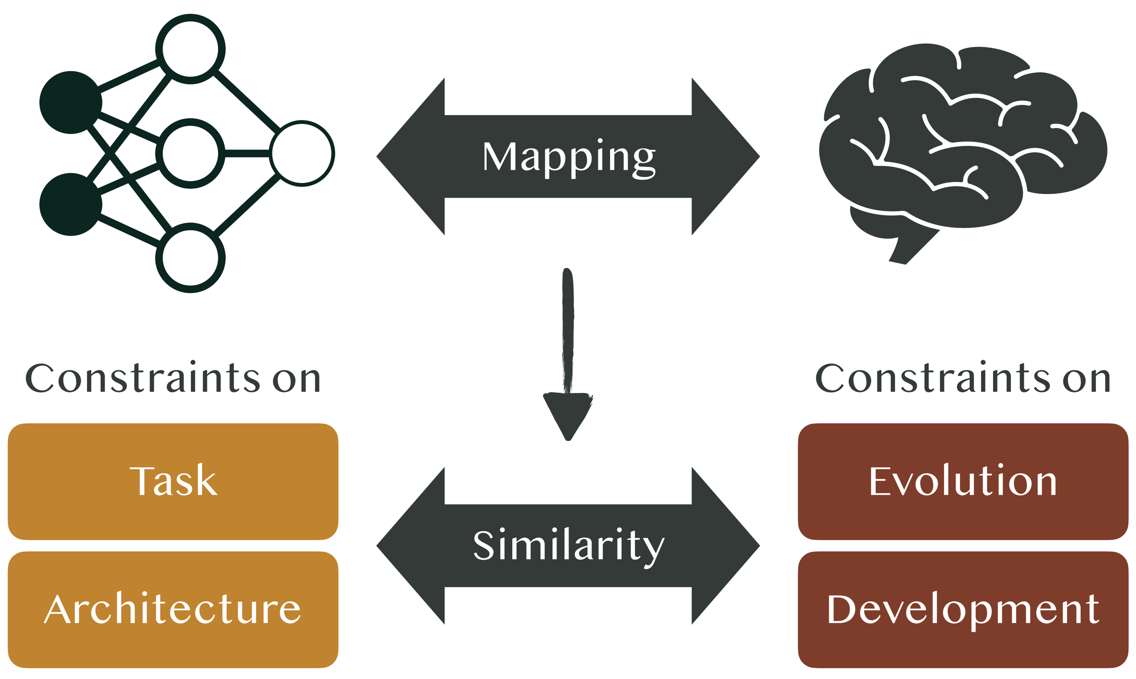
Canonical Dimensions of Vision

Zirui Chen, Michael Bonner
Department of Cognitive Science, Johns Hopkins University

TODO: download link

Motivation

Neuroscientists attributed representational similarity between human brain and deep neural network (DNN) to similarities in the optimization constraints on these systems[1,2].



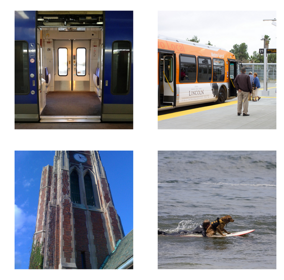
However, studies showed that recent DNNs of different designs do not yield significant variation in brain similarity[3,4,5].

Do DNNs learn constraint-independent, "canonical" features?

