Canonical Dimensions of Vision

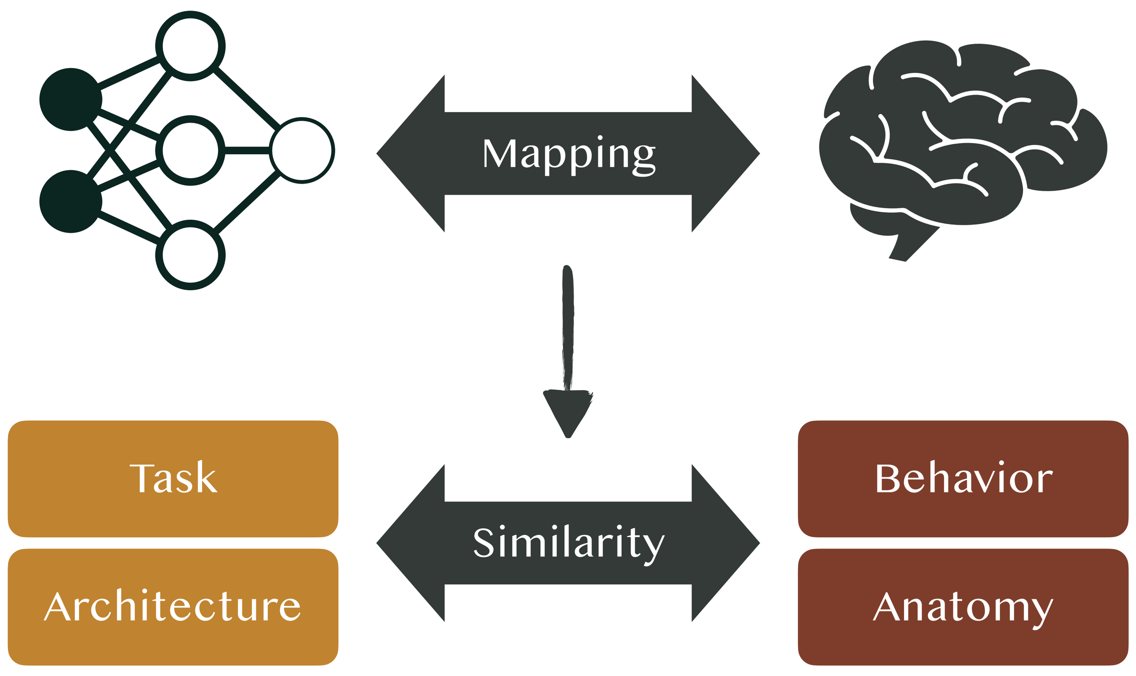
JOHNS HOPKINS
KRIEGER SCHOOL of ARTS & SCIENCES

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

Website Poster Tutorial

Motivation

Representational similarities between deep neural networks (DNN) and brains have been attributed to shared optimization constraints[1,2].



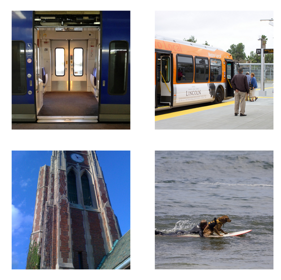
However, DNNs with widely varied designs are all surprisingly similar to the brain[3,4,5].

