#h1 27 - Introduction to Biological Vision

#h2 Photoreceptors, neurons, pathways, and receptive fields

#h3 27.1 - Introduction

#pg To fully appreciate the innovation convolutional neural networks introduced, it is a good idea to look at the machinery they were ultimately designed to mimic: the human visual system.

#im ../assets/figures/027/027-01.png 50 256 Figure 27.1 - Human vision serves as the paradigm for understanding perception. It guides the efforts to build AI that can mimic its abilities.

#h3 27.2 - The Retina

#pg Vision starts when light enters the eye and passes through the retina containing many photoreceptors. In absolute darkness, photoreceptors cells are polarized, meaning there is a difference in voltage inside the cells relative to the outside. Photoreceptors absorb photons, causing a chemical reaction that depolarizes the cell slightly. And the more photons are absorbed, the more the cell depolarizes and the longer it remains in that state. When the stream of photons stops, the cell will return to its original polarized state. The depolarization of the cell reduces the amount of neurotransmitters it releases. In turn, bipolar cells connected to photoreceptors detect the change in neurotransmitters. Already in this stage, we see a form of neural computation arise, as many photoreceptors project to a single bipolar cell. Projecting is the terminology used to describe how one neuron or group of neurons (B) is connected anatomically and functionally to another neuron or groups of neurons A, with activity (and thus information) flowing from A to B. Any activity generated by A has the potential to generate sufficient activity in B. Or, in other words, A projects to B. When many photoreceptors project to single bipolar cells, this is not random. Bipolar cells receive inputs from photoreceptors that are clustered together. And if photoreceptors cluster together, they receive photons coming from roughly the same direction. In other words, those photoreceptors are sensitive to a particular location within the large image of what the eyes are looking at. The bipolar cells thus can be said to already compute spatial averages across the photoreceptor's inputs, presumably to reduce the noise across the photoreceptors. In turn, the bipolar cells project to ganglion cells which also pool the activity from many bipolar cells. However, unlike the simple spatial averaging the bipolar cells performed, the ganglion cells create the fundamental building block of vision: the ON/OFF cells. It is the first time in vision that we see a relative response rather than an absolute one. ON/OFF ganglion cells combine photoreceptors to compare the positive input at the center with the negative input of the area around it (called the surround). Light hitting the surround will make the ganglion cell less active, and light hitting the center will make it more active. To get the maximum response out of the cell, we should only present light to the center, not the surrounding. What does that look like? It will look like a small bright dot on a darker background. For all practical purposes, ganglion cells are dot detectors. The stimulus representation that will yield a maximum response is called the receptive field. Also note that for every ON/OFF center-surround ganglion cell, there is probably an OFF/ON center-surround ganglion cell as well, responding optimally to the same input but inverted.

#im ../assets/figures/027/027-02.png 50 384 Figure 27.2 - The retina already augments incoming information by combining and aggregating signals to enhance and sharpen the response of the photoreceptors.

#h3 27.2 - The visual field and what you think you see but do not.

#pg This large image of what you can see at any given time is called the visual field. As you are reading this, your eyes are continuously looking directly at (called fixating) each word in turn. The fovea is the area directly around your fixation point and has the highest resolution. This higher resolution results from photoreceptors being far more densely packed on the retina at that location. Outside of that is the periphery, where the resolution is quite a bit lower, corresponding with a much lower density of photoreceptors. We typically don't notice this, but the difference is substantial. Outside the fovea, there is hardly any color vision, although we perceive the world in full color. The human visual system is highly adept in filling in the blanks if it doesn't have the necessary information as inputs. Perception generally is a very creative process. The brain often needs to default to the best guess because the incoming data is noisy, incomplete, and ambiguous. To give you at least one dramatic example of this, consider the phenomenon of change blindness. In a simple experiment, unsuspecting subjects were directed to a desk to get a form. They interact with the person at the desk and ask for the form. Supposedly ducking down to grab the form, the person at the desk quickly switches places with a different person hiding behind the desk, who emerges to hand the subject the form. This second person is in no way a lookalike of the first person. They wear a different set of clothing and might even be of another gender or race. Still, most people do not notice that a different person is handing them the form. See if you can spot the difference between the two images flashed below.