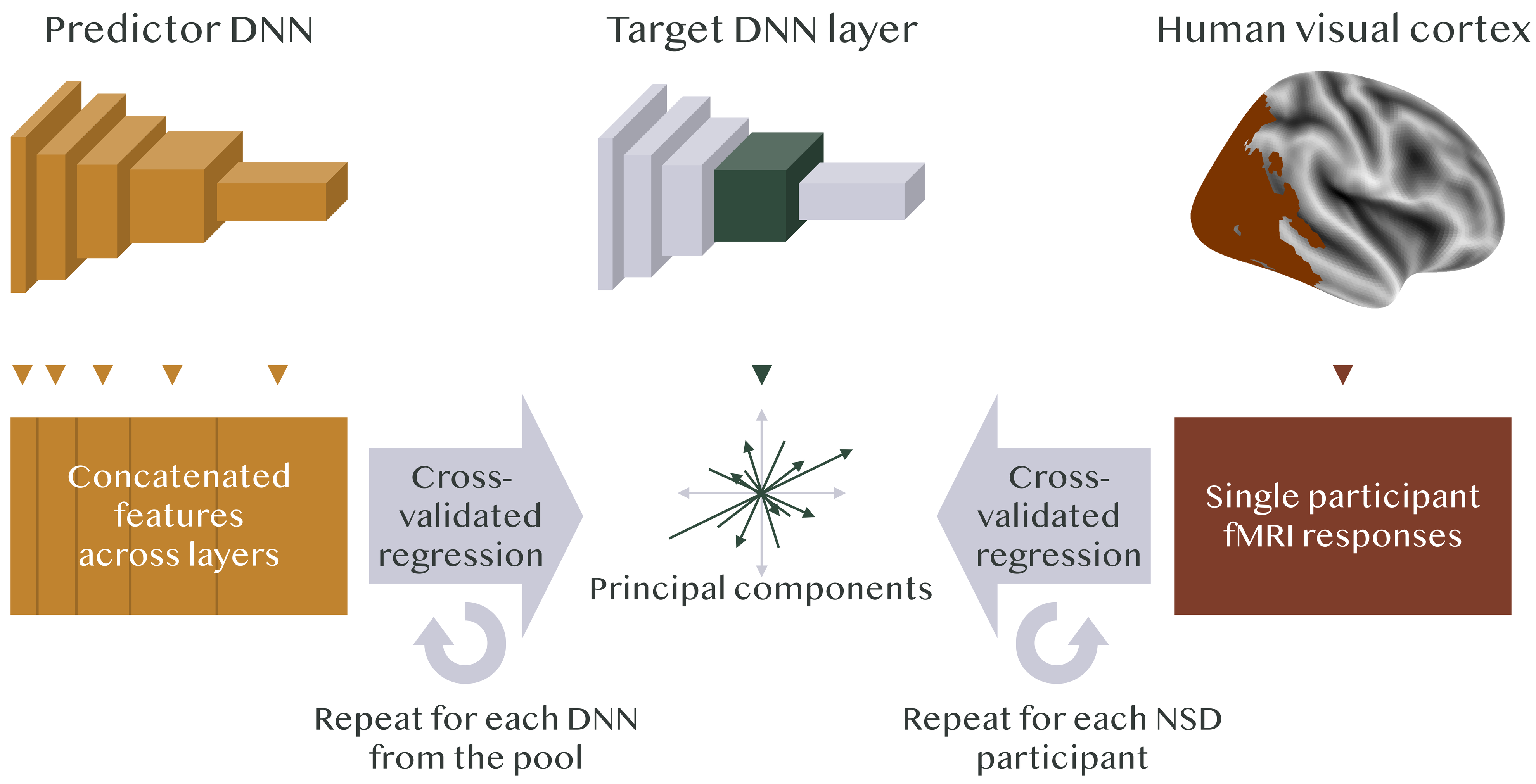
Are these canonical dimensions also encoded in human visual cortex?

Methods

Stimuli and neural data



fMRI responses of 8 participants to 872 color natural scene images from the Natural Scenes Dataset[6] (NSD)



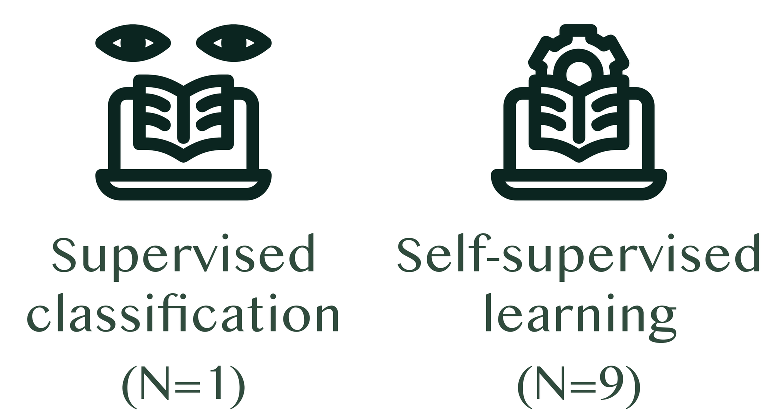
Canonical strength = average prediction accuracy across DNN predictors

