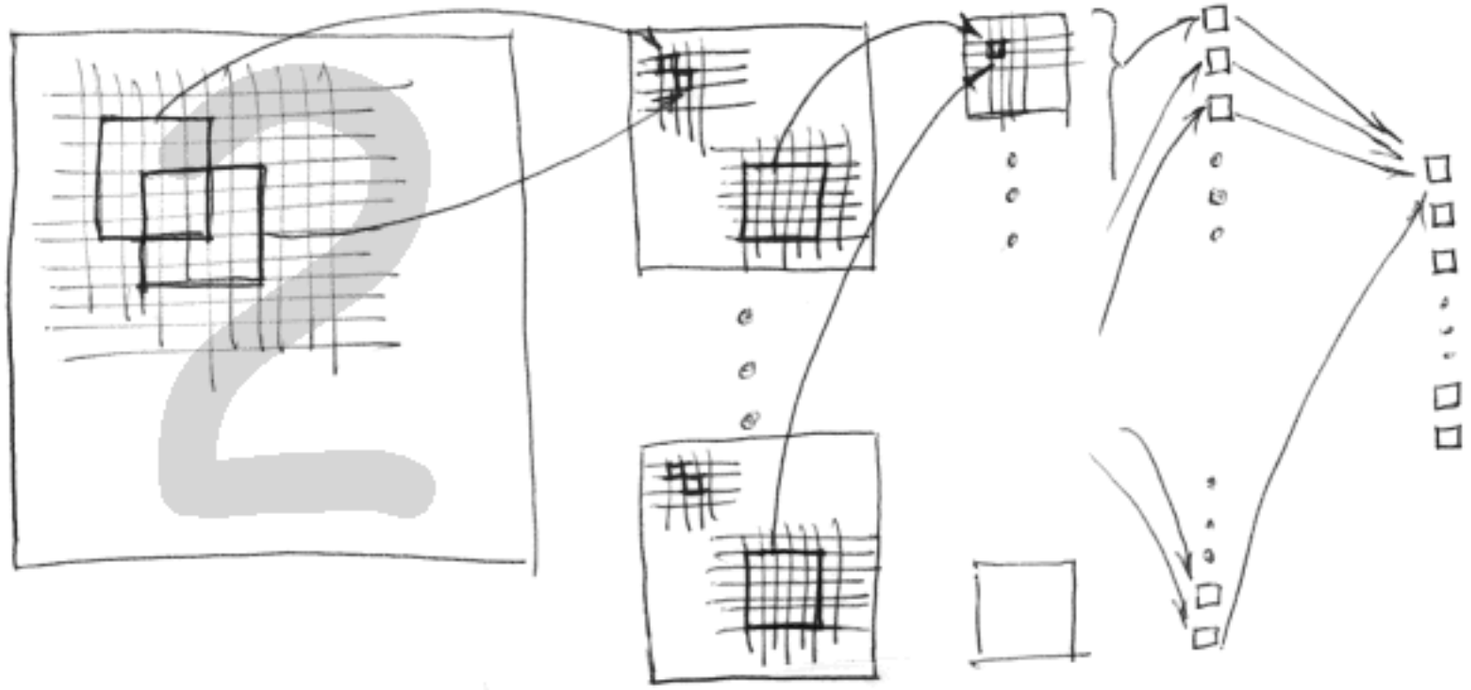








Convolutional Neural Networks

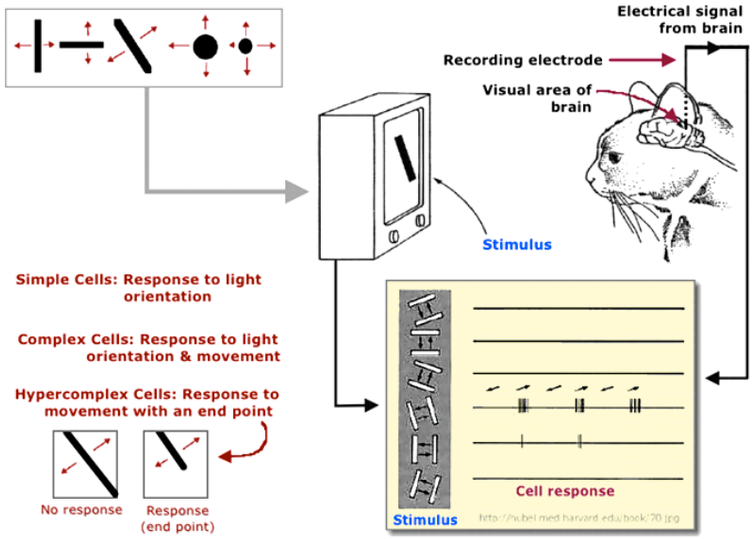


* **Layers** of operations
* Each operation is **local** —
* Uses common **learned parameters** —

