Modeling The Secondary Visual Cortex

Project report by Hannah Machado

Direct Supervisor Dr. Odelia Schwartz

University of Miami

May, 3rd, 2019

Content

Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. . . . . . . . . . . . . . . . . . . . . .3

The Visual Hierarchy . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. . . . . . . . . . . . . . . . . . . . . . . 4

1. Primary visual cortex . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. . . . . . . . . . . . . . . . . . . . . .5
2. Secondary visual cortex . . . . . . . . . . . . . . . . . . . . . . . . .. . . . . . . . . . .. . . . . . . . . . . . . . . . . . . . . .5

Receptive fields . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. . . . . . . . . . . . . . . . . . . . . . 6

A Brief Overview of Deep Learning. . . . . . . . . . . . . . . . . . . . . . . . . .. . . . . . . . . . . . . . . . . . . . . . .7

1. Unsupervised Approach. . . . . . . . . . . . .. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .7
2. Supervised Approach. . . . . . . . . . . . . . . . . . . . . . . . .. . . . . . . . . . .. . . . . . . . . . . . . . . . . .. . . . . .9
3. Comparison. . . . . . . . . . . . . . . . .. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .10

Conclusion . . . . . . . . . . . . .. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. . . . . . . . . . . . . . . . . . . 10

Future Study. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. . . . . . . . . . . . . . . . . . . 11

References . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. . . . . . . . . . . . . . . . . . . 12

