UNIVERSITY OF THESSALY

SCHOOL OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Prediction of the pupil size and neural orientation preference using the neural activity of the visual cortex

Diploma Thesis

Charalampos Peteinarelis

Supervisor: Manolis Vavalis

July 2023

UNIVERSITY OF THESSALY

SCHOOL OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Prediction of the pupil size and neural orientation preference using the neural activity of the visual cortex

Diploma Thesis

Charalampos Peteinarelis

Supervisor: Manolis Vavalis

July 2023

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Πρόβλεψη του μεγέθους της κόρης (οφθαλμού) και της προτίμησης προσανατολισμού νευρώνων αξιοποιώντας την νευρωνική δραστηριότητα του οπτικού φλοιού

Διπλωματική Εργασία

Χαράλαμπος Πετειναρέλης

Επιβλέπων: Μανόλης Βάβαλης

Ιούλιος 2023

Approved by the Examination Committee:

Supervisor **Manolis Vavalis**
Professor, Department of Electrical and Computer Engineering, University of Thessaly

Member **Maria Papadopouli**
Professor, Computer Science Department, University of Crete, Foundation for Research and Technology-Hellas (FORTH)

Member **Ioannis Tsamardinos**
Professor, Computer Science Department, University of Crete

Acknowledgements

I would like to thank Prof. Vavalis (Professor, Electrical and Computer Engineering Department, University of Thessaly) for his availability and help and Prof. Papadopouli (Professor, Computer Science Department, University of Crete / Foundation for Research and Technology-Hellas (FORTH)) for her constant guidance. This thesis was written under Prof. Papadopouli's supervision and is a result of close collaboration with her team in the Foundation for Research and Technology-Hellas (FORTH, Crete) and University of Crete. All data used in this study were collected and provided by the team of Prof. Andreas Tolias (Professor at Department of Neuroscience, Baylor College of Medicine) in collaboration with the team of Prof. Stelios Smyrnakis (Associate Professor of Neurology, Brigham and Women's Hospital, Harvard University). Access to the data was given to the author by Prof. Papadopouli. Special thanks to Dr. Chryssalenia Koumpouzi from Prof. Papadopouli's team at FORTH for her help in practical matters of this thesis.

Finally, I am always thankful to my mother for her support and presence.

DISCLAIMER ON ACADEMIC ETHICS AND INTELLECTUAL PROPERTY RIGHTS

«Being fully aware of the implications of copyright laws, I expressly state that this diploma thesis, as well as the electronic files and source codes developed or modified in the course of this thesis, are solely the product of my personal work and do not infringe any rights of intellectual property, personality and personal data of third parties, do not contain work / contributions of third parties for which the permission of the authors / beneficiaries is required and are not a product of partial or complete plagiarism, while the sources used are limited to the bibliographic references only and meet the rules of scientific citing. The points where I have used ideas, text, files and / or sources of other authors are clearly mentioned in the text with the appropriate citation and the relevant complete reference is included in the bibliographic references section. I also declare that the results of the work have not been used to obtain another degree. I fully, individually and personally undertake all legal and administrative consequences that may arise in the event that it is proven, in the course of time, that this thesis or part of it does not belong to me because it is a product of plagiarism».

The declarant

Charalampos Peteinarelis

Diploma Thesis

Prediction of the pupil size and neural orientation preference using the neural activity of the visual cortex

Charalampos Peteinarelis

Abstract

The development of neuroimaging techniques and the growing popularity and applicability of Machine Learning over the last years open new horizons in Computational Neuroscience. Artificial Intelligence is itself an austere, mathematical paradigm of intelligence that mimics the natural process of pattern recognition with many methods. In this thesis, we explore this perspective by using relatively simple Machine Learning methods to approach commonly studied subjects in Neuroscience, such as the relationship of the primary visual cortex neural activity with behavioral characteristics and the cortical representation of visual stimuli. We used the calcium events (collected with two-photon calcium imaging) of layers 2,3,4 in the V1 area of five mice to predict their pupil size over time, and neuron firing rates to predict the direction of the visual stimulus over time. In pupil size prediction, it appeared that the involvement of pupil history in the model input can drop the Root Mean Squared Error by one order of magnitude, that this does not hold for the involvement of treadmill history, and that prediction of future values is possible yet increasingly sensitive in the long term. We also explored neuron impact in pupil prediction and its possible relationship with firing rates and functional connectivity. Edge Detection and Optical Flow Computer Vision methods for the detection of stimulus direction were also examined and evaluated later on. Lastly, we attempted to predict the direction of stimulus with the use of firing rates per segment by means of classification on three different mice, and find the association between firing rates and neuron orientation preference was briefly studied. The results seem encouraging, however they are preliminary and need to be investigated further.

Keywords:

Computational Neuroscience, Machine Learning, Pupil Size Prediction, Orientation Preference, Primary Visual Cortex

Διπλωματική Εργασία

Πρόβλεψη του μεγέθους της κόρης (οφθαλμού) και της προτίμησης προσανατολισμού νευρώνων αξιοποιώντας την νευρωνική δραστηριότητα του οπτικού φλοιού

Χαράλαμπος Πετειναρέλης

Περίληψη

Η ανάπτυξη των τεχνικών νευροαπεικόνισης και η αυξανόμενη δημοτικότητα της Μηχανικής Μάθησης τα τελευταία χρόνια ανοίγουν νέους ορίζοντες στην Υπολογιστική Νευροεπιστήμη. Η ίδια η Τεχνητή Νοημοσύνη είναι ένα αυστηρό, μαθηματικό πρότυπο νοημοσύνης που μιμείται τη φυσική διαδικασία αναγνώρισης μοτίβων μέσω πολλών μεθόδων. Σε αυτήν την διπλωματική εργασία, εξερευνούμε αυτήν την προοπτική, χρησιμοποιώντας σχετικά απλοϊκές μεθόδους Μηχανικής Μάθησης για να προσεγγίσουμε ευρέως μελετημένα θέματα της Νευροεπιστήμης, όπως η σχέση της νευρωνικής δραστηριότητας του πρωτογενή οπτικού φλοιού με συμπεριφορικά χαρακτηριστικά και με την αναπαράσταση οπτικών ερεθισμάτων στον φλοιό. Χρησιμοποιήσαμε τα ”calcium events” (που συλλέχθηκαν με απεικόνιση ασβεστίου δύο φωτονίων) των στοιβάδων 2,3,4 στην περιοχή V1 πέντε ποντικιών για να προβλέψουμε το μέγεθος της κόρης τους στην πάροδο του χρόνου, και τους ρυθμούς πυροδότησης των νευρώνων για να προβλέψουμε την κατεύθυνση του οπτικού ερεθίσματος στην πάροδο του χρόνου. Στην πρόβλεψη του μεγέθους της κόρης φάνηκε πως η συμμετοχή του ιστορικού της κόρης στην είσοδο του μοντέλου μπορεί να ρίξει τη ρίζα του Μέσου Τετραγωνικού Σφάλματος κατά μία τάξη μεγέθους, ότι αυτό δεν ισχύει με το ιστορικό της κίνησης του ποντικιού και ότι η πρόβλεψη μελλοντικών τιμών μεγέθους της κόρης μπορεί να επιτευχθεί με καλή ακρίβεια, αλλά μεγάλη ευαισθησία μακροπρόθεσμα. Επίσης διερευνήσαμε τον αντίκτυπο των νευρώνων στην πρόβλεψη της κόρης και τη πιθανή σχέση του με τους ρυθμούς πυροδότησης και την λειτουργική συνδεσιμότητα. Μέθοδοι Ανίχνευσης Ακμών και Οπτικής Ροής εξετάστηκαν και αξιολογήθηκαν, επίσης, με σκοπό την ανίχνευση της κατεύθυνσης του ερεθίσματος μέσω Υπολογιστικής Όρασης. Στο τέλος, προσπαθήσαμε να προβλέψουμε την κατεύθυνση του ερεθίσματος μέσω ταξινόμησης σε τρία διαφορετικά ποντίκια με τη χρήση ρυθμών πυροδότησης ανά τμήμα ερεθίσματος, και να βρούμε τη συσχέτιση μεταξύ ρυθμών

πυροδότησης και προτίμησης προσανατολισμού των νευρώνων μελετήθηκαν εν συντομία. Τα αποτελέσματα φαίνονται ενθαρρυντικά, παρ'όλα αυτά είναι πρώιμα και χρήζουν περεταίρω μελέτης.

Λέξεις-κλειδιά:

Υπολογιστική Νευροεπιστήμη, Μηχανική Μάθηση, Πρόβλεψη Μεγέθους Κόρης, Προτίμηση Προσανατολισμού, Πρωτογενής Οπτικός Φλοιός

Table of contents

List of figures

Abbreviations

i.d.	*id est*
w.r.t.	*with respect to*
etc.	*et cetera*

Chapter 1

Introduction and Motivation

From the famous *Turing test* [2] to the fictional *Voight-Kampff* pupillometry test seen in the neo-noir, sci-fi mythology of *"Blade Runner"* [3], many tests and methods have been constructed to empirically approach, define or detect natural intelligence. This endeavor, regardless of its character, which might be of philosophical, fictional or mathematical nature, proves that the extraction of a rigorous definition of intelligence and the mechanisms underneath still remains a task that causes philosophical and scientific confusion; meanwhile, the rapid development of Artificial Intelligence over the course of the last years gradually blurs the distinguishing line between the existence and the lack of "true" intelligence and rises new questions on the definitions and mechanisms behind both Natural and Artificial intelligence. Such an environment dictates the need for interdisciplinary research that addresses problems in both areas and mediates between them. The questions and solutions that arise from the osmosis of these two areas lie in the heart of Computational Neuroscience.

The purpose of Data Science and Artificial Intelligence is to extract patterns and knowledge represented in a system and therefore can be used to analyze, or even reverse engineer complex systems. One such system, that follows a modular, circuit-like network structure is the mammalian brain. Knowledge and patterns can be extracted from treating such a system as a black box and examining its behavioral characteristics (e.g. pupillary or motor responses), but there is also information that can be revealed on its internal mechanics and the connection between the two scopes. Of course, these mechanics themselves can be seen as "black boxes" in the study of a system as complicated as the brain, so it is a problem that must be studied at various levels and granularities.

It is of utmost importance, however, to note that the application of Machine Learning methods, especially when referring to a system like the brain, *does not infer causal laws between internal dynamics and behavioral characteristics* but rather provides patterns and observations that can be useful for the analysis of the system. Artificial Intelligence a mathematical model of Natural Intelligence in many of its parts, that is able to mimic Natural Intelligence and track the course of information in the networks of neurons. The impressive growth of Machine Learning both in research and development and the availability of computational resources combined with the advances in contemporary neuroimaging techniques indicate that now, more than ever, it is the most appropriate time to investigate this perspective.

The motivation between the investigation of the interface between Natural and Artificial Intelligence is dual; we can learn *about* the brain, and we can learn *from* the brain. From the neuroscience side, more can be learned about the representation and processing of information in the labyrinthine, mostly undeciphered dynamics of the brain, knowledge that can be of great importance to medical applications, such as the detection of pathogenic patterns in certain neuron circuits for the detection and prevention of diseases. On the other hand, from the computer science perspective, deeper knowledge of the brain can contribute to the improvement of existing Machine Learning methods (especially as far as generalization and robustness is concerned), or the inspiration of new ones. Thus, although the results of this research field have to be very carefully examined to ensure robustness and the avoidance of any causal inferences, the field itself is a very promising and modern neuroscience-oriented take on the Imitation Game.

1.1 Thesis Subject

This thesis touches certain specific subjects of Computational Neuroscience. It uses computational methods from Machine Learning applied on data collected through experimental procedures in an attempt to study the connection between neuronal activity of the mouse primary visual cortex (V1) and behavioral characteristics such as the size of the pupil, as well as the relationship between the representation of stimulus orientation in V1 and neuronal orientation preference. Our approach mainly includes predictions. We apply regression on the calcium events of neurons in layers L2, L3, L4 to predict the pupil size with respect to time,

as well as classification to predict the direction of visual stimulation in time using the firing rates of neurons in the same layers.

This work on the one hand focuses on how the aforementioned predictions are set up, their evaluation according to baseline models and how they are affected by the inclusion/exclusion of certain variables, while on the other hand attempts to interpret the results by *explaining* the models' performance and exploring the relationships between quantities that describe neuron activity and importance, in order to make biology-related observations. The generalization, though, of specific observations can be erroneous, when the results are not robust (for every mouse/subject, for example) or when simple algorithms are used, that cannot capture complex relationships between given data and target variable. Therefore, one has to be very cautious when interpreting the results of the algorithms used in this study, since they reflect in no way causal mechanisms, but nevertheless automate the process of observation on imaging data, evaluate it with metrics and provide a helpful tool that unveils patterns invisible to the naked human eye.

1.1.1 Contribution

The contribution of this thesis is summed up by the following results/action points:

1. It appears that mouse pupil size can be predicted by calcium events of layers 2,3,4 in V1 with an %RMSE of 11%

2. The inclusion of pupil history is very likely to affect the aforementioned prediction error positively. The same can be claimed for the prediction of future values, too

3. We explored the neurons' predictive power in linear models and its connection to firing rates and degree of connectivity

4. Cortical state can be represented in a low-dimensional space, always according to Principal Component Analysis

5. We employed and implemented Computer Vision Algorithms to quantify the stimulus direction

6. It seems that stimulus direction can be predicted by the firing rates of neurons in layers 2,3,4 in V1

Of course, again, it is important to emphasize that these results are indicative and preliminary and need to be investigated further to be safely generalized to verified biological patterns.

1.2 Thesis Organization

The rest of this thesis is organized as follows. Next we briefly provide a background on the data we use. We locate the area of the brain where they are collected from, we report the techniques that were used experimentally to collect them 2.1 and describe our datasets 2.2. We then review the existing neuroscience-related results 3 as well as the computational techniques that have been employed to solve problems similar to ours 3.1.

Chapter 4 contains the main results of this study. In 4.1 we describe the pupil prediction, the training process, the baseline models used the results yielded and the inclusion/exclusion of certain specific variables. Also neuron impact and its relationships with firing rates and degree of connectivity are explored. 4.2 describes the computer vision methods we used to quantify the orientation of the stimulus. In 4.3 we describe the prediction of stimulus orientation, how we modeled this problem, what were the results and we attempt to correlate the firing rates of neurons with their orientation preference. Our concluding remarks can be found in section 5.

Chapter 2

Background

In 1909, Brodmann [4] used a groundbreaking histological approach to discover that functionality in the cerebral cortex is localized, identifying cytoarchitectonic subdivisions with distinct and specific structure and function. His mapping of the brain, though having undergone some revisions in the following few years, is still widely used today and referred to as ”Brodmann Areas”. This thesis focuses on the Brodmann Area 17 (B17), also known as V1 area, or *primary visual cortex*.

The visual cortex is divided in main areas with structural and functional differences. V1 is the first area in the visual cortex that processes visual information coming from the retinas and acts as a ”sorting” area that passes the information to the rest areas of the cortex for a more fine-detail level of processing. The visual cortex is very well characterized in mammals, and is of sensory character, which means that one can manipulate its input to reverse engineer it and retrieve information on the cortical representation of visual stimuli. Consequently, it offers a cortex paradigm that favors its analysis from the perspective of (computational) neuroscience. This thesis uses data extracted from the V1 area of laboratory mice, which are one of the most characterized mammals, biologically speaking, and its context did not involve any work on the production of these measurements.

The data were collected by the team of Prof. Andreas Tolias (Professor at Department of Neuroscience, Baylor College of Medicine) in collaboration with the team of Prof. Stelios Smyrnakis (Associate Professor of Neurology, Brigham and Women’s Hospital, Harvard University) and their analysis for the development of this thesis was conducted under Prof. Papadopouli’s supervision and in close collaboration with her team in the Foundation for Research and Technology-Hellas (FORTH) and University of Crete.

2.1 Experiments and Data Collection

The data used in this study were collected from experiments on 5 adult mice[1] (that is approximately 10 to 12 weeks of age) head-posted on a treadmill during a quiet wakefulness state with spontaneous walking periods being ignored in the analysis, as the main trends that were examined remained the same after the exclusion of these periods. Mesoscopic two-photon imaging was used as a technique for the collection of data, over 2 weeks after the craniotomy to prevent imaging during inflammation. The two-photon imaging results were yielded at a frequency of $6Hz$ over an approximately $1.2 \times 1.2mm^2$ field of view and the mice were expressing the GCaMP6s fluorescent Ca^{2+} indicator in pyramidal neurons as the F1 cross between BL6-SLC17a7-Cre X Ai162 (JAX stocks #023527 and #031562).

