High-dimensional sampling in random neural networks competes with deep learning models of visual cortex

Atlas Kazemian (akazemi3@jh.edu)
Department of Cognitive Science, Johns Hopkins University
Baltimore, MD 21218 United States of America

Eric Elmoznino (eric.elmoznino@gmail.com)
Department of Cognitive Science, Johns Hopkins University
Baltimore, MD 21218 United States of America

Michael F. Bonner (mfbonner@jh.edu)
Department of Cognitive Science, Johns Hopkins University
Baltimore, MD 21218 United States of America

Abstract

The performance of convolutional neural networks (CNNs) as representational models of visual cortex is thought to be associated with their optimization on ethologically relevant tasks. Contrary to this view, we show that a surprisingly simple statistical principle based on high-dimensional sampling of random features is sufficient to induce brain-like representations in neural network models of visual cortex. Specifically, we constructed CNNs that perform random dimensionality expansion and found that fewer than a thousand features are needed to compete with standard supervised networks at predicting the feature-tuning preferences of primate visual cortex, avoiding the need for massive pre-training or task-specific optimization. Furthermore, we found that when random expansion is followed by dimensionality reduction, the dominant modes of variation correspond to brain-relevant dimensions. In fact, random-expansion CNNs remain competitive with standard pre-trained CNNs even when matching their dimensionalities. Remarkably, this means that brain-relevant dimensions are readily discoverable from the statistics of image activations in random convolutional architectures. These findings reveal the unexpected effectiveness of random expansion in neural network models of vision, and they point toward a simplifying statistical theory of cortical visual representation.

Keywords: convolutional neural networks; deep learning; fMRI; encoding models; random networks

Results

We propose that a neural network with a few design constraints and access to a high dimensional sampling space will be able to predict stimulus-evoked cortical representations without the need for pre-training on a computer vision task (Elmoznino & Bonner, 2022). To demonstrate this, we develop a model using only the following two operations: convolution and max pooling. By stacking these operations in a deep hierarchy, we design a simple 3-layer network that extracts complex features from input images without relying on pre-trained parameters. The first layer of the model convolves images with Gabor filters to efficiently capture low-level stimulus information, as previously done in classic hand-engineered models (Riesenhuber & Poggio, 1999). All the following layers make use of many randomly initialized filters to search for higher-level stimulus-relevant information. To study the effect of the dimensionality of the space considered, we evaluate the encoding performance of models with varying numbers of random filters in their third layer.

We evaluated how well CNNs performed when predicting image-evoked cortical responses in monkey electrophysiology data (Majaj et al, 2015) and human fMRI data (Allen et al, 2021). For our initial investigations, we sought to specifically examine the high-dimensional sampling of features rather than spatial scales. We therefore created CNNs in which the spatial scale of the convolutional filters was fixed but the number of filters varied, and we applied global max pooling to the final convolutional layer to abstract over spatial dimensions. To evaluate performance, feature vectors from the last convolution layer of each model were mapped to cortical responses using a simple regression procedure, which we validated on held-out test data. The encoding score of each model was obtained by measuring the correlation between model-predicted and actual neural responses (Figure 1).

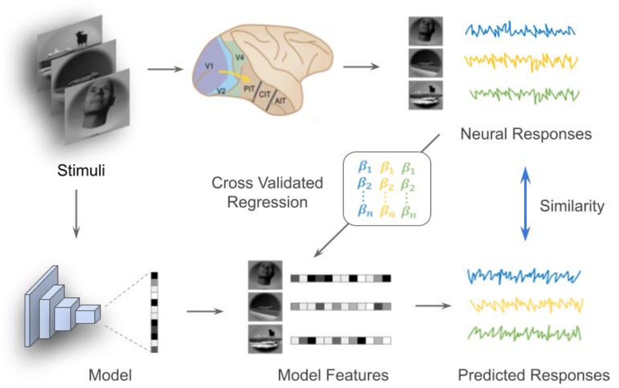


Figure 1: Model evaluation framework

Figure 2 shows the mean voxel-wise encoding score of random-expansion CNNs as a function of the number of random features. For comparison, we also show the performance of layer 5 from ImageNet-pre-trained AlexNet and untrained AlexNet. These results show that in high-level visual regions random CNNs perform strikingly well at predicting image representations, and they improve as the number of features in the expansion layer increases. In follow-up analyses, we found that these results could not be trivially obtained with any high-dimensional set of random features. In fact, when we simply replaced max pooling with average pooling—while keeping the number of features fixed—we found that performance dropped dramatically. Moreover, we found that it is not necessary to retain all features in the expansion layer. As a demonstration, we used principal component analysis to reduce the dimensionality of random CNN to 256 dimensions, which matches the dimensionality of AlexNet layer 5. Remarkably, even this reduced random CNN performs comparably to pre-trained AlexNet.

