Performance Optimization is Insufficient for Building Accurate Models for Neural Representation

Jonathan Yu[1] Qihong Lu[2] Uri Hasson[2] Kenneth A. Norman[2] Jonathan W. Pillow[2]

[1]Princeton University Computer Science Department [2]Princeton Neuroscience Institute

Summary

Previous studies have found a strong correlation between the optimization of the image classification performance of machine learning models and greater visual neural representation (Fig. 1). However, our results indicate that performance optimization may eventually lead to systematic deviation from human brain representation. This suggests that additional constraints informed by visual neuroscience are critical for building better computational models.

Background

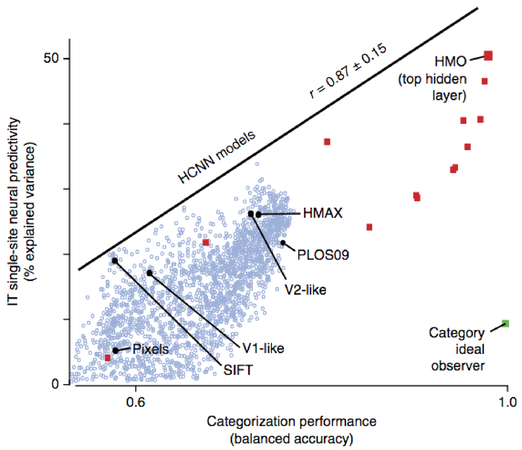

Figure 1: Performance optimization leads to more predictive models of the visual neural pathway [3].

Hypothesis

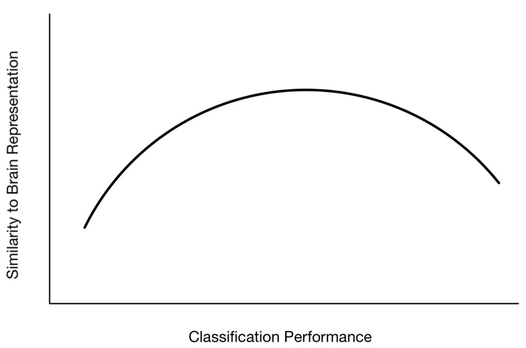

Figure 2: Hypothesized inverted-U-shaped relationship between classification performance and similarity to brain representation.

