Contents

1 Introduction and Recap

For projects, keep your eyes open for things you're interested in working on. We may not get to cover a lot of things, so a lot of room for projects here: E.g., grid and place cells.

We talked about the cerebral cortex – everything else had to keep up with it as it started to grow during evolution. It is spectactularly bigger than a lot of other brains out there. One of the resources you can check is a sheet with a bunch of numbers about the brain. The cortex has 6 layers, and each has functional differences. There are different types, different signals, etc.

We also talked about multiple ways to record the brain: single unit, fMRI, EEG, ECoG, and optogenetics. Optogenetics is very cool – you modify the cells in the brain so that when certain light is shined on them, they fire and you can see them.

2 Vision and Convolutional Nets

2.1 The Eye

Look at a diagram of the eye. We have a lot of photoreceptors. The retina is much denser, and is visible as a dark spot. There are more cells closer to the center of the retina – they are not uniformly distributed.

What is an optimal eye? Best eye has at least one fovia – Olshausen's new paper. Some have two fovias.

If you look beneath what's in the retina, if you look in the microscope, you see something remarkable. There are two kinds of receptor cells: cones and rods. Cones are for color, rods are for seeing in the dark. To see the stars, you need to capture every photon, otherwise you won't be able to see. There are also bipolar cells, which connect to the photoreceptors – they are processing the information.

Remark 2.1. Neanderthals apparently had a larger visual cortex than humans. Apparently insects can also have these properties – it's not just humans.

Back to the eye – there is a procession of information. There are things called retinal ganglia. They take the result of the preprocessing and send it out of the eye. It's interesting to see the stages. How fast does this all happen (each stage)?

So, how do we know all this information? Three people: George Wald, Ragnar Granit, Haldan Keffer Hartline won the 1967 Nobel prize for physiology and chemistry of vision.

Charles Darwin was surprised about the eye – he disbelieved his own theory of natural selection for the evolution of the eye.

2.2 How did the eye evolve?

There's a paper on the web which says the eye would evolve in 100,000 years. But, evolution is a random walk.

There have been many variants of the eye – in the beginning, there were in the skins of organisms. It makes sense to create a cavity for the eye, and then you get to the stage where the cavity almost closes. Then, it can develop some kind of fluid. Then, there is a protolens, and finally a more developed eye. The most developed animal is a squid.

What is the purpose of cavity: In the beginning, protection. And then, focusing on particular things. It helps focus.

An interesting case of eyes: The jumping spider (Olshausen's favorite book, *Animal Eyes*). They have been trying to record from jumping spiders, and now they can. There is an animal with 34 eyes.

What do plants use – they don't have eyes. Cavities don't just provide protection, they also provide more ability to rotate the eye.

If one looks at the phylogenetic tree to vertebrates, you can see the branches which departed before. What they say is that in every one of these animals, during development, you go through the stages – it starts with jawless fishes. In their development in the fetus, they go through certain stages of the eye. In vertebrates, it looks different (look at the chart in the slides to compare).

2.3 The Visual Processing Pipeline

Back to the diagram of the eye: There is a blind spot due to lack of receptors on the retina – you have learned to ignore it, but it is always there (similar to how you learn to see things upside down). So what happens after that? From the two eyes, the nerves cross. The sides of the brain are unified by the corpus callosum. Importantly, the nerves make a stop at the thalamus (lateral geniculate nucleas, or LGN, is here), which acts as a relay station, but much more beyond that. We don't fully understand what the thalamus does. And then what happens after that? It's like a typical deep network. They go to the cerebral cortex – the V1 area. Then from V1, there are the dorsal (the top) and ventral (the bottom) streams.

The dorsal captures “where” information. The ventral captures the “what” pathway – these are approximate.

What arguments are there for why the optical nerve is split up – is there a justification? This is somewhat speculative, but you might want some integration before the nerves pass to the cortex – for instance, eyes track together. You have two eyes for depth perception. According to Wikipedia, depth perception, the crossing of eye streams appears most often in predatory animals who need depth perception to be able to chase prey.

How did Polyak find out the paths – he looked at many dead people (I assume). You can trace neurons all across the brain by staining and a microscope.

Let’s talk about LGN. There are two parts: small cells (ventral path goes through) and large cells (dorsal path goes through). There are a lot of connections between this path and further up. This part helps give you the sense of where your body is, but everything is of course a simplification. What’s the evidence? Hundreds of papers that have been trying to figure these things out. It’s what the field happens to believe. They have some amazing techniques, but some of these items may change. It’s unclear.

One thing I used to think was impossible: How does animal vision – how do we know it’s the same person, from multiple perspectives? It’s sort of amazing. The total time from the path from eye to back of brain is something like 150ms. This is very fast for so many stages. There’s the retina, then the LGN, etc. Probably only 1/10 of this time is actual activation. So the action length is not a problem. But how many stages? How many neurons have to fire? What is the depth? Is it too fast? too slow?

Each neuron firing every 20 ms, total axon distance traveled is about 10 ms, human reaction to a visual stimulus is 250ms, audio is 170ms, touch is 150 ms. Btw, to avoid breaking your car – you need a reaction of 1.5 seconds.

Back to the pathway though - it’s not only in one direction, signals travel both ways. See Jeff Hawkins, “On Intelligence”. He wrote this book in which he says that you know what you expect to see, and you compare it, all of the cells change. They’re not “seeing”, they’re “comparing”. There is another beautiful piece, a philosophical paper, which basically argues that people work on vision too much in isolation – it’s called “The Critique of Pure Vision”.