Convolution

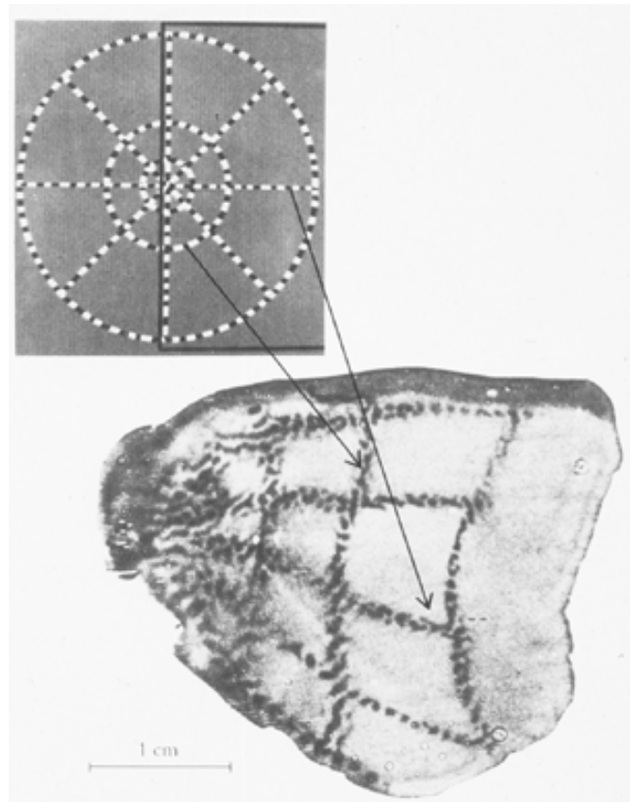
* Each operation by a **artificial neuron**

Why Convolution



Hubel & Wiesel, 1959

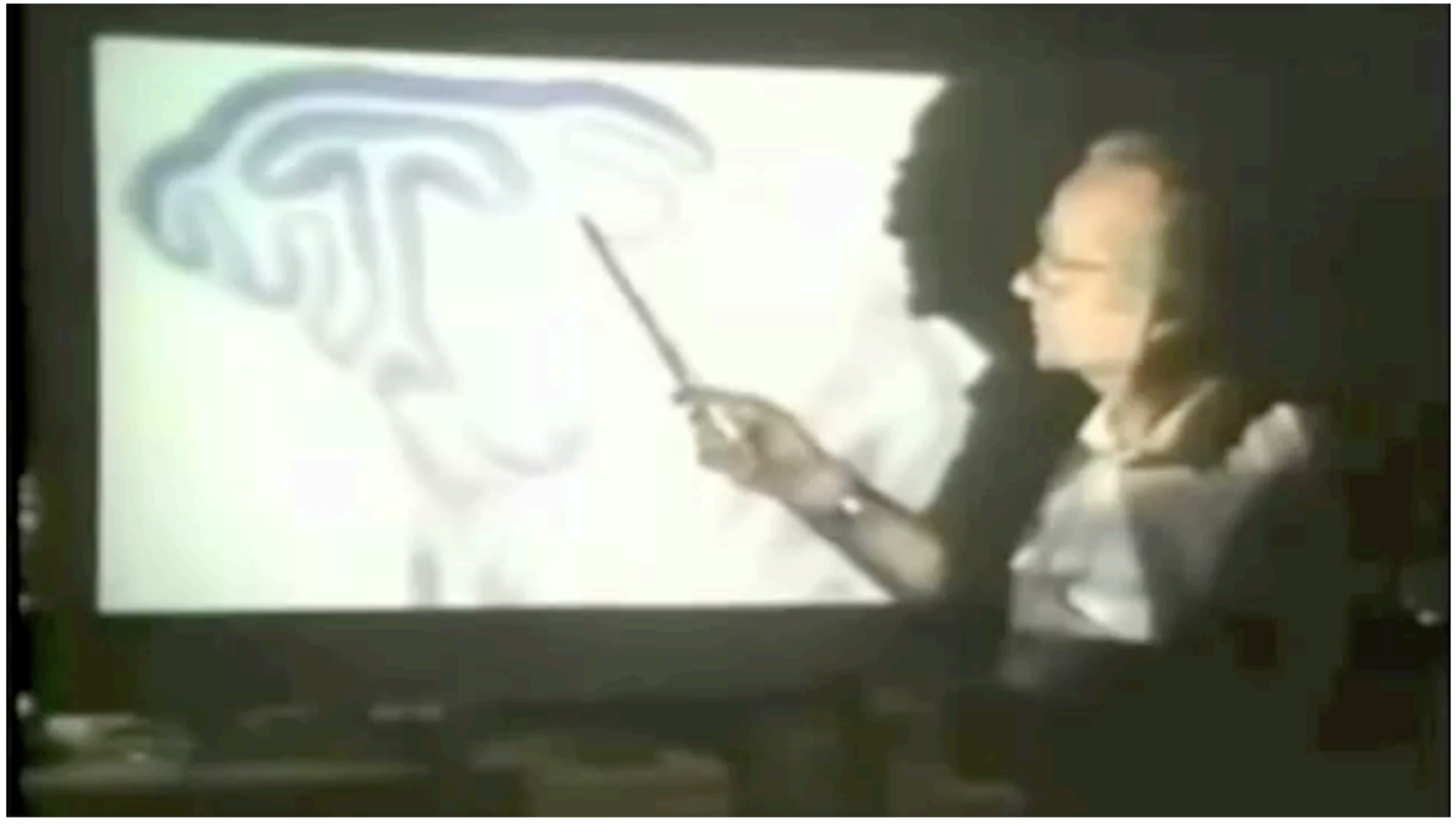
* Receptive fields of visual system are local
* Filters are oriented