#im ../assets/figures/027/027-03.gif 50 384 Figure 27.3 - Change Blindness. Try to detect the change between each two successive images

#h3 27.3 - Lateral Geniculate Nucleus

#pg After these initial stages of detection and local processing, the ganglion cells project to an area quite a bit away in the center of the brain called the thalamus. A likely evolutionary precursor of the cortex around it, the thalamus is located anatomically and functionally at the brain's center. All our sensory systems project to it and through, except for olfaction (an even older system, in evolutionary terms). In addition, together with the hypothalamus near it, it controls many aspects of our physiology. The bundle of all ganglion cells axons that project from the retina towards the thalamus is known as the optic nerve. Since we have two eyes, two optic nerves project towards the thalamus, specifically within the thalamus called the lateral geniculate nucleus (LGN). The LGN is a bilateral structure, meaning that you have two of them, one on each side (left/right) of your thalamus. You might think that the optic nerve from the left eye project to the LGN on the left and the right optic nerve to the right LGN, but it does not. It reveals the clever wiring built into the system. Some of the ganglion cells project to the LGN on the same side, while the others project to the LGN on the other side. More specifically, ganglion cells emerging from the left side of your left eye project to the right LGN, as do ganglion cells emerging from the left side of your right eye. And all the ganglion cells originating from the right side of both your left and right eye project to the left LGN. This clever crossing over regroups the ganglion cells. Before, they were grouped based on eye of origin. Now, they are grouped based on the side of your visual field they originated from. Facing forward, our eyes pretty much see the same thing. However, stereoscopic depth comes from the slight differences in how the eyes see the same thing from a slightly different angle. To use this information means that, at some point, you must have the information of both eyes together in one place. That is one reason for this rewiring. The other point is functional relevance. Again, both eyes see the same thing. What matters is where things are in the visual field because that indicates where things are in space relative to us. It, therefore, makes far more sense to group visual information based on the proximity in space than the eye of origin. You might wonder why the information from the right visual field ends up on the left side of the brain. This crossover is not unique to vision. It happens with all our senses (except olfaction). What you hear with your left ear is initially processed solely by the right side of your brain. In later stages, information is shared between both sides of your brain through the corpus callosum, a massive bundle of nerve fibers projecting in both directions, allowing for information to move between hemispheres (your cortex divides into two hemispheres, the left, and the right).

#im ../assets/figures/027/027-04.png 50 512 Figure 27.4 - The visual pathway. Signals from both eyes are carried through the optic nerves to the lateral geniculate nucleus before projecting to the primary visual cortex

#h3 27.4 - Primary Visual Cortex – Individual neurons

#pg After some additional center-surround processing, neurons in the LGN project to the primary visual cortex, a large stretch of cortex in the back of your head. That's where the cortex takes over and starts building on the initial ON/OFF center-surround inputs it receives. The study of the functional organization of the visual system started with the groundbreaking work by Hubel and Wiesel, which would earn them a Nobel prize in physiology. By carefully recording from a cat's neurons in the feline equivalent of the primary visual cortex, Hubel and Wiesel identified several specialized cells that function as building blocks for more complex cells, able to detect and recognize more complex objects. By shining a bright light on a screen the cat was looking at while simultaneously recording the activity of a neuron, they could map out how the shape, position, orientation, and location of that light on the screen generated more or less activity. The clever bit is that they moved the light freely and were able to adjust its size and shape. Specifically, they could create a rectangular pattern of light, which they could then rotate and move as they wanted. When they did so, they found that some cells were pretty much silent most of the time, except when a) the bar was at a particular position on the screen and b) at a particular angle of rotation. These are known as orientation-selective cells.

#im ../assets/figures/027/027-05.png 50 384 Figure 27.5 - Thorsten Wiesel and David Hubel. Their work on the selectivity of neurons in the visual system of cats earned them the Nobel prize for physiology and medicine.

#br

#pg In addition, the response depended on the region across which a light/dark bar was presented. If the bar was too wide, the cell stopped responding. The cells had inherited the center-surround characteristics of their inputs. Some cells preferred a light patch flanked by darker patches, whereas others preferred the opposite. In machine vision detecting such an abrupt change in the gradient of light is an edge detector, the place where the image suddenly changes in brightness. Simple cells like a pattern of dark/bright but not the other way round. We can combine opposite dark/bright cells by projecting them to a third neuron. Combined in the right way, this third neuron becomes agnostic or invariant to the ordering and will respond to both, provided the pattern of dark/bright is at or close to the cell's preferred orientation.

#im ../assets/figures/027/027-06.png 50 384 Figure 27.6 - Neurons in the primary visual cortex selectively combine inputs originating from the retina into more complex structures like edges

#h3 27.5 - Primary Visual Cortex – Organization

#pg The relative position of the photoreceptors in the retina is preserved through the LGN and into the visual system, starting with the primary visual cortex. That means that neurons in the primary visual cortex that are nearby on the surface of the cortex receive inputs from photoreceptors nearby on the retina. This locality means that the primary visual cortex has a one-to-one mapping of its surface to the visual world and is called a retinotopic map. This map is distorted quite a bit relative to the sensor one finds in a digital camera. For one, the cortex has many folds and is not a flat surface to project on. But more importantly, there is stretching out of the map. Remember that there are far more photoreceptors in the fovea. These higher counts require more neurons and more space in the cortex.