The areas covered by the imaging are the biggest part of dorsal area V1 of the visual cortex and the nearby extrastriate cortex (areas V1, LM, RL and AL) and four planes were sampled that correspond to layer 2, layer 3, layer 4 and layer 5 of area V1. This work focuses on the study of granular layers (layer 4 also seen as L4) and supragranular layers (layers 2,3 also seen as L2/3 or L23). Motion correction was applied on the resulting images as a standard preprocessing, followed by automatic segmentation, deconvolution with the use of the CNMF CaImAn algorithm [5] and appropriate thresholding on the resulting deconvolved signal to obtain *calcium events*. Note that the results are not sensitive to the value of the threshold, and the threshold was ultimately chosen so that the resulting event rates approximate those met in literature.

A total average of 7.76% of neurons in area V1 were ignored in the analysis, comprising all neurons with a calcium event frequency smaller than $0.01Hz$ (0.27% of neurons segmented in V1 of each mouse, on average), as well as the ones $15mm$ away (or less) from the field of view periphery (for the prevention of any potential edge effects caused by incomplete correction of motion artifacts).

[1] We will be referring to these mice as mice 3-7. The number serves strictly as an ID number in the context of this study.

2.2 Datasets Description

As seen above, the activity of the neurons is depicted as calcium events. After thresholding, we end up in a binary event representation, where firing events are represented as events, ”ones” in the dataset, whereas the lack of firing is represented as ”zero”. Therefore, the firing activity of a particular neuron in time is expressed with a vector with its coefficients equal to 0 at the positions that mark the moments of non-firing and equal to 1 at the positions that mark the moments of firing. Note that the length of this vector indicates the recording period *measured in two-photon frames* (for the frequency mentioned above), so each vector coefficient declares whether a specific neuron fired in a specific two-photon frame.

A collection of such vectors for all the recorded neurons of a mouse is called an *eventogram* and it is presented as a $n \times m$ matrix, where n is the total number of two-photon frames and m the number of neurons for the given mouse. It is the input to most of our algorithms.

The pupil data we use are in the form of time series. The pupil time series contains the pupil diameter of a particular mouse as positive floating point numbers in time. Initially, as time series data, pupil data were measuring time in timestamps, but after some preprocessing, two-photon frames were translated to time in timestamps so that each frame can be assigned a pupil size value. At most cases, a frame translated to many timestamps (as a frame lasts longer than a timestamp), in which case the average pupil size value from all the corresponding timestamps was assigned to that frame. This way, time is measured in two-photon frames here too, so pupil size data and neuron firing data are in alignment w.r.t. time. Pupil radius data were collected with the technique of pupil segmentation; an ellipse was fitted on the pupil and then the long radius of the pupil was used. Hence, pupil radius is measured in pixels.

Treadmill data follow a similar structure; they express the velocity of the treadmill the mouse we examine is on, measured in cm/sec, with a signed floating number - the sign determines the direction of the treadmill movement, and when the treadmill is still, the velocity appears equal to 0. Also aligned with the two-photon frame scale. The quantities seen in this paragraph are used as target values in predictions and are normalized for that purpose accordingly.

Other data handled in the scope of this study are neuron firing rates, intra-layer normalized degrees of connectivity, stimulus orientation/direction angles. The calculation of these quantities is discussed later. Firing rates are expressed as positive floating point numbers, and each neuron is assigned exactly one firing rate for each period for which it is examined: if we choose to calculate firing rates for the whole recording period (all two-photon frames of an eventogram) each neuron is assigned one firing rate per eventogram, however firing rate can be calculated for specific segments of the total recording period, in which case the neuron is assigned multiple firing rates, one for each segment.

Intra-layer normalized degrees of connectivity are expressed as positive floating point numbers in $[0, 1]$. The same applies to stimulus orientation and direction angles, with the only difference being that these angles do not belong to any given dataset but are calculated manually with the use of computer vision algorithms and range from 0.0 to 180.0 (degrees) and from 0 to 360.0 (degrees) accordingly. Neuron's direction tuning data follow the same format as calculated directions.

All the results in the context of this study were acquired with the use of custom Python3 code developed with the help of the following libraries and software tools: *Jupyter, pandas, numpy, matplotlib, seaborn, OpenCV, scikit-learn.*

Chapter 3

Literature Review

Neuron activity encodes complex information that can be crucial for the understanding of cortical networks. Theoretically, the number of possible firing rate vectors[1] for a small population of neurons is astronomical, yet, of all possible combinations that could be randomly generated, only a small, certain portion is actually observed and this constraint is expressed even stronger in presence of stimulation [6].

Each work seen in literature employs statistical and computational tools that combine various characteristics (individual neuron activity, population activity, firing rates etc.) of this activity to reveal patterns in cortical functionality. Techniques that provide single-cell resolution can be useful for that cause, however the representation of information is found to be more apparent on the population activity of neurons: neurons are observed to join - promiscuously, to some extent- stimulus-driven or spontaneous functional ensembles [7] that modulate the cortical state. They can be coupled to sensory cortex population activity strongly or weakly, which separates them to choristers and soloists accordingly, a classification that is done w.r.t. anatomical properties of a neuron and explains the majority of pairwise neuron correlations in spontaneous activity [8]. The joint firing of neurons, in areas such as the visual cortex, is also considered responsible for the encoding of visual information [9]. The notion of cortical state itself has been based on population activity, since its defining characteristic is ”the amount of slow fluctuation in the summed activity of a set of local neurons” [10].

Neuronal noise is also a subject of discussion in literature. High trial-to-trial variability in single-cell activity creates noise that is averaged in the level of population activity, but nevertheless suggests that even population activity provides a probabilistic estimation of stimulus

[1] A vector that expresses the firing rates of neurons or neuron groups at a given moment.

information [11]. This variability has been studied through the variance in neuron pairwise correlations, which is found to be affected mostly by spike isolation, tuning distance, cortical distance, joint spike width and joint firing rate, across many timescales [12]. Another information-encoding characteristic of neuron activity that has been previously studied is the presence of functional hierarchy in corticocortical, thalamocortical and corticothalamic connectivity, where higher-order areas (layers L5, L6) are found to ”signal behaviorally relevant changes in image identity” more strongly than low-order ones (layers L2, L3, L4) [13].

Neuron activity and its characteristics are often not studied alone. A big portion of previous work focuses on the connection of cortical state to behavioral characteristics [14][15] [16][17][18] and external stimuli, especially during wakefulness. The most common behaviors studied are locomotion and changes in the pupil size [18] along with general exploratory behaviors [17]. [19] has analyzed the functional connectivity of visual cortex, how the internal state (population activity and pupil size) modulate the neural activity and how the activity is propagated from layer 4 (L4, input layer) to layer 2/3 (L23, output layer) [20] [2].

In [18], where it is demonstrated how locomotion and arousal in mice are observed together but maintain separate roles in V1 activity, changes in pupil dynamics are connected with robust changes in Local Field Potential (LFP) patterns during quiescence, and more specifically it is reported that increased pupillary arousal temporally reduces low-frequency LFP oscillations and suppresses spontaneous firing rates. Changes in pupil are also highly correlated with cortical state, since during dilation, the visual cortex has been found to be in the desynchronized[3] state [17].

3.1 Previous Computational Work

A fair amount of machine learning-oriented work has been published over the last two decades that handles data depicting neuron activity and behavioral characteristics. Firstly, this includes various kinds of neuroimaging data such as ECoG (Electrocorticogram) signals [21], EEG (electroencephalogram) signals and derived features [22], MRI (Magnetic Resonance Imaging) data [23], NIRS (Near Infrared Spectroscopy) data [24] and more often LFP (Local Field Potentials) [18][10][25] and of course two-photon calcium imaging data

[2] This work is under preparation but was extensively discussed to me in the context of my thesis.

[3] According to [10], cortical state can either be classified as synchronized or desynchronized. In desynchronized state population rate fluctuates weakly, whereas in synchronized state it fluctuates more strongly.

[17][19]. These neuronal activity data, in all their different modalities, are collected across literature from the visual cortex of mammals including mice [19], cats [26], macaques [27] and humans [28][29]. Interesting applications that use neuroimaging data from humans range from the study of medical diseases and impairments [28] to BCI (Brain-Computer Interface) applications [22][24][29].

Secondly, measurements of behaviors and external stimuli are also introduced in these *in silico* analyses, and are either combined and correlated with neuron activity or studied independently. As far as behavioral characteristics related with V1 are concerned, computer science-oriented literature mainly includes pupil diameter, both as a characteristic coupled with neuronal activity [21] and as an independent indicator of cognitive processes in humans [30][22][24], as well as non-invasive techniques for pupil measurement, tracking and segmentation [31][32]. Direction and orientation preference is also utilized across literature and is usually coupled with external visual stimulus and neuronal activity to retrieve information on the internal representation of stimuli in the visual cortex [26][27].

The computational part of bibliography is concerned with extracting patterns from the data described above. For that purpose, various machine learning and deep learning methods have been employed and compared on the data described above. Among them, Support Vector Machines stand out [33][28][29], as arguably the most commonly used algorithm for such purposes. Of course, many other supervised and unsupervised methods have been applied. Supervised methods, which are more commonly used, include Artificial Neural Networks like Multi-layer Perceptron networks with various activation functions, Convolutional Neural Networks (CNNs) [31] and Recurrent Neural Networks (RNNs)[26]. Unsupervised methods for the discovery of latent representations [23] include Clustering (hierarchical or graph-based), Gaussian Mixture Models, Decomposition Methods such as Independent Component Analysis (ICA), Principal Component Analysis (PCA) and Matrix Factorization, embedding techniques, Hidden Markov Models and reinforcement learning.

Chapter 4

Pupil Prediction, Stimulus Prediction and Computer Vision Methods

4.1 Prediction of the Pupil Size

The diameter of the pupil has been widely studied as a proxy for the activity of neurons in the visual cortex of mammals. It is one of the most commonly studied behavioral characteristics along with locomotion [18] and general exploratory behaviors [17]. Previous work has utilized the information that changes in the pupil size provide to extract a wide range of knowledge, from the mental or emotional state of humans all the way to the neural activity of rodents and primates, and everything in between.

This study, based on evidence that clearly support the connection between pupil size and mouse visual cortex activity through specific mechanics, aims to ”test” this connection, in computational terms, by attempting to predict the pupil size using the neuronal activity of layers L23, L4 in V1. Purpose of this prediction is to display how the aforementioned connection manifests itself through the application of simple machine learning algorithms and possibly make neuroscience-related observations on the results.

4.1.1 Training Process

By taking a quick glance at related work, we can notice that bibliography truly pays its homage to *Support Vector Machines*; their use is widespread in computational neuroscience [33][28][29] and their mapping and prediction ability has been tested on both structural and functional neuroimaging data, including Local Field Potentials (LFP) and calcium events.

SVMs, in both their regression and classification versions, are able to provide us with a robust prediction, in general, also giving one the ability to perform both linear and non-linear predictions by only adjusting its kernel. Most importantly, it is not a high complexity machine learning algorithm, which makes it *explainable* - at least when using a linear kernel; we are thus able to extract information on the predictive power of each neuron-variable and possibly generalize the conclusions to layer-level or cortex-level observations. This lack of complexity is not only accompanied with a good prediction and explainability, but also with a lack of overfitting: SVMs introduce ”slack variables” into the optimization problem they solve, which prevent them from overfitting. This kind of regularization is also important to our predictions, since we mostly deal with sparse data, which usually make algorithms prone to overfitting.

Therefore, for all predictions related to pupil size in the following sections, we employ *Support Vector Regression*. We use Support Vector Regressors (SVRs) with linear kernels which are trained on the firing data (eventograms) of layers L23 and L4, in all mice, separately or not, in order to predict the pupil size time series, which expresses the pupil diameter of a certain mouse at a given moment - clearly a regression task. We apply k-fold cross validation ($k = 12$), with the leave-one-out strategy, training on $11/12$ of our data and calculating the error with the other $1/12$, for all possible combinations. In all regression tasks involved in this study, the evaluation metric used is the Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\sum_i \frac{(\hat{y_i} - y_i)^2}{N}}$$

where $\hat{y_i}$ denotes the estimated pupil size value at a certain moment, y_i its actual, observed value at that time and N the total number of observations (measurements) on the pupil size. RMSE, besides being a very popular regression evaluation metric is also able to give us an error measured in the same units of measurement as the predicted value, giving us a better perspective on the error in the context of our prediction. At most results though, we use the percent RMSE (% RMSE), which we express as

$$(RMSE)\% = \frac{RMSE}{max\{pupil\} - min\{pupil\}}$$

$RMSE$ being the previously calculated RMSE and $max\{pupil\}, min\{pupil\}$ the maxi-

mum and minimum observed pupil values. The pupil time series data are *standardized* before every fit, as a preprocessing technique, that is they undergo the following transformation:

$$z = \frac{x - \mu}{\sigma}$$

z being the standardized value of the pupil size value to be transformed, x, and μ, σ being the mean and the standard deviation of the pupil size values' set accordingly. This process is equivalent to calculating the values' *z-score*.

4.1.2 Controls: De-synchronization of neurons, Pupil Shuffling, Pupil Shifting

Evaluating the performance of a model, be it a classification or regression model, usually involves the calculation of one or more metrics. However, the exclusive use of metrics for this task is not always the best solution, since at times they can provide a false image of the model's performance. Metrics do not exactly "lie" about the model's performance, but they do have the tendency to reveal only parts of the "truth". This creates the need for baseline models: a baseline model is a model specifically built to provide a prediction which performs very poorly, so that we have a frame of reference for our main model, performance-wise.

The purpose of this study is not mere performance, but the use of prediction to investigate patterns in neuronal activity and generalize them -if possible- to biology-related observations; thus, baseline models, which in the context of this study will be referred to as *"controls"*, or *"control predictions"*, are of great importance, since they are not only contextualizing the performance of our model, but also assess the quality and quantity of information we receive from a prediction. This section's purpose is to present the main controls used for all our predictions. Such controls -like firing event shifting [25], for instance- are widespread in neuroscience and help a neuroscientist assess the quality of information they obtain from *in vivo* or *in silico* experiments.

The main idea behind our controls is *randomness*: after intervening in a specific, random way in the structure and arrangement of our data, we fit our model on them to create a

prediction that virtually is random, and contains no substancial information. One way to do this is to circularly shift the events of each neuron by a random number of positions, so as to understand if a certain synchronization in the firing of a layer's neurons provides useful information to the prediction.

By fitting on the shifted events, we observe that the RMSE is noticeably larger. For example, for both layers L4, L23, the error in control is close to 25%, while for the same data the normal prediction error is around 13%. This means that the inherent synchronization of the neurons indeed provides some useful information to the prediction. Notice that in this control prediction the use of both layers yields the biggest error, followed by L23 and L4 - exactly the opposite order to the initial prediction. This can be possibly explained by the randomness of the control prediction; if the prediction is basically random and does not utilize the information given by the data, more data means larger error.

Another control prediction we can create is fitting on the random shuffling of the target variable's elements. To create this control prediction we randomly shuffled the elements of the pupil size data on which the linear SVR was trained, and tested it with pupil size data that were not shuffled. This way, we trained the SVR while practically canceling any sense of history in the pupil size time series and any connection between a calcium event and its respective pupil size value. The resulting % RMSE was even higher than the previous control, ranging approximately from 22% to 24% across layers.

We also created a last control, the pupil shifting control, that is similar to the last one, in that it serves the purpose of changing the pupil time series elements so that the prediction ignores the actual connections between calcium events and pupil size values, by making randomly false ones. To create this control, we shifted circularly the pupil size data on which the linear SVR was trained and, again, tested it with non-shifted pupil data. The results give us clearly the highest error of both previous controls, across all layers.

After the fit, the difference between the prediction error and the control prediction error is noticeably large, approximately 10% or bigger, depending on the layer we refer to. Notice that for more dense, largely populated layers this difference is more noticeable: layer L4 gives us the smallest difference, followed by L23 which gives a bigger difference, ending up to the case of both layers where the difference is maximized. This leads us to the conclusion that layer density and population -the former more than the latter- are responsible for bigger differences between prediction and control prediction and hence, layers with these character-

istics provide more information to a prediction model.

