The Diversity of Neural Predictivity in Mouse Visual Cortex: Correspondence of Goal Driven ANN (AlexNet) and Regions Via Regularized Multivariate Regression

John Cocjin
Departmet of Bioengineering
Stanford University

Sophia Sanchez
Department of Computer Science
Stanford University

Ramon Iglesias
Department of Civil Engineering
Stanford University

Abstract

Recent studies have shown a marked representational resemblance between convolutional neural networks (CNN) trained to perform computer vision tasks such as object recognition and the mammalian ventral stream. However, the core premise of these studies requires training linear regression models where the number of parameters far exceeds the number of samples. The present study investigates key regression and regularization approaches and tests their performance in rhesus macaque (*Macaca mulatta*) and mouse (*Mus musculus*) neural datasets. While subsampling with regularization did improve variance explained, the within and between region analysis revealed poor performance. While the possibility of measuring across regions is intriguing, significant problems remain with regard to the relatively high number of parameters to samples, which hinders data exploration. Future work may benefit from subsampling, as well as from collecting data with far more image presentations than the dataset utilized in the present study.

1 Introduction

Recent studies have shown a marked representational resemblance between convolutional neural networks (CNN) trained to perform computer vision tasks (particularly object recognition) and the mammalian ventral stream (Cadieu et al. (2014); Cadena et al. (2017); Yamins et al. (2014)). The core premise of these studies is that two representations are *equivalent* if it is possible to map one to the other via a linear transformation. These results open an avenue of research for understanding visual sensory processing in biological brains by emulating the tasks in artificial neural networks.

This line of research necessitates training linear regression models that best fit the available data despite a strong dissimilarity in the dimensions of the two representations. Specifically, neural activations measured from biological brains usually have far fewer cells than their CNN counterparts. This problem is aggravated by the fact that measuring biological neural activations is a work-intensive task. Thus, the sample sizes in question are often much smaller the number of parameters to be fit.

The present study aims to characterize different approaches for addressing the dimensional dissimilarity. In particular, we study how different types of regression and regularization affect the predictive

power of the trained models. We achieve breadth by performing a hyperparameter search and by studying rhesus macaque (*Macaca mulatta*) and mouse (*Mus musculus*) neural datasets.

2 Background and Related Work

Significant progress has been made from the perspective of convolutional neural networks and the theory of mammalian brain modeling. Yamels et al. demonstrated that the outputs of the convolutional layers of an HCNN trained using recognition tasks are predictive of neural responses (Yamins et al. (2014)). Of particular interest is that the last layer of the model was especially predictive of the IT region of the brain, while the intermediate layers were more preditive of V4. This serves as strong evidence for a degree of consistency between biological visual systems and HCNN's.

However, predictivity alone is not sufficient for a strong computational model. Yamins and DiCarlo assert that the ideal computational model of the brain should not simply "predict the stimulus-response relationship of neurons in one final area, such as (in vision) anterior inferior temporal cortex." They argue that it is also necessary for the model to represent intermediate cortical areas (Yamins and DiCarlo (2016)).

In the present study, our primary aim was to characterize which layers of AlexNet best predict which neurons in mouse visual cortex via multivariate regression, and to compare the diversity in layer-neuron profiles between and within regions. In order to address this question, we focused on the critical subtask of solving for regularization across regions, and adressing the problem of a high parameter, low sample space. Specifically, we investigated predictivity, shape, and consistency of our results.

3 Methods

We leveraged an array of existing methods and datasets from the machine learning and neuroscience community to carry out the study. In particular, Section 3.1 briefly describes the two brain datasets we utilized and Section 3.2 covers the CNN model, regression and regularization methods, and the optimization formulation.

3.1 Data

We tested our approach on two datasets, one corresponding to neural activations in macaque and another corresponding to activations in mice. The following subsections describe these datasets.

3.1.1 Allen Institute Mouse Cortex Data

The Allen Institute Brain Observatory dataset all consists of in vivo calcium imaging on transgenic mice expressing GCaMP6f in laminar specific subsets of cortical pyramidal neurons within visual cortical regions. In imaging sessions, mice were shown natural scenes, natural movies, drifting gratings, or locally sparse noise. Mice were on a circular treadmill (running) in the course of the experiments.

