Predicting the Activity of IT Cortex Neurons

Viva Berlenghi, Marcin Kroczak, Sophie Meuwly
NX-414: Brain-like computation and intelligence, EPFL May 7, 2025

The goal of the following project was to predict as accurately as possible the activity of the inferior temporal (IT) cortex in non-human primates. This was done with the data from Majaj et al., containing the activity of 168 IT neurons while looking at 2592 RGB images.

I. Linear Regression from Pixels

We first investigated the performance of Ordinary Least Square (OLS) and Ridge Regression models (1, 2 on table I). The rather poor performance is likely due to the extremely high dimensionality of the input relative to the number of samples. To address this issue, we applied Principal Component Analysis to reduce the feature space to 1000 components. This led to similar performances (3, 4 on table I), which highlights the inherent limitations of raw pixel-based modelling.

II. Task-Driven Approach - ResNet50

It is a well-known fact that biological neural layers encode various structured visual features, like edges, textures, or shapes. To leverage this, we extracted layer-wise activations from a pre-trained ResNet50 and computed the 1000 principal components to feed the regression models (5 on table I).

Predictive performance peaked at `layer3` (see fig.1), indicating that mid-level features in the task-trained ResNet50 carry the most relevant information to predict IT responses. The relevance of this approach was further confirmed after applying the same regression pipeline to a randomly initialized ResNet. This indeed resulted in a near-zero correlation (7 on table I, figure), confirming the importance of task-driven representation learning.

III. Data-Driven Approach - Shallow CNN

This approach tries to learn patterns directly from neural recordings and stimuli.

We trained a custom 2-layer Convolutional Neural Network (CNN) from scratch to predict neural activity. Despite stable training dynamics with low validation loss (0.127), the network failed to generalize, resulting in low explained variance and correlation (8 on table I). This performance gap supports the difficulty of learning biologically aligned features purely from limited neural data, compared to leveraging pretrained task-driven representations.

IV. Best Model Exploration

We evaluated how well features from pretrained ConvNeXt-Base and ResNet50 models predict neural activity to natural images.

To maximize performance, we explored some pre-trained models and fine-tuned regression heads on top of their features. We began with feature extraction.

Linear regression on PCA-compressed features showed promising results (I.9) which were improved by adding a deeper neural network head (I.10).

Pushing further, we revisited ResNet50 by building a Deep ResNet50 Head and evaluating the MSE loss (I.11), then obtaining the best results unfreezing layers 2-4 and applying data-augmentation to the inputs (I.12).

This confirms that representations learned via object recognition (task-driven) can align closely with IT neural activity, especially in mid-level layers.

The final ResNet50 model achieving the best performance has been saved and is available for download and test.

TABLE I: Comparison of models for IT neural activity prediction.

#	Model	Mean EV	Mean Corr.
Week 6 – Linear from Pixels			
1	OLS	−0.037	0.220
2	Ridge (α = 0.1)	−0.921	0.164
3	PCA (1000) + OLS	−0.077	0.216
4	PCA (1000) + Ridge (best α)	−0.088	0.215
Week 6 – Task-driven (ResNet50)			
5	Pretrained ResNet Layer3 + OLS	0.398	0.623
6	Pretrained ResNet Layer3 + Ridge	0.398	0.623
7	Random ResNet Layer3	−537.1	0.034
Week 7 – Data-driven CNN			
8	Shallow CNN	0.108	0.350
Week 8 – ConvNeXt			
9	ConvNeXt (linear)	−0.102	0.381
10	Deep ConvNeXt head	0.253	0.497
Week 8 – ResNet Variants			
11	ResNet50 + Deep Head	0.219	0.482
12	**ResNet50 (Best)**	**0.472**	**0.678**

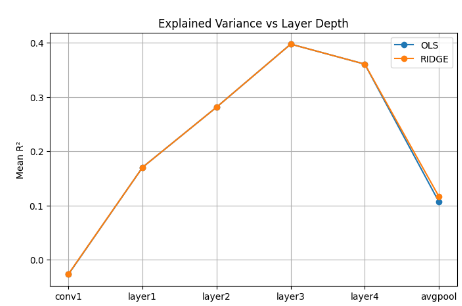

Fig. 1: Plotting the explained variance depending on the depth of the network, making it visibly clear that layer 3 encodes the most information about the data