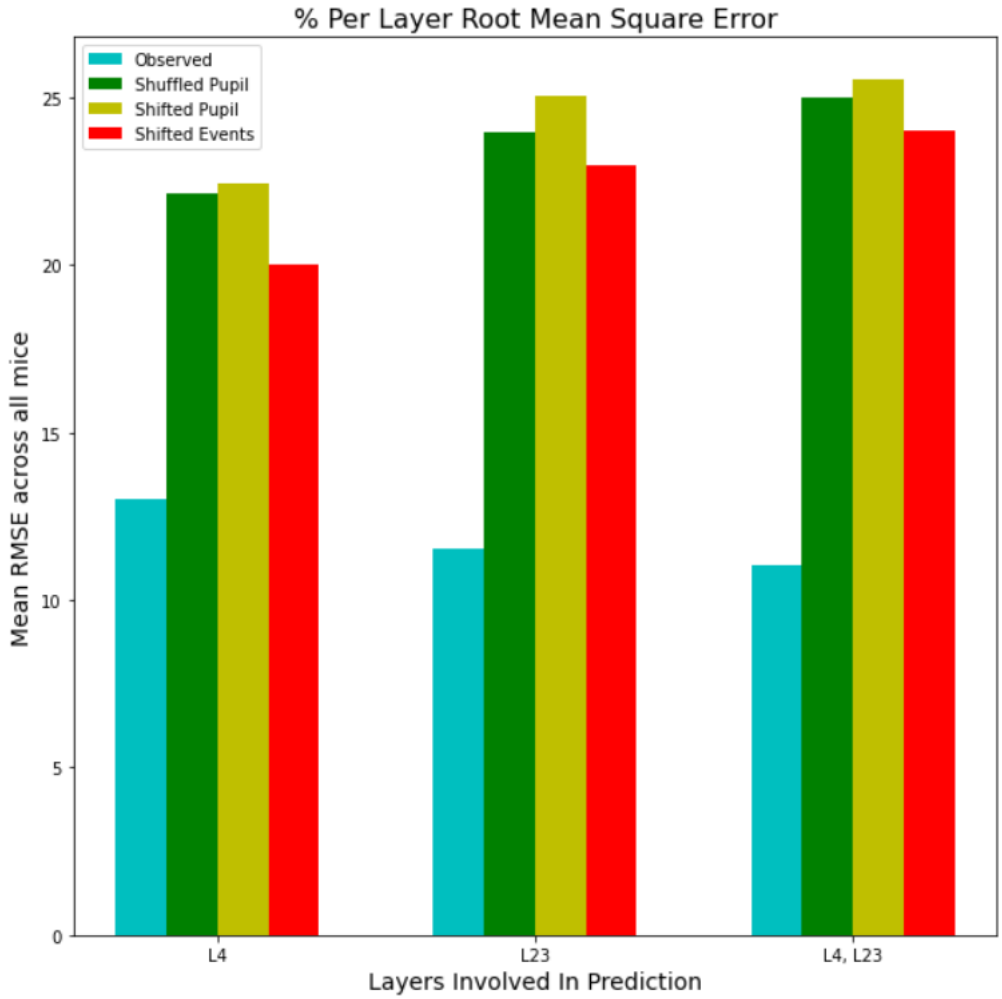

Figure 4.1: *x-axis: 3 different prediction cases, prediction involving layers L4, L23, and both. y-axis: values that represent the total % RMSE (averaged across all mice) in prediction when a particular layer is involved. The standard prediction error for each case is denoted with blue bars, the shuffled pupil control error with green, the shifted pupil control with yellow and the shifted events (de-synchronization of neurons) control with red.*

4.1.3 Neuron importance

Principal Component Analysis (PCA) is possibly the most known dimensionality reduction technique. To reduce the dimension of a dataset PCA captures and quantifies the contained information in each variable according to how much variance the variable is responsible for, and then projects the initial data to a low-dimensional space. In this context, a simple step that we could make in order to identify the main variables that regulate the cortical state

is to apply PCA on the calcium events of layers L23, L4, considering the activity of each neuron over time as a separate variable. By doing so, we will be able to extract exactly how much of the total variance in calcium events is *explained* by each variable we examine and also how is variance distributed over the variables.

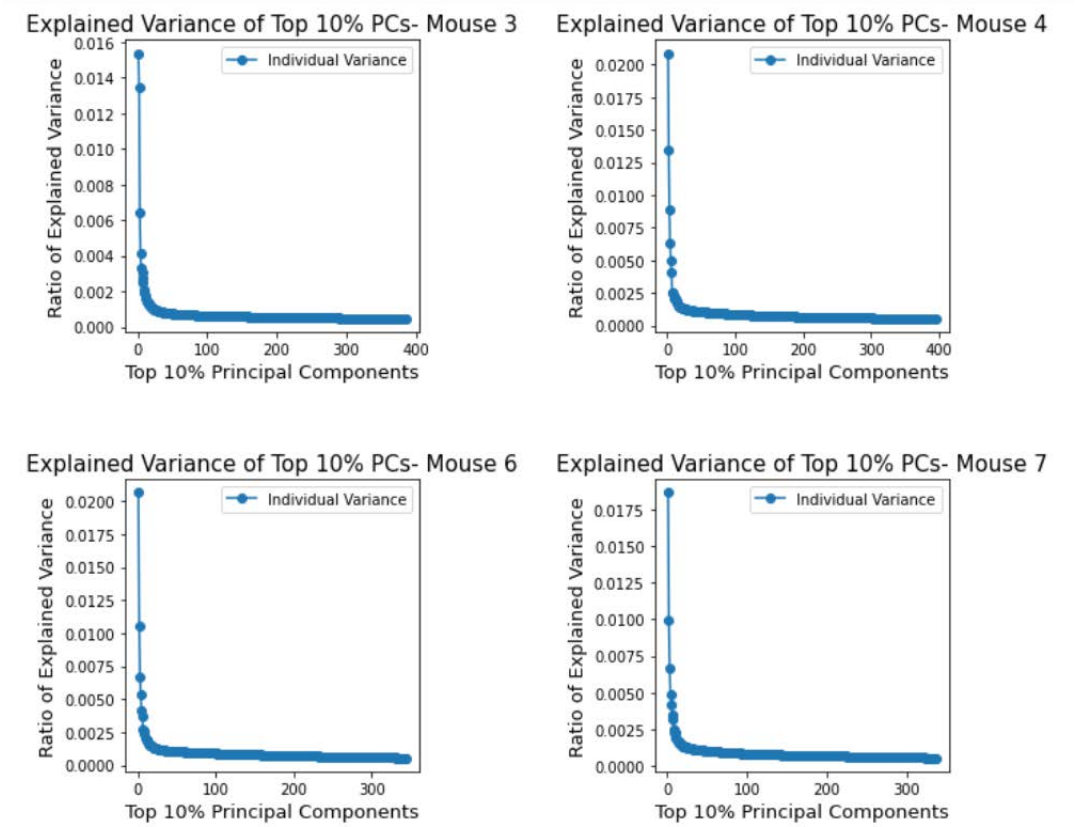

Figure 4.2: *Ratio of total variance explained by the top 10% most important variables (according to PCA) in spontaneous calcium activity of 4 different mice. y-axis: Variance percentage (1.0 is 100% of the variance). x-axis: Principal Components, in descending order of importance.*

The results suggest that explained variance in the spontaneous calcium events of the mice is accumulated in very specific variables for each mouse. This does not necessarily indicate that these variables have an 1-1 correspondence with neurons and their importance can be assigned to neurons. PCA traditionally creates a new space to which the initial data are to be projected, with a minimal loss of information.

A point in that new space, is not a vector that purely expresses neuron activity. However, this space represents information about the cortical activity, and these results indicate that its dimension can be surprisingly smaller than the initial one, with little loss of information. This means that cortical state, under spontaneous circumstances at least, can actually be

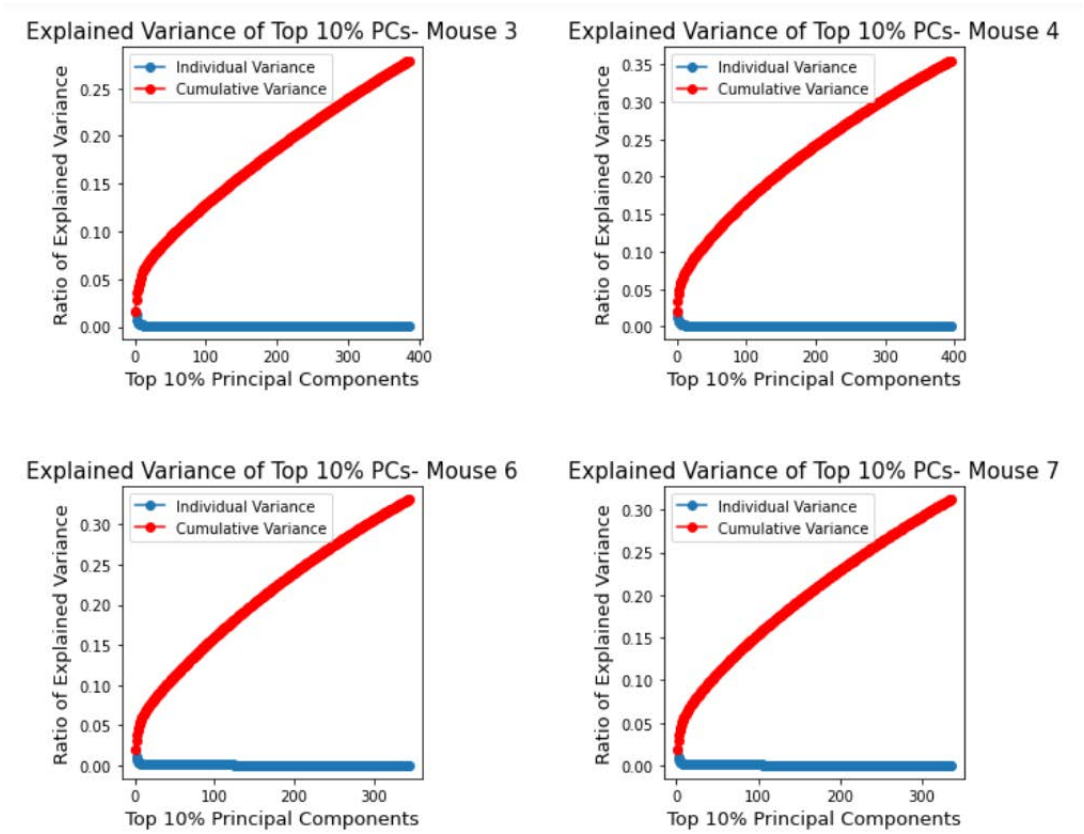

Figure 4.3: *Ratio of total variance explained by the top 10% most important variables (according to PCA) in spontaneous calcium activity of 4 different mice. y-axis (Blue): Variance percentage (1.0 is 100% of the variance). y-axis (Red): Cumulative variance, i.e. the summed variance of the top x variables, where x is the corresponding value in the x axis. x-axis: Principal Components, in descending order of importance.*

successfully represented in a low-dimensional space and its information can be significantly compressed. This result may remind us of previous work that seeks to find *latent representations* of neuron activity such as [34], where it is found that neuron population activity ”has a low-dimensional structure, reflecting underlying constraints on how neurons comodulate their activity”. Of course, for more useful information to be extracted, non-linear and most probably supervised methods have to be applied on individual and population activity, so that more complex relationships can be captured with greater certainty. Our analysis does not prove, but depicts, a very possible low-dimensional underlying structure in spontaneous calcium activity. [1]

[1] We should mention that the results seen here, although having been produced by my code, are not original, but part of a much more detailed Principal Component Analysis on calcium eventograms conducted by Manos Koniotakis, member of Prof. Papadopouli’s team.

4.1.4 History of the Pupil Size

Neuronal activity and behavioral characteristics such as pupil size present spatial and temporal locality when studied as signals [25]. This means that physical proximity (space) and simultaneous firing (time) in neurons increase correlations of these signals with themselves and reveal local relationships. Similarly, as described in [18], pupil size, as a time series, exhibits temporal dynamics, especially during dilation and contraction. This arises the question of how such temporal relationships could influence pupil size prediction. Therefore, we have to measure the importance of the history of the pupil time series and the extent to which it affects prediction. That is, we want to investigate whether the knowledge of previous measurements of the pupil time series can have a positive effect on the prediction of the present and future ones. Firstly, to begin our analysis on pupil time series history, we need a rough measure of how much and how fast pupil size values change with respect to previous values, to detect and quantify any local relationships across time in the time series. To measure this similarity between values at different moments, we calculated the autocorrelation function (ACF) of each series. The autocorrelation function is mathematically defined as:

$$R(\tau) = cor(X_t, X_{t+\tau})$$

where $cor()$ denotes the *Pearson* correlation between two variables, X our time series (pupil time series) as a stochastic process and X_t, $X_{t+\tau}$ its values at times t, $t+\tau$ accordingly.

For each pupil time series of each mouse we calculated and plotted the ACF for 100 steps and then calculated the average ACF across all mice for the same number of steps. The ACF calculates the Pearson correlation between the pupil time series and 100 of each shifted variants individually (each variant being the series shifted by $1, 2, ..., 100$). Note that the study of the pupil time series with autocorrelation makes the assumption that the given series is a random process, pupil size values are independent and identically distributed (*i.i.d.*) and pupil time series is wide sense stationary (*WSS*). The amount of shifting in the calculation of each individual correlation is referred to as “lag”.

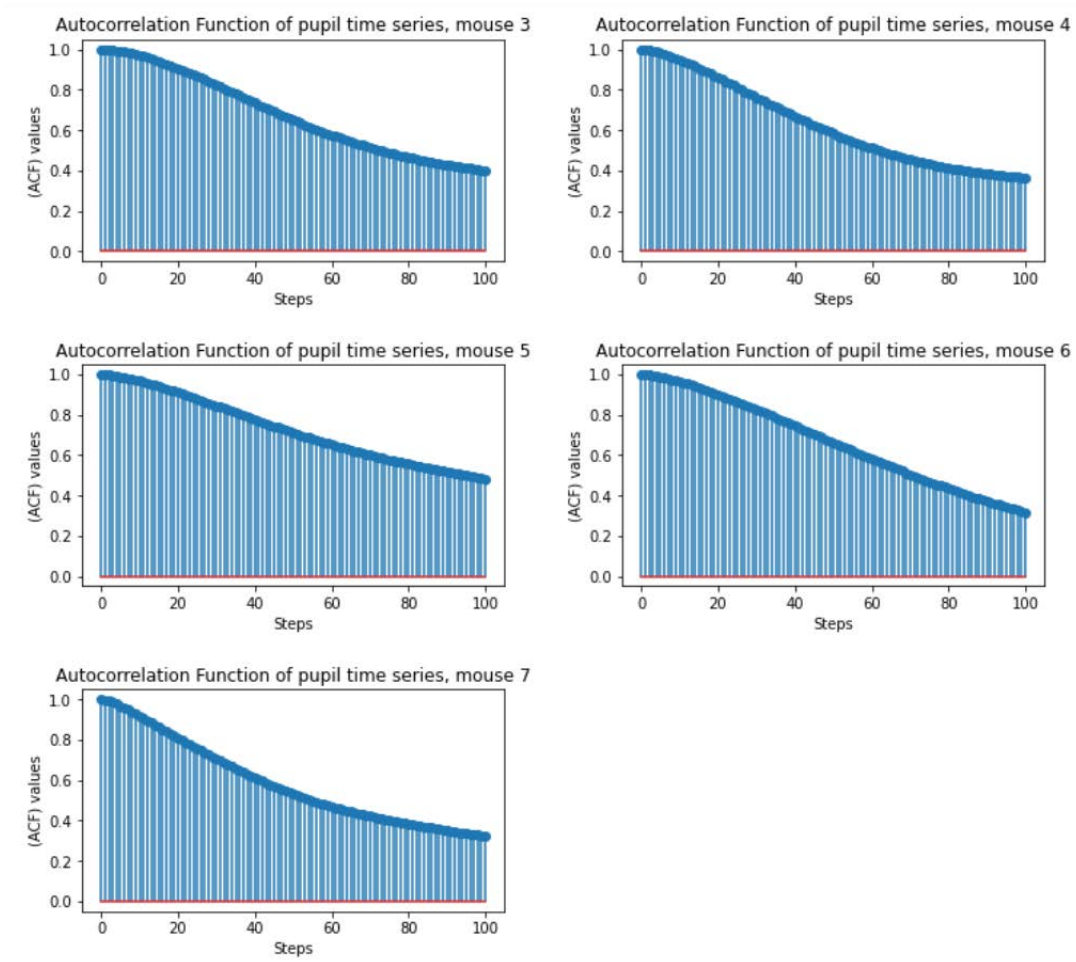

Figure 4.4: *(For mice 3-7): x-axis: lags, the amount of shifting in each correlation. We chose to plot the ACF for lags in [0,100] for all mice. y-axis: the ACF value, calculated at a certain lag.*

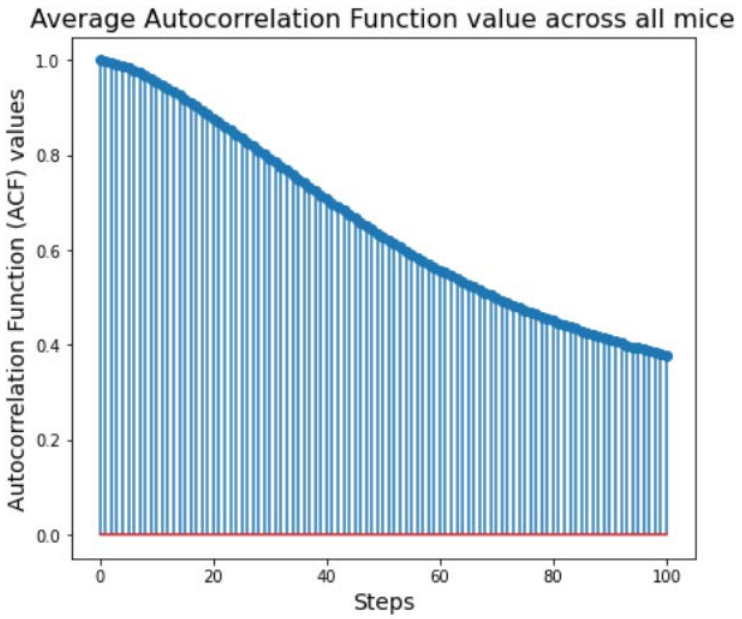

Figure 4.5: *(For all mice): x-axis: lags, the amount of shifting in each correlation. We chose to plot the ACF for lags in the interval [0,100] for all mice. y-axis: the average ACF value, calculated at a certain lag, across all 5 mice.*

Both from the ACF values individual to each mouse and the averaged ACF values we observe a significant autocorrelation that "lasts" for a noticeable number of lags. Specifically, for lags in the interval $[0, 20]$, the ACF is greater than 0.8. This indicates that, there is a window of 20 lags (or steps in the time series) in which any values are significantly highly correlated, statistically speaking. In terms of prediction, this creates a window of up to 20 past pupil size values that can possibly provide useful information for the estimation of the present, or a future pupil size value, and increase the prediction accuracy significantly.

