Brown University

Honors Thesis

Comparative Analysis of CNNs and DoG Filters to Model Mouse Visual Cortex

Michele Winter

Advisor: Dr. Thomas Serre
Reader: Dr. Lucien Bienenstock
Reader: Dr. Derek Stein

In Fulfillment of the
Sc.B in Computational Neuroscience

April 15, 2018

Acknowledgements

Special thanks to my advisor, Dr.Thomas Serre for his guidance and wisdom, and my mentor Dr.Drew Linsley for his support in implementing the ideas in this thesis. Many thanks to my readers, Dr.Lucien Bienenstock and Dr.Derek Stein, and Dean Margaret Chang for her support in completing my Independent Concentration. Finally, many thanks to Rohit Saha, Vijay Veerabadran, Lakshmi Narasimhan Govindarajan, Andreas Karagounis, Tarun Sharma and Benjamin Murphy for their helpful comments.

Contents

List of Figures

1 Introduction

1.1 Motivations for Study in Vision

We rely on our visual system for everyday behavior, navigating through the world and interacting appropriately at every stop along the way. Progress in the study of machine vision has resulted from a boom in industry automation, with businesses aiming to produce self driving cars and other intelligent pattern recognition and classification systems for use in factories for quality assurance. Based on foundations in biological vision, these computational models have reached near human accuracy in image recognition problems. However, many of these models lack the complexity that humans have to be able to learn meaningful representations of stimuli with a fairly small number of training examples.

For example, a popular state-of-the-art model in computer vision published in 2012 - the AlexNet model - with 8 nonlinear layers, requires millions of training examples to be able to accurately classify images from the ImageNet dataset [9]. Yet, this barely measures up to even a human toddler, who is able to generalize to various visual stimuli even though their visual history does not extend to the hundreds of thousands of labeled images needed to train today's best computer vision programs.

Yet, even with an extremely limited understanding of mammalian visual processing, these models are used ubiquitously in industry as exemplified by the automated categorization of numbers in the U.S. postal system as well as facial recognition in modern cameras. My goal in studying mammalian vision is to seek a more biologically dedicated understanding of visual processing that will allow for further insight into the complex processing underlying our ability to understand the environment in a way that is robust to changes that would trick leading vision algorithms.

1.2 Evolution of Vision

Around one third of cortex is either directly or indirectly involved in visual processing. Research in evolutionary biology has revealed the development of vertebrate vision as an elaborate and complex process necessary for survival. Beginning with photosensitivity present in early bacteria, modern mammalian vision has evolved a sensory percept that supports survival. The evolution of visual organelle from light sensitive spots to more complex structures with multiple components has been observed in fossils of cyanobacteria, a bacteria that is believed to have had the first form of vision [7].

Beyond light sensitivity and the ability to discriminate between light intensity, the development of color vision in vertebrates allows for the ability to distinguish between various wavelengths of light. Most mammals are dichromats, with the exception of some primates (including humans), cetaceans, and seals. Dichromacy is determined by the ability to only distinguish between two wavelengths of light due to only having two sets of light sensitive cones. Some primates, including humans, are trichromatic and have three cones that allow for the ability to distinguish between long, medium, and short wavelengths with a visible range between 400 - 700 nm. In contrast, cetaceans and seals are monochromatic and therefore can only distinguish between light intensities.

Vertebrate mammals have pairs of eyes in which light passes through a lens that refracts the waveform to an area in the back of the eye that has a dense collection of light sensitive neurons. As the light reaches the retina, the wavelengths stimulate rod and cone cells that send action potentials through retinal ganglion cells to the brain. This bundle of nerves passes through the optic disk, also known as the blind spot of the eye, and leads toward the lateral geniculate nucleus (LGN) before high level processing in visual cortex. As neural activity passes through to primary visual area, the input is processed and projected to areas of higher visual processing. This hierarchical structure provides the basis for increased complexity in spatial and temporal representations of visual input as information from downstream neurons is categorized and understood. To better understand this phenomenon, scientists study a variety of animal models and apply their findings to homologues in humans.

1.3 Visual Projections: Ventral and Dorsal Pathways

Previous literature has supported the theory of a bidirectional processing pathway in mammalian visual cortex. Numerous primate studies have suggested the existence of a dorsal pathway that allows for temporal processing for motion recognition as well as a ventral pathway that focuses on spatial processing for functions including object recognition. Our ability to deduce motion quickly and accurately has been vital for survival [8].

In non-human primates, studies of neuronal connectivity have suggested for the existence of organized projections from primary visual area to higher visual areas. Specifically, there are relationships between laminar specificity in primary visual area in macaque monkeys (area V1) and neuronal projections to either the dorsal or ventral streams through different cell types [14]. These connections were discovered by staining different types of neurons in primary visual area and observing their expression in secondary visual area. Connectivity results such as this have been vital for understanding the different pathways by which information travels through visual cortex and at which point separation in selectivity occurs.

Similarly, mice demonstrate a hierarchical connectivity that allows for increased complexity in visual representations over time. Scientists used calcium imaging to observe increased complexity in neuronal response of subnetworks in the dorsal and ventral areas of developing mouse brains [15]. Over time, neurons in these areas showed increased selectivity to temporal and spatial features respectively, suggesting a similar progression to what has been found in mammalian brains. Additionally, there is evidence for robust object recognition in the ventral stream homologue of mice [16]. These discoveries provide support for mice as reliable models for the study of aspects of human vision including spatial and temporal processing.

1.4 Current State of Visual Models of Vertebrates: Primate and Rodent

Due to the limitations of non-invasive imaging techniques, human models used in research do not allow for high definition understanding of biological activity at the cellular level. For example, it is difficult to take intracellular recordings of neurons in the motor cortex of a human participant while monitoring their responses to visual stimuli, unless the human is already participating in invasive experimental procedures such as deep brain stimulation.

Functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), and electroencephalography (EEG) are commonly used when studying human visual processing because they are non-invasive. While these imaging methodologies are extremely powerful and allow for monitoring of whole brain activity, they are limited to external recordings of the brain.

Therefore, neural activity from structures deep within the brain is usually washed out by noise and is difficult to analyze. Additionally, because these methods consider the whole brain at a time, it is impossible to discriminate activity on a cell-by-cell basis.

More invasive techniques, such as fluorescent dyes to trace neuronal projections or radioactive tracers for mapping neuronal activity cannot be used in humans.

For a deeper understanding of the cellular organization of visual cortex and the connectivity of neurons through different layers of the brain, these methodologies are extremely important. As a result, scientists have turned to other animal models to study visual processing with the intent to apply discoveries to further understand human visual processing.