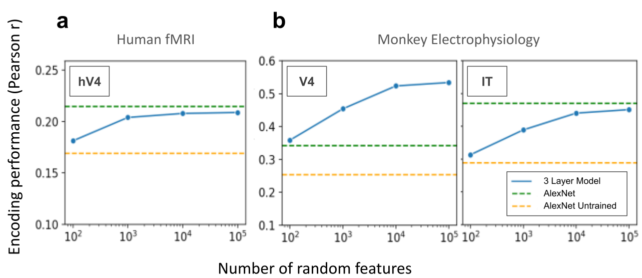


Figure 2: Encoding performance as a function of the number of random features. The performance of pre-trained and untrained AlexNet are shown for comparison. (a) Performance on the Natural Scenes fMRI study (region hV4). Correlations are averaged across voxels and subjects. (b) Performance on the monkey electrophysiology study (regions V4 and IT). Correlations are averaged across units and animals.

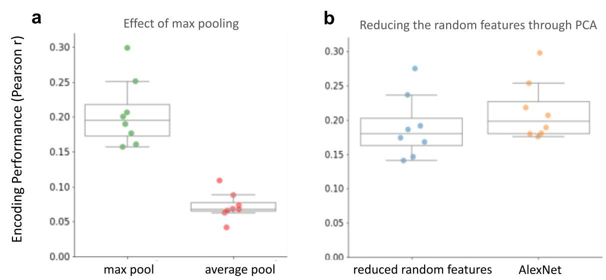


Figure 3: Effects of max pooling and dimensionality reduction. Each point refers to the mean Pearson r correlation across voxels for one subject from the fMRI study (region hV4). (a) There is a strong difference in performance for models with max pooling vs. average pooling . (b) There is little effect on performance when reducing the number of random features through PCA.

Discussion

Our work has several implications for computational neuroscience. First, we show that task optimization is not required for generating brain-like representations in neural networks. Following the development of diverse models and learning objectives in deep learning, it is becoming increasingly evident that encoding models of the visual cortex are degenerate with respect to architectural details (Conwell *et al,* 2021) and training paradigms (Zhuangl *et al,* 2021). Here we show that another factor— task optimization—which is commonly believed to be central to the high performance of these models, does not uniquely explain their success. We have demonstrated this by replacing task optimization with high-dimensional random sampling. Second, by limiting the training process to the last layer's parameters during linear mapping, our approach has significant computational benefits. For example, high-throughput ablation studies can be conducted to explore various components in isolation. For instance, to perform the same study as in Figure 3a using pre-trained models, one would have to re-train all model parameters, which would require massive computational resources. Lastly, we suggest that our approach can aid our understanding of the underlying properties governing the representations of computational models. Given our findings, we suggest shifting the focus away from studying the architectural and training details of any one family of models and towards exploring the nature of general statistical principles that underlie high-performance encoding models of cortical representation.

References

Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., Nau, M., Caron, B., Pestilli, F., Charest, I., Hutchinson, J. B., Naselaris, T., & Kay, K. (2021). A massive 7T fmri dataset to Bridge Cognitive Neuroscience and artificial intelligence. *Nature Neuroscience*, *25*(1), 116–126. https://doi.org/10.1038/s41593-021-00962-x

Conwell, C., Prince, J. S., Alvarez, G. A., & Konkle, T. (2022). Large-scale benchmarking of diverse artificial vision models in prediction of 7T human neuroimaging data. https://doi.org/10.1101/2022.03.28.485868

Elmoznino, E., & Bonner, M. F. (2022). High-performing neural network models of visual cortex benefit from high latent dimensionality. https://doi.org/10.1101/2022.07.13.499969

Frankle, J., & Carbin, M. (2019). The lottery ticket hypothesis: Finding sparse, trainable neural networks. arXiv.org. Retrieved April 3, 2023, from https://arxiv.org/abs/1803.03635

