Early Visual Processing: A Computational Approach to Understanding Primary Visual Cortex

Ryan Moye, Cindy Liang, and Mark V. Albert

Visual Processing

As previously mentioned, scientific research has led to various insights in the field of artificial intelligence (AI). AI models perform a variety of tasks from classification to regression, but one area of particular interest is in vision-related tasks; one of the most common visual models, Convolutional Neural Networks (CNNs), is inspired by research in neuroscience (Lindsay, 2020). Visual processing within the brain is accomplished through a multitude of tasks coordinated mostly by the region of the brain known as the occipital lobe. By understanding and exploring early visual processing in mammals, we are able to advance computational technologies and algorithms.

The occipital lobe is responsible for processing and interpreting the visual data collected by our eyes, which, over the millennia, our brains have adapted to process in a specific and efficient manner (Olshausen & Field, 1996). We can understand sight through a series of analogies that explain how our body functions in a computer-like fashion. The first step in seeing is when our eyes detect light. The cornea works like a camera lens to refract the light to a specific location, in our case the pupil. Our pupils then adjust, much like a camera aperture, to let a certain amount of light enter our eyes. Next, our lens focuses the light into the back of our eyeball in order for the retina to sense the light and convert it to electrochemical

R. Moye (✉) · M. V. Albert
Department of Computer Science and Engineering, University of North Texas,
Denton, TX, USA
e-mail: ryanmoye@my.unt.edu; Mark.Albert@unt.edu

C. Liang
Texas Academy of Mathematics and Science, Denton, TX, USA
e-mail: cindyliang@my.unt.edu

M. V. Albert et al. (eds.), *Bridging Human Intelligence and Artificial Intelligence*, Educational Communications and Technology: Issues and Innovations, https://doi.org/10.1007/978-3-030-84729-6_12

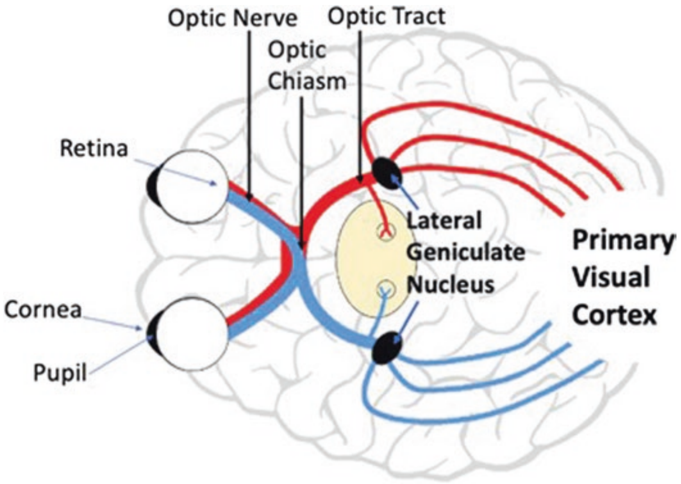

Fig. 1 The anatomy of sight
Light first enters the cornea and is then redirected from the pupil to the retina. From the retina, the light is converted into electrochemical signals which are passed down the optic nerve and split off in the optic chiasm, where the red represents the right side of vision for both eyes and the blue represents the left side. The signals travel down the optic tract and into the lateral geniculate nucleus, then finally to the primary visual cortex for processing.

impulses that are sent back to the occipital lobe via the optic nerve (Albert, 2015). These impulses can be thought of as the "data" which are now being transmitted to the "processor" or occipital lobe via the optic nerve. Once the retina converts the optical signals into electrochemical signals the optic nerve transmits this data to the optic tract, then to the lateral geniculate nucleus (LGN) which finally passes the signal into the visual cortex as shown in Fig. 1. The visual cortex is the part of the occipital lobe which processes visual data. The occipital lobe is split up into six sections, V1–V6, which deal with the various aspects of vision. V1, or the primary visual cortex, is the "sorting algorithm" of the brain. It is the first stop for all visual data to be preprocessed before being sent to V2. V1 processes sight and interprets the information it receives (Albert, 2015), then sorts the data into two categories: the what and the where. Understanding how our brain interprets and receives visual data allows us to more accurately imitate these processes in an algorithmic format.