Macaque monkey and rodent models are alternative models that have been used throughout scientific experiments to study visual processing at both the behavioral and cellular level. Primate visual cortex has been studied extensively, with a large corpus of literature that covers much of the functionality, connectivity, and computations of neurons in visual cortex. Though the mapping from human to non-human primate visual cortex is not one-to-one, it is the most similar with which to compare, and therefore has been a model example for scientific research.

Following the development of techniques specified for rodent neurophysiology, rodents have become a popular model for use in vision research. It is key to note that these mammals do not have a fovea in their retina, meaning that they do not have a point of focus in which they see color. Rather, they possess less visual acuity than primates and distinguish mostly light intensity because of their evolutionary drive to see in the dark [13]. Despite these fundamental differences, the development of genetic markers for tracking connectivity between neurons in mouse cortex has allowed for the discovery of visual pathways that mirror human object recognition; an ability once thought nonexistent in rodents [16]. With the discovery of a more complex visual system than previously assumed, scientists have learned more about the rodent visual system and its homologue in primate visual cortex. Similarities have been discovered between rodents and primates regarding the processing hierarchy of visual information.

Due to the promising nature of this research, fueled by additional discoveries of object processing areas that were once thought nonexistent, research in rodent visual cortex has increased significantly and there is potential for new discoveries to contribute to novel theories in visual processing.

1.5 Allen Brain Observatory Calcium Imaging Data

A leading biological research institute, the Allen Institute for Brain Science has published extensive sets of mouse, human, and primate neural data. Specifically, their data on transgenic mouse lines allows for the tracking of gene expression throughout individual cells as well as across cortex. Neurons in these mice were genetically engineered to bear transgenic lines expressing genetically encoded fluorescent calcium sensors [2]. With this modification, calcium influx in cortex is accompanied by observable transient increases in fluorescence. Using two-photon imaging, the release of calcium which is directly related to neural activity, can be recorded across time in response to a stimulus. With six different gene lines expressed across neurons spanning various layers of the mouse brain, this data allows for the mapping of connectivity between cells as well as differentiation between different neurons.

This data was collected *in vivo* as the mice were presented with various stimuli, including drifting gratings, static gratings, locally sparse noise, natural scenes, and natural movies. The mice were also presented with locally sparse noise stimuli, which allowed for the characterization of their receptive field properties. With the exhaustive stimulus list that was used, the recorded response profiles of these neurons offers the potential for deep insight into the processing hierarchy within mouse visual cortex.

1.6 Project Details

There has been little to no prior research on encoding the activity of neurons at a scale as large as this dataset. By taking advantage of modern advances in algorithmic analysis and computational power, this project attempts to predict neuronal responses. For this project, I attempt to take advantage of the expression activities in the Allen Brain Observatory Calcium Imaging Dataset to localize the responses of individual neurons with laminar and spatial specificity.

To fit the neurons in this dataset, the project uses the approach of neural encoding as opposed to decoding. Decoding, which is currently a popular approach, tries to determine the stimulus from neural activity. By working backwards, the model discriminates between different neural responses to find patterns that are specific to different input stimuli, thereby working in a dimension on the order of the number of stimuli.

Neural encoding takes the opposite approach. This project focuses on models that take in the original natural movies as input and attempt to predict neural responses as output. In this way, the model tries to better fit the predicted neural response, thereby allowing for a dedicated model that should in theory have properties inherent to the neurons it is fitting. Ideally, this would allow for the model to generalize over novel stimuli by producing expected neural activity that would match actual recorded responses.

This learning method is termed representational learning, and has not yet been applied to a set as exhaustive as the Allen Brain Calcium Imaging data [11].

This project will initially compare the fits of different spatiotemporal models to the data, including full 3-dimensional convolutional models as well as frame by frame 2-dimensional models across cells in mouse visual cortex. The responses of all of the neurons in mouse primary visual area will be jointly fit with a single model which will possess mechanisms that will allow it to automatically determine how to separate neuronal responses. This would take advantage of a model fitting procedure that characterizes all neurons across a visual area and becomes tuned to their spatiotemporal separability. In doing so, the differences in the spatial and temporal properties of the neuron receptive fields could be teased apart.

1.6.1 Model Fitting: Difference of Gaussians

Additionally, because of the complex nature of the neuronal profiles in this dataset, a difference of Gaussians (DoG) model was also fit to the data. The DoG model has previously been shown to be representative of neural activity in the Lateral Geniculate Nucleus (LGN) and may better predict the response profiles of neurons in this dataset.

As such, three different variants of the DoG model will be implemented and tested with this data: a Difference of Gaussians model implemented by Antolik et. al [3] with four parameters, a convolutional neural network initialized with DoG kernels, and a convolutional neural network of DoG kernels with learnable parameters. The DoG models are parameterized by center and surround receptive field sizes, and center and surround receptive fields weights.

2 Models

2.1 Difference of Gaussians

The Difference of Gaussians (DoG) model is a popular representation for cells in the Lateral Geniculate Nucleus of the thalamus [3]. It is defined as the difference of two Gaussian distributions, and consequently is an accurate model of the center and surround receptive field components of early visual neurons. It is parameterized by center and surround receptive field weights.

2.2 Convolutional Difference of Gaussians

Allowing for the convolution of DoG filters across the image as well as allowing for the filters to be learned should result in better fits than the aforementioned DoG model. The location invariance afforded by convolution is highly beneficial when processing natural scene stimuli, thus benefiting a convolutional approach. The Convolutional DoG (DoG_conv) model utilizes 64 initialized DoG filters that can then be learned during training. The initial filters can be seen in Figure 1.

2.3 Parameterized Convolutional Difference of Gaussians

Finally, in between the DoG and the convolutional DoG models, a parameterized convolutional DoG (DoG_param_conv) model was tested. While initialized similarly to the convolutional DoG model, the only learnable parameters are two tensors of σ