Results

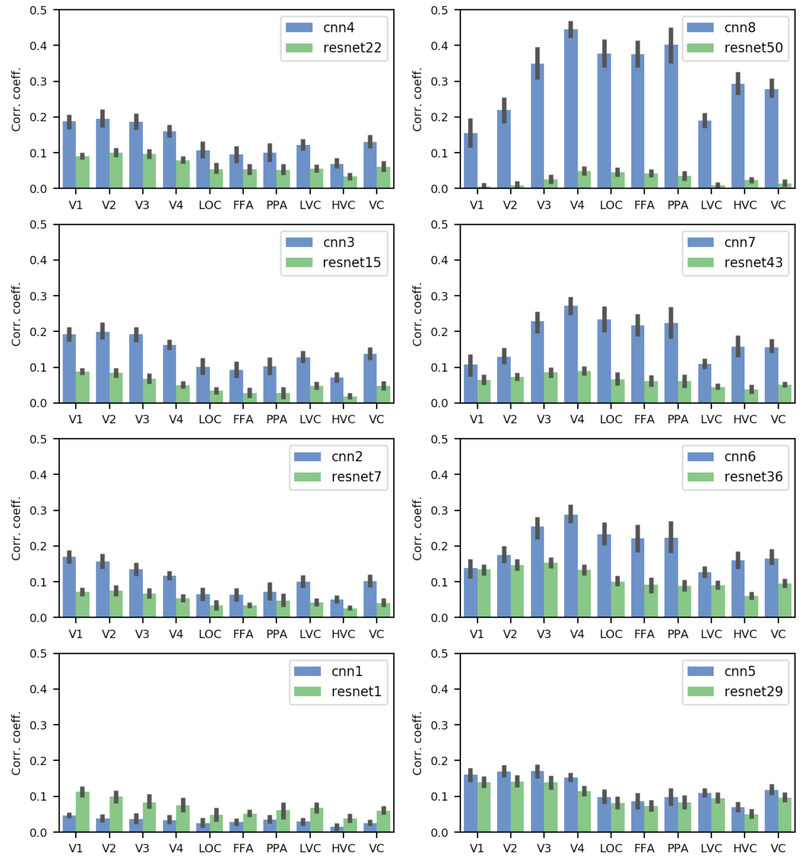

Figure 3: The correlation between fMRI data from selected Regions of Interest and the hidden states of AlexNet (cnn) and Resnet-50 (resnet). The indices represent layer numbers. For ResNet-50, we chose 8 roughly evenly-spaced layers across the network architecture. fMRI data were collected using an experiment in which participants performed a one-back task on ImageNet images [2].

AlexNet vs. ResNet-50

ImageNet-Trained Models

- AlexNet: an 8-layer convolutional neural network
- ResNet-50: a 50-layer residual neural network

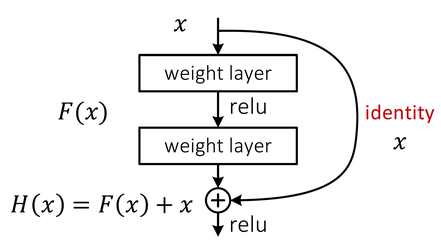

Figure 4: A "skip connection" adds the activity vector of the first layer to the third layer in this residual block.

Classification Performance

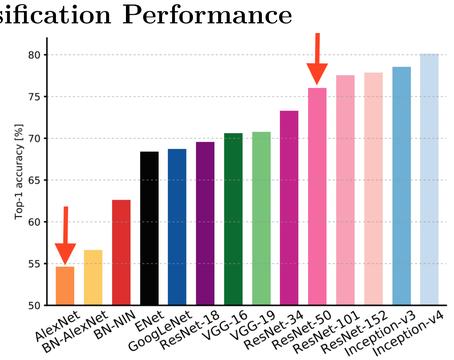

Figure 5: ResNet-50 is significantly more optimized than AlexNet on ImageNet data [1].

Predicting Neural Network Activity

Linear Mapping

- For both models, 1000 random activation units were chosen from 8 evenly-spaced layers.
- Linear models were trained to map the fMRI responses to the hidden state of each unit chosen from AlexNet and ResNet-50 (e.g., Fig. 6).

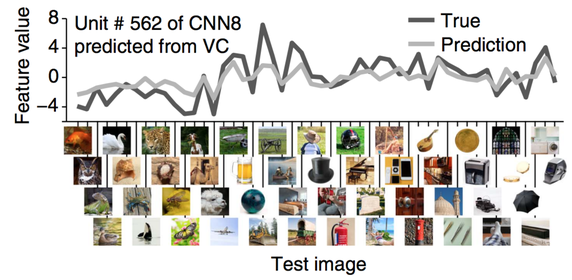

Figure 6: An example activation unit from cnn8 predicted by the fMRI responses from all Regions of Interest combined [2].

- The linear prediction performance of ResNet-50 is significantly worse than that of AlexNet for 7 of 8 layer-to-layer comparisons (Fig. 3).

Conclusion

- ResNet-50 is an inferior model for visual neural representation, in comparison to AlexNet.
- Performance optimization can lead to deviation from brain representation, especially when the model exceeds human-level performance.
- Performance optimization alone is insufficient for building accurate computational models of the brain.

References

[1] A. Canziani, A. Paszke, and E. Culurciello. An analysis of deep neural network models for practical applications. May 2016.

[2] T. Horikawa and Y. Kamitani. Generic decoding of seen and imagined objects using hierarchical visual features. *Nat. Commun.*, 8:15037, May 2017.

[3] D. L. K. Yamins and J. J. DiCarlo. Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.*, 19(3):356–365, Mar. 2016.