Now that we have a basic idea of how the brain processes visual data, let us look at the same topic from a computational standpoint. Computers portray images as a matrix of connected pixels. A single pixel gives no indication as to what we are looking at. However when viewed as a whole, the pixel matrix portrays an image. The same is done for videos, where each frame of the video is an individual pixel matrix that varies slightly over each frame, thus allowing us to see motion and real-time events. Computers are also able to speed up or slow down the frames per second in a video, which produces video content in high definition or slow motion. The interesting aspect of computer vision is how it "sees" pixels. These pixel values are stored in a matrix of digits which the computer can process in order to "see" or interpret the image. From there, we can perform a variety of tasks such as image classification, generation, and upscaling. This understanding of how computers interpret visual data allows us to bridge the gap between computers and an appropriate approximation of human neuroanatomy. For now, we will take a closer look at the evolution of our brain and how we can relate AI to early visual processing.

How Neural Codes Are Represented

Gabor Filters

Simple cells in the primary visual cortex (V1) use an encoding similar to two common linear filter transformations of images merging together. While the pixel-based coding described previously is a common encoding scheme for representing the light intensity of each pixel in an image, it is too localized to represent neural codes. There is a similar issue with Fourier codes as the Fourier transform results in the localized structure of the signal being lost. However, Hubel and Wiesel (1962, 1968) demonstrated that the simple cell responses of the primary visual cortex, when presented with a stimulus, can be approximated by a 2D Gabor wavelet code, similar to the one shown in Fig. 2. This is possible because the wavelet code created by the Gabor function is both localized and periodic, similar to observed simple cell responses. It is important to note that the 2D Gabor wavelet code is an oversimplification and idealization of simple cells and does not fully describe complex cells (Albert & Field, n.d.). Knowing how simple cell responses look allows for the creation of computer-generated models, specifically AI generated receptive fields.

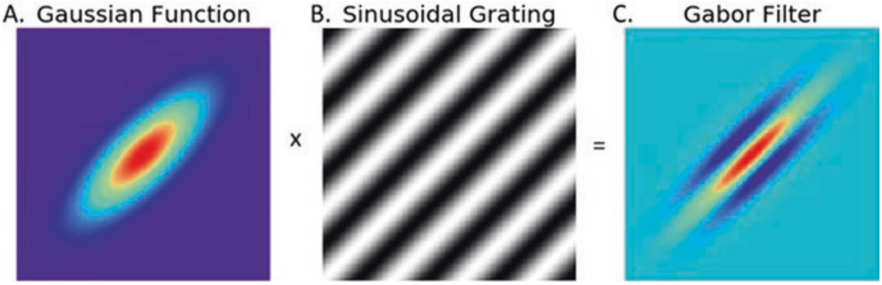

Fig. 2 Modulation of a Gaussian and sinusoidal grating to create a Gabor filter
(**a**) shows a Gaussian function from the top down, (**b**) a sinusoidal grating, and (**c**) the corresponding Gabor filter. Let G = a Gaussian function and S = a sinusoidal grating. Then, G * S = the resultant Gabor Filter. Notice that the sinusoid becomes spatially localized when modulated with the Gaussian function.