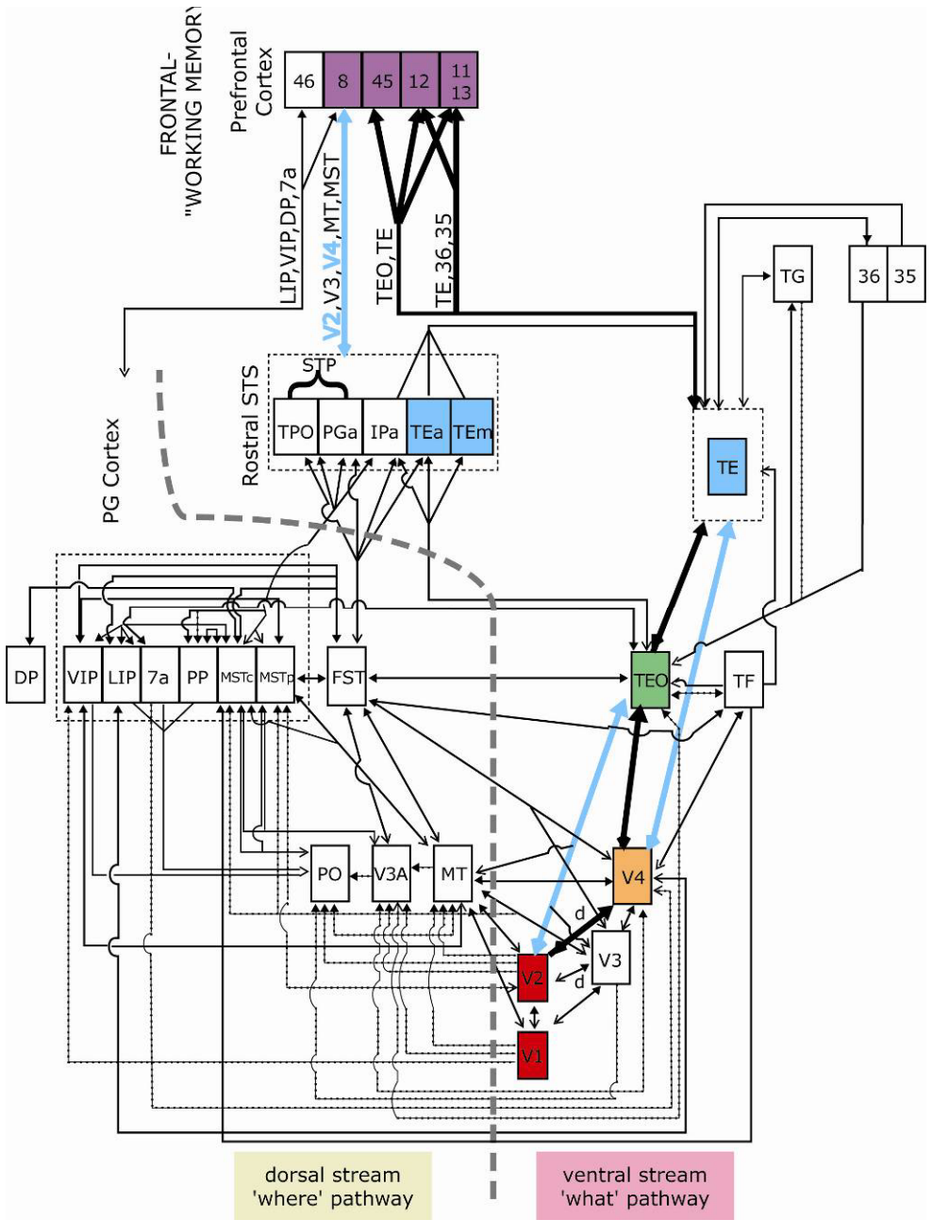
* Topographic Mapping — nearby cells in retina processed by nearby regions in the brain

