CSEP 576
Machine Learning
Neural Networks

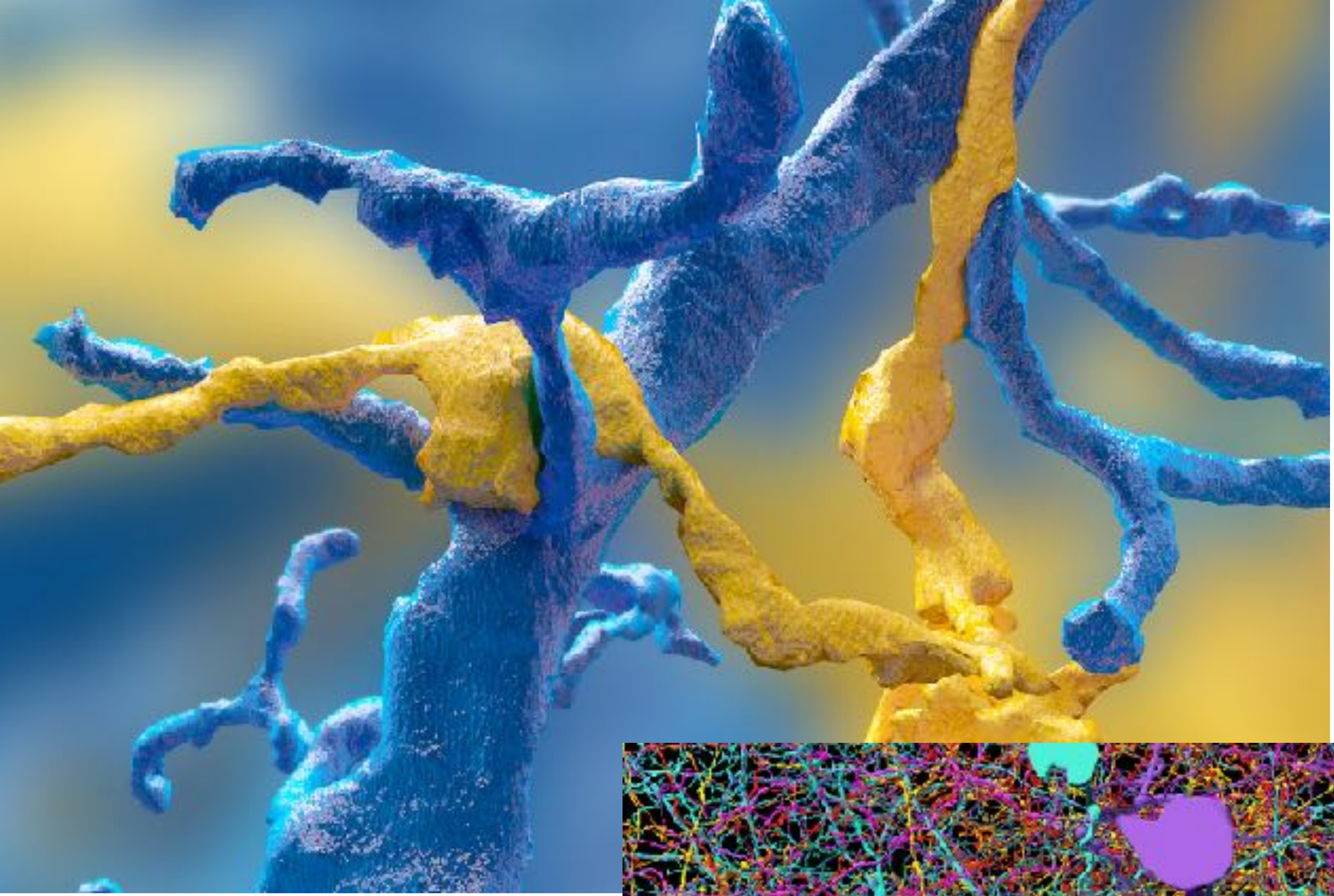
Robert Gens

Today

- Act I
    - Perceptrons
    - Multilayered Perceptrons
- Act II
    - Backpropagation
- Act III
    - Convolutional Neural Networks

