Mini Project (Predicting Neural Activity)

***Abstract*—This project explores the prediction of neural activity from images using task-driven and data-driven approaches. The metrics used during this project to compute the goodness of the model were explained variance (EV) and correlation (Corr) between real and predicted values.**

I. Introduction

We analyze pre-processed neural recordings from rhesus macaques' inferior temporal (IT) cortex during object recognition tasks (Majaj et al., 2015). The data include firing rates (70–170 ms post-stimulus) recorded as monkeys viewed images also shown to humans. We model how visual stimuli drive IT responses based on findings that linear decoders can predict human performance.

II. Predict neural activity from pixels

We first attempted to predict neural activity from raw pixel values using ridge regression. This approach yielded poor results (EV = 0.0953, Corr = 0.2932), likely due to the high dimensionality of the input and the limited expressiveness of a linear model.

We applied PCA to reduce dimensionality and retained the first 1000 components (explaining 98% of the variance), but this only slightly affected performance (EV = 0.0948, Corr = 0.2922). We then performed cross-validation to optimize the regularization parameter, with the best result at $\alpha = 5$. Retraining with this value led to a small improvement (EV = 0.0968, Corr = 0.2931).

These results show that pixel-based features, even after dimensionality reduction and tuning, fail to capture the structure needed to predict IT neural activity effectively.

III. Task-driven approach

To evaluate whether task-driven representations align with biological neural responses, we used a pre-trained ResNet50 and extracted layer-wise activations for the stimulus images. Each set of features was reduced via PCA (first 1000 components) and used to predict IT neural activity with ridge regression.

Prediction performance varied across layers: conv1 performed poorly, while layer3 achieved the best results (EV = 0.2598, Corr = 0.5213), suggesting that mid-level features best match IT responses. Deeper layers like layer4 and avgpool showed a decline, likely due to increased abstraction and loss of task-irrelevant variance.

The pretrained ResNet50 consistently outperformed the randomly initialized model across all layers, with neuron-wise metrics showing higher explained variance and correlations. Overall, mid-level layers of the trained model best predicted neural activity, suggesting these representations align closely with brain function.

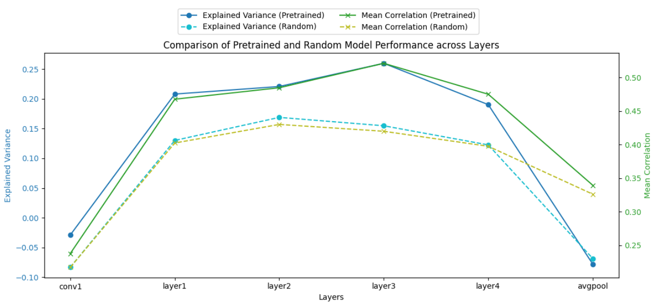

Figure 1: Prediction performance across ResNet50 layers using explained variance and correlation.

IV. Data-driven approach

To assess whether neural activity could be predicted by learning directly from data, we trained a shallow CNN end-to-end on the stimulus images and IT responses. The model reached an overall explained variance of 0.1732 and correlation of 0.4002, outperforming pixel-based approaches but falling short of the best task-driven result.

Although it did not match the peak performance of ResNet50's layer3 (layer 3: explained variance of 0.2598 and correlation of 0.5213), the results show that learning directly from neural data allows the model to capture meaningful structure, even without prior training on external tasks.

V. Final challenge: Toward an optimal model

To improve on the shallow CNN, we tested several optimizations. Adjusting normalization and the learning rate scheduler modestly improved performance (EV = 0.1961, Corr = 0.4362). However, reducing the dropout rate to 0.2 led to a small decline (EV = 0.1801, Corr = 0.4196) and a more unstable training. Adding L2 regularization (weight_decay) helped stabilize training and slightly improved performance (EV = 0.2081, Corr = 0.4526).

We then added a convolutional layer, which led to another slight improvement in terms of our correlation metric (EV = 0.2043, Corr = 0.4620), but the model remained limited. This motivated us to explore fine-tuning pre-trained models.

Among CNNs, our optimized EfficientNet-B5 performed best (EV = 0.4642, Corr = 0.6706), followed by ResNet50 (EV = 0.4231, Corr = 0.6379). In contrast, vision transformers underperformed: for example, ViT Base gave weak results (EV = 0.0569, Corr = 0.2196).

We also trained multi-objective models combining classification and neural prediction using ResNet50. Giving equal weight to both yielded moderate results (EV = 0.1988, Corr = 0.4348). Prioritizing the neural objective (weight = 1 vs. 0.3) led to a clear improvement (EV = 0.3156, Corr = 0.5506), suggesting that direct optimization for neural alignment is beneficial for the model, but combining classification is not (ResNet50 trained on neural prediction gave better results than any multi-objective model).

