Mini-project report: Predicting IT neural activity

CÉLIA BENQUET, NICOLAS REATEGUI
NX-414 - Brain-like computation and intelligence, EPFL, Switzerland

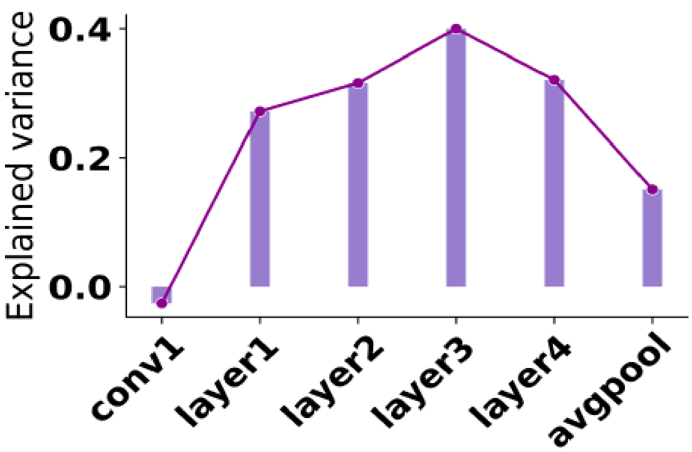

Figure 1: Variation in neural prediction variation across the activations of a ResNet-50 pre-trained model.

In the following project, we aim to predict the inferior temporal (IT) neural activity of non-human primates while they are presented with different natural images. We compared linear, task-driven, and data-driven models and found that a data-driven model fine-tuned on a pre-trained ResNet-152 model could explain neural variance the best.

a) Task-driven models: By initially training a hierarchical convolutional neural network on an image classification task, given the proper inductive bias, the network can develop activations similar to areas of the visual cortex. Hence, for all models tested, we train a linear regression on the first thousand principal components (PCs) of each activation layer. First, we extracted the activations from a ResNet-50 architecture. The pre-trained model performs better than a randomly-initialized model (Table I-Task-driven) and layer 3 of the network explains the most neural variance (see Figure 1). To boost performances, we fine-tuned the ResNet-50 model to classify the images from our specific dataset and extracted the fine-tuned activation layers to predict neural activity. Performance (not shown) did not improve much. The small performance increase might be due to the pre-trained model having learned natural image representations that can already generalize well on unseen categories. Next, we considered Schrimpf and colleagues' work on developing Brain-Score — a composite of neural and behavioral benchmarks evaluating how similar artificial neural networks are to the brain's mechanisms for core object recognition. We selected DenseNet-169 [1], which reaches the highest Brain-Score of all evaluated models [2]. Performances (see Table I-Ours) are better than the ones obtained with ResNet-50.

b) Data-driven models: We first implemented a simple CNN (see Table I–Data-driven) to predict neural activity from the full images. Then, to increase performances, we explored different pre-trained state-of-the-art computer vision models, i.e., ResNet [3] (18, 50), DenseNet [1] (121, 169), VGGNet [4] (VGG16) and EfficientNet [5] (B6) ranging from 11M up to 138M parameters. As those models are originally trained on a classification task on ImageNet, we replaced the last linear layer so that it predicts spike activation rather than image classes. For all models, results were better when fine-tuning the entire model rather than only the output layer. On top of that, we performed a grid search on the learning rate and optimizer, used a `OneCycle` learning rate scheduler with a warm-up to improve convergence, and implemented an early stopping scheme to stop the training after 20 consecutive epochs showing no improvement.

c) Hybrid approach: We also tested fine-tuning Resnet-50 on both the image classification and the spike prediction tasks. For that, we customize the loss so that it is the sum of the cross-entropy and mean standard error losses and replaced the original output layer with two custom output layers where one predicts one of the 65 image classes from our dataset and the other predicts spike neuron activity. Although this new architecture was able to converge we weren't able to improve our previous results.

	Model	R^2 **score** (↑)	**MSE** (↓)	**Exp.** σ^2 (↑)
Linear regression	Linear regression + PCA	-0.090	0.149	-0.086
	Ridge regression + PCA	-0.090	0.149	-0.086
Task-driven	Random ResNet-50 - Layer 3	0.133	0.116	0.137
	Pre-trained ResNet-50 - Layer 3	0.401	0.0784	0.404
Data-driven	2-layers CNN	0.238	0.102	0.244
	3-layers CNN	0.251	0.099	0.259
Ours	ResNet-152 (Data-Driven)	**0.440**	**0.063**	**0.453**
	DenseNet-169 (Task-Driven)	0.422	0.0721	0.424

TABLE I

COMPARISON TABLE. WE COMPARE PERFORMANCES BY EVALUATING THE COEFFICIENT OF DETERMINATION (R^2), MEAN STANDARD ERROR (MSE), AND NEURAL EXPLAINED VARIANCE (EXP. σ^2). THE BEST PERFORMING MODEL WAS OBTAINED BY FINE-TUNING A PRE-TRAINED DENSENET-169 MODEL, LEARNING RATE 1E-03, 96 EPOCHS, SCHEDULER `ONECYCLELR`, OPTIMIZER `NADAM`

References

[1] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," arXiv, Aug. 2016.

[2] M. Schrimpf, J. Kubilius, H. Hong, N. J. Majaj, R. Rajalingham, E. B. Issa, K. Kar, P. Bashivan, J. Prescott-Roy, F. Geiger, K. Schmidt, D. L. K. Yamins, and J. J. DiCarlo, "Brain-Score: Which Artificial Neural Network for Object Recognition is most Brain-Like?," bioRxiv, p. 407007, Jan. 2020.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," arXiv, Dec. 2015.

[4] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv, Sept. 2014.

[5] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," arXiv, May 2019.