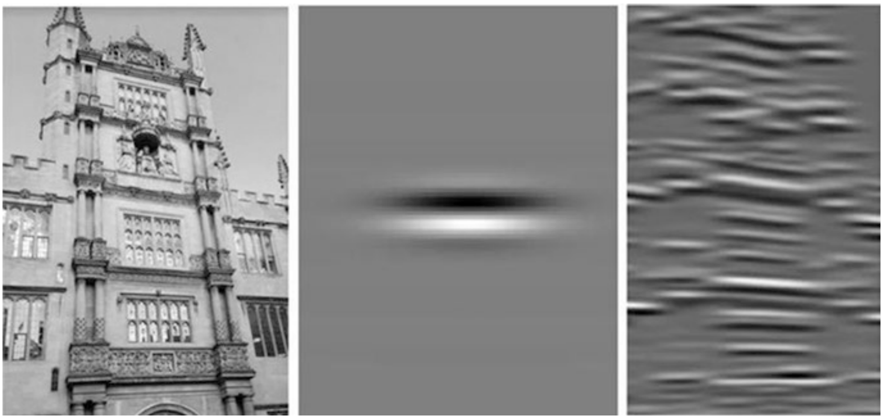

Fig. 3 Gabor filter applied to a natural image
This building (left) represents a natural image as previously discussed. When a horizontal Gabor filter (middle) is applied to the picture on the left, the resultant image (right) is produced. We can observe that (right) still maintains the basic structure of the building, but only the horizontal surfaces are prevalent in the image.

How Gabor Filters Work

Gabor filters are simply sinusoids multiplied by a Gaussian function (Prasad & Domke, 2005) as shown in Fig. 2. Gabor filters respond to oriented lines (in our case a 2D Gabor filter responds to the oriented lines in an image) and are orientation-sensitive. This means that a picture of a building (Fig. 3), when convolved with a horizontal Gabor filter, would result in an image that shows the horizontal structures of the building while ignoring the other orientations (vertical, diagonal, etc.). That is to say the Gabor filter will only give a strong response if its direction matches the direction of the lines on the building. This further illustrates why buildings with straight lines could be called "natural images," which will be discussed in further detail in the Natural vs. Non-natural Images section.

Efficient Coding Hypothesis

Upon receiving sensory stimuli (for the purpose of this chapter we will only consider visual data), neurons inside the brain communicate and relay messages to each other by action potential spikes (Olshausen & Field, 1996). Neurons require energy in order to produce these neural codes. Thus, if our brains required a neuron to represent every potential scene we might encounter, there would need to be an infinite number of neurons in our brain, which is impossible. A solution to this problem would be to have a set number of neurons that communicate details of an image by using many different complicated patterns. However, this idea is also implausible because solely relying on these patterns to interpret stimuli would still be very energy-consuming if we need a different set of neurons for each pattern that exists. Therefore, to conserve energy, a goal in early sensory processing is to reduce redundancy (Field, 1987). The Efficient Coding Hypothesis states that by utilizing sparse coding, more images can be interpreted with a limited number of neurons in an energy efficient way (Albert & Field, n.d.).

Over time, animals have evolved to use sparse coding of sensory stimuli as a way to increase metabolic efficiency. Sparse coding allows for less neuronal firing, which reduces energy consumption upon receiving external stimuli from natural sources. While the efficient coding hypothesis was derived from a neuroscience standpoint, extensive research has been done to understand this concept from a computational view. Olshausen and Field (1996) showed how simple cells can be represented with the use of sparse coding, which allowed computationally derived neural filters to be.

Although the efficient coding hypothesis models neuron behavior in a detailed way, neural codes are actually more complicated and provide more information than the individual firing rate of a neuron (Albert & Field, n.d.). There is an ongoing debate as to whether neural coding is a form of rate coding in which the average firing rate of a neuron is accounted for or temporal coding in that the relative timing of each neural spike matters. Since neurons have high frequency fluctuations with respect to their firing rates, we can surmise that the fluctuations are either noise, or potentially carry information. Rate coding suggests that this is noise, while temporal coding suggests that the relative timing of the neural spike also carries relevant information. Thus, while the efficient coding hypothesis gives us an understanding of the genetic algorithm our brains use to produce neural codes, there is still more to the process as it is more complex than a firing rate.