Hubel & Wiesel

* Won Nobel Prize for this work



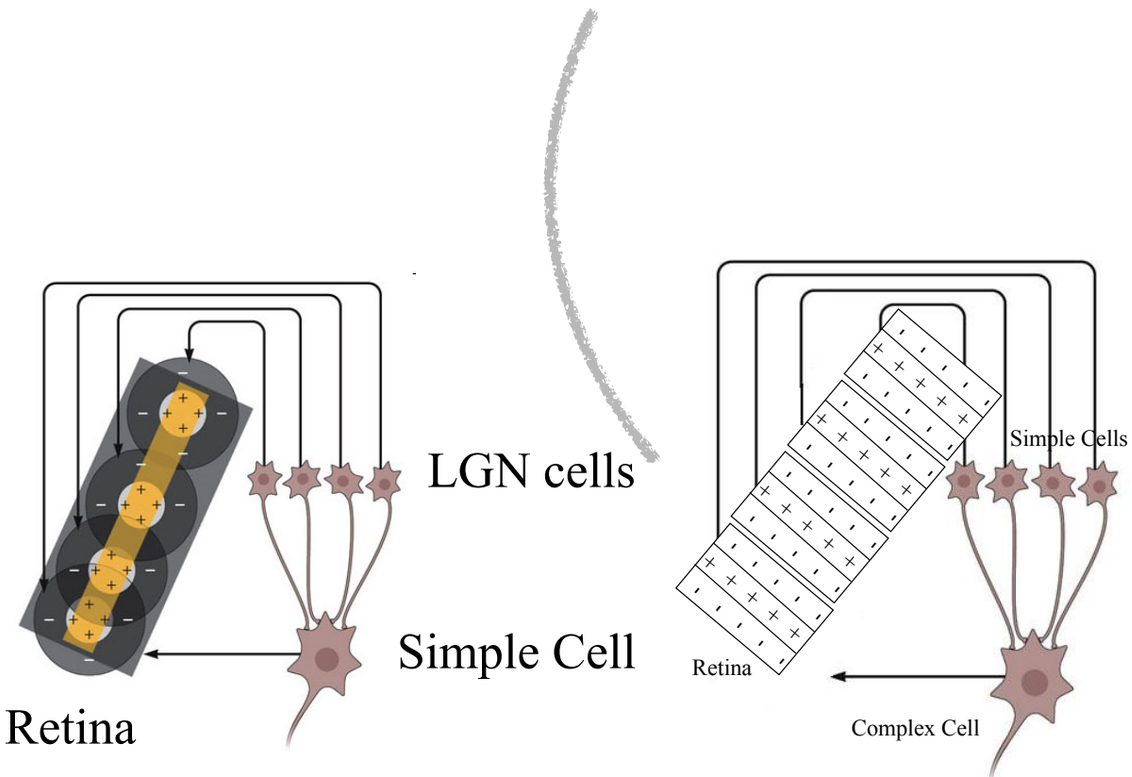
Why Layers



Modified from (Ungerleider & VanEssen)