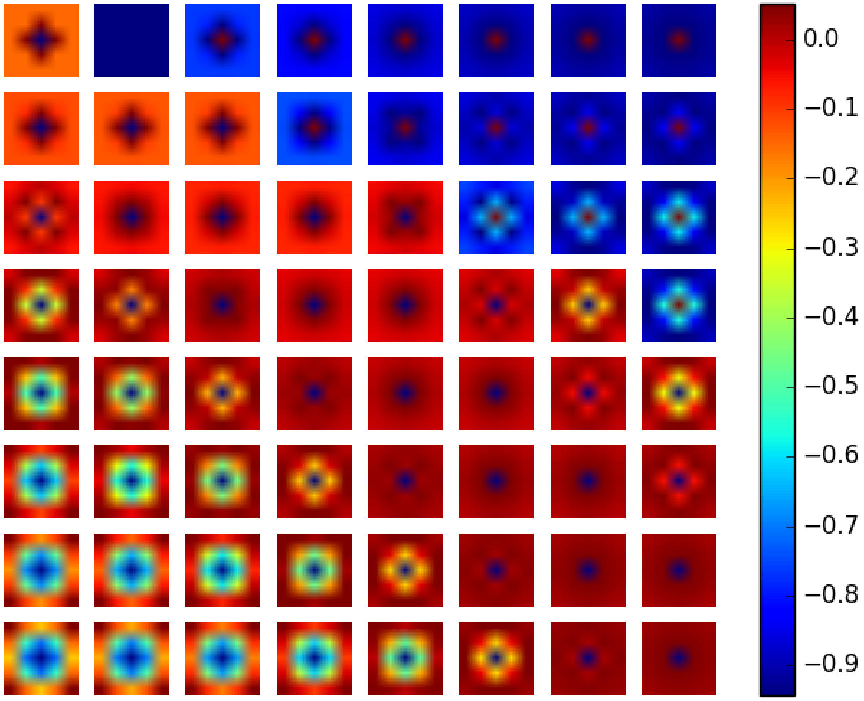


Figure 1: Visualization of the initial filters for the convolutional DoG model.

values determining the standard deviations of the Gaussians that are being subtracted from one another. As these parameters are updated during training, a new filter bank of 64 kernels is created.

3 Experiments

3.1 Fitting Models to $\frac{\partial F}{F}$

Due to the inherent noisiness in fluorescence traces of cells, a common practice is to work with changes in calcium fluorescence over time, $\frac{\partial F}{F}$. Because the baseline for raw fluorescence of cells can drift during the time of recording, the change in fluorescence over a time window better represents a normalized change in calcium fluorescence, and therefore cell activity over time.

3.2 Fitting Models to Fluorescence Trace

Rather than trying to fit the changes in fluorescence trace over time, I next tried to work with the raw fluorescence trace. Raw fluorescence trace is a secondary measure for neuronal activity because it quantifies the amount of calcium present in the cortex. This calcium fluorescence is propagated through GCaMP6, a fluorescent protein that is expressed in targeted neurons. When these neurons fire action potentials, their fluorescent proteins can be observed with two-photon imaging. The intensity of the fluorescence determines the level of activity in the neuron at a certain time. Inherently, fluorescence traces are noisy - they can have changing baselines because of calcium saturation or differences in cell morphology and enzymes that take up calcium at different rates. Action potentials can also be overpowered by a slowly decaying calcium fluorescence signal from a previous powerful burst of activity. To get a better sense of the activity of the neuron over time, I attempted to use a method to deconvolve the fluorescence trace into its underlying spike trace over time, using an algorithm known as OASIS [6].

3.2.1 Deconvolving the Fluorescence Trace with OASIS

OASIS stands for an Online Active Set method to Infer Spikes, and utilizes an autoregressive function to predict spikes from calcium fluorescence traces. At every time step, it looks at the previous time step to make a prediction about spike probability. This results in a spike probability trace.

After the spike probabilities are deduced, spike rate over time can be derived. The fluorescence responses are imaged at 30 fps, and so by averaging over five frames, one can attribute a firing rate at each frame. This is done by using the spike probabilities to count the number of spikes in the window and dividing to get a spikes per second ratio. By tallying up the number of cells with similar maximum spike rates throughout the stimulus presentation, one can get a good idea of the level of activity of the cells.

4 Results

4.1 OASIS Deconvolution

Due to the variability within the data, deconvolving the fluorescence traces of the neuronal responses was not helpful in better predicting the data. In Figures 2 and 3, it can be see that due to inherent variability in cell responses to the stimulus as well as blatant lack of response of cells to entire stimulus sets, spiking recovery from the calcium imaging did not enhance the quality of the data to be fit. As a result, rather than working with deconvolved fluorescence traces, the decision was made to fit $\frac{\partial F}{F}$ signals instead. Figures 2 and 3 are of cells responding to natural movie stimuli.

4.2 Initial Model Results

Initial attempts at fitting the data to complex spatiotemporal models did not result in noteworthy results. An explanation for this could have been that these models were too complex to learn from the small number of training examples and the inherent noisiness of the neuronal profiles in this dataset. This was analyzed further by calculating the reliability of each cell's response over many repetitions of a stimulus set.

