Predicting IT Neuron Activity

Grosjean Barbara, Ingster Abigaïl
NX-414 Brain-like computation and intelligence, EPFL

This project investigates various modeling techniques to analyze the relationship between neural data from the inferior temporal (IT) cortex of non-human primates and the visual stimuli encountered during object recognition tasks. The datasets utilized are sourced from Majaj et al. (Journal of Neuroscience, 2015), and have undergone preprocessing, encompassing neural recordings obtained via multielectrode arrays implanted in the IT cortex of monkeys (168 neurons), alongside images of objects set against a natural landscape background (RGB channels, 224×224 pixels).

I. Linear regression from the pixels

We employed ridge regression to linearly regress neural data from input pixels and optimized the regularization parameter through $k-$fold cross-validation. So as to reduce overfitting and given the overparameterized nature of the problem ($3 \times 224 \times 224$ features in to 168 features out), we extracted the first 1000 principal components (PCs) of the flattened pixels and used them as input features instead of the raw pixels. Such a low-complexity model is unable to generalize to unseen examples as it has learnt to fit the noise present in the training data.

II. Task-driven approach

The intuition behind this approach is that brain areas are optimized to perform specific tasks, thereby defining a utilitarian normative optimization goal for the brain. The IT cortex being an area involved in visual processing, by optimizing a neural network for object recognition, we could achieve operations similar to those of IT neurons. To this end, we leveraged a pre-trained convolutional network incorporating residual connections (ResNet50, PyTorch), extracted the activations of some of its layers to use them as input features for a ridge regression model employing IT neurons as targets. Analogously to Section 1, we used the 1000 first PCs of the extracted layer activations as input features. Such a procedure is called *linear probing*. As depicted in Fig. 1, the layer enabling the best reconstruction of IT neural signals is the second convolutionl layer of the ResNet, which suggests some mirroring of the visual stream hierarchical organization by the ResNet's architecture. However, the depth of best predicting layer highly depends on the regularization factor (λ) selected for the ridge regression. By setting a more stringent upper limit on λ, we find that the third convolutional layer emerges as the most effective for prediction. As presented in Table I, the performance of the ResNet's untrained (random) version is poorer, but sheds light on the portion of performance attributable to ResNet's architecture, independent of the optimization objective.

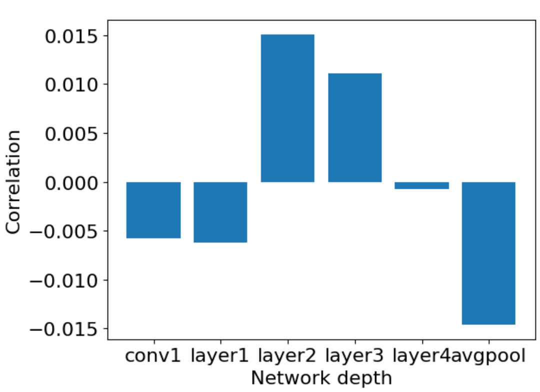

Fig. 1. Mean correlation between ResNet's layer activations and IT neurons.

III. Data-driven approach

Data-driven approaches involve directly optimizing a predictive model of neural data. For this purpose, we used a neural network comprising 3 convolutional layers incorporating batch normalization and variance-preserving initialization, followed by 2 linear layers. Despite validation loss tracking during training and hyperparameter tuning with optuna, an automatic hyperparameter optimization tool, our performance is unexpetedly very low (*cf* Table I).

IV. Exploration

A. Dual-task architecture

Drawing inspiration from the study conducted by Kell et al. (Neuron, 2018), we co-trained a neural network to simultaneously carry out object classification and neural activity prediction. After sharing initial stimulus preprocessing layers (3-layer-deep fully ConvNet) intended to capture pertinent stimulus representations, we allocated a distinct network branch for each task.

B. Autoencoder approach

The idea is to train a neural network to encode the visual stimulus into neuronal spikes, such that it can be reconstructed via a decoder from this neural activity. Such a goal ensures that the information preserved in the bottleneck encapsulates a representation of the stimulus that is as concise and comprehensive as possible. We performed linear probing on the activations of the bottleneck to regress IT neuron spikes.

C. Combining data- and task-driven approaches

The idea is to leverage the representations learnt by the ResNet during task-driven optimization to extract more relevant features compared to raw pixels. The activations of the second (resp. third) convolutional layer, i.e. the layer exhibiting the highest correlation with IT spikes (*cf* Fig. 1), were used as input features for a convolutional neural network optimized at predicting IT neural data in a task-driven approach. A higher performance is reached when training our ConvNet on top of the third layer compared to layer 2 (*cf.* Table I).

TABLE I
Prediction performances of different models

	Explained Variance	**Correlation**
Ridge regression	-0.0899	0.2115
Task-driven approach (layer 2, trained)	1.2716e-08	0.0151
Task-driven approach (layer 2, random)	1.99e-05	0.00736
Task-driven approach (layer 3, trained)	-0.0044	0.0149
Task-driven approach (layer 3, random)	-0.0402	-0.0039
Data-driven approach	0.0012	0.0769
Dual-task architecture	0.0408	0.2312
Combined approach (layer 2)	0.1901	0.4397
Combined approach (layer 3)	**0.2990**	**0.5587**
Autoencoder architecture	0.0912	0.3085

In summary and as presented in Table I, while task- and data-driven methods unexpectedly performed very poorly, the combined approach emerged as the most effective for predicting IT neural activity, although sacrificing interpretability in comparison to the task-driven approach.