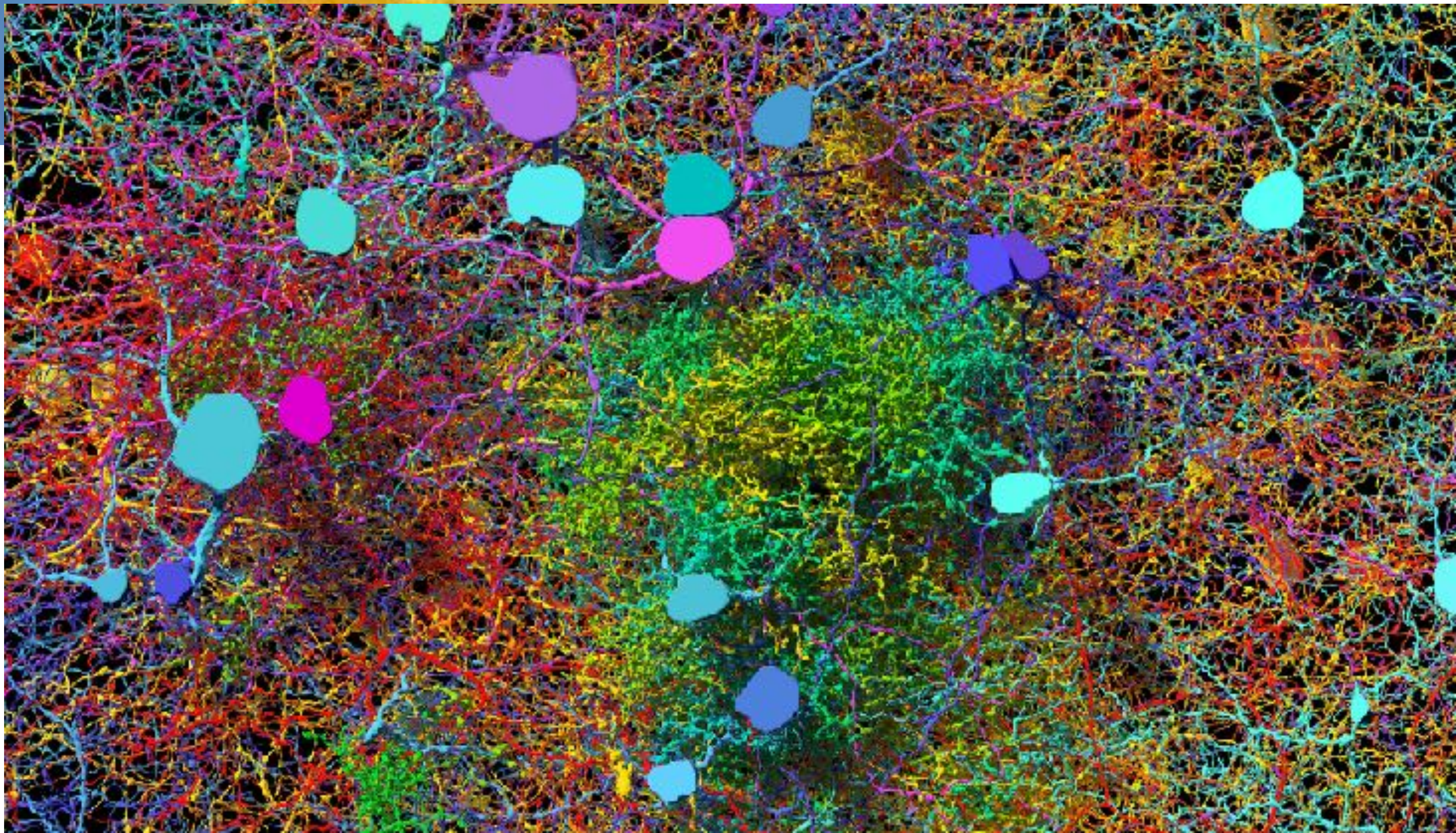
Neural Networks

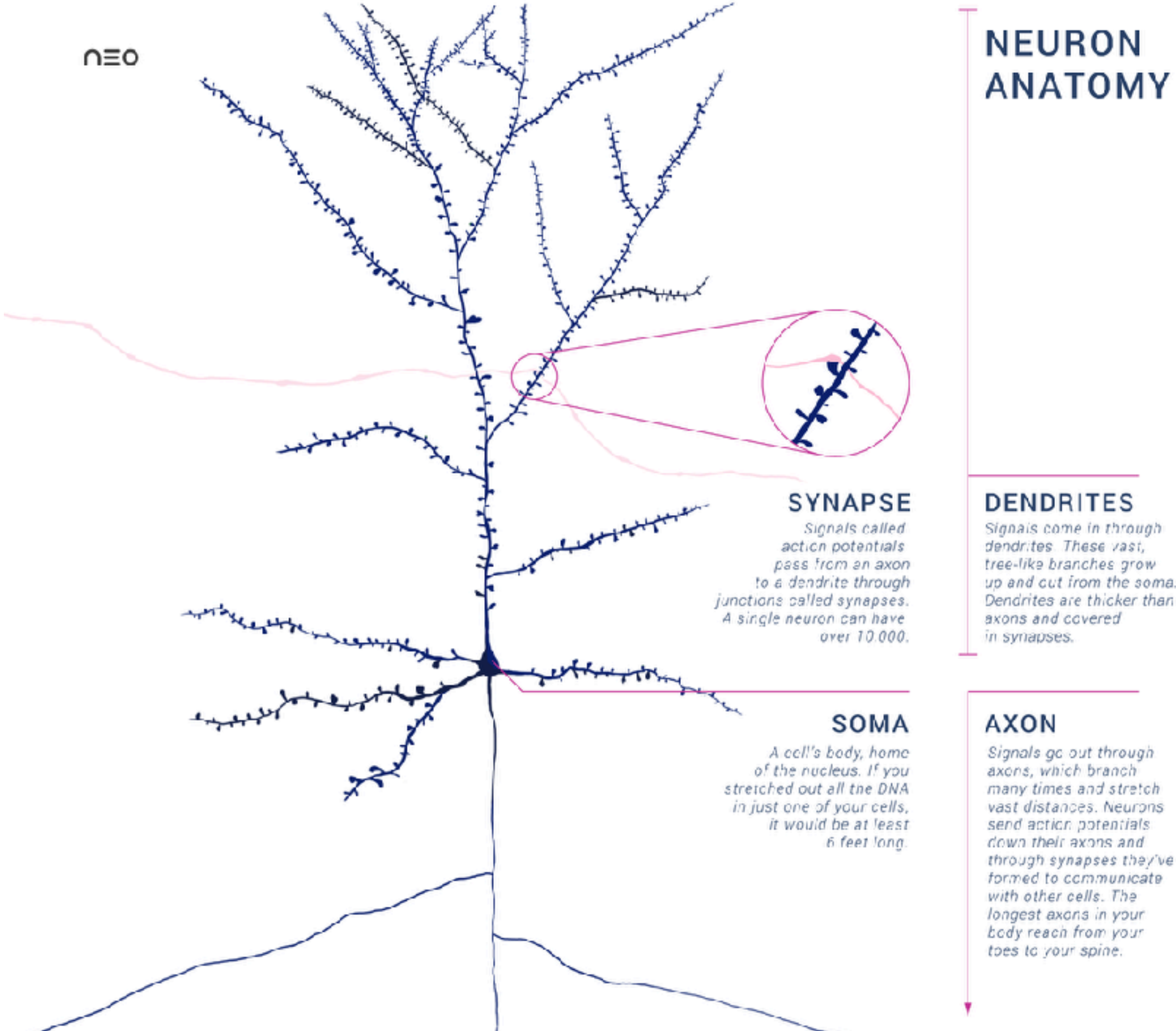
Inspired by the human brain:	**GV100 Volta**
• Neuron switching time ~ .001 second	10^{-9} second
• Number of neurons ~ 10^{11}	10^{10} transistors
• Connections per neuron ~ 10^4	2
• Total synapses ~ $10^{15} \rightarrow 10^{14}$	
• Scene recognition time ~ .1 second	
• All in parallel	
• Total power ~ 13 Watts	300 Watts