Finally, we examined the pretrained weights of various models, following a task-driven approach similar to Week 6's analysis of ResNet-50. Among pretrained models, the best layer of DenseNet-201 achieved EV = 0.3929 and Corr = 0.6253, slightly outperforming ResNet-152 (EV = 0.3838, Corr = 0.6096) and PNASNet-5 Large (EV = 0.3812, Corr = 0.6179), and significantly outperforming ResNet-50, our original model (EV = 0.2598, Corr = 0.5213).

When fine-tuned on neural activity, **DenseNet-201 achieved EV = 0.4937 and Corr = 0.6971**. This relatively strong performance, even with linear regression on pretrained layers, suggests that DenseNet's architecture inherently aligns with neural representations, possibly due to structural similarities with brain activity patterns.

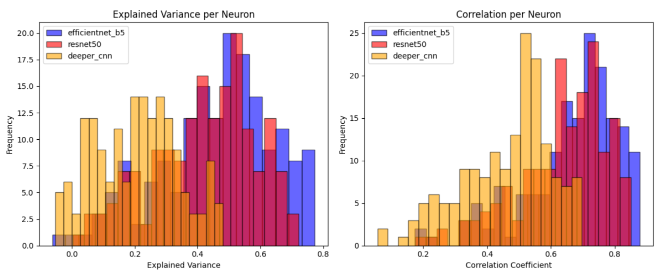

Figure 2: Comparison of performance between some models of the CNN family. DeeperCNN is the best of our shallow CNN models. Overall, EfficientNet-B5 gives the best performance.

VI. Conclusion

Neural activity in the visual cortex can be predicted from both a task-driven and a data-driven approach, or both. Among all models tested, the task-driven approach, finetuned **DenseNet-201 model achieved the highest performance**.

References

Majaj, Najib J. et al. (2015). "Simple learned weighted sums of inferior temporal neuronal firing rates accurately predict human core object recognition performance". In: *The Journal of Neuroscience* 35.39, pp. 13402–13418. DOI: 10.1523/JNEUROSCI.5181-14.2015.

Appendix

Name	EV Score	Correlation Score
Task driven: pretrained layers		
ResNet50 layer 3	0.2598	0.5213
DenseNet201 denseblock4	0.3929	0.6253
ResNet152 layer3	0.3838	0.6096
PNASNet-5 Large cell 5	0.3820	0.6171
Shallow CNN		
Week 7 base	0.1732	0.4002
Improved version	0.2081	0.4526
Deeper CNN	0.2043	0.4620
Finetune existing model		
Densenet201	**0.4937**	**0.6971**
Pnasnet5large	0.4894	0.6890
EfficientNet-B5	0.4642	0.6706
ResNet50	0.4231	0.6379
Multi objective models		
Equal weights objectives (1 against 1)	0.1988	0.4348
Prioritize neural objective (1 against 0.3)	0.3156	0.5506

Table I: EV and Correlation Scores by Section

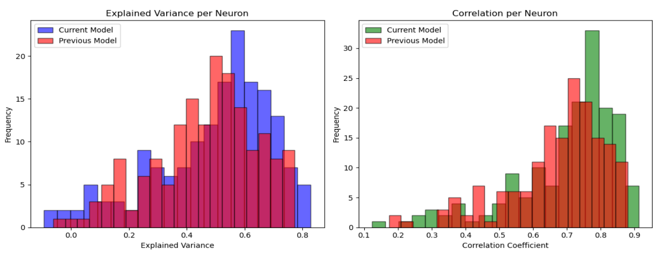

Figure 3: Comparison of performance between EfficientnetB5 (in red) and DenseNet (in blue/red). The latter outperforms the first. DenseNet is our best model!

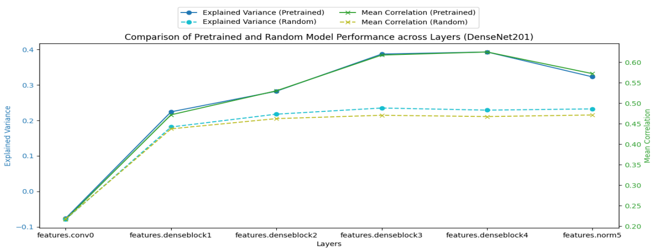

Figure 4: Comparison of different layers and for pretrained/randomly initialized weights of the DenseNet model. We notice that the layers 3 and 4 of the model are the best at predicting neural activity.

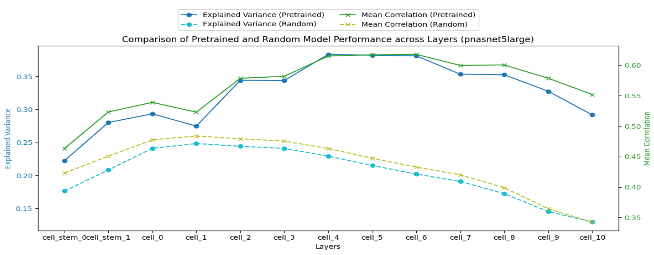

Figure 5: Comparison of different layers and for pretrained/randomly initialized weights of the pnasnet5large model. We notice that the layers around the middle of the model are the best at predicting neural activity.