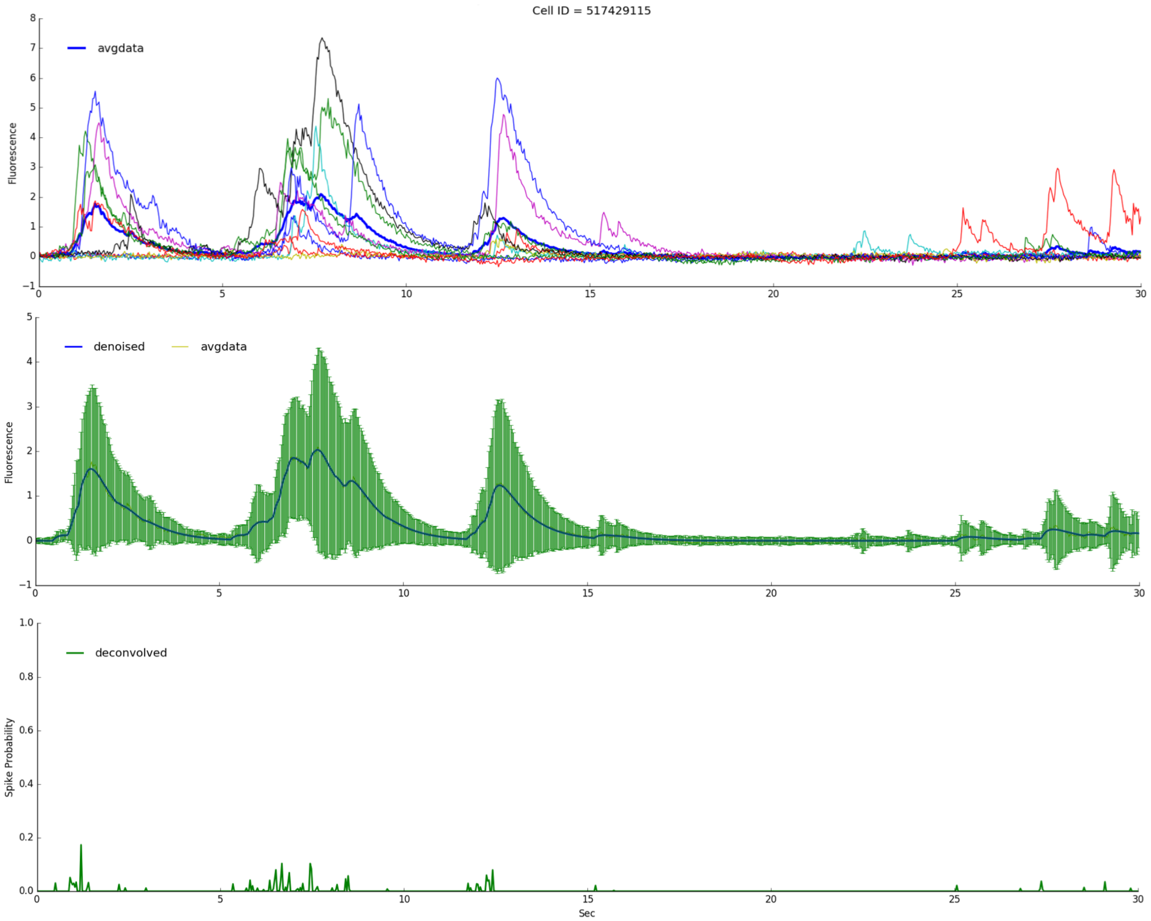


Figure 2: Example of cell with reliable responses over repetitions and a discernible signal. The top plot shows a raw fluorescence trace for multiple repetitions of the same stimulus over 30 seconds. The middle plot shows the denoised trace over the repetitions with standard deviation bars in green. The bottom plot shows the outcome of the OASIS deconvolution: spike probabilities at specific time steps.

4.3 Individual Cell Reliability

Inter-repetition Pearson correlation was used as a measure of neuronal response reliability within different cortical areas.

As can be seen in Figure 4, the distribution of cell reliability is very broad - the majority of their correlations throughout all of the datasets, when averaged over two halves of repetition sets, is less than 0.6, a widely accepted baseline for a reliable test. A large swath of scientific research has been dedicated to neurons with reliable signals and therefore there is considerable bias in the field to cells that are highly responsive. Therefore, it is understandable that large-scale recordings such as the one used for this dataset would collect information from cells that have minimal signal.

Figure 4 shows different datasets that encompass homologues to human visual areas in mice: Primary Visual Area (VISp), Secondary/Lateral Visual Area (VISl), Antero-medial Visual Area (VISam), and Posterior-medial Visual Area (VISpm). VISam and VISpm are homologues to human visual MT area, with receptive fields that are selective for both spatial and temporal components of the visual field [5].

Figure 5 shows the resulting correlation of fits of the three different DoG models to a collection of all neurons from the datasets shown in Figure 4 that demonstrated a reliability coefficient $r_{SB} \geq 0.8$. The distributions of the three histograms point to the convolutional DoG model to have a higher mean accuracy of $r_{dog\_conv} = 0.2118774$, compared to that of the simple DoG model $r_{dog} = 0.1714848$. It has a similar mean accuracy to the parameterized convolutional DoG model $r_{dog\_param\_conv} = 0.21227963$. This is understandable because of the convolution operation in the DoG_conv model, which allows for it to learn to extract meaningful features from the input stimulus that are translation invariant. Meanwhile, the parameterized convolutional DoG model was expected to lie between the original DoG and the convolutional DoG model because its parameters - two vectors of the standard deviations of gaussians that are subtracted from one another - constrain the convolutional filters to DoG kernels, limiting the features that can be extracted from the stimulus with each filter.

A visualization of the filters before and after training in Figure 6 show the differentiation of the filters after training. The filters from the convolutional DoG model seem to be developing orientation selectivity as opposed to the parameterized one, which was constrained to always be a DoG kernel.

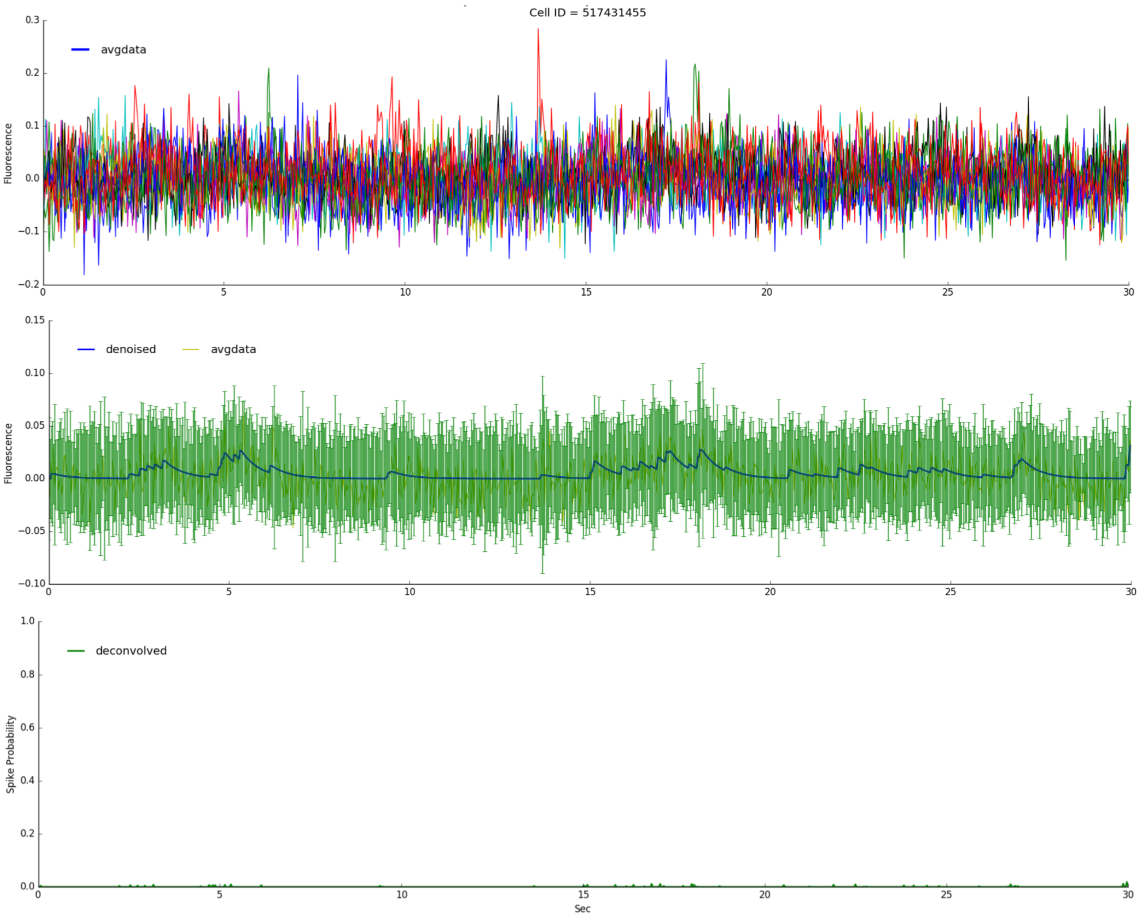


Figure 3: Example of cell with reliable responses over repetitions but not signal in response to the stimulus. The top plot shows a raw fluorescence trace for multiple repetitions of the same stimulus over 30 seconds. The middle plot shows the denoised trace over the repetitions with standard deviation bars in green. The bottom plot shows the outcome of the OASIS deconvolution: spike probabilities at specific time steps.