Majaj, N. J., Hong, H., Solomon, E. A., & DiCarlo, J. J. (2015). Simple learned weighted sums of inferior temporal neuronal firing rates accurately predict human core object recognition performance. The Journal of Neuroscience, 35(39), 13402–13418. https://doi.org/10.1523/jneurosci.5181-14.2015

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*(11), 1019–1025. https://doi.org/10.1038/14819

Zhuang, C., Yan, S., Nayebi, A., Schrimpf, M., Frank, M. C., DiCarlo, J. J., & Yamins, D. L. (2020). Unsupervised neural network models of the ventral visual stream. https://doi.org/10.1101/2020.06.16.155556

High dimensional sampling in random neural networks competes with deep learning models of visual cortex

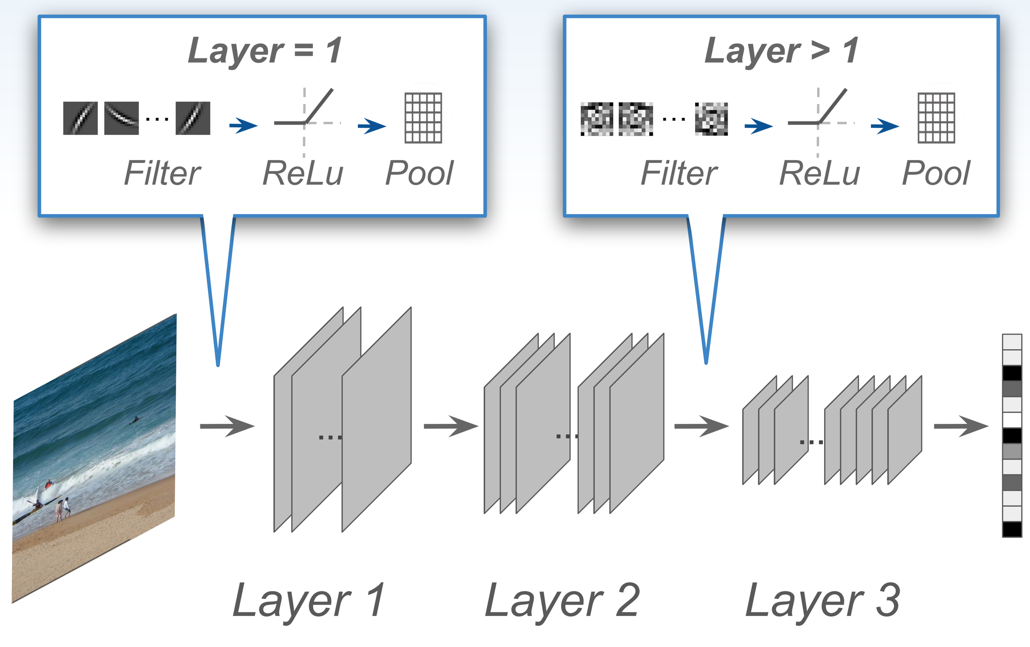
Atlas Kazemian, Eric Elmoznino & Michael F. Bonner