The natural stimuli are taken from a library of 118 natural scenes (Berkeley Segmentation Dataset, van Hateren Natural Image Dataset, McGill Calibrated Colour Image Database). Each stimulus was presented for 250 ms in random sequence 50 times. Natural stimuli presentations were collected in 216 sessions (as per Allen terminology, experiment containers), spread among the following visual regions: VISal, VISam, VISl, VISp, VISpm, VISrl.

3.1.2 V4 and IT Data in Rhesus Macaque

Additionally, we used the macaque neural activation dataset used in Yamins et al. (2014). The dataset consists of neuronal activity was in the V4 and IT of rhesus macaque (*Macaca mulatta*). The stimulus set consisted of 5,760 natural images and 64 3-dimensional objects across 8 categories (animals, boats, cars, chairs, faces, fruits, planes, and tables). Stimuli were superimposed on randomly selected background photography. Images were presented according to three levels of variation (low, medium, and high) in terms of object position, pose, and size. Complete methodology can be found in Yamins et al. (2014).

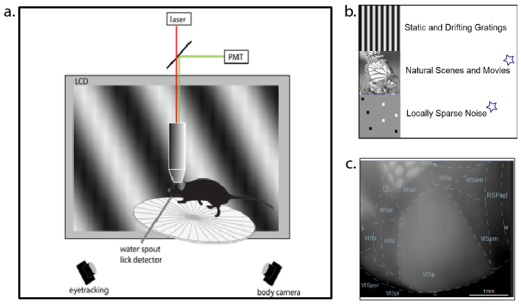

Figure 1: (A) The Allen institute Brain Observatory conducted 2-photon calcium imaging experiments on mice while they viewed images or movies and ran on a circular treadmill. (B) Each stimulus fell into one of three categories: gratings, natural stimuli, or locally sparse noise. (C) Imaging was collected on the regions depicted.

Figure 2: Example images from the natural stimulus dataset

3.2 Regressing Neural Data

The regression strategy is as follows. For each of the images in the stimulus dataset, we forward pass the trained (artificial) neural network to obtain the latent representations at each of the hidden layers. We use these representations as feature vectors to regress the (mammalian) neural activations.

We are interested in evaluating whether the artificial network representations are equivalent (or at least similar) to the mammalian neural representations. Thus, we limit our regression to linear models. That is, our hypothesis is that for two representations to be *sufficiently similar* it must be possible to map them to each other via a linear transformation.

In particular, we study three regression approaches:

- **Simple Fully Connected (Simple).** In this approach, there is a trainable parameter (weight) W_{ijkl} connecting the artificial neural activation r_{ijk} at row i, column j, and depth k. That is, for a mammalian activation neuron, we have

$$\hat{y}_l = \sum_{i,j,k} W_{ijkl} r_{ijk} + b_l \tag{1}$$

- **Spatial Fully Connected (Spatial).** Similar to Simple, this regressor fully connects the inputs to the outputs. However, it leverages a kernel K of dimensions $H \times W \times D \times N_{\text{m}}$ to compute the transformation. Thus, it is possible to impose regularizers that leverage the spatial structure of the representation (see Section 3.2.2). Formally:

$$\hat{y}_l = \sum_{i,j,k} K_{ijkl} r_{ijk} + b_l \tag{2}$$

- **Factored Fully Connected (Factored).** While Spatial uses a single kernel to fully connect input to output, performs a depth-separable convolution over the H, W dimensions with a common spatial kernel, then takes inner product over the D dimension with a feature kernel. Regularizations may apply separately to the spatial mask M and the feature weights W.