Natural Vs. Non-natural Images

Animals' and humans' adaptation to an efficient coding paradigm has resulted in an evolutionarily fit processing of sensory stimuli. This efficient coding has enabled our brains to process natural stimuli while firing a limited number of neurons, thus conserving energy (Barlow, 1961). For our purposes, natural stimuli are anything that humans and animals have seen for many generations such as Fig. 4 (a and b). These images can range from bodies of water and mountain ranges to manmade structures such as bridges and statues. However, these are only a few examples of natural images. To further elaborate on this, natural images are virtually any landscape or any scenery that has existed for long enough for the process of evolution to take place. We consider any structure that has straight edges (horizontal, vertical, diagonal, etc.) to be "natural" as well. Therefore skyscrapers, houses, cars, and other manmade structures also constitute "natural" imagery. Figure 3 shows an example of a manmade structure convolved with a horizontal Gabor filter. Defining what constitutes a "natural" image helps us understand what visual data applies to the efficient coding hypothesis and how we can model this concept computationally.

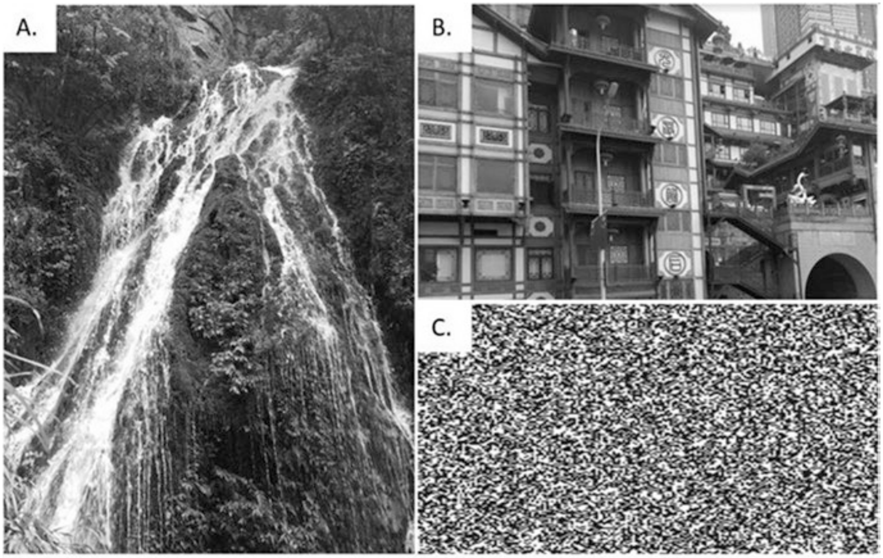

Fig. 4 Example of natural and non-natural images
(**a**) shows a natural scene in which brains have evolved to efficiently encode. (**b**) is a manmade structure, yet because it is composed of horizontal and vertical lines we still consider it to be a natural image. (**c**) shows static, of which our brains have not evolved to process efficiently.

On the other hand, non-natural images are like those in Fig. 4 (c) or psychedelic imagery. Unlike natural images, non-natural images do not have hard edges or structures similar to those seen in nature. QR codes are also good examples of non-natural stimuli, along with TV static that is pictured in Fig. 4 (c). While it may seem counterintuitive, a cloudless, blue sky is also harder for our brains to efficiently process as it is missing any of the hard edges needed for Gabor filters to apply, and thus are also considered non-natural. Essentially, our brains have developed patterns to recognize natural scenes and stimuli, and thus fire a limited, or sparse, amount of neurons in response. However, we have no evolutionary traits that help us process non-natural stimuli (Olshausen & Field, 2000). Thus, these concepts are important to note as the efficient coding hypothesis can only be applied to natural images.

ICA

As mentioned, the primary visual cortex responds to natural stimuli in a sparse manner, so to replicate this in an algorithmic manner we need a way to find the representation of the stimuli. Independent component analysis (ICA) accomplishes this goal as shown by Hyvärinen & Oja (2000). ICA is an unsupervised learning approach that, similar to principal component analysis (PCA), uses linear transformations to re-represent the data with a new set of coordinates, where each representation is as statistically