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

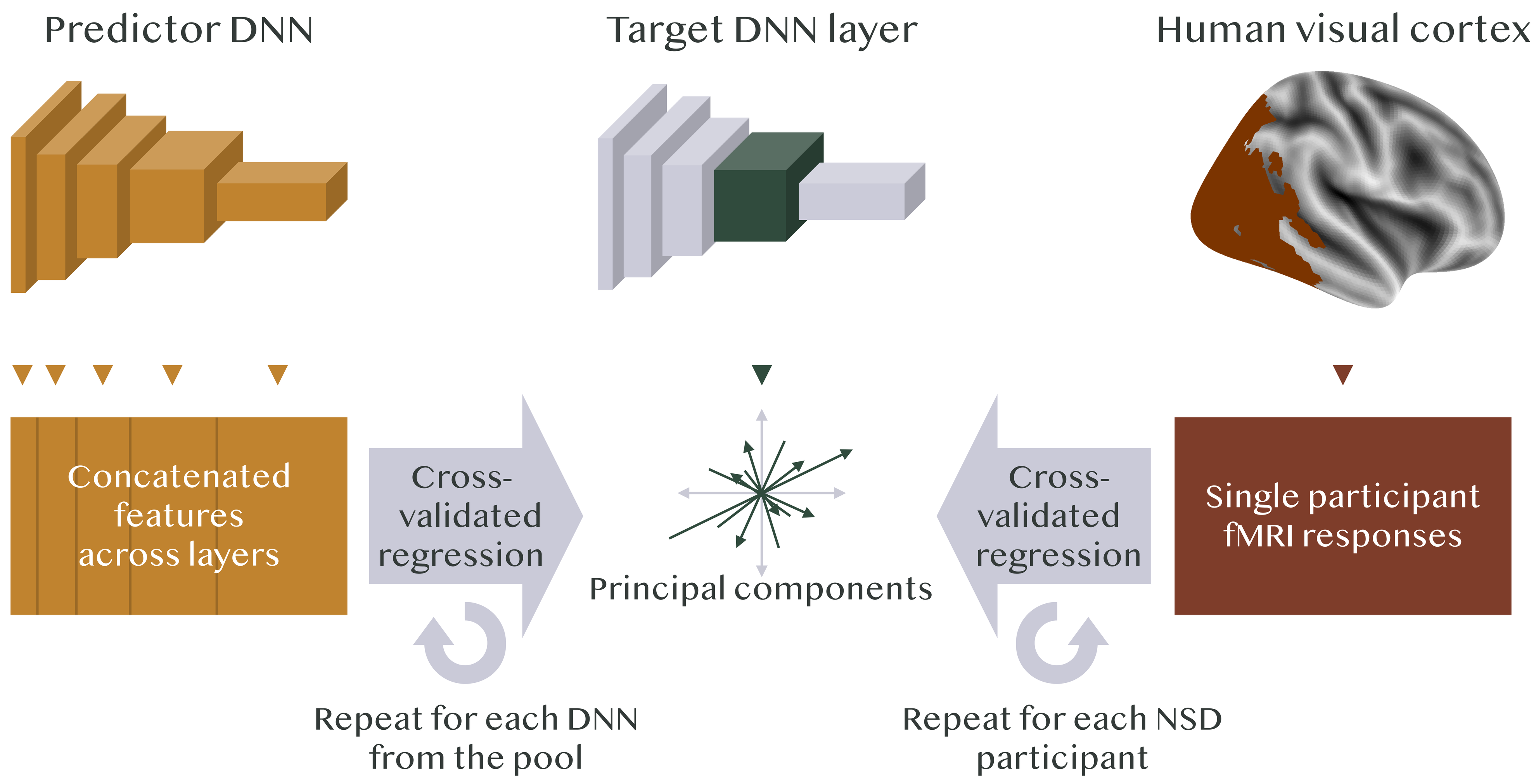
Are these canonical dimensions also encoded in human visual cortex?

Methods

Stimuli and neural data



Natural Scenes Dataset[6] (NSD)
fMRI responses of 8 participants to 872 natural scenes



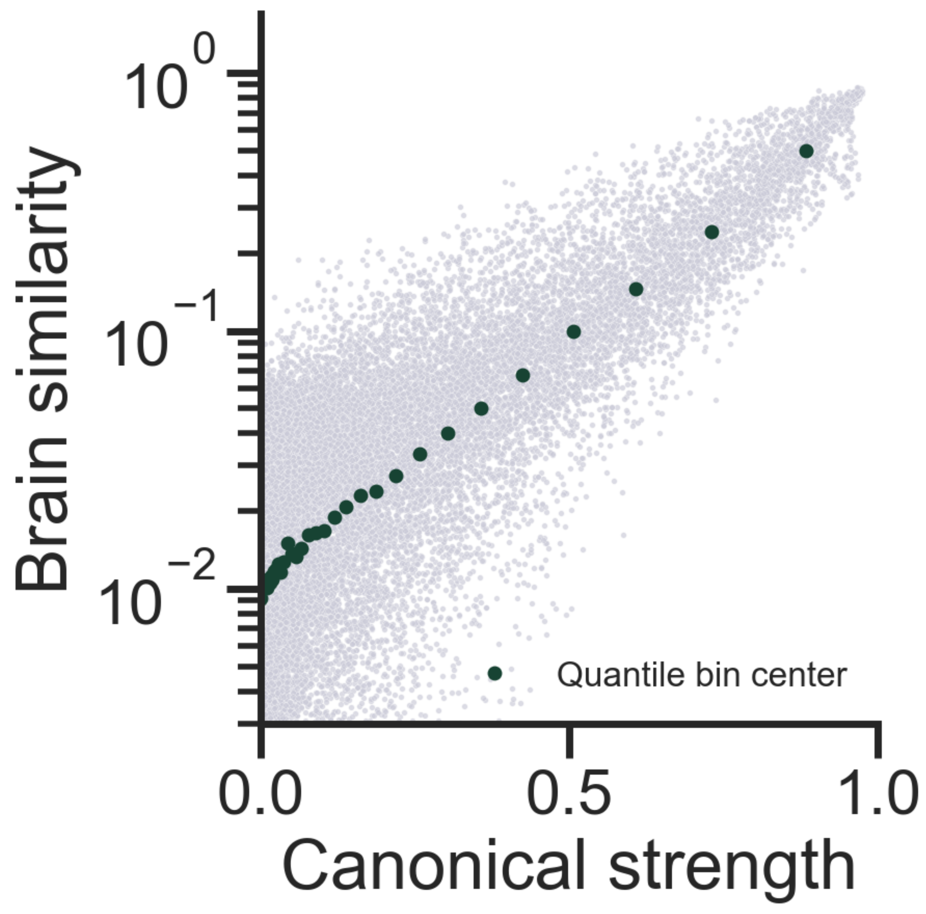
Canonical strength = average prediction accuracy across DNN predictors