#im ../assets/figures/027/027-07.png 50 384 Figure 27.7 - The primary visual cortex (and several cortical areas beyond it) has a retinotopic map of our visual world. What is close together in our visual world is processed closer together in cortical space. The image shows the result of a typical retinotopic mapping experiment. Human subjects were shown high-contrast stimuli differing in eccentricity and polar angle. The resulting activity associated with such stimuli shows a clear radial (eccentricity) and polar organization.

#h3 27.6 - Beyond the Primary Visual Cortex

#pg The primary visual cortex, also known as V1, primarily projects to two areas. The first is the secondary visual cortex, aptly called V2. In addition, it projects to an area called MT (short for mediotemporal visual area). V2, in turn, projects to V3, V3 into V4. All these areas are still retinotopically organized, but the receptive fields of what activates neurons are getting bigger. In other words, the neurons become less sensitive to where inputs are located within the visual field. At the same, we also see an increase in the complexity of what neurons are optimally tuned for (i.e., elicit the largest response). Whereas V1 concerns itself with lines, corners, and edges, V2 already shows sensitivity to complex patterns and textures.

#im ../assets/figures/027/027-08.png 50 512 Figure 27.8 - After the primary visual cortex (V1), neural signals based on visual inputs diverge into a host of secondary and tertiary areas, some of which have precise specializations like color, motion, and depth, to name a few. (A) the spatial organization of these areas on the cortical surface of a macaque monkey. (B) Schematic view of the connections between the areas shown in (A).

#br

#pg This increase in complexity and receptive field size continues throughout the visual system. From V4, some neurons project to area IT, where selectivity to whole objects begins to develop. In humans, area IT consists of many different subregions. One of them, the fusiform face area, specializes in detecting human faces but not much else. The route from V1 to IT is referred to as the ventral pathway and appears to specialize in object detection and recognition. The other route, which has V1 project through V2, V3, and V7 and then into the parietal cortex, is used to encode the relative spatial configuration of objects around us. Finally, the direct connection between V1 and area MT is highly specialized in detecting all kinds of motion of objects in our visual field. After some time, it becomes harder to strictly refer to the areas as purely visual, as other sensory modalities can also change the activity of the neurons, and it closely follows the blurred conceptual line from perception to cognition. Through a series of projections, starting from the retina, the selectivity or tuning of neurons becomes more abstract and conceptual.

#im ../assets/figures/027/027-09.png 50 256 Figure 27.9 - A more schematic version of the visual pathways and their predominant function. Sometimes referred to a the dorsal and ventral routes, sometimes as the 'what' and 'where', the visual system analyzing shape, form, color through one route, whereas the other route is more involved in depth, perspective, relative position and size. This is gross oversimplification, but experimental evidence does support these two modes of vision: the 'where' and the 'what'.

#h3 27.7 – Conclusion

#pg We took this brief detour into the neuroscience of biological vision as a helpful preamble for the next chapter. We will explore deep belief neural networks designed to handle image data as inputs. Half a century of research into the visual systems of various mammals revealed a highly complex yet well-organized and hierarchical system. In this system, visual inputs are processed through a sequence of cortical areas that encode our visual world from basic and highly localized elements such as lines and edges to highly invariant representations of complete objects. Neurons at the early stages of visual processing only respond to spots of light, or lines and edges of a specific orientation, at a particular position within our visual field. In contrast, neurons at the end of this processing stream are responsive only to, for example, human faces, regardless of their position, orientation or size these faces occupy in our visual field. As we will see in the next chapter, a similar hierarchical representation appears in artificial neural networks when we use a specific type of network architecture: a convolutional deep belief neural network (cDBNN). It provides strong evidence for the idea that we can best understand the behavior of biological neural networks processing visual information as convolution from a mathematical and computational point of view. In fact, when we visualize the very first layer of a CBNN, we see patterns of selectivity (the inputs that create the strongest activation in a node) that match the selectivity of neurons in the first layers of visual processing in the brain, found empirically through careful recordings of these neurons (see Fig, 27.9). And we can take this one step further still. Biological visual systems have evolved, and artificial neural networks for computer vision are designed to remove the redundancy of the highly correlated inputs, creating evermore compact, invariant, and abstract representations. This decorrelation of inputs is highly effective and mathematically similar to the methodologies we explored in the chapters on dimensionality reduction and manifold learning.