* Visual system has layers of processing

* Hierarchy of cells

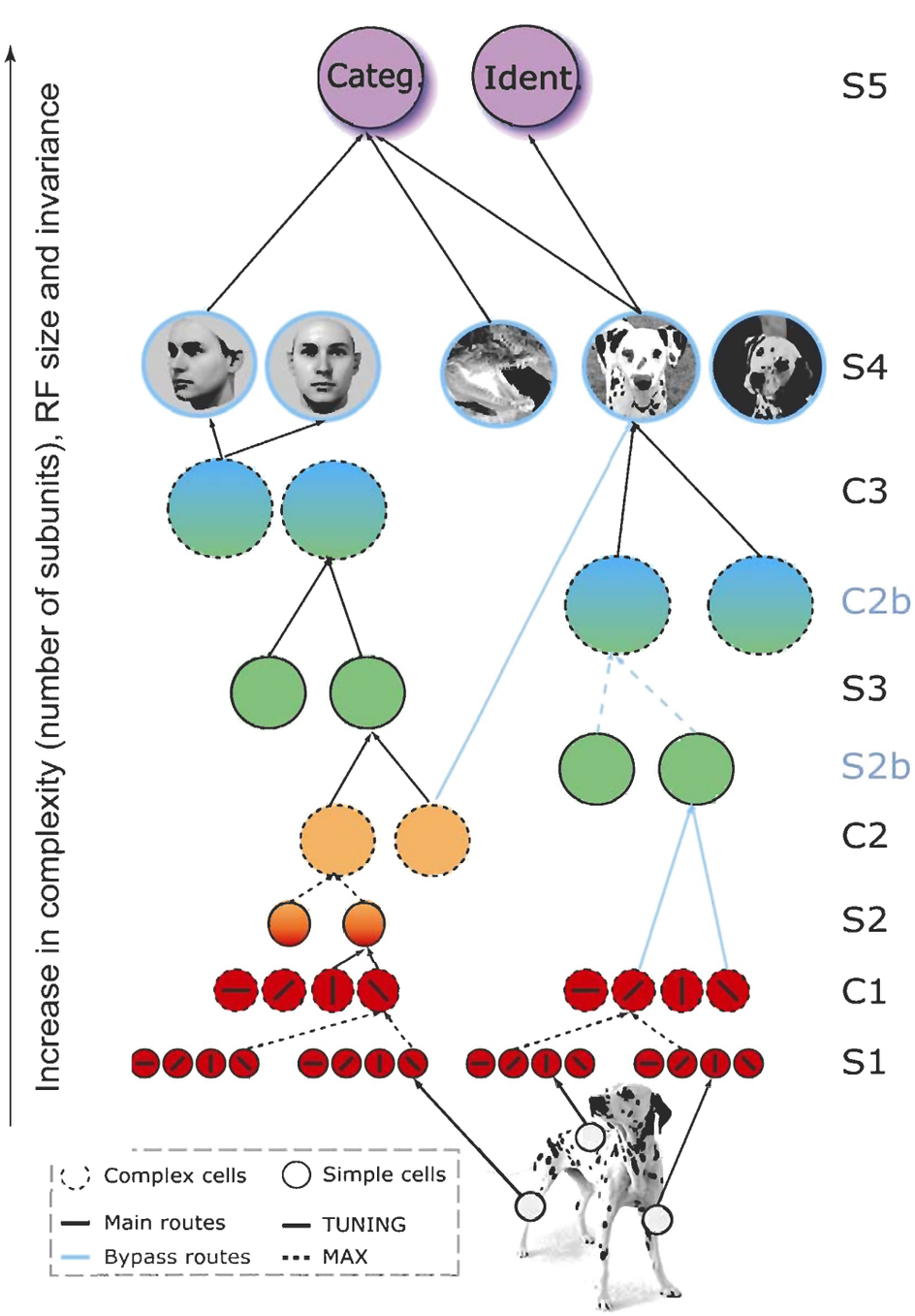


Visual Processing Model

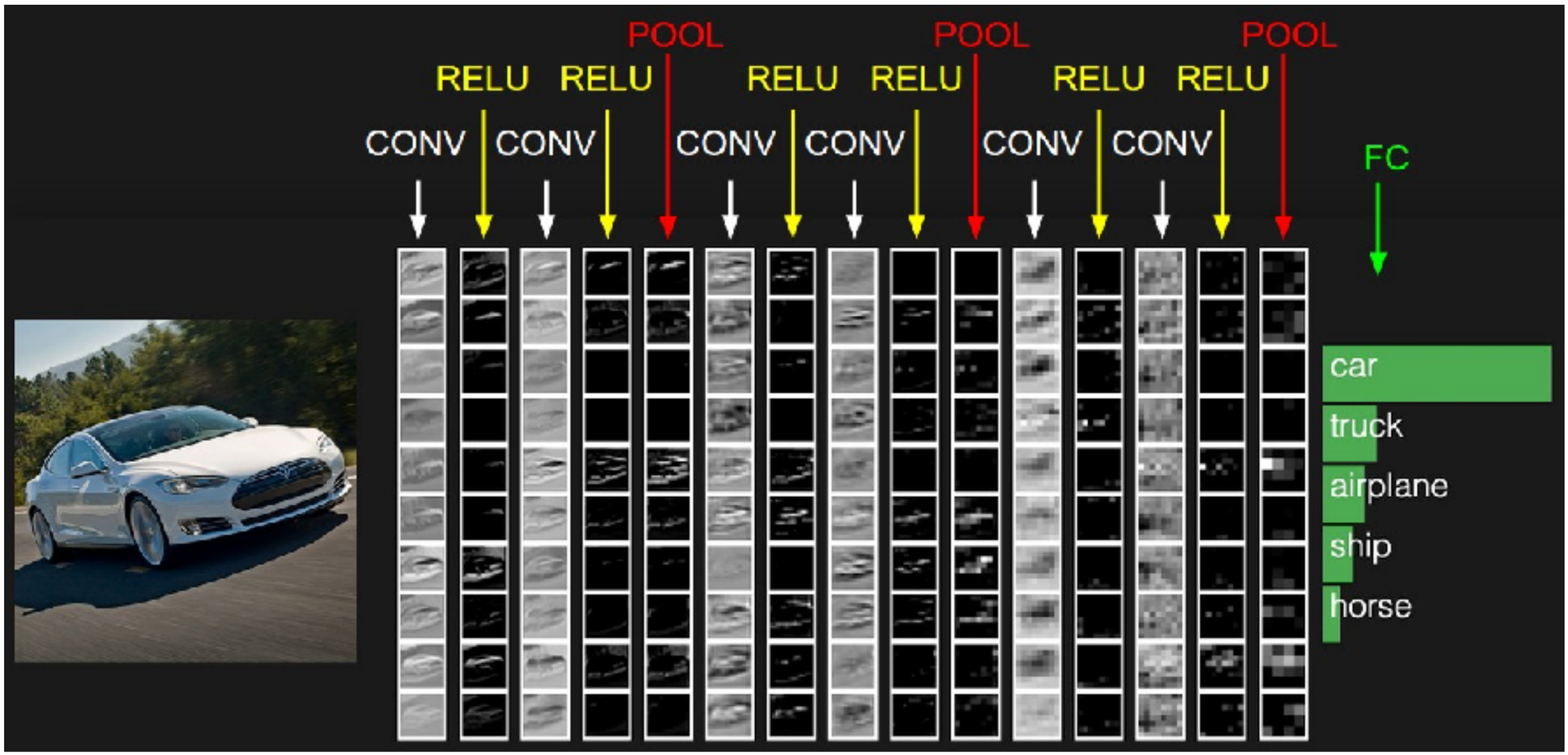
Model layers	Corresponding brain area (tentative)	RF sizes	Number units	
classifier	PFC		1.0 10^0	
S4	AIT	>4.4°	1.5 10^2	~ 5,000 subunits
C3	PIT - AIT	>4.4°	2.5 10^3	
C2b	PIT	>4.4°	2.5 10^3	
S3	PIT	1.2°- 3.2°	7.4 10^4	~ 100 subunits
S2b	V4 - PIT	0.9°- 4.4°	1.0 10^7	~ 100 subunits
C2	V4	1.1°- 3.0°	2.8 10^5	
S2	V2 - V4	0.6°- 2.4°	1.0 10^7	~ 10 subunits
C1	V1 - V2	0.4°- 1.6°	1.2 10^4	
S1	V1 - V2	0.2°- 1.1°	1.6 10^6	

Supervised task-dependent learning

Unsupervised task-independent learning



Convolutional NN



* Parameters of CONV layers are **learned** from data

Emergence of simple cell properties

- Batch of images each with its labels (car, truck, airplane...)
- Loss how well does predicted label match actual label
- Minimize loss
- Emergence of simple cell properties
  - Hubel and Weisel showed that cat responded to oriented bars
  - Learned visual fields of CNN are invariably oriented fiters