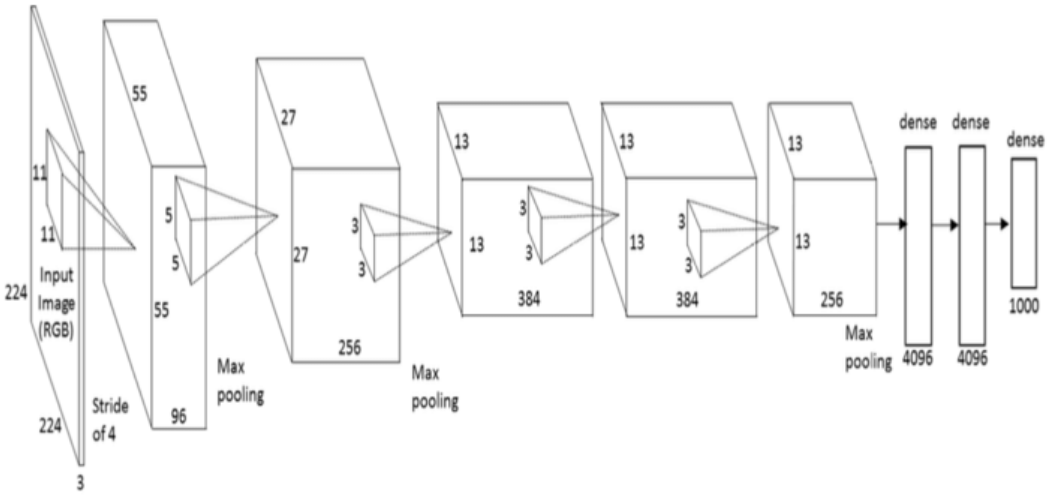

Figure 3: AlexNet architecture. For regressing the

Klindt et al. (2017). Formally,

$$\hat{y}_l = \sum_{i,j,k} M_{ijl} W_{kl} r_{ijk} + b_l \tag{3}$$

3.2.1 AlexNet

We based our analysis using a pretrained network using the architecture and training procedure of the original AlexNet Krizhevsky et al. (2012). Figure 3 shows its architecture. Particularly, we used the convolutional layers to regress the neural data, and disregarded the fully convolutional layers.

3.2.2 Regularization

We implemented mainly three types of regularization: L1, L2, and Laplacian. L1 tends to produce sparse results, while L2 penalizes large weights. Both are widely used in the field. Formally, they are given by:

$$\mathcal{L}_{l1} = \lambda_{l1} \sum_{i,j,k,l} |W_{ijkl}| \text{ and } \quad \mathcal{L}_{l2} = \lambda_{l2} \sum_{i,j,k,l} W_{ijkl}^2 \,. \tag{4}$$

The Laplacian regularizer, adapted from Klindt et al. (2017), imposes a smoothness regularizer on the kernels of the Factored and Spatial Layers, and therefore, it was not implemented for Fully Connected Layers. Formally, the Laplacian regularizer is

$$\mathcal{L}_{\text{laplace}} = \lambda_{\text{laplace}} \sum_{i,j,k,l} (W_{:,:,kl} * L)_{ij}^2 \text{ where } \quad L = \begin{bmatrix} 0.5 & 1 & 0.5 \\ 1.0 & -6 & 1 \\ 0.5 & 1 & 0.5 \end{bmatrix} . \tag{5}$$

3.2.3 Optimization and Hyperparameter Search

Our goal is to find linear transformation that minimizes the squared difference between the predicted mamalian neural respones and the artificial ones. That is,

$$\mathcal{L}_{\text{regression}} = \sum_{b,l} (y_l^b - \hat{y}_l^b)^2 \,, \tag{6}$$

where b denotes the sample stimulus index. Thus, our loss becomes

$$\mathcal{L} = \mathcal{L}_{\text{regression}} + \mathcal{L}_{l1} + \mathcal{L}_{l2} + \mathcal{L}_{\text{laplace}} \tag{7}$$

To find the best regressors, we performed a grid search over the regularizing weights and the regressor parameters. For each combination, we performed Stochastic Gradient Descent using the Adam optimizer. To test their performance, we split the data into train (85%) and validation (15%) and compared on both the regression loss, and the average explained variance.

Table 1: Hyperparameters explored. Spatial factoring has a spatial term included whereas simple does not.