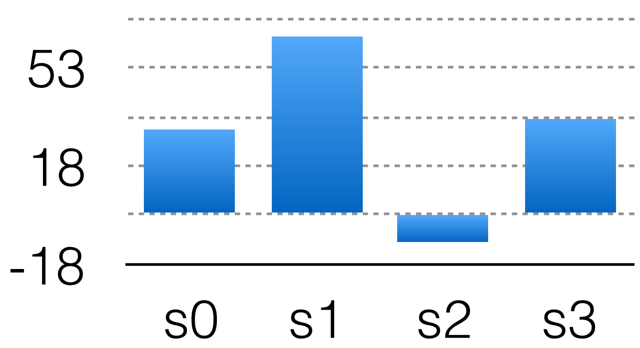
Synapse by Seung Lab, licensed by CC 2.0

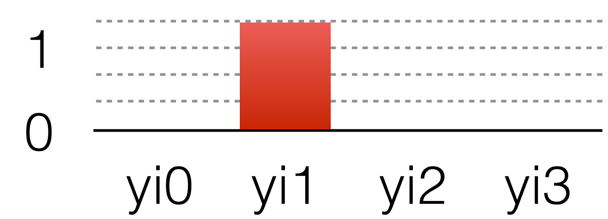
Dense Carpet of Neurons by Seung Lab, licensed by CC 2.0



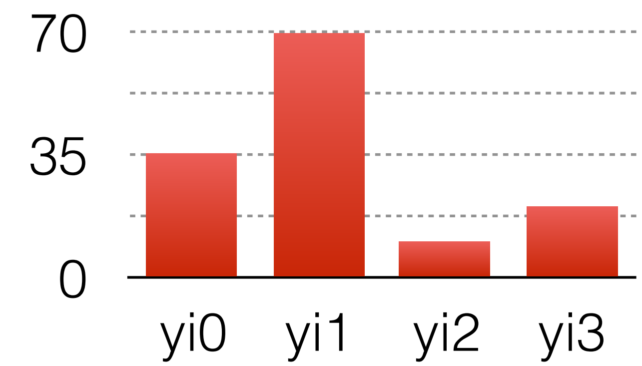

NEURON ANATOMY
SYNAPSE
Signals called action potentials pass from an axon to a dendrite through junctions called synapses. A single neuron can have over 10,000.
DENDRITES
Signals come in through dendrites. These vast, tree-like branches grow up and out from the soma. Dendrites are thicker than axons and covered in synapses.
SOMA
A cell's body, home of the nucleus. If you stretched out all the DNA in just one of your cells, it would be at least 6 feet long.
AXON
Signals go out through axons, which branch many times and stretch vast distances. Neurons send action potentials down their axons and through synapses they've formed to communicate with other cells. The longest axons in your body reach from your toes to your spine.

Loss functions





Classification



Regression

- Classification (y_i is label)
  - Softmax cross-entropy
    $$L_i = -ln(\frac{e^{s_{y_i}}}{\sum_j e^{s_j}})$$
  - Multiclass hinge loss
    $$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$
- Regression (y_i is vector)
  - L1/L2 distance
    $$L_i = \sum_d |s_d - y_{i_d}|^2$$
  - Robust norms
    $$L_i = \sum_d p(s_d - y_{i_d})$$

Backpropagation Algorithm

Initialize all weights to small random numbers

Until convergence:

    For each training example:

        1. Input example to network

        2. For each unit from bottom to top (**forward pass**):

            Compute output from inputs

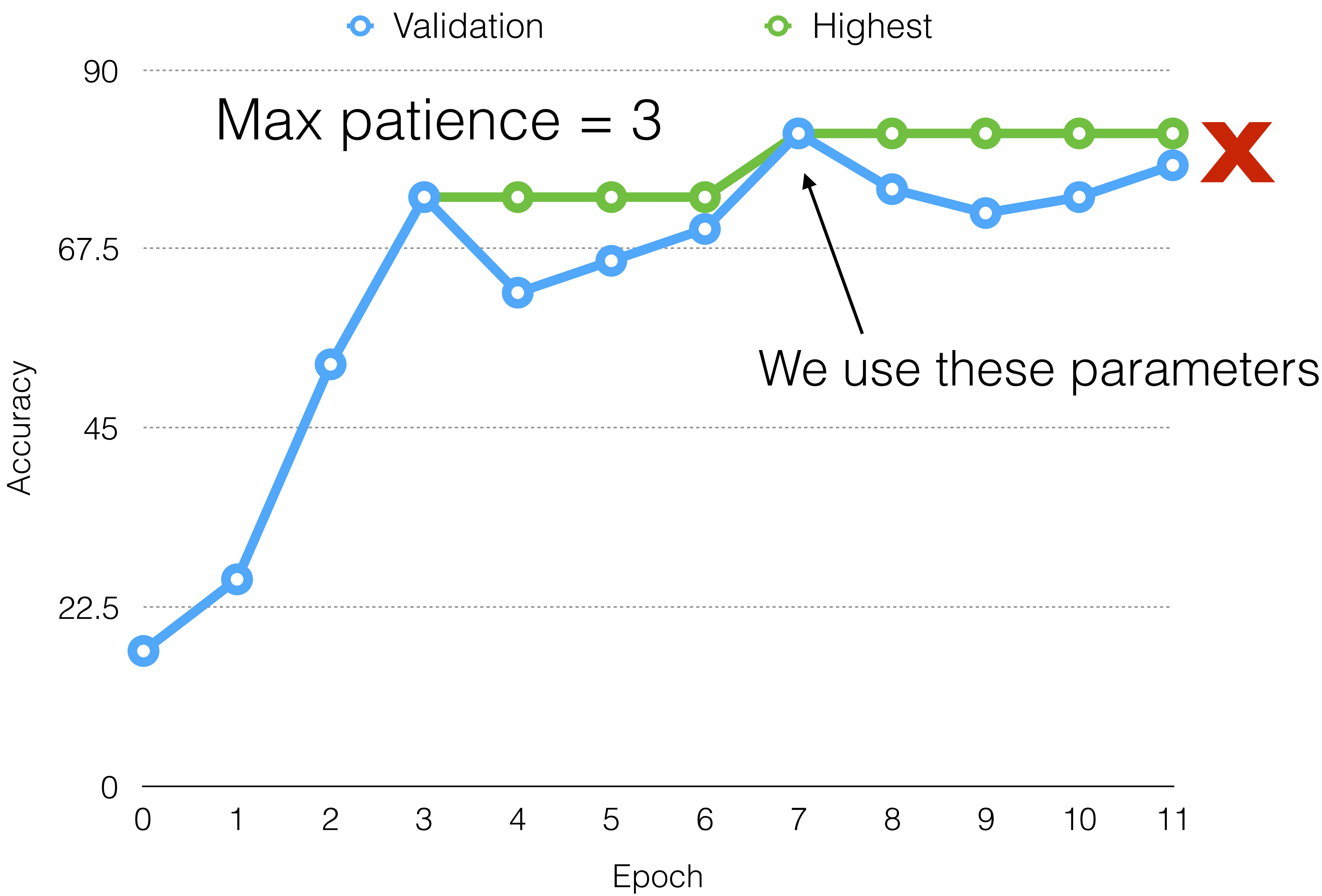
        3. For each unit from top to bottom (**backward pass**):