MOTIVATION

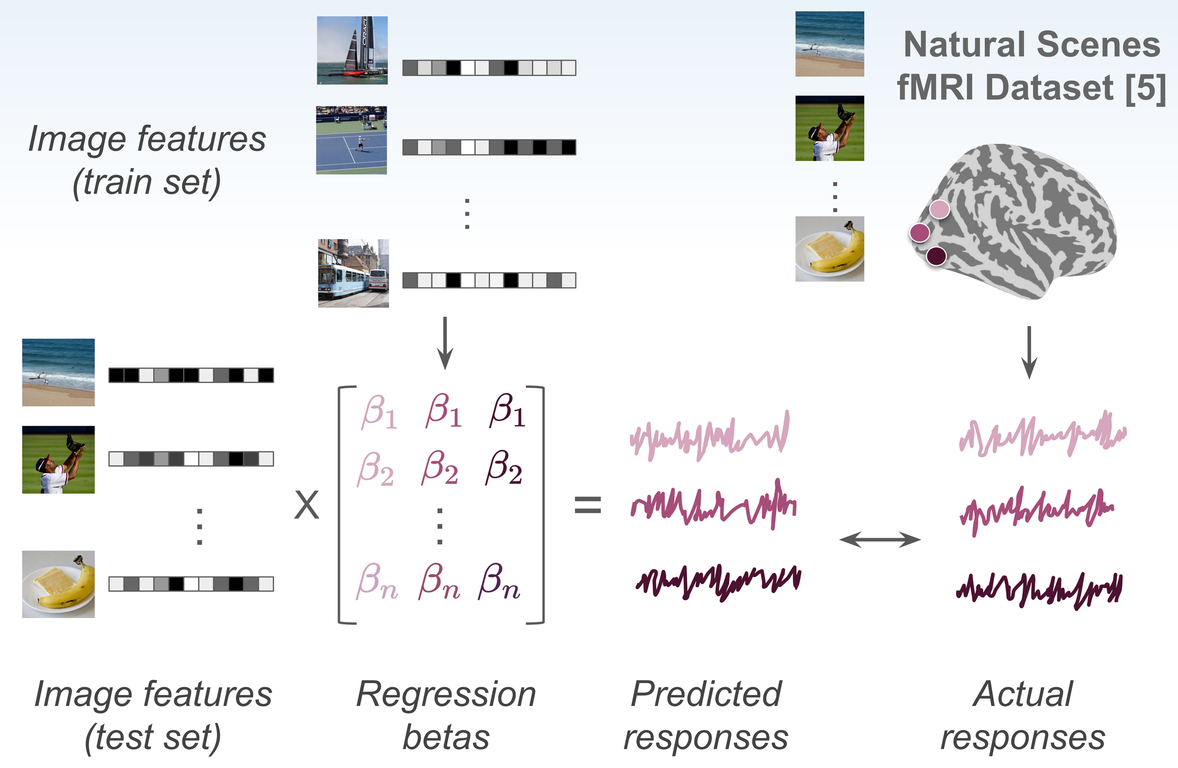
- The performance of convolutional neural networks (CNNs) as models of visual cortex is thought to be associated with task optimization [1-3].
- We hypothesize that simpler statistical principles (such as representational richness [4]) are responsible for much of this performance.
- To test this, we develop a learning-free CNN based on high dimensional random sampling and compare its performance with a pre-trained CNN.