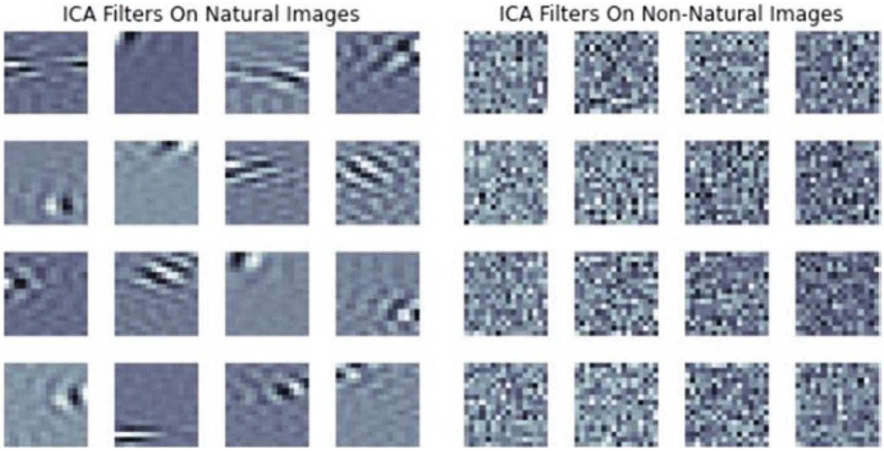

Fig. 5 ICA-encoded filters of grayscale images
Filters produced by ICA when viewing natural image patches left. Filters produced by ICA when viewing non-natural image patches right

independent as possible. ICA achieves this by searching for components that lead to the least Gaussian data distributions. As the central limit theorem suggests, the sum of a set of random variables goes toward a Gaussian distribution (Bell & Sejnowski, 1997). Thus, the original factors of the data will be less Gaussian than a random selection. The combinations of mixed components that are least Gaussian will generally draw out the original sources if it is possible to detect them through a linear combination.

To apply ICA to early visual processing, we can extract small patches as input from a natural image (small being on the order of 8x8 to 32x32 pixels) such that we are unable to tell what we are seeing anymore. Once we pass these patches through ICA, the resultant output closely resembles the neurally encoded filters of the brain. That is to say, the ICA-encoded filters are Gabor-like and represent the simple cell filters as described in the Gabor Filters section. The output from applying ICA to natural and non-natural images is shown in Fig. 5. Urs, Behpour, Georgaras, and Albert (2020) created an accessible Jupyter notebook that allows users to play around with ICA and to see how natural vs. non-natural images produce neural-like and non-neural-like filters respectively. Further research to apply these filters to current models is under way and the potential applications to computer vision looks promising.

Applications of Neural Modeling

Decoding V1 to See What Images Are Being Perceived

There are currently many studies that focus on the applications of 2D Gabor wavelet codes for visual processing. One study further elaborates on the ability of 2D Gabor wavelet codes to reconstruct images using a technique called the Bayesian decoder

(Naselaris, Prenger, Kay, Oliver, & Gallant, 2009), which utilizes FMRI signals to predict the image that one is seeing. FMRI focuses on a signal stimulus that researchers can easily decode and interpret and images are able to be reconstructed by using Gabor codes to analyze early visual areas. As mentioned earlier in this chapter, our brains have adapted to become familiar with natural images. Natural images are important to this field of study because they are relevant for subjective processes and daily perception such as imagery and dreaming (Naselaris et al., 2009). This study combines knowledge of V1 evolution as well as the structural components of Gabor wavelet codes to explain the process in which our brain executes image analysis. This ability to recreate images from neural signals utilizing Gabor wavelet codes demonstrates the current state of technology, but also shows the potential for future developments.

Furthermore, an elaborate study has been made on the vision of monkeys at Cornell University to further investigate Gabor wavelet codes and the primary visual cortex of primate brains (Kindel, Christensen, & Zylberberg, 2017). Researchers built a convolutional neural network to predict the V1 activity produced by natural images. This network uses Gabor wavelets to model simple cell response to visual information. Scientists were able to discover that the rates of firing neurons can be depicted fairly accurately by the network. This process also allowed scientists to identify image features that caused the neurons to spike. The researchers suggest that one potential application of the knowledge accumulated from this study would be that sight could potentially be restored to the blind (Kindel et al., 2017). The example given was a camera to brain translator which could be implemented to feed images into the researchers' neural networks directly from brain activity. Through the advancements in computational neuroscience and AI, modern healthcare is continuing to progress in great strides as illustrated above.