            Compute gradient of inputs from gradient of outputs

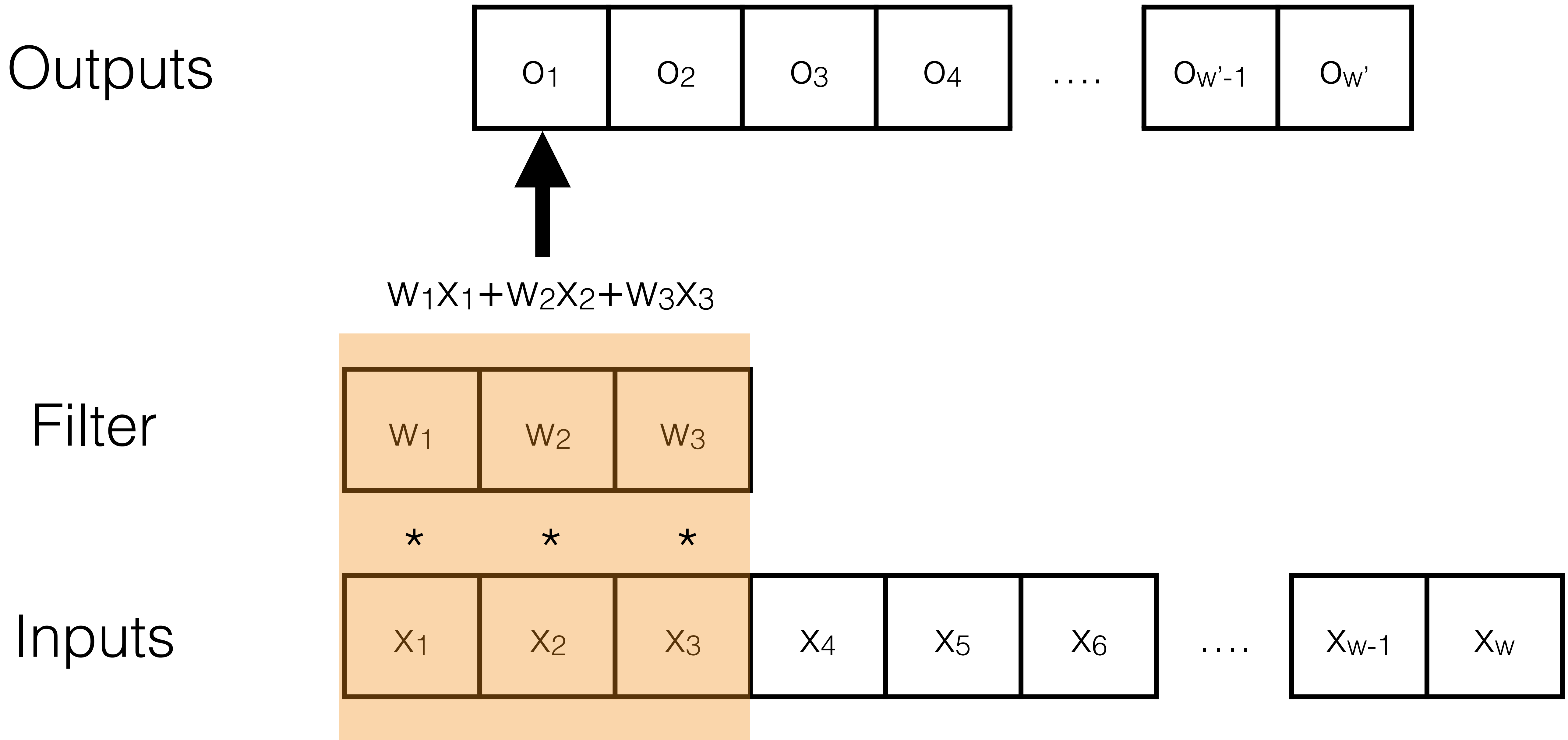
        4. Update each network weight

$$w_i \leftarrow w_i - \eta \frac{\partial E}{\partial w_i}$$

Early Stopping



Gradient of 1D Conv



Given $\frac{\partial L}{\partial o_i}$ **, what are** $\frac{\partial L}{\partial w_i}$ **and** $\frac{\partial L}{\partial x_i}$ **?**