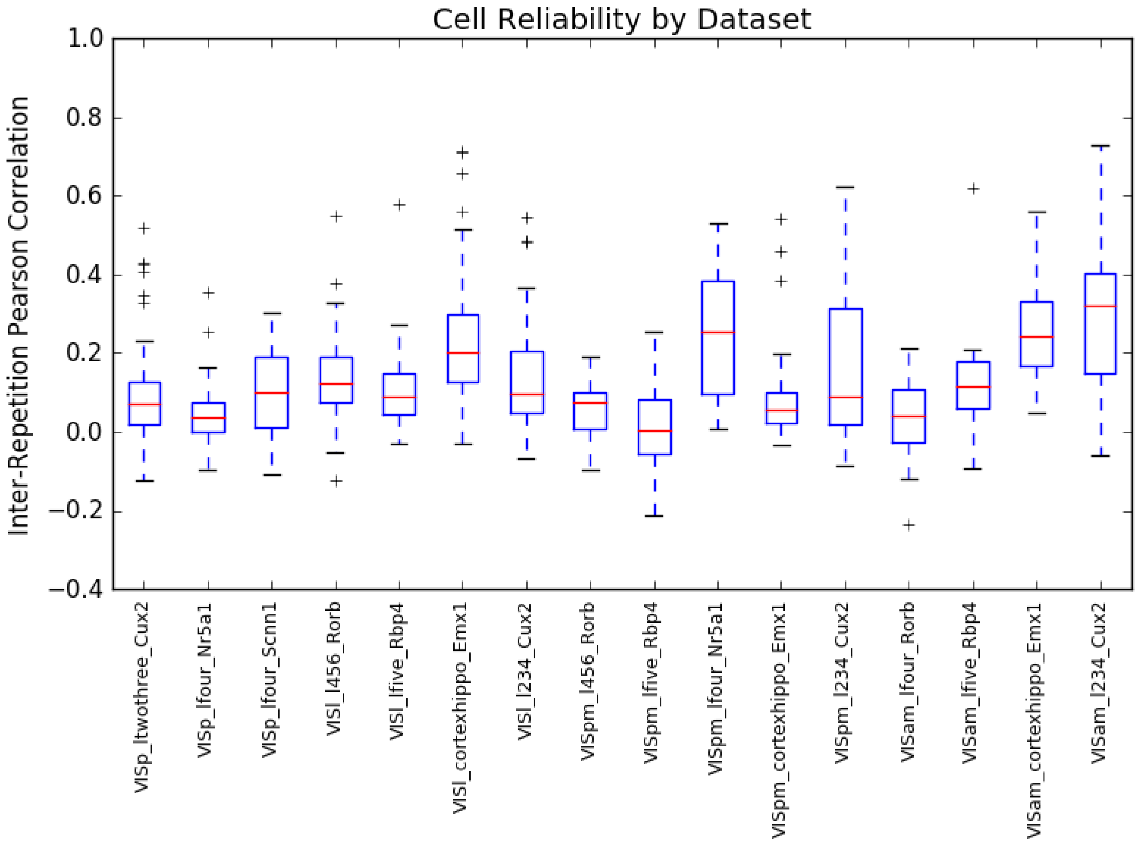


Figure 4: Boxplot of cell reliability as determined by the inter-repetition Pearson correlation for individual cells in different datasets across various visual areas and cortical layers.