Name	Factoring	Weight Decay	L1	Laplacian	Subsampling
s1	simple	1	1	0	Yes
s2	simple	10	1	0	Yes
s3	simple	1	10	0	Yes
s4	simple	10	10	0	Yes
sp1	spatial	0	0	1	No

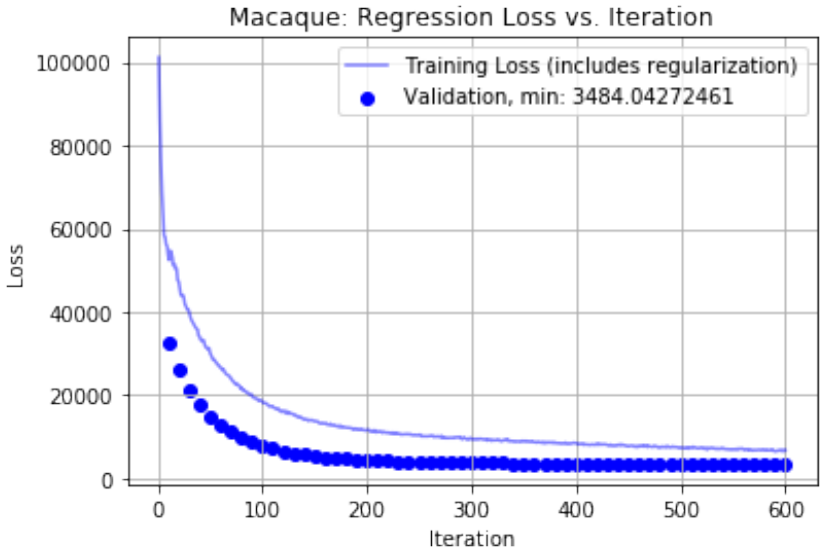

Figure 4: Training and validation loss for the macaque dataset by iteration.

4 Results

4.1 Benchmarking predictivity with our regression approaches on Macaque IT

Due to the tricky form of the loss function, we solved our linear regressions by stochastic gradient descent. We first characterize the iterations necessary to reach assignment 1 level scores (0.25) across macaque IT neurons. We tested both subsampling and full data approaches. After 30 epochs, subsampling approaches were capable of reaching our expected explained variance range whereas full data approaches, even across our gridding, only reached 0 explained variance. Due to this, as well as computational limitations, we restrict subsequents analyses to a single full data weighting and across subsampled formulations.

4.2 Characterizing predictivity across the gridding regime in Mouse

We next examined the capacity of our linear regressions to predict neurons across mouse visual regions. For all parameter regimes examined, 75 percent of neurons have a variance explained below 0. Only two parameter regimes (s2 and s4) yield max variance explained above 0 (0.609 and 0.504 respectively). For across and within region comparisons, we consider the most positive regime – s2.

4.3 AlexNet layer to visual region correspondence

We characterize how well AlexNet layers predict neurons in each region. As can be seen in the following figure, the all regions share a similar prediction profile, with shifts in the (very bad) variance explained.

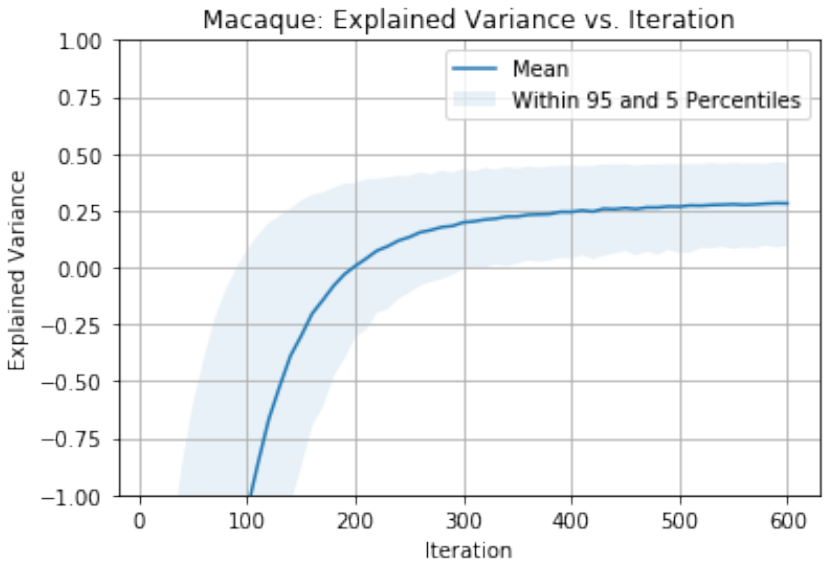

Figure 5: Average explained variance by iteration in the macaque set. The shaded area is covers the range between the 5 and 95 percentiles of explained variance for the individual cells.