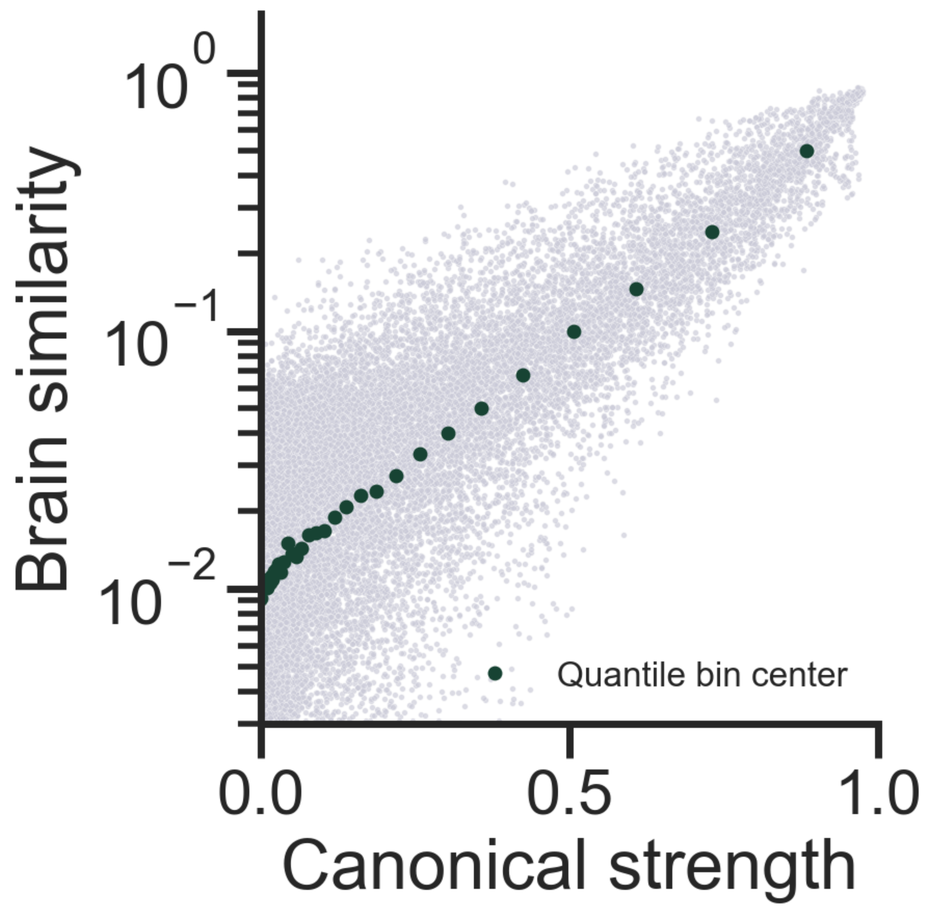
Brain similarity = average prediction accuracy across NSD participant predictors

Results & Discussion

1 Are biological-relevant visual features constrained by **training tasks**?

- DISTINCT tasks
- Same architecture (Resnet-50[7])
- Same training data (ImageNet[8])
- 106,889 features