There is a shallow hierarchy in audio apparently.

2.4 Computation in the Visual Pathway

So what is the computation going on in the visual pathway?

2.4.1 Hubel-Wiesel Cat Experiment

This starts out with the Hubel-Wiesel experiment. They recorded cat vision over LGN. They changed stimulus parameters, such as angle or horizontal offset, and calculated firing rates, receptive fields of each neuron as a function of the parameters.

Why is this important? The eye never stays still. There is a very regular 6 Hz motion where we focus on different parts of the visual scene. There are microcicades. So, Hubel and Wiesel have been doing experiments for months, and nothing has happened. They were showing all kinds of interesting things to the cat, and they were just getting noise. Then, they were showing some slides to the cat. And the slide projector broke. What the cat saw was just a horizontal black line. And then the neurons started going crazy. That's when they realized the visual cortex is lines. It was an incredible moment for neuroscience. We see all these complicated things – first, the cortex finds the lines, then we create everything else. Lines is the first thing we compute. Line needs computation – it takes a lot of sensory neurons to detect a line. So this is a really amazing moment. Make sure to watch the video of their experiments, detecting different line orientations in one location. These are the simple cells. Other kinds, complex cells, have one direction, but any spatial location.

So, Hubel and Wiesel essentially pointed out the importance of edge detection (1962). They got a Nobel Prize (1980) after they published these results and thought for three years to come up with a conjecture about how this works. I'll show you the conjecture, but let's think about it first.

You have a lot of wires coming out of a particular neuron telling you a particular neuron sees light. The question is, how are you going to set things up so that you can detect lines? After you can do this, how can you make detection spatially invariant? Any thoughts? You could have a bunch of ANDs for simple cells (for a set of neurons), and ORs across the simple cells to get the complex cells. That sounds great; let's go deeper. What happens is from cells in the LGN that happen to reflect visual photoreceptors in a line, you get an input to another neuron. If all of them fire, then this new neuron fires. Maybe you have some other cells which are inhibitory. The discussion section of Hubel and Wiesel 1962 is very interesting speculation. The conjunction of a bunch of LGN cells might be a line – that's how simple cells can be formed.

But, it also fires if you put a big square over the whole thing – so you need inhibition, there must be more to it. Also note that there are only a few simple cells in V1. There are others that are phase invariant and so on (look in the Fourier domain, invariant to signals with different spectra).

One thing that is sort of amazing about all of this: How do the LGN cells know they're collinear? Even if they understand, how come they send their signals to the same new cell? One might learn as a child to use these things correctly. In the beginning we don't understand they're collinear, and we could learn the notion of collinearity? But how about the actions? How do they all end up there? The LGN preserves the retina map, so this might be an answer to that question.

This idea was considered a huge turning point in all of neuroscience. As you can see, this is very computational. We're asking what kind of circuit could do this.

How far removed from the rods/cones are these simple cells – at least 4 -6 steps. That could give insights into what computational insights into what is going on underneath – when is the earliest the lines can be detected from the information in the computational pathway? It's interesting in convolutional networks – already at the second layer or so you can have lines/edges. The complex cells is the inspiration for neural nets. That's where this is going.

2.4.2 The Computer Scientists' Perspective

Meanwhile, since the 1940s, computer scientists were also interested in neurons. There was the McCulloch-Pitts neuron in the 40s, in the 50s there was the Rosenblatt perceptron. Then, Minsky and Papert noticed that the perceptron cannot recognize exclusive OR – so that's a disaster. This caused the first AI winter – they wrote a book saying multi-layer neurons are not a good idea. In 1975, Fukushima came up with an interesting neural net for vision – the Cognitron. But there was a lot of criticism: It can recognize patterns in only some locations. Then they looked at Hubel-Wiesel paper and had an idea. Five years later, he came up with the Neocognitron, as in neocortex – he was inspired by the complex cells of Hubel and Wiesel. In some sense, he combined his cognitron with convolutional networks. That was the first instance of convolutional networks. For a while, he was the champion for handwritten digit recognition. This basically says that every cell receives input from an input – it scans the current layer to produce the output for the next layer.

2.4.3 Modern High-Dimensional Models

How are weights in modern convolutional nets learned? Gradient descent typically.

Gradient descent is a very common sense algorithm. However, what is known theoretically about gradient descent? A slightly less trivial theorem – If f is a convex function, and L is a bound on the eigenvalues of the Hessian, then setting $\alpha = 1/L$ guarantees logarithmic convergence to the global minimum. If nonconvex, no such thing can be guaranteed.

However, this doesn't seem to matter – for instance, Olshausen and Field do the optimization anyways on some nonconvex objective. There is lately a lot of mathematical work on why it does not matter that things are nonconvex in high dimensions – gradient descent still works well.

Why have we believed this ? Intuition from low dimensions isn't so good – we can't think well in high dimensions! For instance, what do you think the volume of a high-dimensional ball is as the dimension increases? It goes to zero – the ℓ_2 ball vanishes, if the radius is kept fixed.

This means that all the volume/mass is in the periphery – that's what high dimensions look like.

Stochastic gradient descent is when you sample a function to optimize from the data (in batches, for instance). This is what has conquered the world. This is more computationally efficient than iterating over all the data.