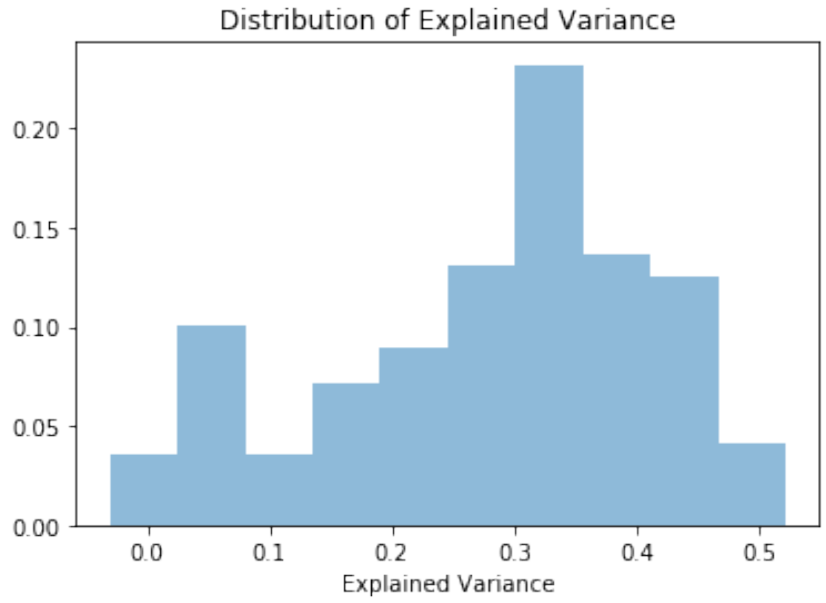

Figure 6: Explained variance distribution at the end of the training.

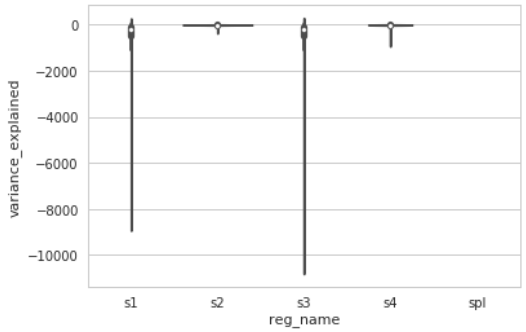

Figure 7: Violin plot of explained variance according to regularization.

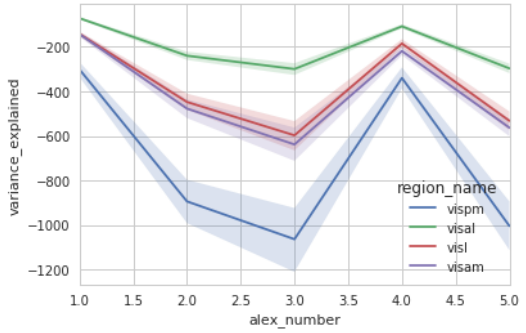

Figure 8: Variance explained according to AlexNet layer for each region (VISal, VISam, VISl, VISpm).

4.4 Consistency of predictivity

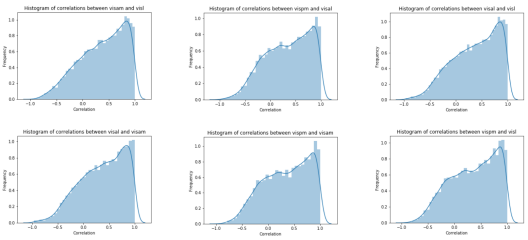

Figure 9: Histogram of correlations between cells of brain regions (VISal, VISam, VISl, VISpm).

We were interested in the consistency of AlexNet predictivity profiles within and between mouse visual regions. We first consider the correlation in AlexNet profiles among neurons within a region, as seen in 9. The within region profiles have a common negative skew distribution, with most mass at high correlation values. This indicates relatively consistent profiles on the whole. The lack of apparent on-diagonal block corroborates consistency 10. Correlation between neurons of different regions yield similar distributions. The hierarchical clustering ordering further shuffles neurons of different regions. The on-diagonal block structure indicates different dominant profiles.