Usually, the past of a time series can provide useful information for the prediction of future values, or for the estimation of the present one. In time series forecasting, regardless of the machine learning algorithm used for the problem at hand, it is a frequently used tactic to include past values of the quantity we want to predict into our data, as extra features; these features are widely known as *lag features*.

Applying the same rationale here, to test if pupil history is indeed important to the prediction, we included past pupil size values in the initial calcium events and performed the standard prediction and all the previously described controls. We also performed a prediction solely on pupil history data. In all cases, for the prediction of each pupil size value at time n, all values from $n-1$ to $n-10$ were available in the dataset as individual features, thus creating a window size of 10 in the pupil time series' past.

We notice a very significant drop in %RMSE: the new RMSE is approximately 1%, an order of magnitude smaller than the prediction that did not involve pupil size history. At the same time, the control prediction errors remain the same. This leads to even larger differences between standard and control prediction and reassures us of the prediction's quality.

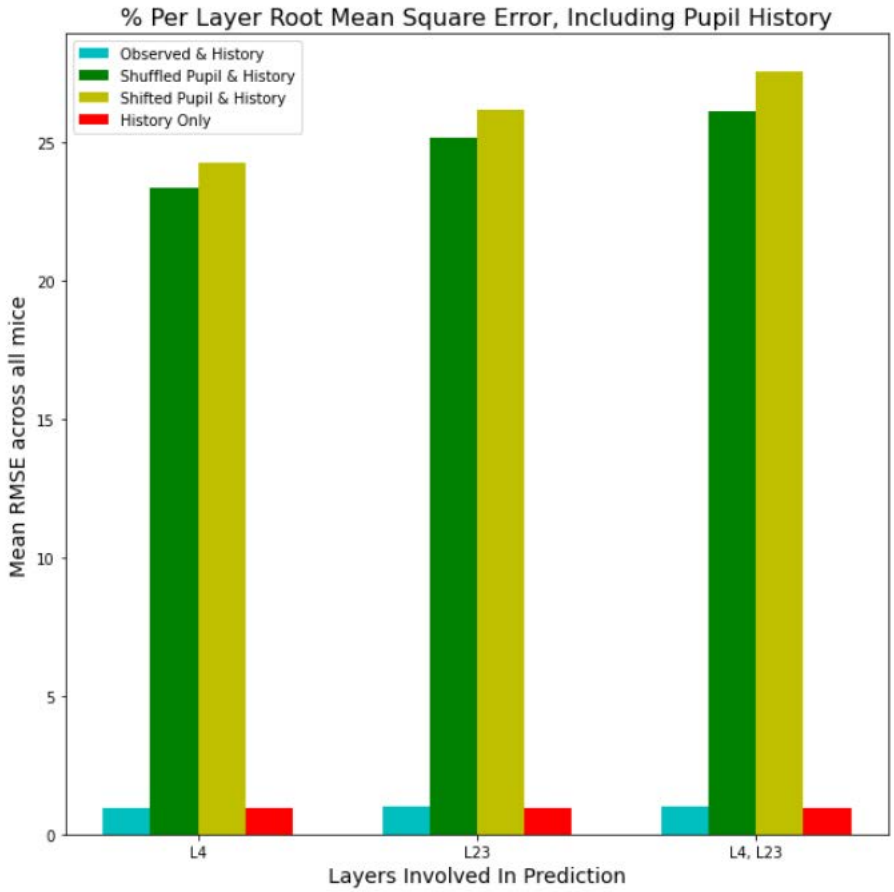

Figure 4.6: *x-axis: 3 different prediction cases, prediction involving layers L4, L23, and both. y-axis: values that represent the total % RMSE (averaged across all mice) in prediction when a particular layer is involved. The prediction that uses observed events and previous pupil size values is denoted with blue bars, the shuffled pupil control with green (same data), the shifted pupil control with yellow (same data), and the prediction that uses only previous pupil size values with red.*

Hence, pupil size history does have a significant impact on the prediction. Also note that the prediction error resulting from the case where only the pupil history is used is very similar to the one resulting from the case where both history and eventograms are used. This indicates that, in machine learning terms, pupil history is clearly more responsible for the error drop than the firing events themselves, and has such a big effect in prediction that it monopolizes it in a way.

The window size determines the number of past pupil measurements that are taken into consideration when predicting the present pupil size value. To understand if and how this number affects prediction accuracy, we considered 3 different window sizes: 5, 10, 20. For each window size, we performed the standard prediction (no controls) on layers L4, L23, and both, separately. This means that each prediction of a pupil size value at a certain timestamp is informed by the last 5, 10, or 20 pupil size values.

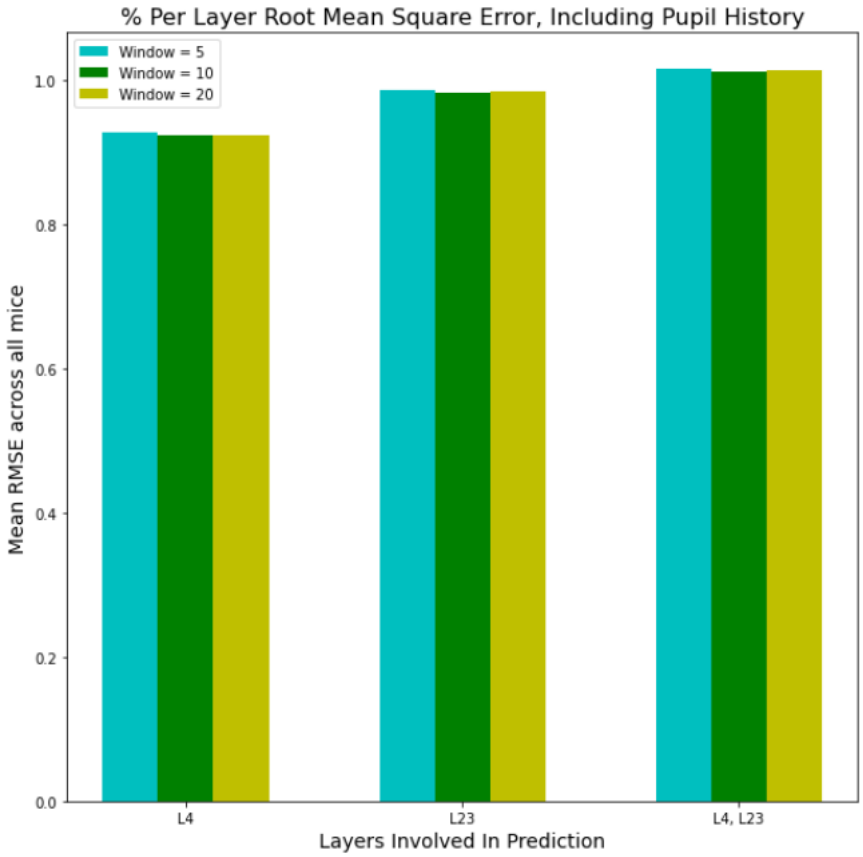

Figure 4.7: *x-axis: 3 different prediction cases, prediction involving layers L4, L23, and both. y-axis: values that represent the total % RMSE (averaged across all mice) in prediction when a particular layer is involved. The prediction that uses observed events and pupil history with a window size of 5, is denoted with blue bars, the one with window size = 10 with green and the one with window size = 20 with yellow.*

We observed no significant difference in prediction accuracy in any layer. Taking all layers into consideration, a window size of 10 gives the best results, followed by the window size = 20, followed by the window size = 5. However, the differences could be characterized as almost negligible, since they are less than 0.1%. We chose not to investigate further cases, since pupil size values further in the past can contribute even less to the prediction, as the autocorrelation function suggests.

4.1.5 Treadmill velocity history and prediction of future values

Behavioral characteristics span across many external reactions of a lab mouse that can be monitored with non-invasive methods, including whisking, running and moving, or the position and radius of the eye's pupil. Changes in these behaviors are often observed together, correlated and reflect changes in the firing rates of V1 neurons and thus in the cortical state [18]. Consequently, it will be interesting to investigate if a characteristic such as the history of the velocity of the treadmill that each mouse is on, can have a significant impact on the prediction. To do that, we included past treadmill velocity values in the initial calcium events and performed the standard prediction and all the previously described controls.

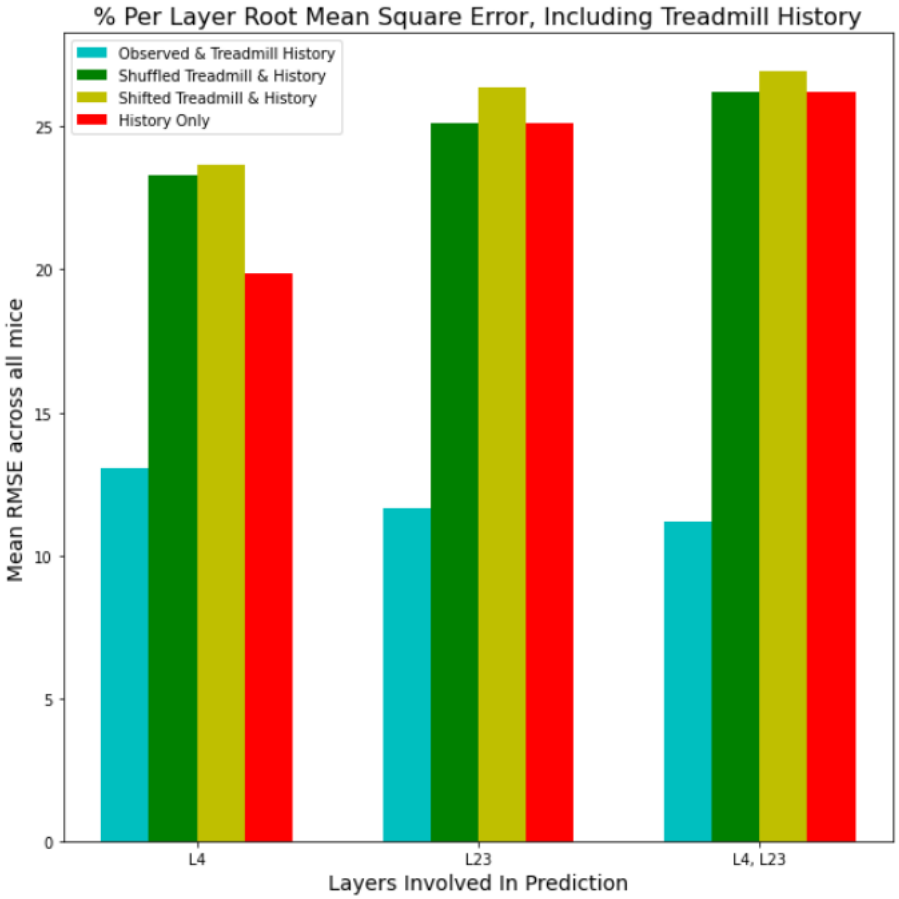

Figure 4.8: *x-axis: 3 different prediction cases, prediction involving layers L4, L23, and both. y-axis: values that represent the total % RMSE (averaged across all mice) in prediction when a particular layer is involved. The prediction that uses observed events and previous treadmill velocity values is denoted with blue bars, the shuffled pupil control with green (same data), the shifted pupil control with yellow (same data), and the prediction that uses only previous treadmill velocity with red.*

We also performed a prediction solely on treadmill history data. In all cases, for the prediction of each pupil size value at time n, all treadmill velocity values from $n-1$ to $n-10$ were available in the dataset, thus creating a window size of 10 in the treadmill time series' past. Note that each past time step is included as an extra feature and also that the velocity taken into account is absolute (there is no difference between a forwards or a backwards movement of the treadmill, just the speed at which that happens).

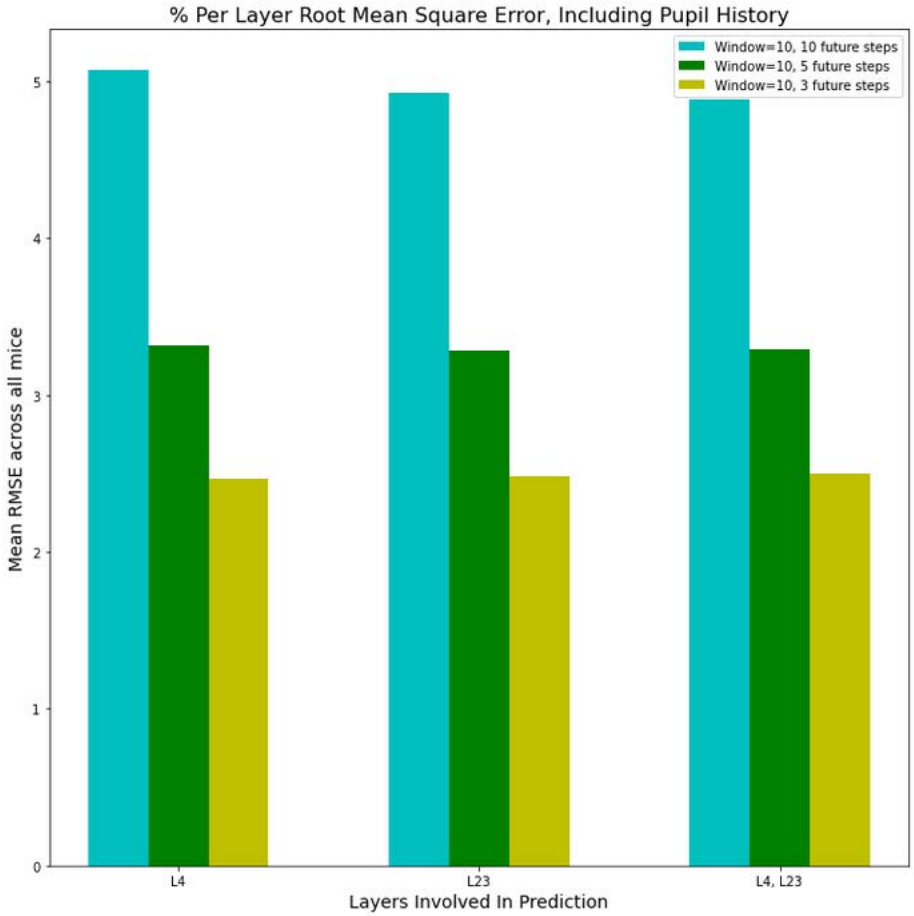

Figure 4.9: *Prediction of future values of the pupil size. x-axis: 3 different prediction cases, prediction involving layers L4, L23, and both. y-axis: values that represent the total % RMSE (averaged across all mice) in prediction when a particular layer is involved. The prediction that uses observed events and predicts pupil history 5 values ahead is denoted with blue bars, the one that predicts 10 values ahead with green and the one that predicts 20 values ahead with yellow.*

So far, we have tested the ability of our model to predict the present pupil size value, using adjacent past ones. Purpose of this section is to use the same window we have established in the previous sections, but use older values of the pupil history in that window, or differently put, see if a certain window of values can help us predict future values of the pupil time series.