METHODS

Expansion model architecture

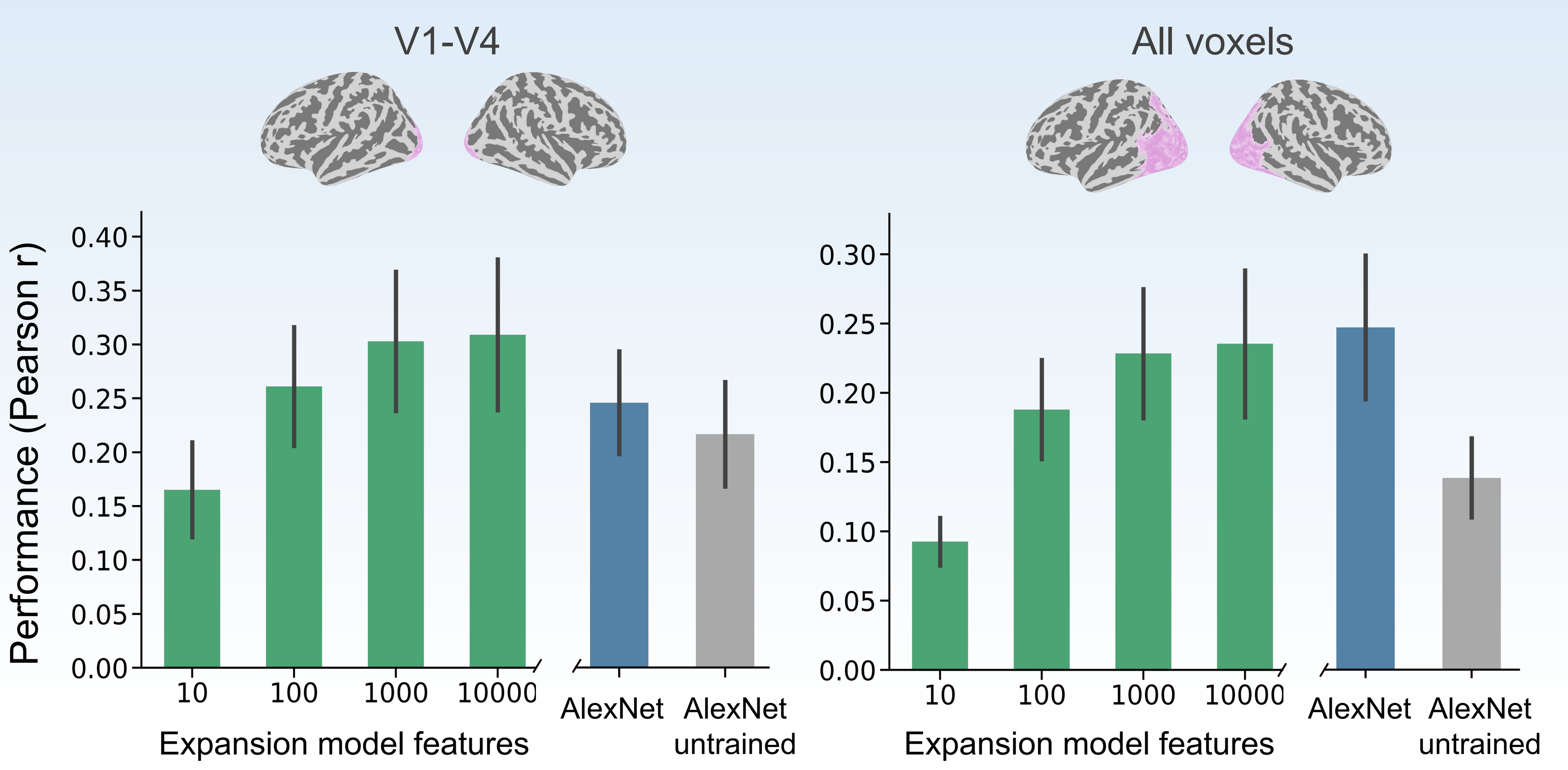


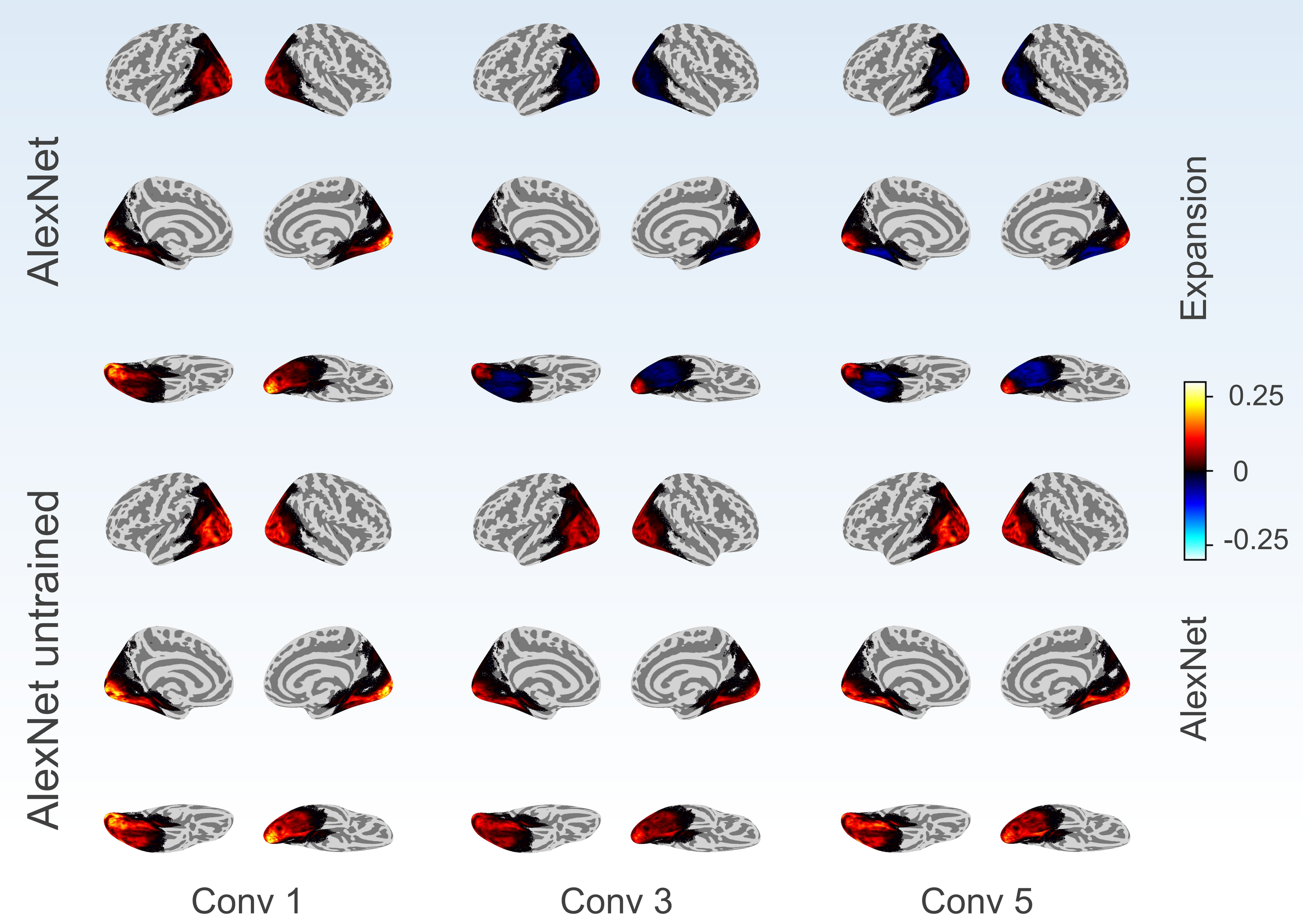
Mapping procedure



RESULTS

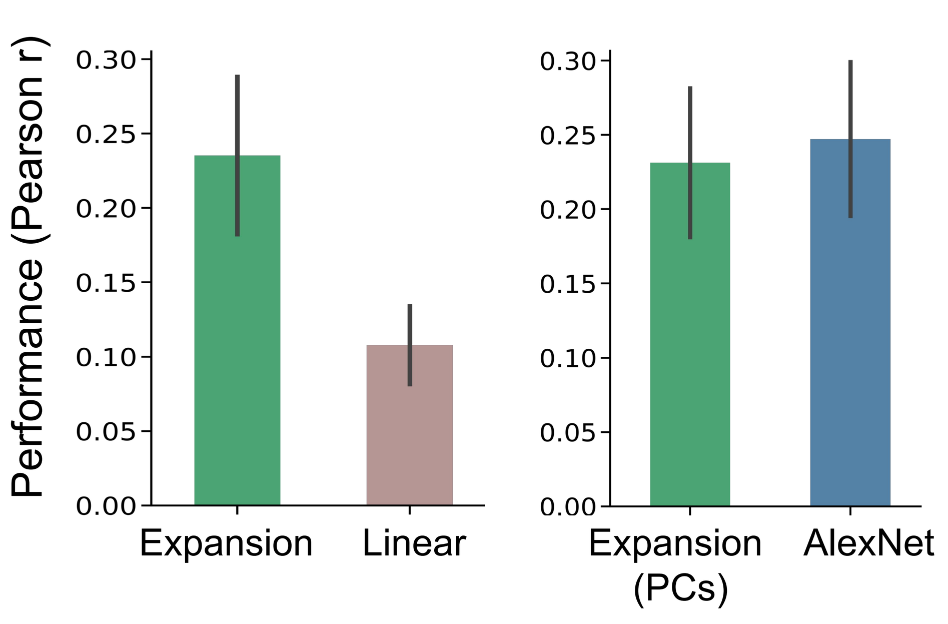