5 Conclusions

Here, we sought to explore the variation of AlexNet predictivity profiles among neurons within and between visual regions of the mouse visual cortex. The Allen Brain Data set appeared as a possible trove for examining between region differences provided the number of animals and regions collected from. However, we were posed with the difficulty of mapping a large number of artificial network activations with an order of magnitude smaller sampling in images presented. We attempted to

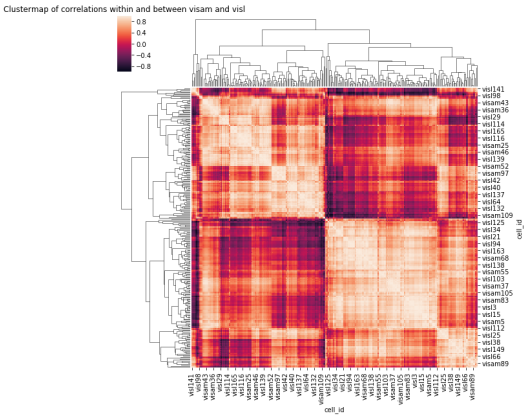

Figure 10: Clustermap of correlations within and between VISam and VISl. See APPENDIX for additional clustermaps.

address this problem with subsampling on the artificial neurons as well as regularization. While subsampling with regularization did aid in improving our metric of performance, variance explained, we still were left with poor performance. From there, we carried on the motions of a within and between region analysis. We found that, in this regime of poor performance, predictivity varied among the regions but with similar AlexNet profiles. There further appeared to be profile segregation among the region. It is difficult (and probably ill advised) to interpret these findings with regards to neural computational mechanisms. In future work, we would hope to improve our image sampling with an appropriate dataset that has far more image presentations than that contained here.

Acknowledgments

The authors would like to thank Dr. Dan Yamins, Damian Mrowca, Aran Nayebi, Daniel Bear, and the members of the Stanford NeuroAILab for their assistance during the course of this research.

References

Cadena, S. A., G. H. Denfield, E. Y. Walker, L. A. Gatys, A. S. Tolias, M. Bethge, and A. S. Ecker, "Deep convolutional models improve predictions of macaque v1 responses to natural images," *bioRxiv*, (2017), p. 201764.

Cadieu, C. F., H. Hong, D. L. Yamins, N. Pinto, D. Ardila, E. A. Solomon, N. J. Majaj, and J. J. DiCarlo, "Deep neural networks rival the representation of primate it cortex for core visual object recognition," *PLoS computational biology*, vol. 10 (2014), p. e1003963.

Klindt, D., A. S. Ecker, T. Euler, and M. Bethge, "Neural system identification for large populations separating" what" and" where"," in *Advances in Neural Information Processing Systems*, pp. 3508–3518, 2017.

Krizhevsky, A., I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

Yamins, D. L., and J. J. DiCarlo, "Using goal-driven deep learning models to understand sensory cortex," *Nature neuroscience*, vol. 19 (2016), pp. 356–365.

Yamins, D. L., H. Hong, C. F. Cadieu, E. A. Solomon, D. Seibert, and J. J. DiCarlo, "Performance-optimized hierarchical models predict neural responses in higher visual cortex," *Proceedings of the National Academy of Sciences*, vol. 111 (2014), pp. 8619–8624.