To do that, we kept a window size of 10 that contains values 3,5 and 10 time steps before the present one.

As one would expect, the % RMSE in prediction drops as we shift the window towards the pupil's past. The decrease in the error seems to be more steep than a linear one, indicating that % RMSE is not simply a linear function of that shift, but a more complex one (e.g. a power law), that introduces bigger sensitivity with respect to the shift - although the plot above might seem misleading as for whether the relationship described is linear, the reader is a few simple calculations away from noticing that a change in the shifting does not yield proportional change in %RMSE.

4.1.6 Stimuli Data

So far, we have concerned ourselves with the prediction on spontaneous data, where the mouse in question received no visual stimulus. In order to study better the way neuronal activity encodes visual stimuli and the relationship between neuronal activity under stimuli and pupil size, we performed our prediction on stimuli data. To maintain the reference frame the spontaneous prediction provides, we performed our prediction on the stimuli data in exactly the same fashion, also carrying out the shuffled pupil time series and shifted pupil time series control predictions.

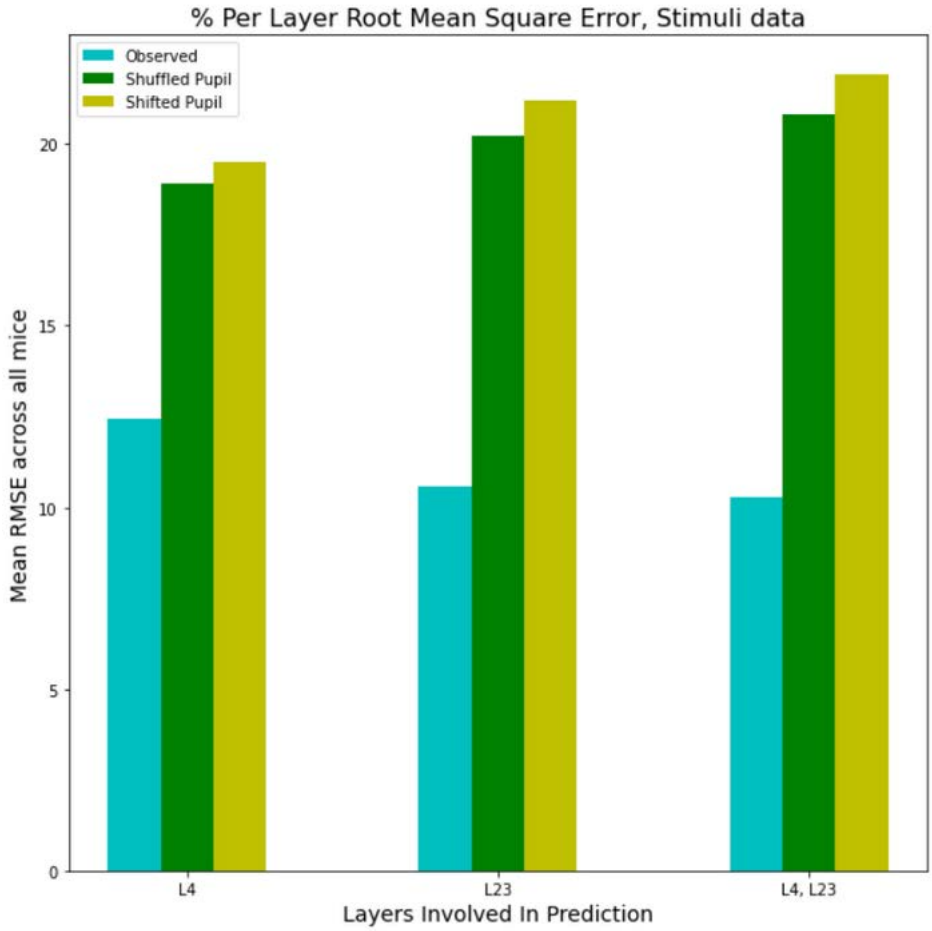

Figure 4.10: *x-axis: 3 different prediction cases, prediction involving layers L4, L23, and both. y-axis: values that represent the total % RMSE (averaged across all mice) in prediction when a particular layer is involved. The prediction that uses observed stimuli events to predict pupil size under stimulation is denoted with blue bars, the shuffled pupil control with green (same data) and the shifted pupil control with yellow (same data).*

We notice that the stimuli prediction does not differ significantly from the spontaneous prediction in terms of error. The prediction with observed data yields almost the same %RMSE on each layer. Similarly with the spontaneous prediction with the observed data, the lower

%RMSE is achieved when both layers L4, L23 are included and is approximately equal to 11%. The only noticeable difference with the spontaneous prediction is that the %RMSE in both controls is a little lower, around 20% on average; but although the error difference between control prediction and prediction with observed data quantifies in some way the amount of information our data provide to predict pupil size, this error gap remains large enough to give us a meaningful stimuli prediction. Now, in order to investigate if the inclusion of pupil size history plays an important role in stimuli prediction too, we included it in prediction, following the exact same process we followed with the spontaneous prediction and also ran the control predictions that are directly associated with pupil history: the shuffled pupil and shifted pupil controls.

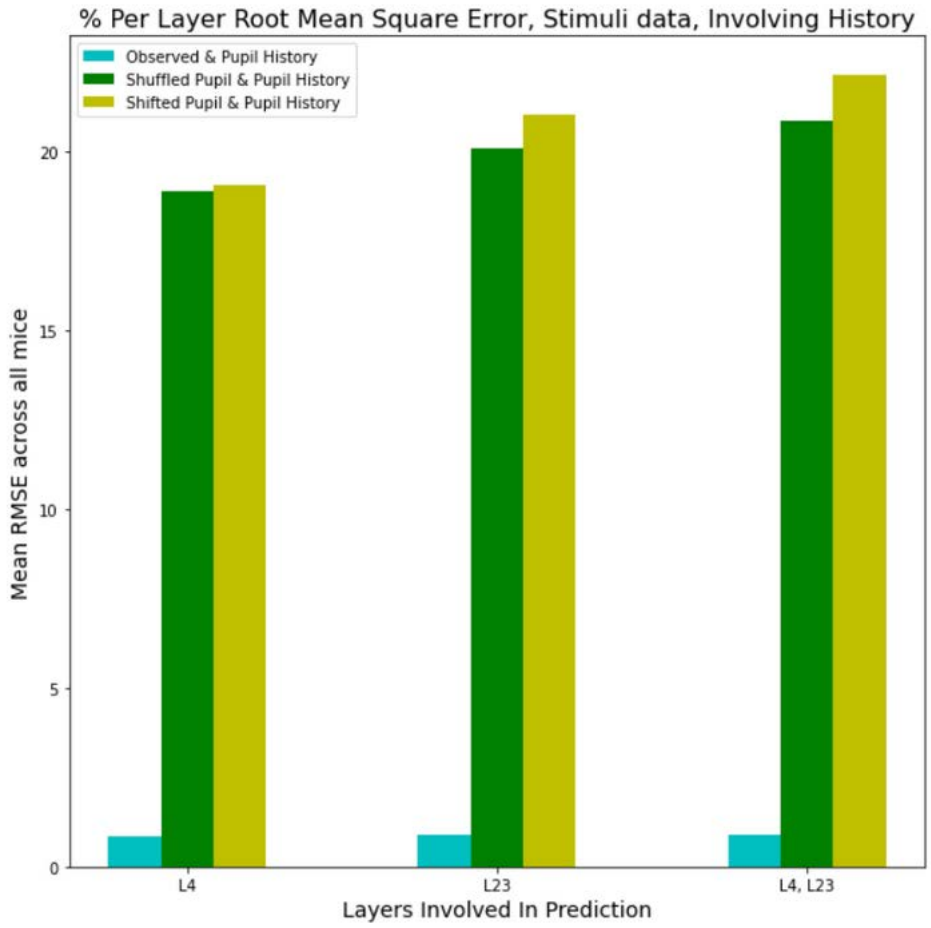

Figure 4.11: *x-axis: 3 different prediction cases, prediction involving layers L4, L23, and both. y-axis: values that represent the total %RMSE (averaged across all mice) in prediction when a particular layer is involved. The prediction that uses observed stimuli events and stimuli pupil history to predict pupil size under stimulation is denoted with blue bars, the shuffled pupil control with green (same data) and the shifted pupil control with yellow (same data).*

The plot above is very similar to the spontaneous prediction; the %RMSE we achieve is approximately 1% which indicates that the involvement of pupil history can drastically reduce the error, which in turn supports the argument that the pupil time series history can catalyze pupil size prediction with or without the presence of visual stimulus.

4.1.7 Firing Rate

Across literature, some common metrics are employed to measure, describe and study neuronal activity. A very commonly met metric, is *firing rate*. It is a basic measure of general neuronal activity and strongly connected with *cortical state*, as it is understood for spontaneous activity; in turn, cortical state itself has been correlated with changes in sensory responses [18][10]. Thus, since the use of firing rate can be a very useful tool in our analysis, not only because it can describe the cortical state of the primary visual cortex, but also because of its connection with behavioral responses, it is necessary that we include it in our study and see how it is related to our predictions.

In order to do that, we have to define firing rate, intuitively and mathematically. The firing rate, as expected, expresses the frequency at which a firing of a particular neuron is observed. This does not imply that the events of firing are of periodic nature but provides a simple, high-level statistic that gives us a measure of a neuron's activity during a recording period. We define the firing rate of a specific neuron algebraically as follows:

$$f = \frac{S}{N \cdot T_s}$$

S being the total number of events observed for a neuron during the recording period, N being the total number of frames in the eventogram and T_s being the sampling period (inverse to the sampling frequency, which for our data is $6.30072Hz$).

Our goal in this section is to understand the relationship between the prediction power -or neuron impact, which, as we mentioned before, quantifies the ”importance” of a neuron- and the firing rate of neurons (if any). To achieve that, we followed our standard prediction process, trained the linear kernel SVR on the eventogram of each mouse and, by using k-fold cross validation ($k = 12$) with no shuffling we assigned an impact to each neuron, equal to its average coefficient across all folds. We also calculated the firing rate of each neuron. We started by investigating the neurons of layer L4, for each mouse separately.

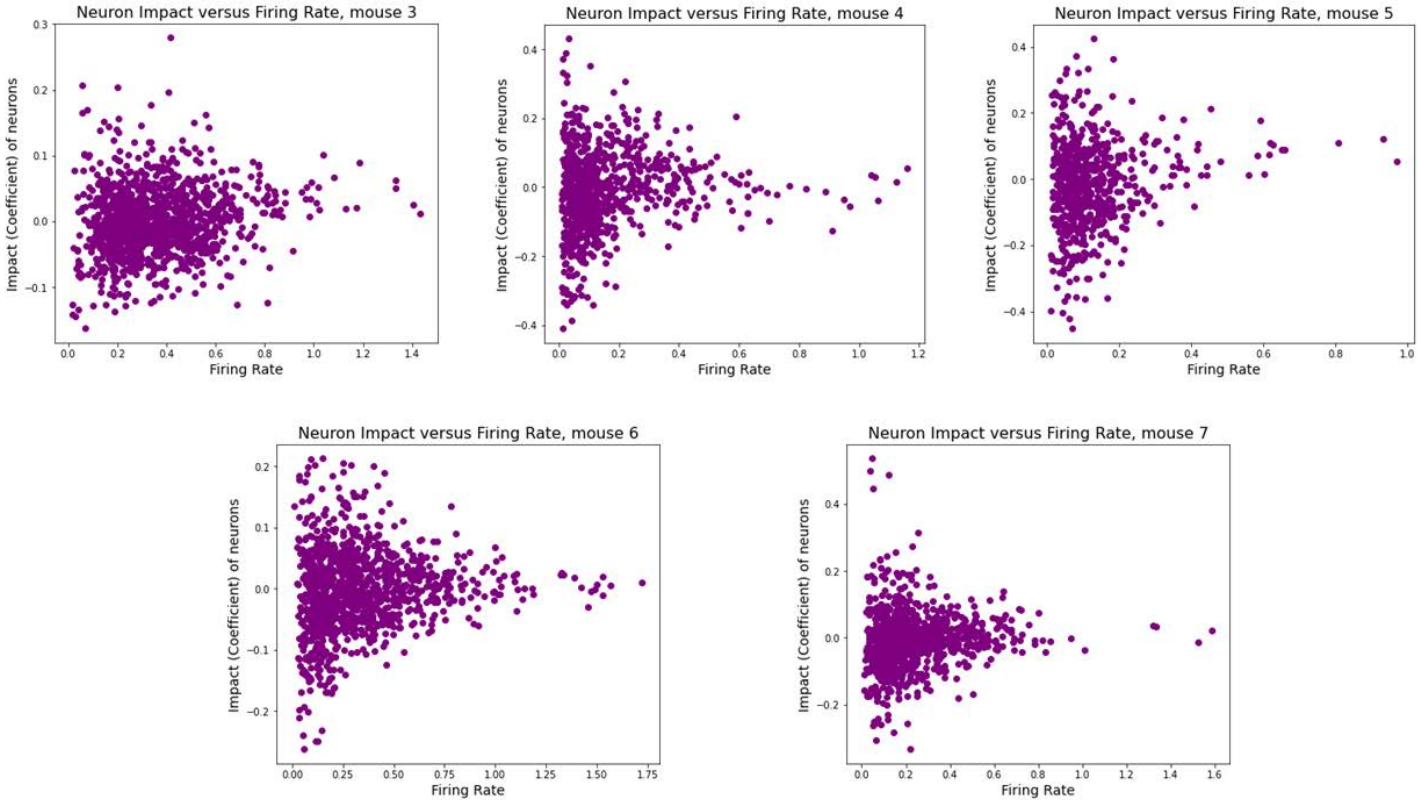

Figure 4.12: *x-axis: the firing rate of neurons in the L4 layer, for mice 3-7. y-axis: the mean prediction coefficient of each neuron of mice 3-7 obtained by the k-fold cross validation process; Each coefficient was averaged across all 12 folds. Coefficients are also referred to as impacts, that is they express the importance of a neuron in prediction.*

From the scatterplots of the majority of the mice, we can detect a pattern in the distributions of impacts across different values of firing rates. We notice that the vast majority of neurons with big impacts seem to have a low firing rate - please note that a “big impact” can be a strongly negative coefficient, meaning that it is the absolute value of the coefficient that

determines an impactful neuron. This result indicates that, roughly, the pupil size is modulated mainly by neurons that fire rarely. This pattern is also supported by the results in layer L23. By following the same training rationale, we get the following results:

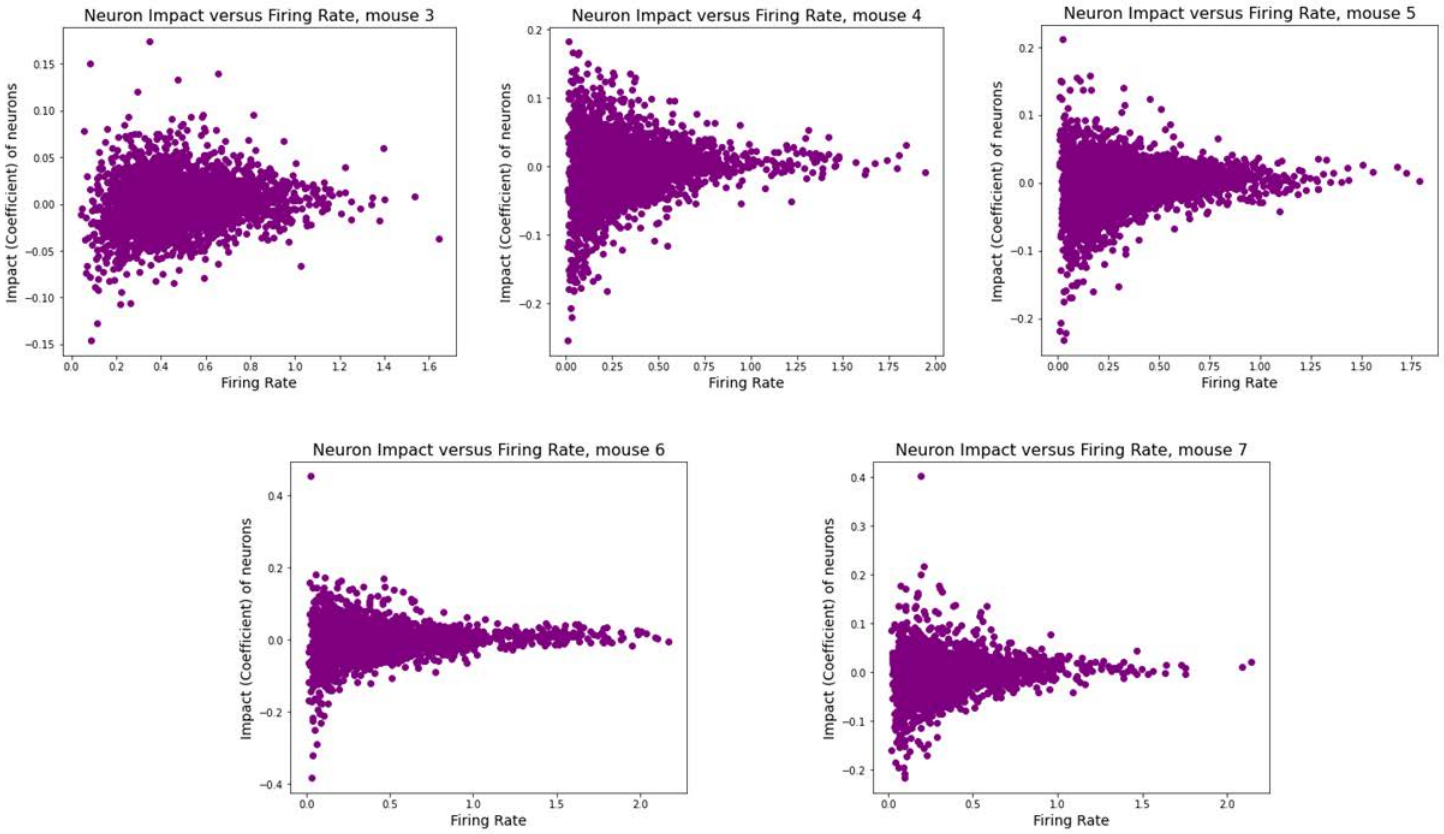

Figure 4.13: *x-axis: the firing rate of neurons in the L23 layer, for mice 3-7. y-axis: the mean prediction coefficient of each neuron of mice 3-7 obtained by the k-fold cross validation process; Each coefficient was averaged across all 12 folds. Coefficients are also referred to as impacts, that is they express the importance of a neuron in prediction.*

4.2 Algorithms for the Stimulus Orientation/Direction

During certain periods, the five mice examined in this work are exposed to visual stimulus in the form of video that features certain patterns moving towards various orientations. In order to understand how this visual stimulus is represented by the activity of the neurons, we first need to find a deterministic method, an automated way to calculate quantities that express this movement of the stimulus with numbers and label each different kind of movement orientation. The following section describes our thought process in detecting the changing angle of the moving patterns of the stimulus, provides a mathematical intuition on the main algorithms used for that purpose and a conclusion to which one is the best.

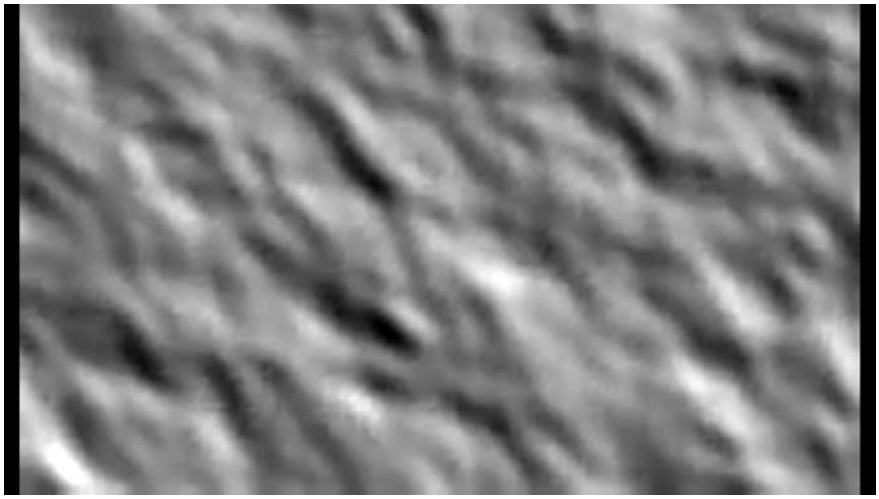

Figure 4.14: *Example frame from the visual stimulus that was used on the mice. It features a wave-like movement of the 2D plane in greyscale. Each mouse was shown a total of 240 videos that lasted individually around 15 seconds, accumulating to 1 hour of stimulus in total. Each video comprises 900 frames and 16 segments, each segment defining a period in the video where the direction of movement is held constant. Each segment shows a different direction. The frames were produced by Prof. Andreas Tolias's team (Andreas Tolias Lab).*

4.2.1 The Farnebäck Algorithm

At first, in an attempt to find a suitable computer vision algorithm that captures the moving patterns' angle of direction and magnitude, we narrowed our research down to *optical flow* methods [35]. Basically, optical flow quantifies the frame-to-frame changes in terms of displacement and direction (angle) of displacement of particular points of interest[2] in the 2D

[2]"Points of interest" is used to denote objects, or whatever an algorithm perceives as an object, or part of an object using certain criteria (edges, pattern complexity etc.), features which are considered as key areas for the

plane. It is the pattern of apparent motion of these points, or mathematically, a vector field that contains their displacement vectors. Given this information, and the wave-like motions we wanted to detect in the stimulus, we considered optical flow was a good place to start.

We are now going to briefly analyze the theoretical context optical flow methods that perform *gradient-based* motion estimation[3].

In these methods the information of the video is described by a function $I(x, y, t)$.I expresses the total brightness, in greyscale in our case, x, y the coordinates of an object in the 2D plane and t, time, or better yet, the frame. In our method two assumptions are made: firstly, that neighboring pixels have similar motion and secondly we make the *brightness constancy assumption*, i.e. we assume that the brightness of a moving object does not change across frames. Mathematically:

$$I(x + u, y + v, t + 1) = I(x, y, t).$$

According to Taylor's theorem, we can say that for a continuous, n times differentiable function f in $[x, x + h]$, we can roughly make the following approximation:

$$f(x + h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + ... =$$

$$f(x) + hf'(x) + O(h^2) \approx$$

$$f(x) + hf'(x)$$

So we apply the same rationale to I to get a multi-variable version of this first-order approximation:

$$I(x + u, y + v, t + 1) \approx$$

$$I(x, y, t) + u\frac{\partial I}{\partial x} + v\frac{\partial I}{\partial y} + \frac{\partial I}{\partial t} =$$

$$I(x, y, t) + uI_x + vI_y + I_t,$$

I_x, I_y, I_t being partial derivatives of function I w.r.t. its variables.

detection or measurement of various properties.

[3]Process that uses the discrete gradients of the image to extract information on motion.

From the first-order approximation and the brightness constancy assumption we get:

$$I(x, y, t) + uI_x + vI_y + I_t = 0$$

The above equation is called the gradient constraint equation, or the Optical Flow equation for pixel (x, y). Note that we cannot solve this equation for just one pixel, because both u and v are unknown.

Next, we apply what is called *Area-Based Regression*. Suppose that the gradient constraint equation holds for an area of pixels. For each pixel the equation is different. For 2 pixels we can express the equations in matrix form:

$$\begin{bmatrix} I_x(x_1, y_1, t) & I_y(x_1, y_1, t) \\ I_x(x_2, y_2, t) & I_y(x_2, y_2, t) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} + \begin{bmatrix} I_t(x_1, y_1, t) \\ I_t(x_2, y_2, t) \end{bmatrix} = 0$$

Especially when we use more than 2 pixels, we create an over-determined linear system - we cannot satisfy exactly all constraints. It is more sensible to define a metric to minimize, in an attempt to equally satisfy (to the extent possible) all constraints - of course, special weighting can be performed to prioritize certain constraints. That said, we define the least squares velocity estimate:

$$E(u, v) = \sum_{x,y} \left[uI_x(x, y, t) + vI_y(x, y, t) + I_t(x, y, t)\right]^2$$

To minimize it, we differentiate and set to zero:

$$\nabla E = 0 \implies$$
$$\frac{\partial E}{\partial u} = 2 \sum_{x,y} \left[uI_x^2 + vI_xI_y + I_xI_t\right] = 0,$$
$$\frac{\partial E}{\partial v} = 2 \sum_{x,y} \left[vI_y^2 + uI_xI_y + I_yI_t\right] = 0$$

In a matrix form, this is (solving for u, v):

$$\begin{bmatrix} \sum_i {I_{x_i}}^2 & \sum_i I_{x_i} I_{y_i} \\ \sum_i I_{x_i} I_{y_i} & \sum_i {I_{y_i}}^2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} + \begin{bmatrix} \sum_i I_{x_i} I_{t_i} \\ \sum_i I_{y_i} I_{t_i} \end{bmatrix} \implies$$

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum_i {I_{x_i}}^2 & \sum_i I_{x_i} I_{y_i} \\ \sum_i I_{x_i} I_{y_i} & \sum_i {I_{y_i}}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i I_{x_i} I_{t_i} \\ -\sum_i I_{y_i} I_{t_i} \end{bmatrix},$$

i denoting a pixel. We have now obtained a general solution for the motion estimation problem. If we use the above method for a 3×3 patch around the point of interest, we will end up with 9 pixels, and the same number of equations. This is known as the *Lukas-Kanade* method [35]. Lukas-Kanade provides a solution using a small area of interest, computes optical flow for a sparse feature set, finding what is called *Sparse Optical Flow*.

For a more sophisticated approach, we need to find the *Dense Optical Flow* of the video. A popular and effective optical flow method to do that is the Farnebäck algorithm [36]. Farnebäck uses a more dense approach, meaning that the process we described above is performed on all or most areas of a frame. Also, Farnebäck uses a *quadratic* Taylor polynomial approximation, meaning that the first-order approximation seen in the previous steps is replaced with a second-order approximation for more accuracy.

Of course, to proceed and use the algorithm, the next step was to validate its functionality. We tested it on a sufficient number of video segments of different directions with the purpose of extracting their angle of movement and validate the results. After this validation process, the algorithm appeared to have a particular quirk: despite the fact that most directions were detected flawlessly with minimal or no tuning, we noticed that it is not able to detect angles $< 47^\circ$ and $> 314^\circ$ (degrees) correctly, and labels them inaccurately.

Problems such as determining the direction of movement in a particular video of certain morphology are usually ill-defined, meaning that there is no proof for the existence of a unique solution to the problem, or any ”key and lock” relationships between problems and solutions, but only purpose-specific methods to solve the problems. In our case, despite the algorithm’s performance in all angle intervals besides the one referred above, we had to consider it unsuitable, as it would definitely cost a significant amount of accuracy later on.

4.2.2 Manual Methods

After the failure of the Farnebäck algorithm we decided to develop a more ”manual” process to identify the direction of the stimulus at a given instant. First of all, it should be mentioned that in this method, we make a clear distinction between *orientation* and *direction*. As orientation we perceive the waves' front, the axis they move against (perpendicularly) to and is expressed by an angle (w.r.t. the horizontal axis) between $0°$ and $180°$. Direction on the other hand, is perceived as the axis the stimulus moves along, while also depicting which way it moves, from frame to frame. It is expressed by an angle (w.r.t. the horizontal axis) between $0°$ and $360°$. Note that orientation is always perpendicular to direction. Also note that direction is not static information, it is a pattern extracted by the comparison of two consecutive frames, as opposed to orientation, which can be understood by looking at a single frame.

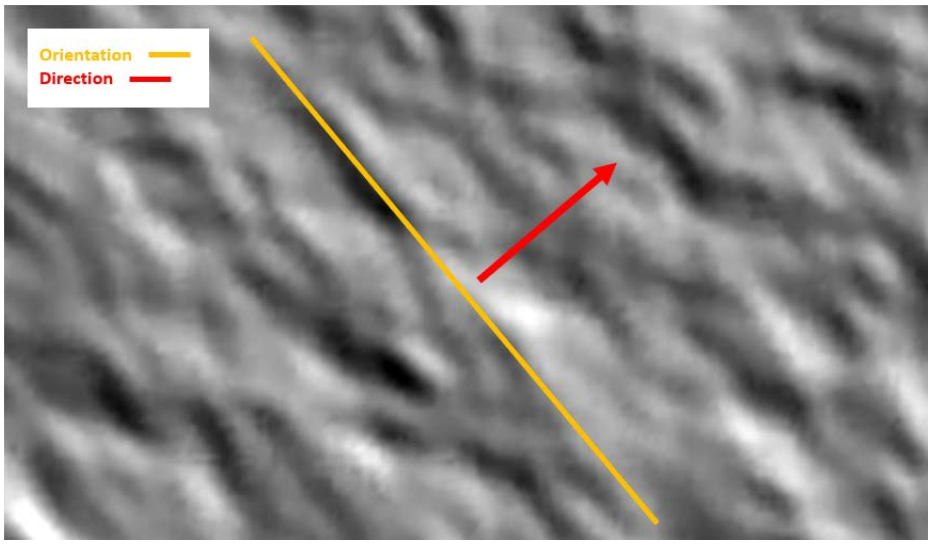

Figure 4.15: *Example that clarifies the difference between orientation and direction of a frame. The frames were produced by Prof. Andreas Tolias's team (Andreas Tolias Lab).*

The first step of our manual method to detect orientation and direction, is to perform *edge detection* on every given frame. Intuitively, as edges of an image are considered all the areas were sudden changes in brightness occur. For that purpose, we employed the Canny Algorithm [37]. In Canny edge detection, the abrupt changes in brightness are located with the use of the discrete partial derivatives of the image. In this model, the mathematical representation

of the image is the same as before, with the difference that we wipe out the variable of time. So we use $I(x, y)$ to represent the image and $I_{i,j}$ to represent the pixel at position i, j. That said, the partial derivatives of the image are expressed as:

$$\frac{\partial I}{\partial x} \approx \frac{1}{2\epsilon}((I_{i+1,j+1} - I_{i,j+1}) + (I_{i+1,j} - I_{i,j}))$$
$$\frac{\partial I}{\partial y} \approx \frac{1}{2\epsilon}((I_{i+1,j+1} - I_{i+1,j}) + (I_{i,j+1} - I_{i,j}))$$

ϵ being the pixel width. Note that this is actually an approximation; the expressions seen above are also known as Finite Difference Approximators. This differentiation is usually computationally implemented by a 2D convolution between the image and a filter kernel. Although a study of those filters is outside the scope of this analysis, we will however mention that we use a Sobel filter.

The norm [4] of the gradient of an image, $||\nabla I(x, y)||$ quantifies how abrupt the changes in brightness are. With some appropriate thresholding, we can detect the image areas where it is high enough, so as to detect the location of the edges, that is, the maxima of $||\nabla I(x, y)||$ indicate edges. In a similar way, the *Laplacian* $\nabla^2 I(x, y)$ of the image is also used for the detection of edges. Its finite difference approximations are:

$$\frac{\partial^2 I}{\partial x^2} \approx \frac{1}{\epsilon^2}(I_{i-1,j} - 2I_{i,j} + I_{i+1,j})$$
$$\frac{\partial^2 I}{\partial y^2} \approx \frac{1}{\epsilon^2}(I_{i,j-1} - 2I_{i,j} + I_{i,j+1})$$

so that the Laplacian is expressed as:

$$\nabla^2 I = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$$

The zero crossings of the Laplacian, and not its maxima, help us locate an edge. The simple gradient can provide the location, the magnitude and the orientation of an edge, while the Laplacian can only locate an edge, but more accurately and efficiently. The Canny algorithm combines both quantities to yield optimal results.

[4] in most applications it is the Euclidean norm $||.||_2$

The first step of the algorithm is to convolute the image with a 2D Gaussian filter for smoothing and noise reduction, to obtain $n_\sigma * I$, n_σ being the filter of standard deviation σ, and $*$ denoting 2D convolution. Then the gradient of the image is obtained, $\nabla n_\sigma * I$ and its norm is computed for every pixel. After that, the orientation of the gradient, $\hat{n} = \frac{\nabla n_\sigma * I}{||\nabla n_\sigma * I||}$, is obtained and the Laplacian is computed across $\hat{n}$, for every pixel. The points were the Laplacian becomes zero, give us the edges.

This concludes our brief theoretical explanation on the first step of our method, edge detection. Of course, all computer vision-related processes described are carried out automatically with the use of the OpenCV library. The next step is to convert the resulting image to a binary image, where each pixel can have only two possible brightness values, maximum brightness value (white) and minimum brightness value (black).