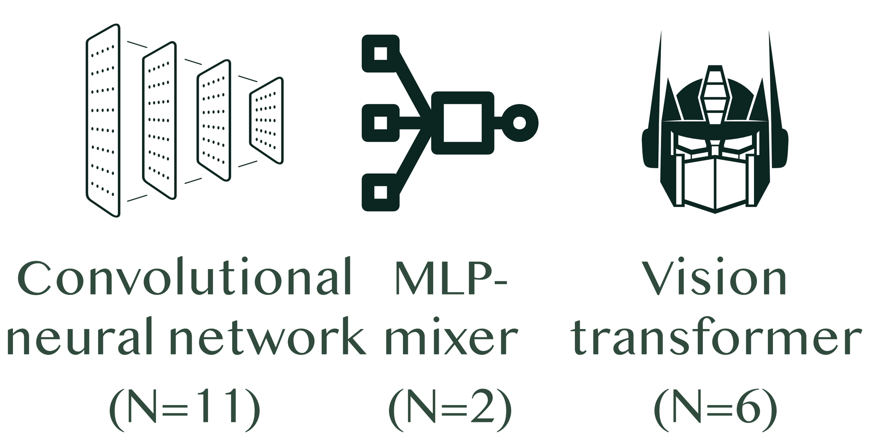


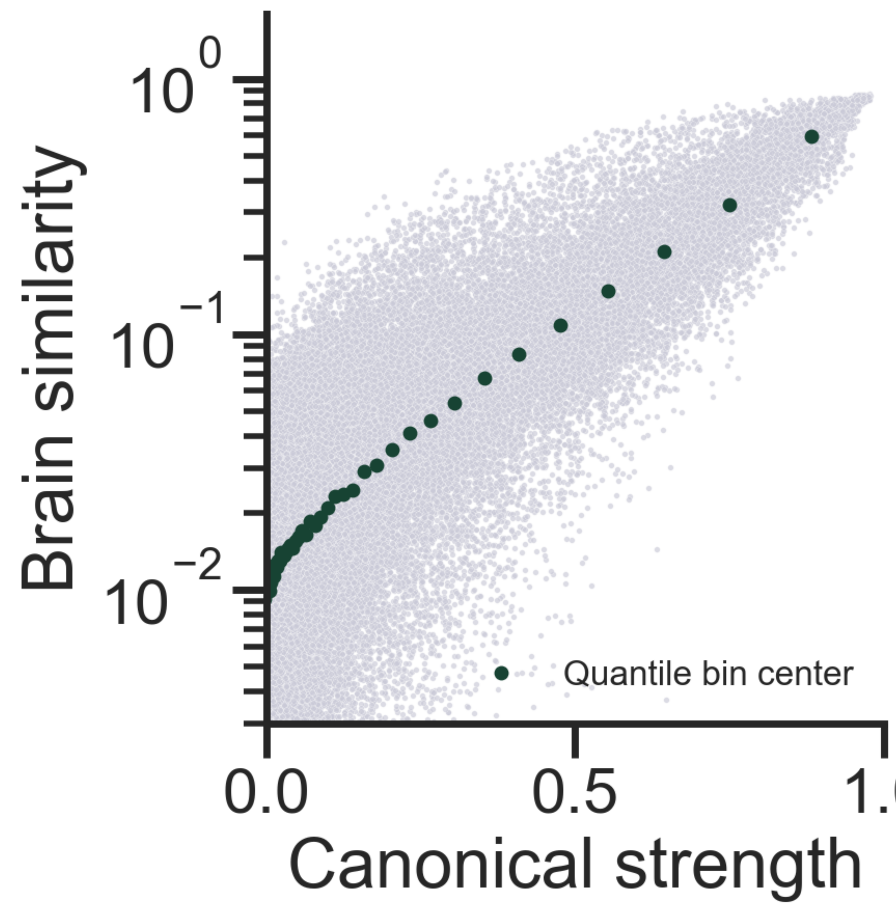


DNNs learn canonical dimensions independent to training tasks.