The Expansion model competes with a pre-trained CNN at modeling visual cortex



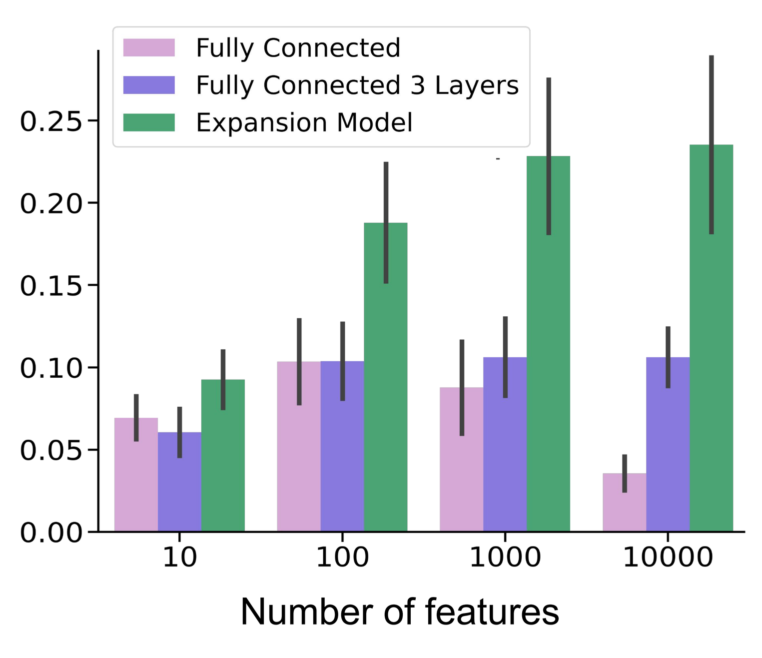


Q What factors drive encoding performance?