Next, we apply the Hough transform [38] on the binary image. It is a widely used transform in boundary detection applications, but we utilize its properties for a slightly different purpose. What the Hough transform does, essentially, is to *fit* many small, elementary lines (or polynomials in the general case), to the extracted edges. The representation of the line though, is not cartesian ($ax + by + c = 0$), but *polar*:

$$xsin\theta + ycos\theta + \rho = 0$$

This automatically projects the line from the coordinate system of a, b, c into the polar parameter space of $\rho, \theta \in \mathbb{R}$. The fitting process attempts to minimize the root mean square distance between the elementary lines and the actual edges:

$$E = \frac{1}{N} \sum_i (x_i sin\theta + y_i cos\theta + \rho)^2$$

N being the total number of points and i the point index. As also seen in Farnebäck, this problem naturally becomes a Least Squares problem for a great number of points. Of course, this optimization problem becomes enriched with the tradeoff between *elasticity* and *smoothness*, i.e. how flexible the total contour line[5] is in fitting and how much it "resists" in exhibiting discontinuities and roughness, accordingly. In a good solution, these are two constraints that have to be equally satisfied, yet they cannot be fully satisfied at the same time.

[5]The total shape made of all the elementary lines we mentioned that fits onto the edges.

By obtaining a polar representation of the video's edges, and particularly their θ parameter, we can infer their orientation. Of course, different elementary lines fitted on different edges have slightly different orientation, and for that reason we average all the calculated θ parameters into one angle $\bar{\theta}$, which gives us the total, average orientation of the stimulus at a given frame.

Given the orientations of all frames, we can determine the direction towards which the stimulus moves by comparing consecutive frames of a segment. This "comparison" is achieved with the orientation-dependent subtraction of the two consecutive frames. If the orientation of the two consecutive frames[6] is close to horizontal (a decision made with thresholding), we calculate the horizontal differences between the frames; this is done by calculating two quantities, the difference between the first frame and an the second frame shifted upwards (by one pixel is our default amount) and the difference between the first frame and an the second frame shifted downwards. These differences basically represent the two main possible directions of the relative movement between the two frames and can help us estimate that direction, if we also consider that orientation and direction always form a $90°$ angle. The following algorithmic scheme sums up our method:

Algorithm 1 Orientation and Direction Detection Scheme

```
Define Canny edge detection parameters
Define Hough transform parameters
for Every frame in a video do
    Convert frame to grayscale
    Apply Canny edge detection
    Convert edge image to binary image
    Apply Hough transform to binary image, extract θ̄ (Orientation)
    if θ̄ ∈ [45°, 135°](horizontal) then
        Calculate vertical differences (upwards, downwards)
    else
        Calculate horizontal differences (right, left)
    end if
    Infer frame direction from differences
end for
```

[6] assuming they have the same orientation because they belong to the same segment

Of course, the algorithm works accordingly for the vertical orientation case, where the differences between the first frame and an the second frame shifted to the right/left are computed. The present method provides a good estimation on orientation and on direction without exhibiting problems in certain degree regions, as the Farnebäck algorithm unfortunately did. Thus, we considered it a reliable tool to quantify orientation and direction in the stimulus, that requires minimal or no tuning.

4.3 Neuronal Orientation and Direction Preference

In 1968, Hubel and Wiesel [1] demonstrated that cells in the primary visual cortex of macaques and spider monkeys exhibit selectivity regarding the orientation and direction of visual stimuli and, according to the different levels of selectivity observed in neurons, they classified them into three main categories: *simple, complex* and *hypercomplex cells*. Since this study handles neurons from V1, this implies that many of them are orientation-selective or direction-selective and consequently have an increased firing response to specific stimulus orientations/directions, and a very constrained one to every other orientation/direction.

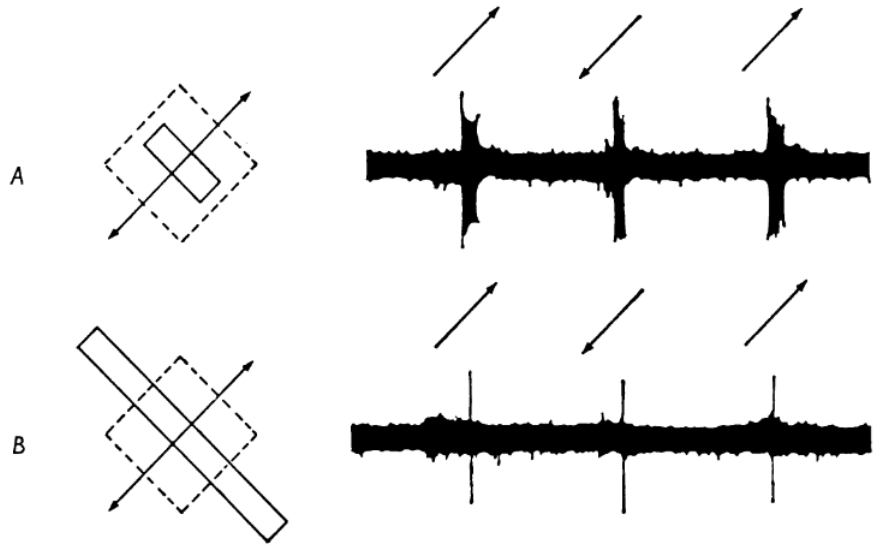

Figure 4.16: *From [1]. Hubel and Wiesel tracked the responses of neurons to the varying orientation, thickness, length of a line stimulus (whether it was a slit of light, a dark bar or the edge if a thick bar, within an "activating region"). Here the action potentials of a hypercomplex cell (recorded from right striate cortex, layer L2) are demonstrated. A: stimulus of left eye by moving slit within activating region. B: Similar stimulation with slit extending beyond activating region.*

This concept, among others, is a motivation for us to investigate the representation of visual stimuli in the visual cortex by examining the patterning in firing under various orientations and directions of the stimulus. As it has been observed on the auditory cortex of mice [6] (which can possibly be generalized to all somatosensory cortices), neurons exhibit a consistent *"temporal patterning"* in both spontaneous and stimulated states, at the same time that firing rates vary, depending on the state. This means, that the information of a stimulus is

most probably represented in the cortex by the variations of firing rates it produces, and not by the relative timing of neuron events. Therefore, it is more wise for us to express neuronal activity with firing rates per each stimulus, than to use raw events, from now on.

In an attempt to reveal how well the frequency of firing represents visual stimulus in layers L23, L4, we predicted the direction of the stimulus at a given time using firing rates, calculated for each neuron and each segment separately. The problem of direction prediction was modeled as a classification task: by splitting the interval $[0°, 360°)$ into equally-sized bins of various lengths, we modeled the problem as a multi-class classification, where each bin represents a class. For better data representation, the bins had to be *zero-centered*, i.e. the first bin had to be from $-L/2$ degrees to $L/2$ degrees where L is the bin width, so that the algorithm captures the cyclicity of the data in a simple way - otherwise, $359°$ would be considered far from $0°$, which is not.

To choose an algorithm, we compared Support Vector Machines with ensemble methods, and more particularly a Bagging Classifier versus a Support Vector Classifier. The estimator of the Bagging Classifier was the decision tree (i.e. the Bagging Classifier behaved as a random forest, basically), the SVM's kernel was a radial basis function kernel, and the comparison was performed on a $90°$ binning. The training process was similar to the prediction of the pupil, the models were trained with k-fold cross validation (k=12) and fitted on each fold separately. The metric used for evaluation was classification accuracy, and total accuracy of a prediction in one mouse was calculated as the mean accuracy of all folds.

The baseline model (control) which helps us assess all predictions, is equivalent to the assignment of a random class to each entry. The classes, though, are not drawn from a uniform distribution, but from the empirical distribution that they actually follow in the dataset - this way, we created a more strict control prediction for this classification task. To the author's surprise, the SVM with the rbf kernel exhibited a much better performance than the Bagging Classifier, with the SVM achieving an accuracy of approximately 95% and the Bagging Classifier an accuracy of approximately 85%. Thus, we chose to use the SVM for the prediction.

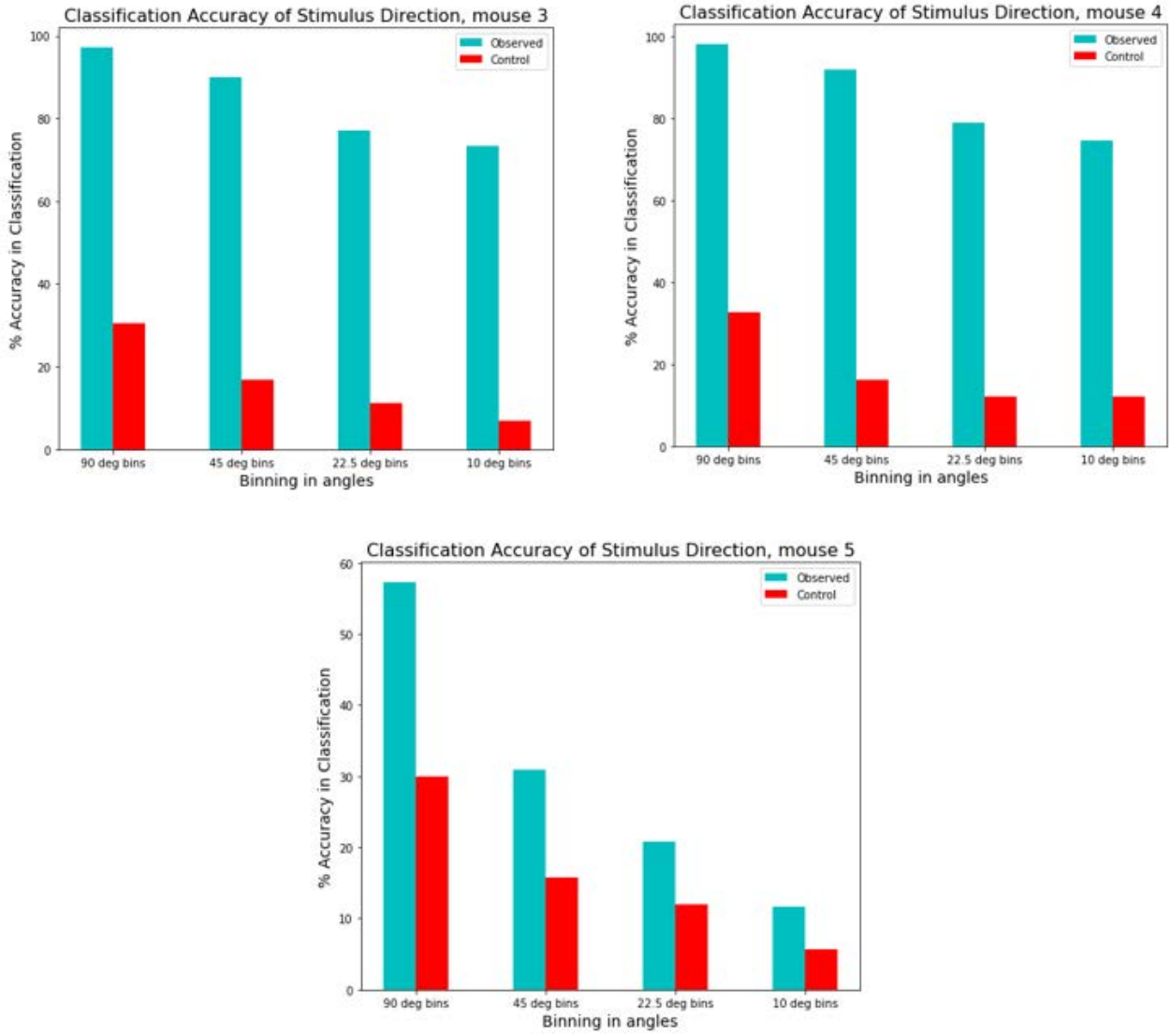

Figure 4.17: *Classification of stimulus direction for 3 different mice. y-axis: Prediction accuracy (blue) and control prediction accuracy (red). x-axis: Bin width, 4 different cases:* $90°, 45°, 22.5°$ *and* $10°$.

The same training rationale was applied to fit the radial basis function SVM on 3 different mice. Various binnings were examined in the fit, including $90°-, 45°-, 22.5°-$ and $10°-$ degree bins. For mice 3 and 4 the prediction accuracy as well as its distance from the control prediction were satisfying. Prediction in mouse 5 did not bring such an optimistic result; further searching has to be done in order to grasp why mouse 5 exhibits this outlying behavior, why it is confined to mouse 5 and if it is due to characteristics of the experimental data gathering procedure. Whatever the case might be, it gives us a ”floor” for the worst-case prediction on a single mouse.

To continue our analysis and investigate how orientation/direction preference is mani-

fested in neuron activity, we had to connect neurons' firing rates with their respective direction preferences. To do so, we expressed direction preference in zero-centered bins of $22.5°$, dividing the $360°$ to 16 equal, centered parts. For each different direction preference bin (that is, for all neurons that "prefer" a direction that falls into that bin), we calculated the firing rate per segment of each individual neuron and found the average firing rate of all neurons with a certain direction preference, across all segments. So, we plotted the mean firing rates distributions of all the neurons participating in each bin, for all bins, in violin plots. This left us with one violin plot for every possible direction preference, given the $22.5°$ binning.

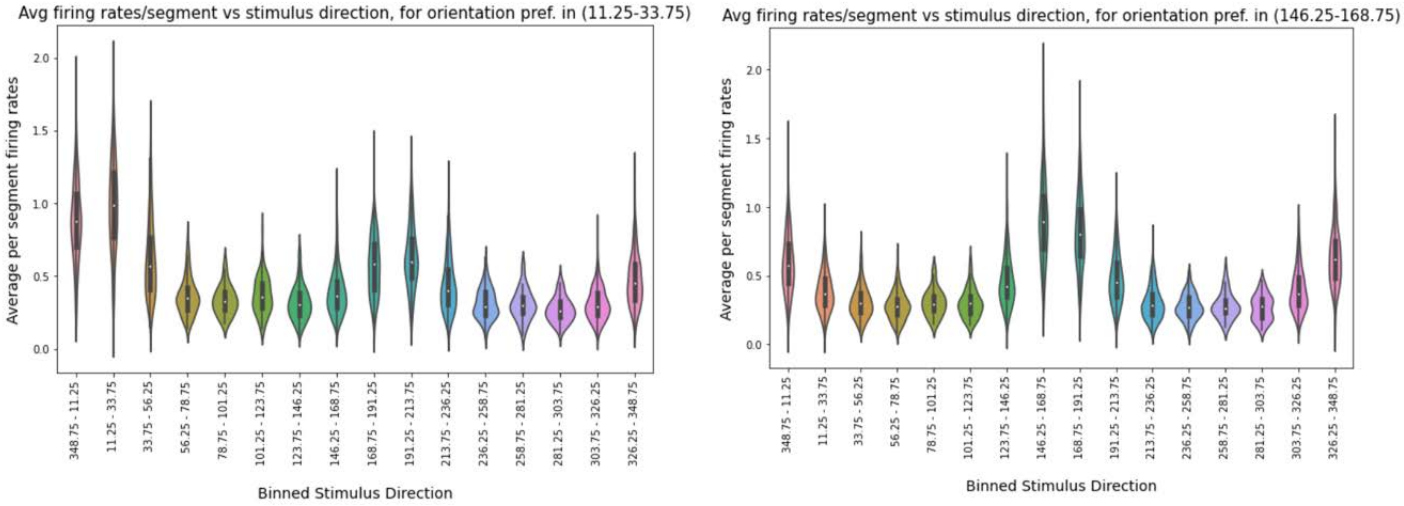

Figure 4.18: *Examples of two violin plots showing the distribution of average firing rate of neurons in layers L4, L23 of mouse 3 with direction preference in* $11.25° - 33.75°$ *(left) and* $146.25° - 168.75°$ *(right). y-axis: average firing rates of the neurons indicated by the plot title, calculated in the period of a segment, for each segment orientation presented to the mouse separately. x-axis: all possible segment directions in* $22.5°$ *bins.*

The results revealed interesting patterns in a case study of mouse 3. From each specific "violin", we can understand the shape of the distribution of firing rates in each direction bin. At the direction resolution a $22.5°$ binning provides, we noticed that the distributions describing the average firing rates of neurons in their preferred direction bins exhibit big standard deviations, high-valued means and a more uniform-like distribution across possible values. On the other hand, distributions describing the average firing rates of the same neurons outside their preferred direction bins exhibited distributions with mostly small standard deviations, a low mean value, while all the firing rates were accumulated in a relatively small,

confined range of values, this way giving short and wide ”violins”.