2 Are biological-relevant visual features constrained by **architectures**?

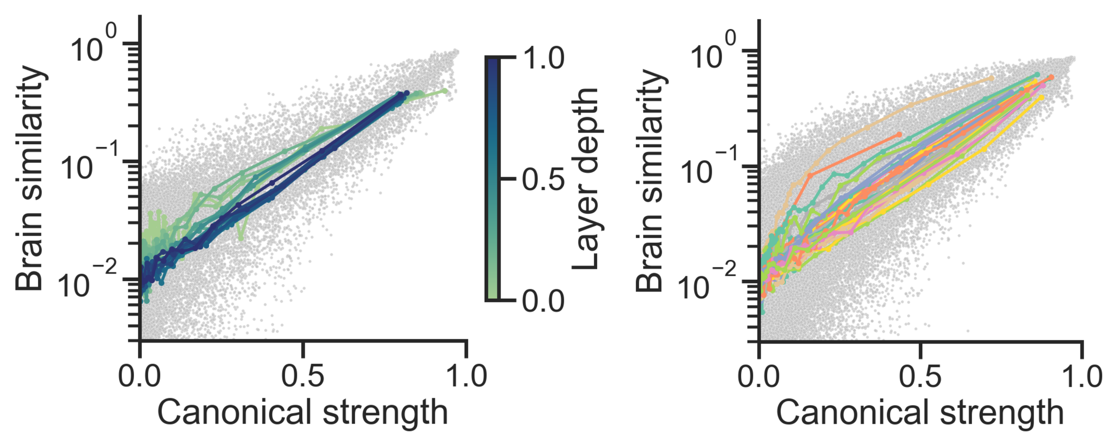
- DISTINCT architectures
- Same task (object classification)
- Same training data (ImageNet[8])
- 217,879 features





DNNs learn canonical dimensions independent to architectures.

3 Is the effect confounded by specific layer depth or models?



Consistent effect across all layers and models.

4 What are the most canonical features?

Takeaway:

- Biologically relevant visual features are generically learnable and are independent to constraints on task or architecture.
- There are unifying statistical principles across biological and artificial visual system which drives the learning and extraction of the canonical dimensions.

References

1. Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014, May). Performanceoptimized hierarchical models predict neural responses in higher visual cortex. Proceedings of the National Academy of Sciences, 111(23), 8619–8624. Retrieved from https://doi.org/10.1073/pnas.1403112111 doi: 10.1073/pnas.1403112111
2. Zhuang, C., Yan, S., Nayebi, A., Schrimpf, M., Frank, M. C., DiCarlo, J. J., & Yamins, D. L. K. (2021, January). Unsupervised neural network models of the ventral visual stream. Proceedings of the National Academy of Sciences, 118(3). Retrieved from https://doi.org/10.1073/pnas.2014196118 doi: 10.1073/pnas.2014196118
3. Conwell, C., Prince, J. S., Alvarez, G. A., & Konkle, T. (2022, March). Large-scale benchmarking of diverse artificial vision models in prediction of 7t human neuroimaging data. Retrieved from https://doi.org/10.1101/2022.03.28.485868 doi: 10.1101/2022.03.28.485868
4. Storrs, K. R., Kietzmann, T. C., Walther, A., Mehrer, J., & Kriegeskorte, N. (2021, August). Diverse deep neural networks all predict human inferior temporal cortex well, after training and fitting. Journal of Cognitive Neuroscience, 1–21. Retrieved from https://doi.org/10.1162/jocna01755 doi: 10.1162/jocna01755
5. Kriegeskorte, N. (2015, November). Deep neural networks: A new framework for modeling biological vision and brain information processing. Annual Review of Vision Science, 1(1), 417–446. Retrieved from https://doi.org/10.1146/annurev-vision-082114-035447 doi: 10.1146/annurev-vision-082114-035447
6. Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., . . . Kay, K. (2021, December). A massive 7t fMRI dataset to bridge cognitive neuroscience and artificial intelligence. Nature Neuroscience, 25(1), 116–126. Retrieved from https://doi.org/10.1038/s41593-021-00962-x doi: 10.1038/s41593-021-00962-x
7. He, K., Zhang, X., Ren, S., & Sun, J. (2016, June). Deep residual learning for image recognition. In Proceedings of the ieee conference on computer vision and pattern recognition (cvpr).
8. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., . . . Fei-Fei, L. (2015, April). ImageNet large scale visual recognition challenge. International Journal of Computer Vision, 115(3), 211–252. Retrieved from https://doi.org/10.1007/s11263-015-0816-y doi: 10.1007/s11263-015-0816-y