#im ../assets/figures/027/027-10.png 50 256 Figure 27.10 - Low-level receptive fields. Think of these receptive fields (whether they are derived or recorded experimentally) as the optimal image that evokes the largest response from a particular basis function (A), a node in an ANN (B), or an actual neuron in the primary visual cortex of a macaque monkey (C). (A) By mapping small patches of natural images onto a set of mathematically defined basis functions, we find we can map such patches onto patterns of light and dark with varying spatial frequencies and orientations. (B) When we visualize the properties of nodes in a convolutional neural network trained to classify natural images, we find patches that show a remarkable resemblance to the theoretical basis function needed to encode natural images. This example also includes filters sensitive to edges that vary in hue, not just brightness. The stimuli used in (A) were all black and white, so no basis functions were derived with a sensitivity to color boundaries. (C) Reconstructed receptive fields of actual V1 neurons of a macaque monkey. Mapping entire receptive fields using electrophysiological recordings (electrodes implanted directly into the visual cortex) is extremely difficult, as it requires high precision and large quantities of data to be collected from a single cell while the animal is awake, alert, and shown visual stimuli on a computer monitor. Thus, compared to theoretical and artificially created receptive fields, receptive fields reconstructed from actual recordings are quite a bit noisier. Nevertheless, these four individual neurons clearly have receptive fields capable of detecting oriented edges in the image this particular animal was shown.

#h3 27.8 - Demo notebooks

#pg Only one notebook accompanies this mostly theoretical chapter: a simulation of how we can use a technique called reverse correlation to efficiently reconstruct a neuron's receptive field.

#h3 27.9 – References

#bs

#be

#be

#bp Marr, D. (2010). Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. United Kingdom: MIT Press.

#bp Siegelbaum, S. A., Mack, S. H., Koester, J. D., Kandel, E. R. (2021). Principles of Neural Science, Sixth Edition. United States: McGraw-Hill Education.

#bp Deco, G., Rolls, E., Rolls, E. T. (2002). Computational Neuroscience of Vision. United Kingdom: OUP Oxford.

#bp Warland, D., De Ruyter Van Steveninck, R., Bialek, W., Rieke, F. (1999). Spikes: Exploring the Neural Code. United Kingdom: MIT Press.

#bp Brodmann, K. (2006). Brodmann's localization in the cerebral cortex the principles of comparative localization in the cerebral cortex based on cytoarchitectonics. Springer US.

#bp Wiesel, T. N. (2004). Brain and Visual Perception: The Story of a 25-Year Collaboration. United Kingdom: Oxford University Press, USA.

#bp Hubel, D. H., Wiesel, T. N. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. The Journal of Physiology, 160(1), 106.

#bp Hubel, D. H., Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. The Journal of Physiology, 195(1), 215-243.

#bp Dayan, P., Abbott, L. F. (2005). Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems. United Kingdom: MIT Press.

#bp Levitt, J. B., Kiper, D. C., Movshon, J. A. (1994). Receptive fields and functional architecture of macaque V2. Journal of Neurophysiology, 71(6), 2517-2542.

#bp Movshon J. A. (1974). Proceedings: Velocity preferences of simple and complex cells in the cat's striate cortex. The Journal of Physiology, 242(2), 121–123.

#bp Pasternak, T., Movshon, J. A., Merigan, W. H. (1981). Creation of direction selectivity in adult strobe-reared cats. Nature, 292(5826), 834–836.

#bp Movshon, J.A. Landy, M.S. (1991). Computational Models of Visual Processing. UK: Bradford.

#bp Brouwer, G. J., Heeger, D. J. (2013). Categorical clustering of the neural representation of color. The Journal of Neuroscience, 33(39), 15454–15465.

#bp Brouwer, G. J., Heeger, D. J. (2009). Decoding and reconstructing color from responses in human visual cortex. The Journal of Neuroscience, 29(44), 13992–14003.

#bp Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P., Movshon, J. A. (2013). A functional and perceptual signature of the second visual area in primates. Nature Neuroscience, 16(7), 974–981.

#bp Kanwisher N. (2017). The Quest for the FFA and Where It Led. The Journal of Neuroscience 37(5), 1056–1061.

#bp Kanwisher, N., Stanley, D., Harris, A. (1999). The fusiform face area is selective for faces not animals. Neuroreport, 10(1), 183–187.

#bp Simons, D. J., Levin, D. T. (1997). Change blindness. Trends in cognitive sciences, 1(7), 261-267.

#bp Quiroga, R., Reddy, L., Kreiman, G. et al. Invariant visual representation by single neurons in the human brain. Nature 435, 1102–1107 (2005).

#be