Conclusion

AI and technology have long been modeled after our understanding of neuroanatomy, and this chapter bridges some of the gaps between the two subjects. As illustrated in the previous section, the applications of 2D Gabor wavelet codes are immense, and are capable of representing simple cells found in the brain. Upon further development, these codes have the potential to give sight to the blind (Kindel et al., 2017), as well as other visual impairments people may experience. With new knowledge of 2D Gabor wavelet codes, engineers and researchers can find ways to further enhance their respective fields. Algorithms, such as ICA, have been shown to produce Gabor-like filters similar to those observed in the brain (Hyvärinen & Oja, 2000). These oriented and bandpassed neural filters help the brain interpret visual data; however, they only apply to "natural" images. As stated by the efficient coding hypothesis, one of the main goals in early visual processing is to reduce redundancy in neuronal firing, and thus process data in a metabolically efficient manner. Continued studies on these topics have allowed researchers to apply computational methods to early visual processing and in turn learn more about the field of computer vision. The ability of computers to model processes that traditionally only occurred within the primary visual cortex of human or animal brains illuminates how human intelligence can inspire artificial intelligence and culminates in new and exciting research.

References

Albert, M. V. (2015). *The Brain Geography Mini-Course: a neuroscience outreach effort*. Retrieved from https://ecommons.luc.edu/cgi/viewcontent.cgi?article=1106&context=cs_facpubs

Albert, M. V., & Field, D. J. (n.d.). Neural Representation/Coding. *Encyclopedia of Perception*. Retrieved from https://doi.org/https://doi.org/10.4135/9781412972000.n205

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. *Sensory Communication, 1*, 217–234.

Bell, A. J., & Sejnowski, T. J. (1997). The 'independent components' of natural scenes are edge filters. *Vision Research, 37*(23), 3327–3338.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America. A, Optics and Image Science, 4*(12), 2379–2394.

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*, 106–154.

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

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks: The Official Journal of the International Neural Network Society, 13*(4–5), 411–430.

Kindel, W. F., Christensen, E. D., & Zylberberg, J. (2017, June 19). *Using deep learning to reveal the neural code for images in primary visual cortex. arXiv [q-bio.NC]*. Retrieved from http://arxiv.org/abs/1706.06208

Lindsay, G. W. (2020). Convolutional neural networks as a model of the visual system: Past, present, and future. *Journal of Cognitive Neuroscience*, 1–15.

Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., & Gallant, J. L. (2009). Bayesian reconstruction of natural images from human brain activity. *Neuron*. Retrieved from https://www.sciencedirect.com/science/article/pii/S0896627309006850

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature, 381*(6583), 607–609.

Olshausen, B. A., & Field, D. J. (2000). Vision and the coding of natural images: The human brain may hold the secrets to the best image-compression algorithms. *American Scientist, 88*(3), 238–245.

Prasad, V. S. N., & Domke, J. (2005, 2005). Gabor filter visualization. *Journal of the Atmospheric Sciences, 13*.

Urs, N., Behpour, S., Georgaras, A., & Albert, M. V. (2020). Unsupervised learning in images and audio to produce neural receptive fields: A primer and accessible notebook. *Artificial Intelligence Review*.

Ryan Moye is a software engineer who obtained his master's degree in artificial intelligence from the University of North Texas. His primary research interests include automation, computational neuroscience and efficient coding techniques.

Cindy Liang is currently a student on the computer science track at the Texas Academy of Mathematics and Science. Cindy is interested in learning about the applications of computer science and AI to the biomedical field.

Mark V. Albert professional goal in life is to leverage machine learning to automate the collection and inference of clinically useful health information to improve clinical research. His projects in wearable sensor analytics have improved the measurement of health outcomes for individuals with Parkinson's disease, stroke, and transfemoral amputations with a variety of additional populations and contexts including children with cerebral palsy as well as healthy toddler activity tracking. Current projects include video-based activity tracking and mobile robotic platforms, all in an effort to improve measures of clinical outcomes to justify therapeutic interventions.