5.1 APPENDIX

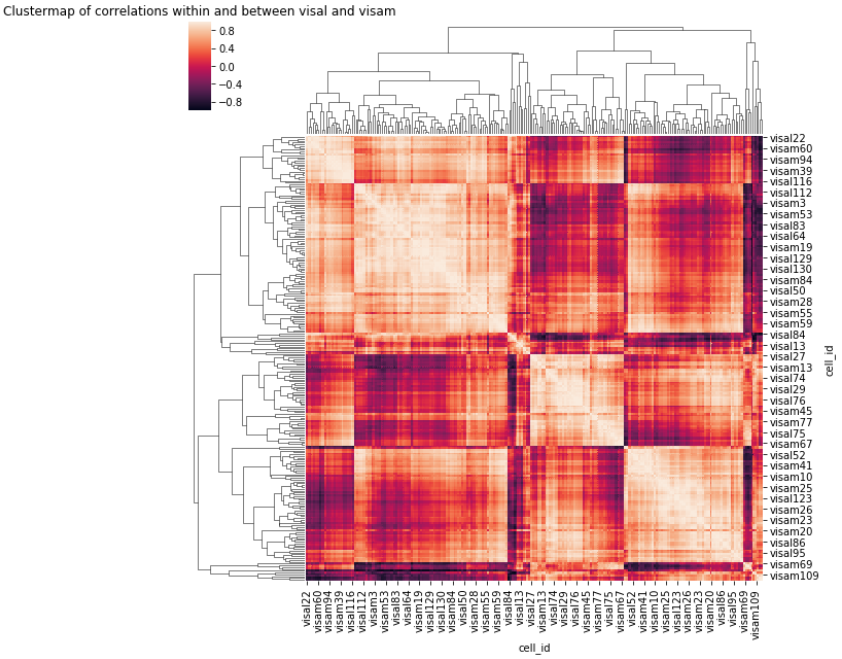

Figure 11: Clustermap of correlations within and between VISal and VISam.

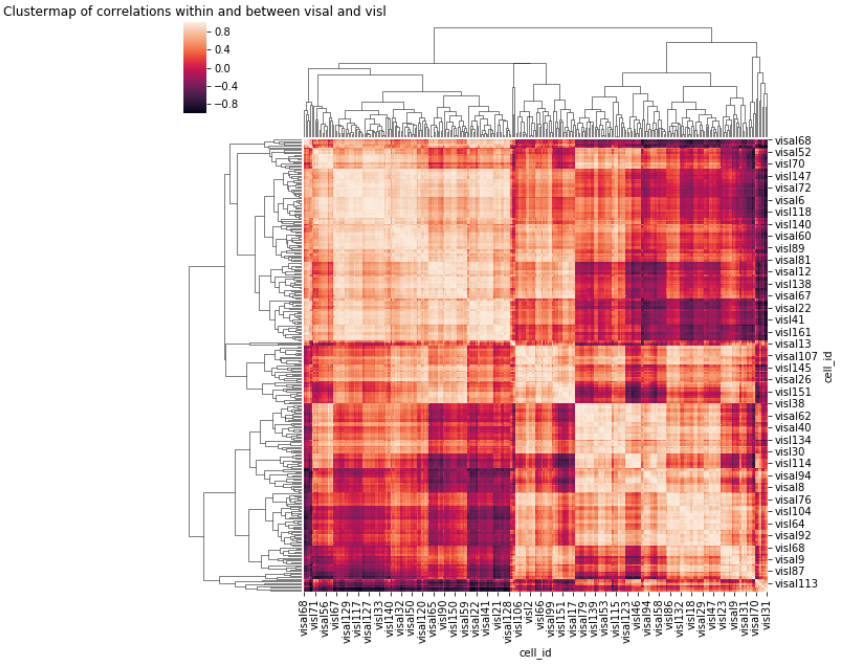

Figure 12: Clustermap of correlations within and between VISal and VISl.

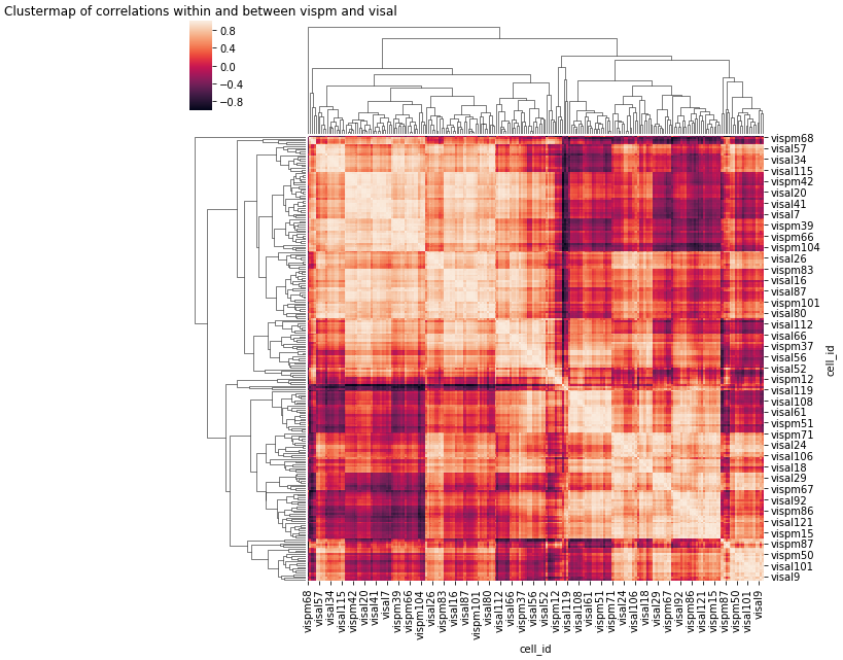

Figure 13: Clustermap of correlations within and between VISpm and VISal.