Although this observation has to be verified across various mice and different binnings to be robust, it reveals intuitive patterns. After all, the calculated average firing rate is a good indicator of *population activity*. Neurons with a certain direction preference, have unanimously low firing rates when a stimulus of a different direction is shown, meaning that their activity is suppressed during these segments. At the same time, the activity of the same neurons is on average higher and more diverse when a segment of their preference is presented. This observation suggests that, most probably, direction preference is depicted in the firing frequency of a neuron population with a certain preference and, if anything, proves the usefulness and informativeness of firing rates as an indicator of neuronal activity for direction-specific analyses.

Chapter 5

Synopsis and Prospects

Computational Neuroscience is a new, challenging and fascinating field of research. Although Neuroscience, as a purely medical/biological discipline, has been gradually developing its foundations for more than 100 years, *Computational* Neuroscience is only a few decades old, as is Computer Science. It is a multidisciplinary field that draws knowledge from Biology and methods from Computer Science and Mathematics, and appears demanding to the researcher, as they have to use elements from both disciplines, but fruitful and promising, as it can bring the best out of both of them. Areas of the brain, when studied at the functional network level, provide, through individual and ensemble firings, patterns whose complex information can be captured by Machine Learning and Data Science methods. This way we can automate and evaluate our observations on neuroimaging data to extract the largest amount of information possible about an area of interest. This thesis, was an attempt to perform this extraction of information on the activity of the mouse primary visual cortex to study its relationships with pupil size and the direction of visual stimuli, by using simple Machine Learning methods, be it predictions, or other tools. Ultimately, the purpose of this work was to characterize these predictions and use the results to make, to some extent, neuroscience-oriented observations, while also avoiding to infer causal relationships.

5.1 Conclusions

Pupil size seems to be forecastable through calcium events of layers 2,3,4 in V1 of mice with an %RMSE of approximately 11% when all layers are involved. The history of pupil is likely to affect pupil size prediction error significantly - specifically, the %RMSE error can drop to approximately 1% when the 10 past pupil size values are included in the model's input. The same improvement is observed in the prediction of future pupil values too, for values not too far in the future. PCA suggests that the cortical state of layers L23, L4 can most probably be represented in a small-dimensional space with very little loss of information. It appears that stimulus direction can be predicted with good classification accuracy (close to 80% with an rbf kernel and zero-centered, 22.5-degree bins) by the firing rates of neurons in of layers 2,3,4 in V1 of mice 3,4 - prediction in mouse 5 exhibits bad behavior for reasons that are to be investigated. Orientation preference of neurons possibly regulates the distribution of average per segment firing rates, as a case study of mouse 3 showed.

5.2 Future Work

As we have previously emphasized, the results in the context of this study are of preliminary nature and they need to undergo further investigation and verification to be conclusive, from the neuroscience point of view. The first steps we would follow towards this direction are the following. For the sake of robustness, all results have to be averaged across all mice, where it has not been done. The use of more sophisticated algorithms, might take a toll on the explainability of a prediction, but also will make the coefficients more meaningful and reduce the possibility of them being statistical artifacts.

Instead of PCA, it would be also beneficial of our analysis if we applied supervised methods, so as to capture the explained variance of each neuron *w.r.t. the pupil size*. In stimulus direction prediction, the prediction can be modeled as a regression task too, possibly with better results; in that case the representation of angles will be given by the tuple $(sin\theta, cos\theta)$, to represent their circular nature. Finally, the strange behavior of mouse 5 in stimulus direction prediction has to be studied and attributed to either "anomalies" in the experimental collection of data or poor quality prediction.

Bibliography

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

[2] A. M. TURING. I.—computing machinery and intelligence. *Mind*, LIX:433–460, 10 1950.

[3] Ridley Scott. *Blade Runner [Film]*. The Ladd Company, Shaw Brothers, 1982.

[4] Korbinian Brodmann. *Vergleichende Lokalisationslehre der Großhirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Johann Ambrosius Barth, 1909.

[5] Eftychios A. Pnevmatikakis et al. Simultaneous denoising, deconvolution, and demixing of calcium imaging data. *Neuron*, 89:285–299, 2016.

[6] Artur Luczak, Peter Barthó, and Kenneth D. Harris. Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron*, 62:413–425, 5 2009.

[7] Jae Eun Kang Miller, Inbal Ayzenshtat, Luis Carrillo-Reid, and Rafael Yuste. Visual stimuli recruit intrinsically generated cortical ensembles. *Proceedings of the National Academy of Sciences of the United States of America*, 111:E4053–E4061, 9 2014.

[8] Michael Okun et al. Diverse coupling of neurons to populations in sensory cortex. *Nature 2015 521:7553*, 521:511–515, 4 2015.

[9] I. Smyrnakis, M. Papadopouli, G. Pallagina, and S. Smirnakis. Information capacity of a stochastically responding neuron assembly. *Neurocomputing*, 436:22–34, 5 2021.

[10] Kenneth D. Harris and Alexander Thiele. Cortical state and attention. *Nature Reviews Neuroscience*, 12:509–523, 9 2011.

[11] Bruno B. Averbeck, Peter E. Latham, and Alexandre Pouget. Neural correlations, population coding and computation. *Nature Reviews Neuroscience 2006 7:5*, 7:358–366, 5 2006.

[12] David P.A. Schulz, Maneesh Sahani, and Matteo Carandini. Five key factors determining pairwise correlations in visual cortex. *Journal of Neurophysiology*, 114:1022–1033, 5 2015.

[13] Joshua H. Siegle et al. Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature 2021 592:7852*, 592:86–92, 1 2021.

[14] C. Bennett, S. Arroyo, and S. Hestrin. Subthreshold mechanisms underlying state-dependent modulation of visual responses. *Neuron*, 80:350–357, 2013.

[15] Y. Fu, J.M. Tucciarone, J.S. Espinosa, N. Sheng, D.P. Darcy, R.A. Nicoll, Z.J. Huang, and M.P. Stryker. A cortical circuit for gain control by behavioral state. *Cell*, 156:1139–1152, 2014.

[16] Carsen Stringer, Michalis Michaelos, and Marius Pachitariu. High precision coding in visual cortex. *bioRxiv*, page 679324, 11 2019.

[17] Jacob Reimer et al. Pupil fluctuations track fast switching of cortical states during quiet wakefulness. *Neuron*, 84:355–362, 2014.

[18] Martin Vinck, Renata Batista-Brito, Ulf Knoblich, and Jessica Cardin. Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. *Neuron*, 5 2015.

[19] Maria Papadopouli et al. Networking within and across layers in area v1: Identifying neuronal modules of communication (under preparation).

[20] Maria Papadopouli. FUNCTIONALLY CONNECTED NEURONAL MODULES IN AREA V1, Invited Speaker Abstract. *AREADNE Research in Encoding and Decoding of Neural Ensembles*, 2022.

[21] Özge Yüzgeç, Mario Prsa, Robert Zimmermann, and Daniel Huber. Pupil size coupling to cortical states protects the stability of deep sleep via parasympathetic modulation. *Current Biology*, 28:392–400.e3, 2 2018.

[22] Andreas Duenser. Combining eeg with pupillometry to improve cognitive workload detection. 2016.

[23] Meenakshi Khosla, Keith Jamison, Gia H. Ngo, Amy Kuceyeski, and Mert R. Sabuncu. Machine learning in resting-state fmri analysis. *Magnetic Resonance Imaging*, 64:101–121, 12 2019.

[24] Berdakh Abibullaev and Jinung An. Classification of frontal cortex haemodynamic responses during cognitive tasks using wavelet transforms and machine learning algorithms. *Medical Engineering Physics*, 34:1394–1410, 12 2012.

[25] Alain Destexhe, Diego Contreras, and Mircea Steriade. Spatiotemporal analysis of local field potentials and unit discharges in cat cerebral cortex during natural wake and sleep states. *Journal of Neuroscience*, 19:4595–4608, 6 1999.

[26] Danko Nikolić, Stefan Hä Usler, Wolf Singer, and Wolfgang Maass. Distributed fading memory for stimulus properties in the primary visual cortex. *PLOS Biology*, 7:e1000260, 12 2009.

[27] William F. Kindel, Elijah D. Christensen, and Joel Zylberberg. Using deep learning to probe the neural code for images in primary visual cortex. *Journal of Vision*, 19:29–29, 4 2019.

[28] Song Pan, Serdar Iplikci, Kevin Warwick, and Tipu Z Aziz. Parkinson's disease tremor classification – a comparison between support vector machines and neural networks. *Expert Systems with Applications*, 39:10764–10771, 2012.

[29] Kyung Hwan Kim, Sung Shin Kim, and Sung June Kim. Improvement of spike train decoder under spike detection and classification errors using support vector machine. *Medical and Biological Engineering and Computing*, 44:124–130, 2006.

[30] Joanne C. Van Slooten, Sara Jahfari, Tomas Knapen, and Jan Theeuwes. How pupil responses track value-based decision-making during and after reinforcement learning. *PLOS Computational Biology*, 14:e1006632, 11 2018.

[31] Allen Jong Woei Whang et al. Pupil size prediction techniques based on convolution neural network. *Sensors 2021, Vol. 21, Page 4965*, 21:4965, 7 2021.

[32] Alejandro Lara-Doña, Sonia Torres-Sanchez, Blanca Priego-Torres, Esther Berrocoso, and Daniel Sanchez-Morillo. Automated mouse pupil size measurement system to assess locus coeruleus activity with a deep learning-based approach. *Sensors 2021, Vol. 21, Page 7106*, 21:7106, 10 2021.

[33] Graziella Orrù, William Pettersson-Yeo, Andre F. Marquand, Giuseppe Sartori, and Andrea Mechelli. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review. *Neuroscience Biobehavioral Reviews*, 36:1140–1152, 4 2012.

[34] Marcus A. Triplett, Zac Pujic, Biao Sun, Lilach Avitan, and Geoffrey J. Goodhill. Model-based decoupling of evoked and spontaneous neural activity in calcium imaging data. *PLOS Computational Biology*, 16:e1008330, 11 2020.

[35] Optical flow tutorial, *OpenCV Documentation*.

[36] Gunnar Farnebäck. Two-frame motion estimation based on polynomial expansion. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2749:363–370, 2003.

[37] John Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8:679–698, 1986.

[38] Richard O Duda and Peter E Hart. Use of the hough transformation to detect lines and curves in pictures. *Comm. ACM*, 15:11–15, 1972.

[39] Catherine S. Cutts and Stephen J. Eglen. Detecting pairwise correlations in spike trains: An objective comparison of methods and application to the study of retinal waves. *Journal of Neuroscience*, 34(43):14288–14303, 2014.

[40] Paolo Bonifazi et al. Gabaergic hub neurons orchestrate synchrony in developing hippocampal networks. *Science (New York, N.Y.)*, 326:1419–1424, 5 2009.

Appendix A

Degree of Connectivity

The following appendix describes how (Normalized) Degree of Connectivity is introduced in a study such as this one. We included it in the context of a preliminary analysis of our data, but most results seemed rather unfruitful. In this appendix, we will briefly present the tools we used to include Degree of Connectivity into our analysis. The appandix refers to how we related it to neuron impact, but paints the picture of how it should be calculated independently, too.

A tool usually employed in computational neuroscience to model neuron networks is Graph Theory. Thus, looking ourselves at our data with a graph theoretical approach, we included metrics such as the degree of connectivity of a neuron in our analysis. In Graph Theory, the degree, or degree of connectivity of a directed graph's vertex is defined as the number of incoming edges connected to the vertex. In the scope of this thesis, the neuron's network is a directed graph. The edges are assigned a certain *statistical significance* so that they can be separated into significant and non significant ones. To quantify statistical significance, we first need to find a measure of temporal correlation between the firings of two neurons participating in an edge. A correlation index robust to various firing rates of the two neurons is the Spike Time Tiling Coefficient (STTC) [39]. Considering two neurons A,B, "traditional" STTC introduces T_A as the fraction of the total recording time which lies within a time window $\pm\Delta t$ far from any spike of A. T_B is defined accordingly. Also, it incorporates P_A (P_B, accordingly) as the proportion of A's (B's) firings that lie within Δt *before or after*

each firing of B (A), to the total firings of A (B), and is defined as:

$$STTC = \frac{P_A - T_B}{1 - P_A T_B} + \frac{P_B - T_A}{1 - P_B T_A}$$

We chose to use a slightly modified version of STTC [19] that maintains a directional character and yields different coefficients for the edges $(A \to B)$ and $(B \to A)$. Hence, to define $STTC_{AB}$, we first define the proportion of A's firings that lie within Δt *before* each firing of B, to the total firings of A as P_A^{B-}. With the same rationale, P_B^{A+} is the proportion of B's firings within Δt *after* each firing of A, to the total firings of B. In this metric, T_A, T_B naturally correspond to T_{A+}, T_{B-} when talking about $STTC_{AB}$. This way we express the modified coefficient as:

$$STTC_{AB} = \frac{P_A^{B-} - T_{B-}}{1 - P_A^{B-} T_{B-}} + \frac{P_B^{A+} - T_{A+}}{1 - P_B^{A+} T_{A+}}$$

All quantities calculated express the STTC for the edge $(A \to B)$. The reader can now understand how $STTC_{BA}$ would be calculated. With this metric, we have achieved to find a measure the correlation of the two neurons in terms of firing, this way quantifying their synchrony. Now, intuitively, to quantify the significance of edge $(A \to B)$ we have to create a control to which $STTC_{AB}$ will be compared to. This is achieved by shifting the events of the neurons independently by random numbers drawn from a uniform distribution, creating a random distribution of STTCs, called the *null* STTC distribution. The only thing left to calculate statistical significance for edge $(A \to B)$, or any edge really, is to measure the difference between the observed distribution of STTCs and the null distribution of STTCs. This is achieved for every individual edge through the calculation of its z-score:

$$z = \frac{STTC - \mu_{null}}{\sigma_{null}}$$

where $STTC$ is the STTC of the edge in question and $\mu_{null}, \sigma_{null}$ the mean and standard deviation of the null distribution. Now that we have quantified statistical significance for all the edges through their z-score, we are able to define the neurons' Degree of Connectivity.

For our purposes we will be using the Normalized Degree of Connectivity (NDoC), which is calculated as follows: for each reference neuron r, we take the number of neurons that have

statistically significant incoming edge to r (i.e. z-score > 4) and divide it with the number of all the neurons of the layer where the edge is coming from (for $inter - layer$ computation) or with all the neurons minus 1 (for $intra - layer$ computation).

Let us define the NDoC mathematically. Let r be the neuron of interest in our network, E_L the set of all connections within a layer and $Z(e)$ the z-score of an edge $e \in E_L$, then $E_{\text{in}}(r) = \{(u, r) \mid (u, r) \in E_L,\ Z((u, r)) > 4\}$ will be the set of statistically significant incoming edges to r, u denoting an neighboring neuron and (u, r) its connection to r. Thus, the normalized degree of connectivity will be:

$$D(r) = \frac{|E_{\text{in}}(r)|}{|E_L| - 1},$$

where $|.|$ denotes the cardinality of a set. Note that this formula performs an *intra-layer* computation, since that is the one we will be using.

To investigate if there is any relationship between a neuron's impact in prediction and its normalized degree of connectivity, we worked on each layer and mouse separately: similarly with before, for each mouse, we trained a linear SVR on its calcium events using k-fold cross validation. We used 12 folds, without shuffling, and calculated the mean impact (coefficient) of each neuron across all folds for every mouse, and plotted the results in scatterplots. Note again, that by the term impact we mean the coefficient that is assigned to a neuron after an algorithm is fit on the eventograms and more specifically, in this case, the average coefficient across all folds.

Neuronal connectivity is an important higher-order statistic that can be connected to the importance of neurons. Intuitively, one would expect NDoC to be proportional to a neuron's impact; highly connected neurons are expected to have bigger power over the regulation of the cortical state, behavioral characteristics and thus our prediction, and this would be the case in a simple graph model, but the aforementioned relationships are far more intricate.

The lack of a crystal-clear pattern in the preliminary analysis does not indicate the failure of NDoC as a metric to study neuronal impacts, but suggests that, for the impacts to be related to other metrics, more complex, non-linear methods have to be used - methods that capture nonlinearities in data can assign more representative coefficients to neurons so that we can quantify their impact better. Even then, one might find disappointing results, if they adopt a connectivity-implies-importance rationale, since this might not be the case. Regardless,

the information the NDoC provides might be indirect at cases, however it classifies neurons according to graph theoretical and cytoarchitectural properties in a way that makes cortical networks easier to study, and, if combined with other neuron properties, can lead to interesting conclusions and theories, such as the existence of *hub neurons* [40].