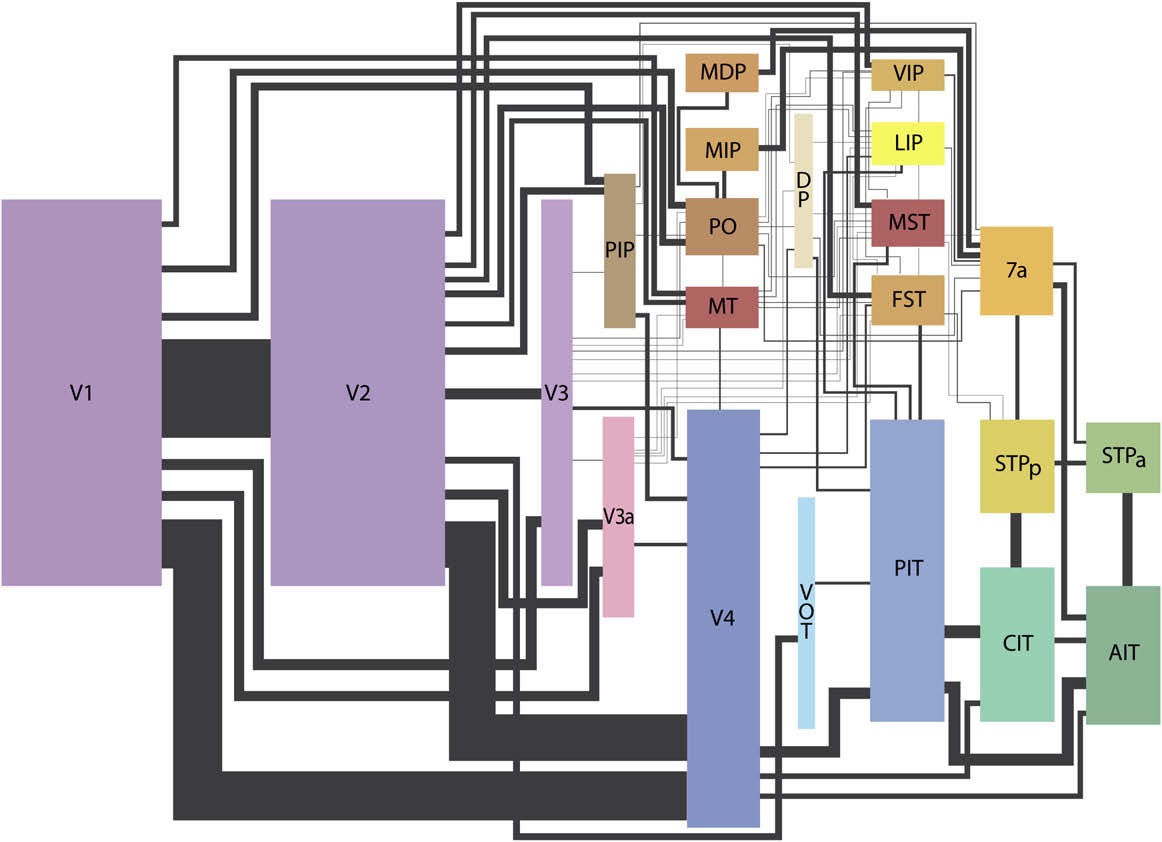
**Introduction**

The ability to see and perceive stems from a complex decoding system performed by the human brain. Once presented with stimuli, information travels to the visual cortex through a network of neurons, and it is interpreted by electrical synapses as a way to make sense of the world around us. This intricate network is divided in a hierarchy of cortical levels, meaning that different level areas process different complexities of space that together decode the specific objects and its surroundings. The Primary Visual Cortex, also called V1, is considered to be the first cortical area of visual processing. It is located within the occipital lobe, and its neurons are driven by orientation. V1 neurons essentially deconstruct a scene according to the spatial location of oriented components (continuing puzzle). The primary visual cortex has been extensively studied and its visual processing role is well understood when compared to higher visual cortical areas. Recent advances in deep learning and artificial neural networks have triggered an era where world tasks engage with computing power, significantly improving computational models of the brain.

Despite the fact that V1 has been extensively studied, little is known about role of the secondary visual cortex, V2, in visual processing. Neuron specificity in V2 is not clearly defined, but the area is nonetheless essential to vision. Merigan et al. (1993), evidenced that an injury in that area significantly affected the ability of their trial monkeys to perform complex spatial tasks. Even though the role of V2 is visual processing is not yet clearly understood, it is without a doubt important. The aim of this project is to get a better understanding of the secondary visual cortex, biologically speaking, and study existing computational models and their accuracy with respect to similarites to the brain.

Throught this paper, it is important to understand the concept of deep learning, which refers to a machine learning subset within the context of artificial intelligence that constitutes a network with the ability to recognize and learn unsupervised or supervised standpoints and from labeled or unlabeled (unstructured) data representation. Deep learning represents an important phase in the move towards the development of autonomous computer systems. At the code of the deep learning networks lies the concept of artificial neural networks. Unsupervised and supervised approaches have been modeled as attempts to obtain a model of V2, and this paper explores the concept of the deep neural network in regard to conceptual capabilities within the secondary visual cortex.

**The Visual Hierarchy**



(Image from Wallisch and Movshon 2008; After Felleman and Van Essen, 1991)

The visual system is an intricate network of neurons that encode and decode stimuli into the world we perceive. It is noticeable, from the graph above, that the system increases in complexity as the processing levels escalate. Specifically, the rectangles in this picture represent each layer within the visual cortex, and the shape size is in accordance with the total surface area designated to that specific cortex. The primary visual cortex is the lowest layer of visual processing (V1), followed by the secondary visual cortex (V2). Together, they take up a lot more grey matter, which consists of neurons, an white matter, which are connections, than the rest of the visual cortex. This highlights their significance in visual processing. Currently, a lot more is known about processing in V1 than V2. This brief overview leads the way into exploring current techniques that aim to model V2 and how well their performance can be ranked when compared to the brain.

1. **Primary visual cortex**

Located in the occipital lobe, the primary visual cortex (V1) has been extensively studied, and is the first cortical level of visual processing. It receives sensory input from the thalamus and is highly responsible for processing simple visual features, such as edge orientation within a small spatial region. Higher areas process progressively more complex visual features in larger portions of the visual space. Computationally speaking, V1 has a well-defined map of spatial processing, and much research has been done on the area to evidence how neurons process information. On the above graph, each area is represented according to total surface area occupied in the brain.

1. **Secondary Visual Cortex**

Following V1, the next level of visual cortical processing is the secondary visual cortex. Unlike the primary cortex, V2 is poorly understood. Specific excitatory stimulus for neurons in this area is yet to be discovered, but there is some experimental data on selectivity for angle, texture, and figure-ground data. V2 is the second layer of visual processing and also the second major area of the visual cortex. It receives direct information from V1 and sends it further into higher cortical areas. Understanding what happens at the secondary level can bring insights into even higher levels, and perhaps a deeper understanding on how information is visually interpreted.

**Receptive fields**

Receptive fields are regions on the visual field upon which a stimulus can modify the firing of the neuron and obtain a response. Computationally speaking, a receptive field is a filter that captures the stimulus specifications and generates responses accordingly. They are often assumed linear. Examples include the retina, skin, tongue, and other body parts. The retinal ganglion cells, or eye cells, offer the simplest receptive field display. They mainly consist of an on-center, off-surrounding, or off-center on-surrounding structure. What this means is that when presented with a stimulus in the center, an on-center cell will fire, while an off-center cell will be inhibited. V1 and V2 receptive fields, however, start to become more complex. Hubel and Wiesel (1959) classified three types of cells within the striate cortex. The first type is a simple cell, which has elongated or rectangular receptive fields with activated or deactivated regions. They are selective for certain orientations and phases and also have center and surround regions in order to activate the neuron.