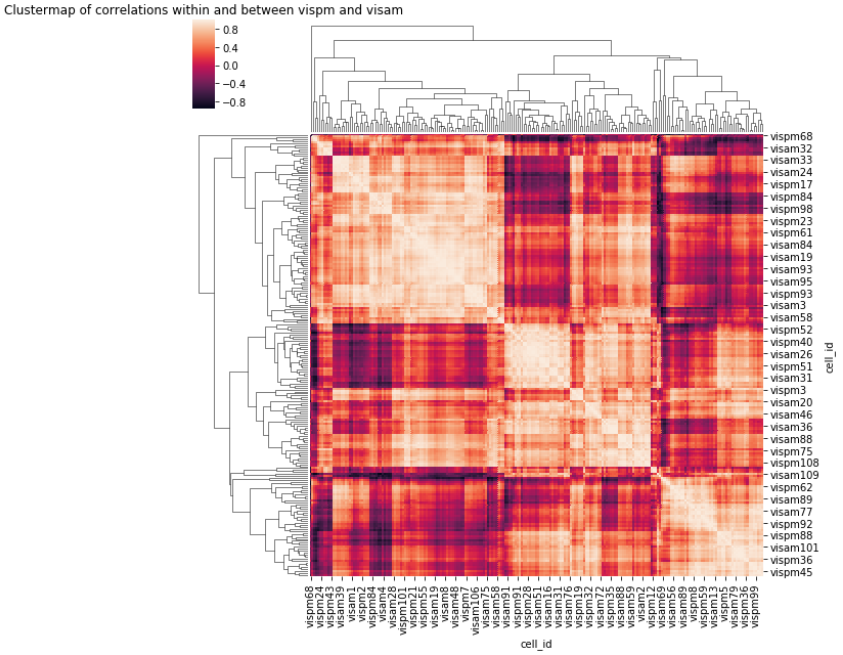

Figure 14: Clustermap of correlations within and between VISam.

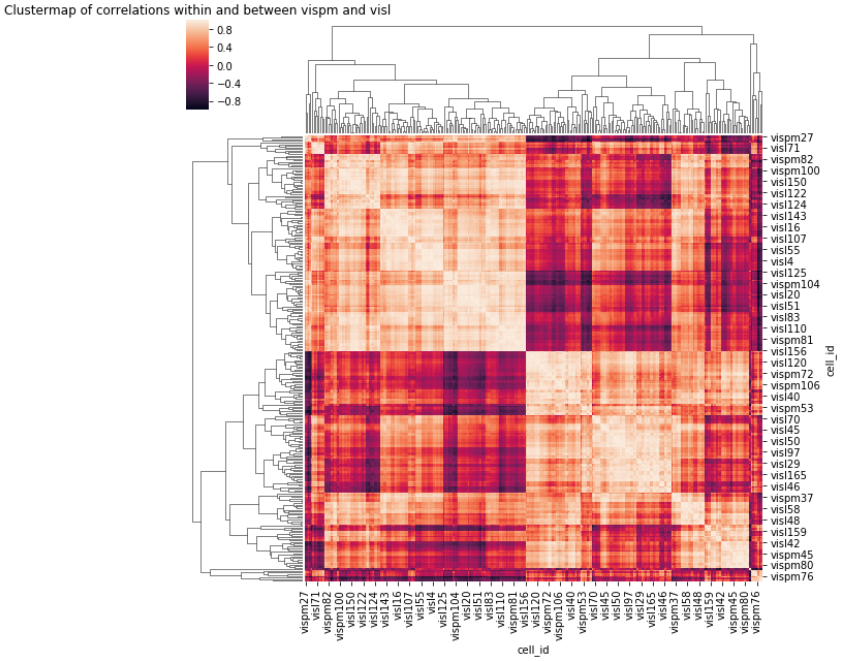

Figure 15: Clustermap of correlations within and between VISl.