High dimensional regression *does* ***not*** *drive performance*

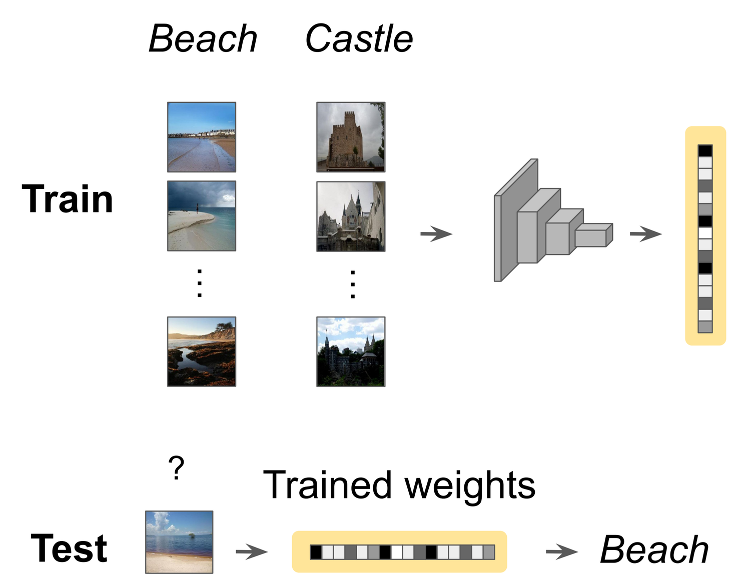


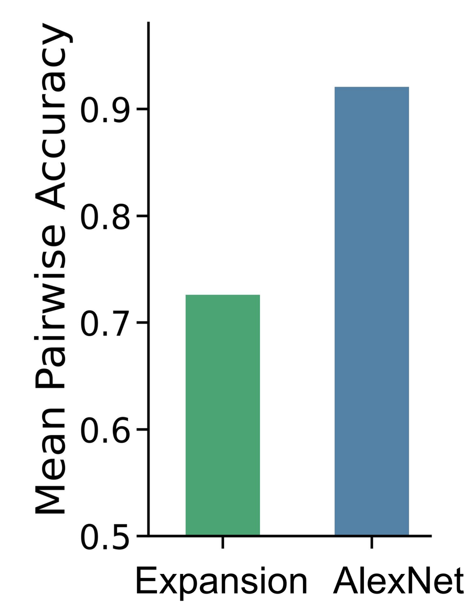
Convolutional architecture *is needed*



Q Is the model good at image classification?

Not all high performing models of visual cortex are good at computer vision tasks





CONCLUSION

- An untrained model that expands the dimensionality of representations competes with a pre-trained CNN at modeling visual responses.
- The performance of the model is mainly associated with the convolutional architecture and high dimensional random sampling.
- The model's low image classification performance shows that this metric does not always correlated with brain-similarity.
- Our results suggest that there may be simplifying statistical principles underlying the image representations of deep neural networks and visual cortex.

REFERENCES

[1] Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, *111*(23), 8619–8624. https://doi.org/10.1073/pnas.1403112111

[2] Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, *19*(3), 356–365. https://doi.org/10.1038/nn.4244

[3] Cao, R., & Yamins, D. (2021b). Explanatory models in neuroscience: Part 2 – constraint-based intelligibility. arXiv. Retrieved from https://arxiv.org/abs/2104.01489 doi: 10.48550/ ARXIV.2104.01489

[4] Elmoznino, E., & Bonner, M. F. (2022). *High-Performing Neural Network Models of Visual Cortex Benefit from High Latent Dimensionality*. https://doi.org/10.1101/2022.07.13.499969

[5] Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., Nau, M., Caron, B.,Pestilli, F., Charest, I., Hutchinson, J. B., Naselaris, T., & Kay, K. (2021). A massive 7T fmri dataset to Bridge Cognitive Neuroscience and artificial intelligence. Nature Neuroscience, 25(1), 116–126. H ttps://doi.org/10.1038/s41593-021-00962-x