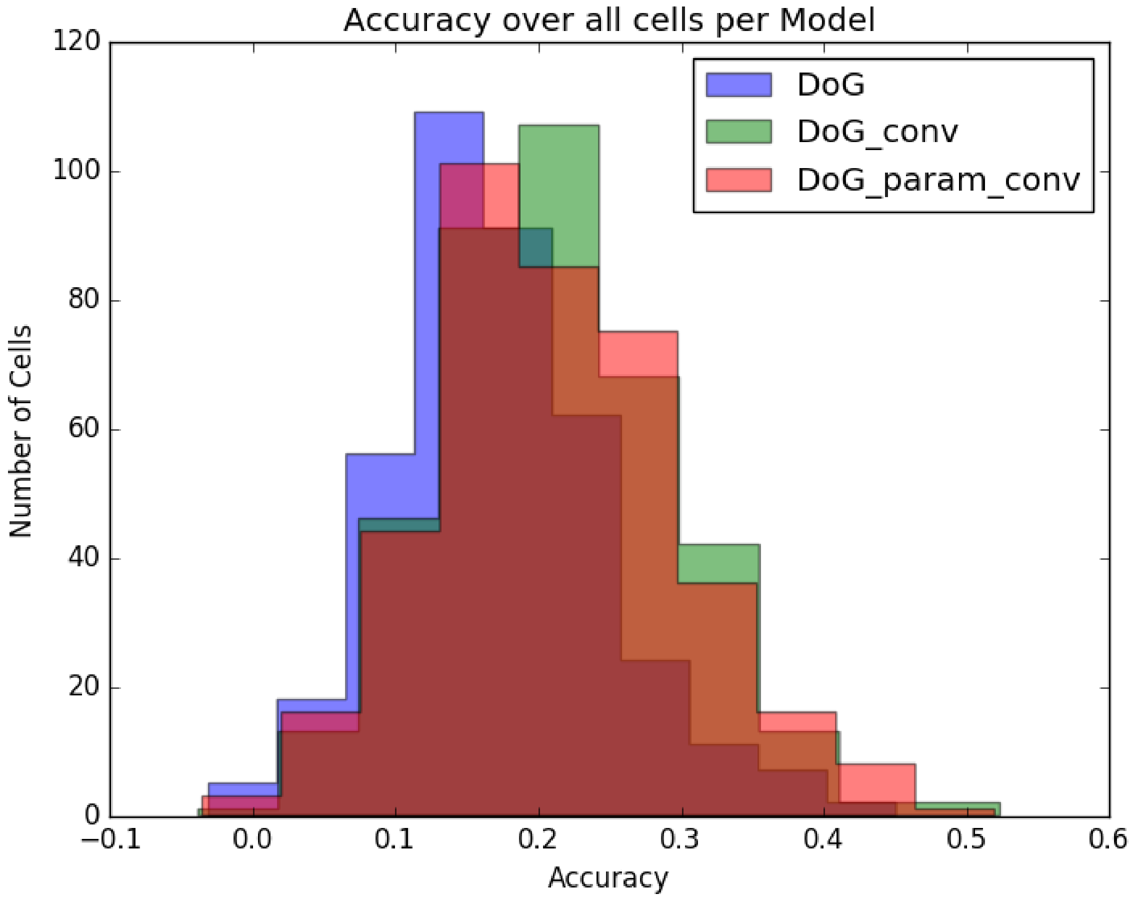
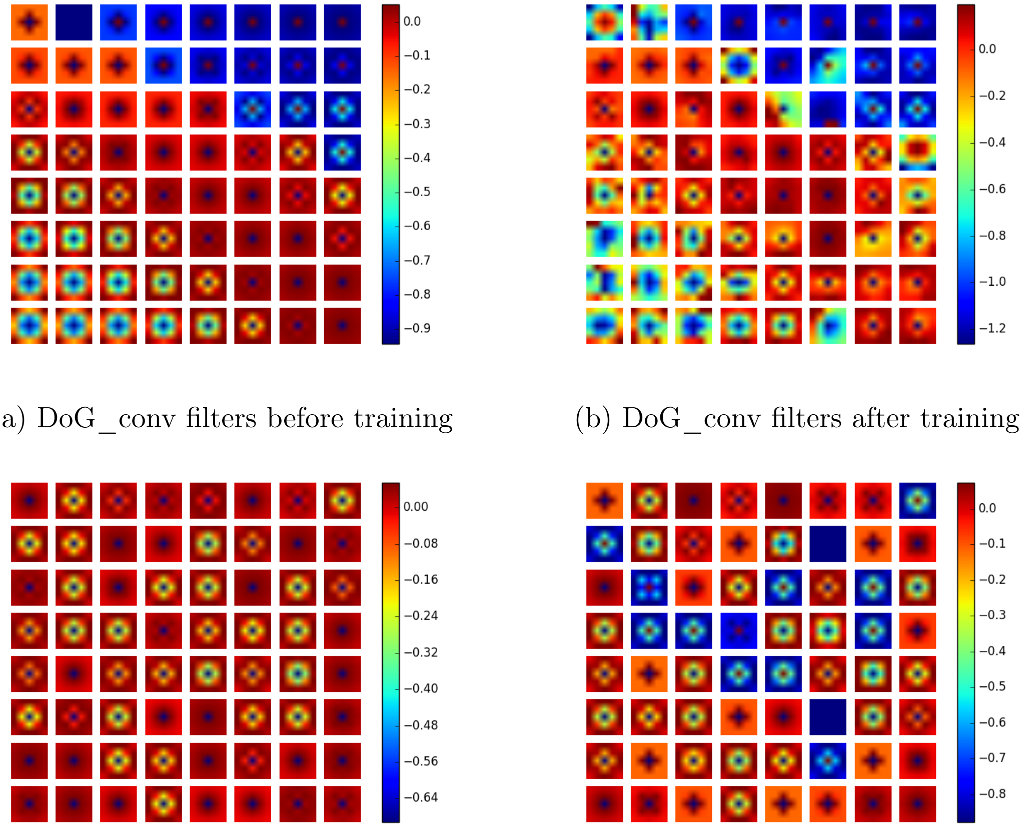


Figure 5: Histogram showing accuracy per cell for each model. This histogram was aggregated over all visual areas and layers where inter-repetition Pearson correlation was positive. The number of cells represented in this figure is 385. Accuracy was calculated as the Pearson correlation between the model's predicted neural activity and ground truth.



(a) DoG_conv filters before training

(b) DoG_conv filters after training

(c) DoG_param_conv filters before training

(d) DoG_param_conv filters after training

Figure 6: Visualizations of DoG kernels before and after training.

5 Methods

5.1 Dataset Creation

The Allen Brain Observatory API was used to query for cell data in relation to specific stimuli, laminar specificity, and visual area. Each dataset was created by querying for a specific stimulus: such as natural movie one, natural movie two, or natural scenes. Then, individual cells, marked by their cell ID number, were filtered by a measure of their reliability, the inter-repetition Pearson correlation calculated for each cell. Cells with positive correlation coefficients in a specific dataset were kept and the rest were removed; this was in an attempt to increase the quality of the data to be fit. Neural responses were ordered by stimulus according to time stamps relating to stimulus presentation, and then averaged over repetitions. This averaged response was then split into a 95% training set and a 5% validation set.

5.2 Model Implementation

The models applied in this paper were implemented and trained in TensorFlow, an open source software library for deep learning released in 2015 by Google [1]. Images and labels were packaged into TFRecord format during dataset creation to speed up training, and analysis was done using Python libraries including Numpy and SciPy [12]. Experiments were run with the Adam optimizer on a Pearson dissimilarity loss and an initial learning rate of 0.003.

6 Discussion

The original goal of building a map of neurons with different spatiotemporal receptive fields was far from met. Due to the nature of the dataset, along with difficulties in implementation, I did not find evidence to support the claim that the Separable Gated Recurrent Unit would provide better predictions of these neural responses.

Concurrently, it was discovered that calcium imaging fluorescence deconvolution is an ongoing field of research with open questions on accurate spiking activity recovery methods. While calcium imaging allows for large scale recordings over time, its signal was unreliable in this dataset.

Instead, I tried to fit different variants of the Difference of Gaussians model introduced by Antolik et. al. in order to determine if convolution or parameterized convolution could contribute to a better fit to neural data [3]. After experiments on the datasets shown above, there was a large discrepancy between the cells within a dataset which were better fit by the three different models. Using a convolutional model with DoG filters resulted in an increase in prediction accuracy on the calcium imaging data. However, this increase was small, which can be explained by the relatively low inter-repetition Pearson correlation for the cells in the dataset. This low score of reliability speaks to the large amount of noise in neural responses in this dataset which makes it difficult to fit the data with accuracy higher than 0.4 or 0.5.

6.1 Further Directions

To reiterate earlier goals that were not met, by finding models that reliably predict neural activity, one could build a computational map of visual cortex. Isolating the models that best represent the underlying computations of neurons in different visual areas would allow for better understanding of the visual processing hierarchy.

6.1.1 Neuron Subspace Dimensionality

A potential goal for this project could have been to answer the question, What is the subspace of the neurons in mouse visual cortex: what is the dimensionality that best fits these neurons?