The second type of cell, a complex cell, has a larger receptive field, that is phase invariant, and they may respond to a specific stimulus anywhere within the field. Thus, they have no obvious center and surround and the image may need to be correctly oriented so as to excite the cell. Finally, the third type is hypercomplex cells. This third type of cell is not intricately discussed in this paper, given that it goes beyond the scope of the research. Briefly, besides having the same specifications as the previous cells, hypercomplex cells are also selective for a certain length of contour, meaning that they are only activated if the image is of a certain length.

**A brief overview of deep learning**

Deep learning refers to an increasingly popular subset of the concept of machine learning. Deep learning models are constructed from neural networks that allow for the taking and processing of inputs through artificial neural network engines by means of weights adjusted appropriately during training (Williford & Von Der Heydt, 2016). Organizations that employ such systems must have the technical expertise to train them in accordance to the nature of data to be analyzed or simulated.

Deep neural networks therefore, are an impotent source of insights to be used by companies as a source of competitive advantage. There is a striking similarity and correspondence between deep neural network and the human brain. Laskar et al. (2018) explained how deep neural network works to model secondary visual cortex (V2). They noted that deep convolutional neural networks (CNNs) trained on scenes and objects have a fascinating ability to predict certain response properties of visual cortical neurons. The computations and factors that results to such ability and the role of various intermediate processing stages to explain changes that result across regions of the cortical hierarchy are not clearly understood. Deep convolutional neural networks (CNNs) are effective for capturing changes developing across early areas of cortex. The modelling of the deep convolutional neural networks (CNNs) using aspects of secondary visual cortex, V2, makes it easier to understand the computations that result to hierarchical processing in the human and machine brain. It is possible to carry out various modeling exercises to explain the deep neural networks and their close relationship in functionality to the human cortical neurons for vision.

1. **Unsupervised approach**

The work of Hosoya and Hyvarinen (2015) presented a Hierarchical Statistical Model of Natural Images to explain the Tuning Properties in V2. Their work employed an unsupervised training method to conduct the experiment. They noted that previous theoretical and experimental studies have succeeded in demonstrating a close relationship between neural representation in V1 and natural image statistics. The receptive field properties in V1 were highlighted and classified as similar to both, simple and complex cells, and sparse coding of natural images. From there, efforts to interpret V2 started to form with scientific reasoning. In order to address the behavior of V2, an unsupervised training method was used. A sparse coding model was elaborated to take input of the output data of a fixed model with V1 properties. The data was then fed with a large number of natural image patches that acted as its input. Upon training, the model showed response compatible both qualitatively and quantitative with three main neurophysiological outcomes on macaque V2. The similarities included both homogenous and heterogenous integration of the data’s local orientation. The data indicated a wide range of angle selectivity biased in regards to their sensitivities. Elements of length and width suppression were also observed. The use of unsupervised approach indicated results that were consistent with the notion that V2 employs a sparse code of natural images.

Sparse coding: One-layer vs multilayer

The work of Hosoya and Hyvarinen (2015) employed multilayer sparse coding technique. Sparse representation is a crucial concept in image and signal processing that has therefore led to remarkable developments in regards to applications. In its most basic interpretation, the sparse representation model assumes that a natural image or signal is represented by a vector x that may be conceptualized as a linear combination of a few atoms and columns from a larger matrix referred to as a dictionary (Aberdam et al., 2018). A one-layer sparse coding is simpler than a multi-layer one which takes more time and consumes more resources during analysis.

1. **Supervised approach**

Unlike unsupervised learning, supervision requires input patterns and associated desired outputs included in the data. Realistically speaking, a supervised approach to learn natural phenomena is often unavailable. Most natural data is unlabeled. The work of Kriegeskorte (2015) explored deep neural networks as a new methodology for modelling biological vision and brain information processing. The research provided deeper insights by exploring all the aspects of DNPs. Kriegeskorte (2015) further indicated how supervised approach could be used to reach specific goals in deep neural analysis. Using a supervised approach, it is possible to carry out various simulations on the image to achieve high visual outcomes that previously had not been thought of or modelled.