There are about 70,000 cells in the Allen Brain Observatory Calcium Imaging dataset. Mathematically, it is necessary to have at least as many equations as there are unknowns in order to determine the unknowns. By this rule, the models begin with a weight for each neuron it is trying to fit. However, this is computationally inefficient. Fitting a model across every neuron is a computationally intensive task.

A computationally efficient approach can increase the model's prediction accuracy by minimizing the number of weights learned. Additionally, understanding of this subspace of weights and whether or not it corresponds to specific neurons that code for more significant responses can lead to further conclusions about the response profiles of these neurons.

6.1.2 Spatiotemporal Model Fitting

Today, most neural network models take as input 2-dimensional grayscale images and learn a 2-dimensional kernel. These inputs are static representations of the visual field, limited to an observation at one time point. Although this is important for object localization and recognition, the discovery of neurons with preferred stimuli in the spatiotemporal domain (i.e., with an orientation and time component, such as a drifting bar) leads to the belief that a model limited solely to the spatial domain is not sufficient for an accurate prediction of neural activity [17]. With a 2-dimensional kernel, most of the temporal information encoded in a spatiotemporal neuron is overlooked.

From primate literature, studies have shown that there are neurons in primary visual area that respond maximally to stimuli with combined spatial and temporal components. To learn more about the response profiles of these neurons, a different methodology must be used. Instead, a 3-dimensional kernel applied to a 3-dimensional stimulus would learn a representation that is more similar to the actual neurons in mouse visual cortex. This model, applied to the natural video stimuli of frames over time, should result in better neural response predictions.

6.1.3 Recurrent Neural Network

Unlike a simple feedforward Convolutional Neural Network, what sets a Recurrent Neural Network (RNN) apart is its use of a hidden state that is recurrently modified over time. In the below figure, the input x is convolved with a weight matrix U that is incorporated into the hidden state s per time step. Simultaneously, there is a hidden-facing weight matrix W that is convolved with the hidden state s to determine the amount of the previous state that is to be incorporated into the current state. The output o is the result of s convolved with another matrix V. For each time step t, there is a different output o_t. Because all of the parameters U, V, W remain the same throughout time and are applied at each time step, this model is recurrent.

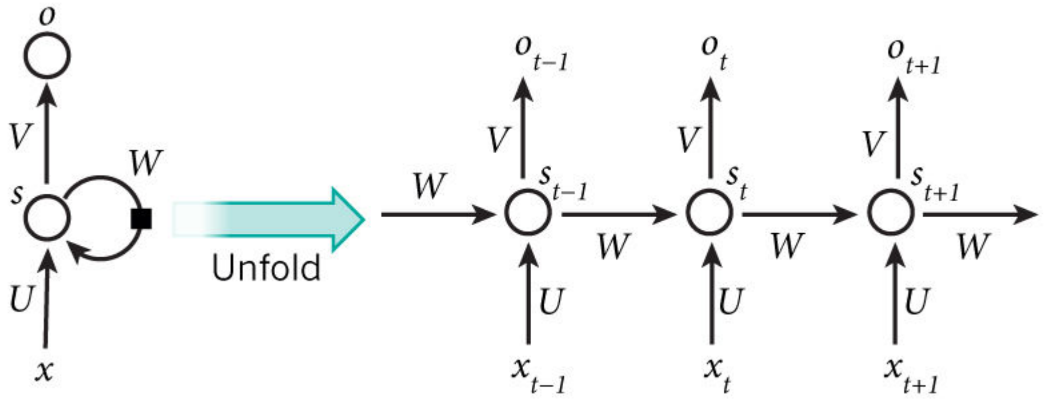


Figure 7: Visualization of a RNN unfolded through time. [10]

RNNs are extremely useful in deep learning because of their ability to learn datasets with a small number of parameters as well as their ability to respond to changing inputs over time. However, these models suffer from a learning problem known as vanishing or exploding gradients. As the weight matrices are updated through gradient descent, which is a method by which the parameters are optimized to learn a specific output by being reduced in the negative direction of their gradient, the values in the matrices can become arbitrarily large or small. To mitigate this problem while training a model of this type, one tactic is to gate the outputs of the parameters with nonlinear functions. An example of a model that does this is the Gated Recurrent Unit.

6.1.4 Gated Recurrent Unit

A Gated Recurrent Unit (GRU) is a model with a gate on a RNN in an effort to mitigate vanishing gradients during training in order to learn long term dependencies. In the example below, there exist two parameter matrices, that when convolved with the hidden state or the input state, undergo a nonlinear gating that determines the amount of information that is incorporated from the hidden state.

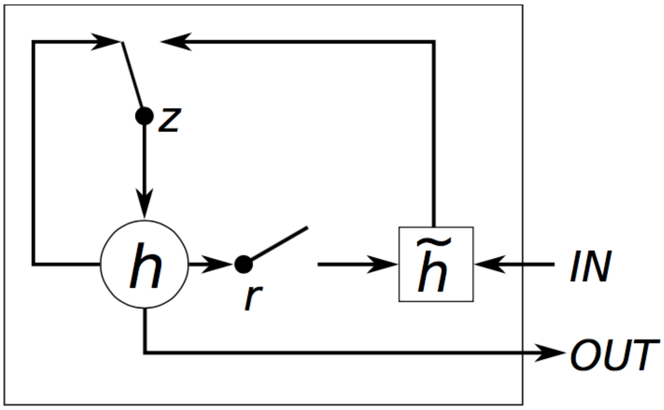


Figure 8: Gated Recurrent Unit. [4]

A typical Gated Recurrent Unit will have the following structure:

$$z = \sigma_g(x_t U_z + h_{t-1} W_z)$$

$$r = \sigma_g(x_t U_r + h_{t-1} W_r)$$

$$h_t = z \circ h_{t-1} + (1 - z) \circ \sigma_h(x_t U_h + W_h(s_{t-1} \circ r))$$

Where z determines the amount of information that will be incorporated into the current state, r determines the amount of information that will be remembered from the current state, and h_t is the output of the state $\vec{h}$ at time t. The σ_s represents a sigmoid gating function, the σ_h represents a hyperbolic tangent, and the $\circ$ indicates a Hadamard product or a dot product.

6.1.5 Separable Gated Recurrent Unit

To incorporate separability between space and time, the Gated Recurrent Unit model was modified to have two gates regulating its hidden state. An input- facing sigmoid gate along with a hidden state-facing sigmoid gate both regulate the hidden state and allow for spatial and temporal information to independently be incorporated into the model.

In the GRU example above, there is a slight modification:

$$
\begin{aligned}
z &= \sigma_g(x_t U_z + h_{t-1} W_z + b_z) \\
r_x &= \sigma_g(x_t U_r + b_{r_x}) \\
r_h &= \sigma_g(h_{t-1} W_r + b_{r_h}) \\
h_t &= z \circ h_{t-1} + (1 - z) \circ \sigma_h(U_h x_t \circ r_x + W_h(r_h \circ h_{t-1}) + b_h)
\end{aligned}
$$

Where z determines the amount of information that will be incorporated into the current state, r_x determines the amount of information from the spatial component that will be remembered from the current state, r_h determines the amount of information from the temporal component that will be remembered from the current state, and h_t is the output of the state $\vec{h}$ at time t. The σ_s represents a sigmoid gating function, the σ_h represents a hyperbolic tangent, the $\circ$ indicates a Hadamard product or a dot product, and the b terms indicate bias values.

A strong benefit of this approach is that the flexibility between spatial and temporal domains is gained without an increase in the number of parameters that are used. There are still only two weight matrices W, U that are trained in this model.

6.1.6 Discrimination of Separable and Entangled Response Profiles

As noted earlier, neurons in primary visual area in both primates and mice have been shown to encode spatial, temporal, and spatiotemporal representations. Though these neurons project to different pathways of visual processing like in previously studied primate models, neurons in mouse primary visual cortex are not organized into columns by shared orientation tuning [17]. Instead, they seem to be scattered throughout the primary visual cortex randomly.

One goal of this project would be to determine the ability to tell whether or not a neuron's response is separable (i.e., have individual space and time components that do not depend on each other) or a product of space and time (i.e., "entangled" in spatial and temporal components). A range of models are fit to these neurons to determine which spatiotemporal models best capture this activity. Through learning which neurons are separable and non separable, a map of the mouse visual cortex is being created, indicating which neurons potentially project to the dorsal and ventral streams.

To do this, the project takes advantage of a recurrent neural network architecture, allowing for the inclusion of short term memory into the model. An additional parameter considering the response of the model to the stimulus in the previous time step allows for a spatiotemporal kernel to be learned from the video stimuli. In this way, the spatial component of the video image is incorporated with information from the previous time step.

To tackle the problem of separability, a sigmoid gating function is included in the model. By forcing inputs from both the spatial and temporal components of the input to either 0 or 1, this function acts as a threshold that will either activate or inactivate the two respective components of the recurrent network. If a stimulus induces a low spatial response from a component of the model, but the temporal response is strong enough, that individual component, representing the activity of one neuron, will be demonstrating a temporal response profile. If the opposite is true, a spatial neuron will be modeled. Finally, if both components are allowed through by the gating mechanism, an "entangled" neuronal response will be modeled. In this way, the model will learn to discriminate between the response profiles of the different neurons in mouse primary visual cortex.

After this differentiation is learned, the next step could be to build a map of separable vs entangle neurons across primary visual area. Using the transgenic lines expressed by these neurons, it could also be possible to map these characteristics with laminar specificity.

References

[1] Martín Abadi et al. "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems". In: (). URL: https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/45166.pdf.

[2] Allen Institute for Brain Science. *Allen Brain Atlas: Mouse Connectivity*. 2015.

[3] Ján Antolík et al. "Model Constrained by Visual Hierarchy Improves Prediction of Neural Responses to Natural Scenes". In: *PLOS Computational Biology* 12.6 (June 2016). Ed. by Matthias Bethge, e1004927. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1004927. URL: http://dx.plos.org/10.1371/journal.pcbi.1004927.

[4] Junyoung Chung, Caglar Gulcehre, and Kyunghyun Cho. "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling". In: (). URL: https://arxiv.org/pdf/1412.3555.pdf.

[5] Kathleen Esfahany et al. "Organization of Neural Population Code in Mouse Visual System". In: (). DOI: 10.1101/220558. URL: https://www.biorxiv.org/content/biorxiv/early/2018/03/25/220558.full.pdf.

[6] Johannes Friedrich, Pengcheng Zhou, and Liam Paninski. "Fast online deconvolution of calcium imaging data". In: *PLOS Computational Biology* 13.3 (Mar. 2017). Ed. by Joshua Vogelstein, e1005423. ISSN: 1553-7358. DOI: 10.1371/

journal.pcbi.1005423. URL: http://dx.plos.org/10.1371/journal.pcbi.1005423.

[7] Walter J. Gehring. "The evolution of vision". In: *Wiley Interdisciplinary Reviews: Developmental Biology* 3.1 (2014), pp. 1–40. ISSN: 17597684. DOI: 10.1002/wdev.96.

[8] Zoe Kourtzi, Bart Krekelberg, and Richard J A van Wezel. *Linking form and motion in the primate brain.* 2008. DOI: 10.1016/j.tics.2008.02.013.

[9] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: (). URL: https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

[10] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". In: *Nature* 521.7553 (May 2015), pp. 436–444. ISSN: 0028-0836. DOI: 10.1038/nature14539. URL: http://www.nature.com/articles/nature14539.

[11] Chee-Kit Looi et al. "Representational Learning". In: *Encyclopedia of the Sciences of Learning.* Boston, MA: Springer US, 2012, pp. 2832–2835. DOI: 10.1007/978-1-4419-1428-6{_}524. URL: http://www.springerlink.com/index/10.1007/978-1-4419-1428-6_524.

[12] Travis E. Oliphant. “Python for Scientific Computing”. In: *Computing in Science & Engineering* 9.3 (2007), pp. 10–20. issn: 1521-9615. doi: 10.1109/MCSE.2007.58. url: http://ieeexplore.ieee.org/document/4160250/.

[13] Nicole C. Rust. “Do rats see like we see?” In: *eLife* (2017). issn: 2050084X. doi: 10.7554/eLife.26401.

[14] L. C. Sincich and Jonathan C Horton. “Divided by Cytochrome Oxidase: A Map of the Projections from V1 to V2 in Macaques”. In: *Science* 295.5560 (Mar. 2002), pp. 1734–1737. issn: 00368075. doi: 10.1126/science.1067902. url: http://www.ncbi.nlm.nih.gov/pubmed/11872845%20http://www.sciencemag.org/cgi/doi/10.1126/science.1067902.

[15] I T Smith et al. “Stream-dependent development of higher visual cortical areas”. In: *Nat.Neurosci.* (2017).

[16] Sina Tafazoli et al. “Emergence of transformation-tolerant representations of visual objects in rat lateral extrastriate cortex”. In: *eLife* (2017). issn: 2050084X. doi: 10.7554/eLife.22794.

[17] Samme Vreysen et al. “Dynamics of spatial frequency tuning in mouse visual cortex.” In: *Journal of Neurophysiology* (2012). issn: 1522-1598. doi: 10.1152/jn.00022.2012.