Supervised neural network learning algorithm can account for error derivatives with respect to the specific weights. While such can be achieved, recent advances in the domain of neural network modelling has allowed for major strides in computer vision. The advances in technology have made it possible for computer systems to attain human-level visual recognition abilities through artificial systems. Such artificial neural networks mimic the brain and their methods of computation can be effectively implemented in normal biological neurons. The primate visual hierarchy can be mimicked through convolutional feedforward networks. However, in order to achieve specific engineering goals, more detailed and accurate systems and methodologies are desired (Freeman et al., 2013). The achievement of biologically faithful feedforward and recurrent computational models can only be achieved using supervised approaches to sparse modelling. The approach has objectivity and is task based. The implication is that supervised systems can be optimized to the end to achieve new heights in system and process cognition, reaction and control.

1. **Comparison**

A comparison of the supervised and unsupervised systems reveals different applicability. It is generally thought that the unsupervised approach shapes early areas well, whereas higher areas are better modeled by supervised networks. While the unsupervised system is good for making discovery in data sources, the supervised method makes it possible to attain specific goals and objectives while encouraging optimization in machine vision. Therefore, the two approaches have different aims and objectives. In a world that is increasingly more competitive and efficient, the supervised systems are preferrable. However, within the domain of scientific discovery, the non-objective but discovery-oriented unsupervised sparse coding methodology is preferred.

A network may, however, do an excellent job in object recognition and be able to recognize images, but that does not mean it works like the brain. Many advancements have been recently made in object recognition and computational models, but there is no measurement or comparison to what extent they are similar to what the brain actually. For a general exploration of the powers of deep learning algorithms, unsupervised sparse coding may be used. Such a process may lead to massive quantities of resources and time dedication to a process without a definite objective. A more befitting process would be pegged on efficiency in resources and time management. Such a high efficiency level can only be achieved through supervised systems.

**Conclusion**

Understanding the visual cortex and how the eyes and brain is an astoundingly difficult task. Modeling that same approach is evidently an intricate task. A supervised approach may be better than an unsupervised one to obtain concrete outcomes due to its ability to optimize processes and objectives. Additionally, the ability to reduce error rates in the process of data analysis and recognition makes it a better choice in a world of efficiency at the core of every process.

The design of better facial recognition systems can therefore be achieved using such a supervised system. Perhaps a better approach would be to use a hybrid system employing both supervised and unsupervised modules for better outcomes. The danger of using hybrid system includes a general loss of objectivity. Organizations and scientists should therefore dedicate more resources towards the design of more efficient systems based on supervised sparse coding algorithms that are known to produce better results across the spetrum.

**Future Study**

Upon continuity, a future study design to gain more insights on until what extent the present neural networks represent the brain is to run stimuli with Hosoya model and a supervised learning network and then compare the results to what the brain actually responds to. The brain reponses would be obtained by showing images to the brain and data to the model, and from there it is possible to analyse how well the machine can predict responses compared to the brain. Ideally, brain responses would be obtained from fMRI and electrophysiology. In addition, the stimuli should be comprised of more than just angle stimuli, given that it is not known to what necessarily V2 neurons respond to. Texture and figure-ground data, along with angle stimuli, would be a wise choice, given that data has been found for possible activation of V2 nurons with these stimuli. This comprehensive study can potentially compare and rank the best approaches to model the brain and therefore provide a better direction to future studies.

**References**

Aberdam, A., Sulam, J., & Elad, M. (2019). Multi-Layer Sparse Coding: The Holistic Way. *SIAM Journal on Mathematics of Data Science*, *1*(1), 46-77.

Boynton, G. (2004). Visual Cortex: The Continuing Puzzle of Area V2. *Current Biology, 14*

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.

G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. Improving

neural networks by preventing coadaptation of feature detectors. arXiv:1207.0580, 2012.

Hosoya, H., & Hyvärinen, A. (2015). A hierarchical statistical model of natural images explains tuning properties in V2. *Journal of Neuroscience*, *35*(29), 10412-10428

Ito, M., & Komatsu, H. (2004). Representation of Angles Embedded within Contour Stimuli in

Area V2 of Macaque Monkey. *Journal of Neurosci*. 13, 3313-3324

Kriegeskorte, N. (2015). Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annual review of vision science*, *1*, 417-446.

Merigan, W.H., Nealey, T.A., and Maunsell, J.H. (1993). Visual effects of lesions of cortical

area V2 in macaques. *Journal of Neurosci*. 13, 3180-3191

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

Wallisch, P., & Movshon, J. (2008, October 19). Structure and Function Come Unglued in the Visual Cortex. *Neuron 60*

Williford, J. R., & Von Der Heydt, R. (2016). Figure-ground organization in visual cortex for natural scenes. *eNeuro*, *3*(6)

Yamins, D., & DiCarlo, J. (2016, February 23). Using goal-driven deep learning models to understand sensory cortex. *Nature America, 13(3)*