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)
- Same training data (ImageNet)
- 106,889 features



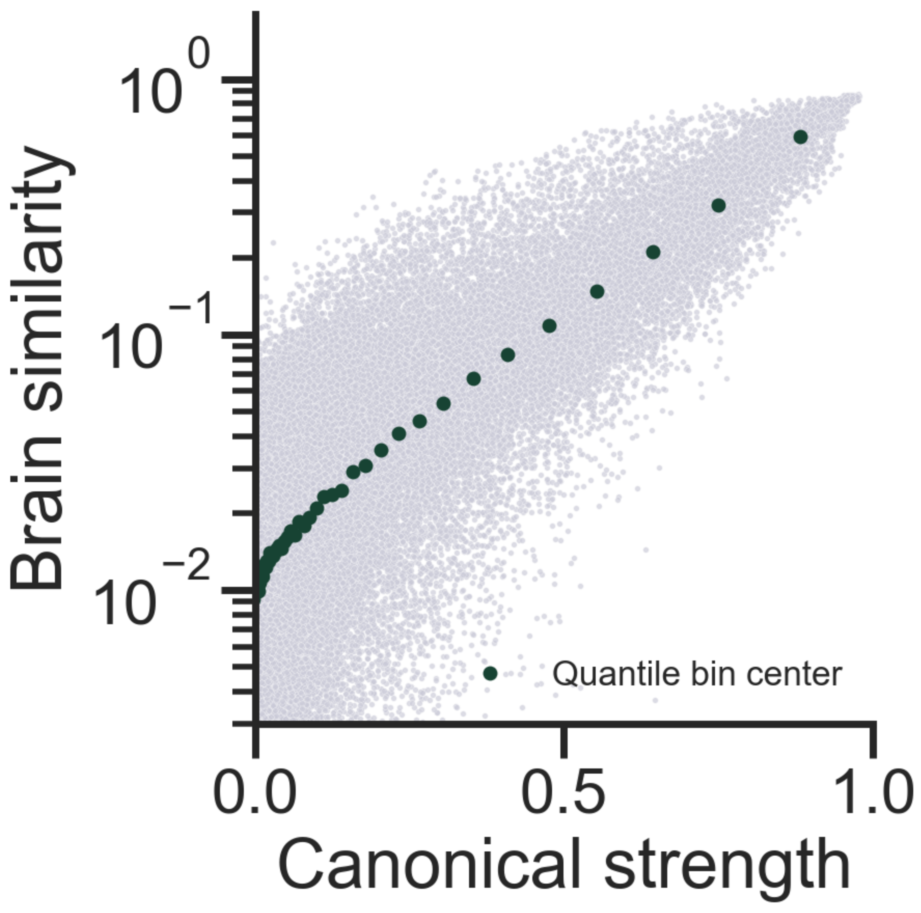
Supervised classification (N=1)

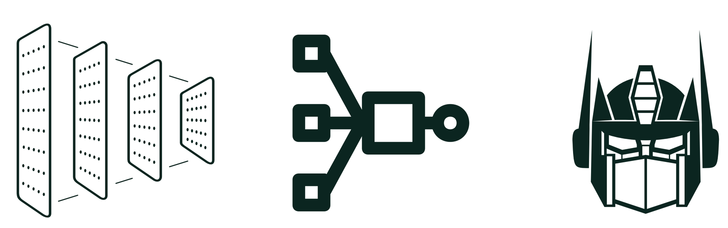
Self-supervised learning (N=9)

DNNs learn canonical dimensions independent of training tasks.

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

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





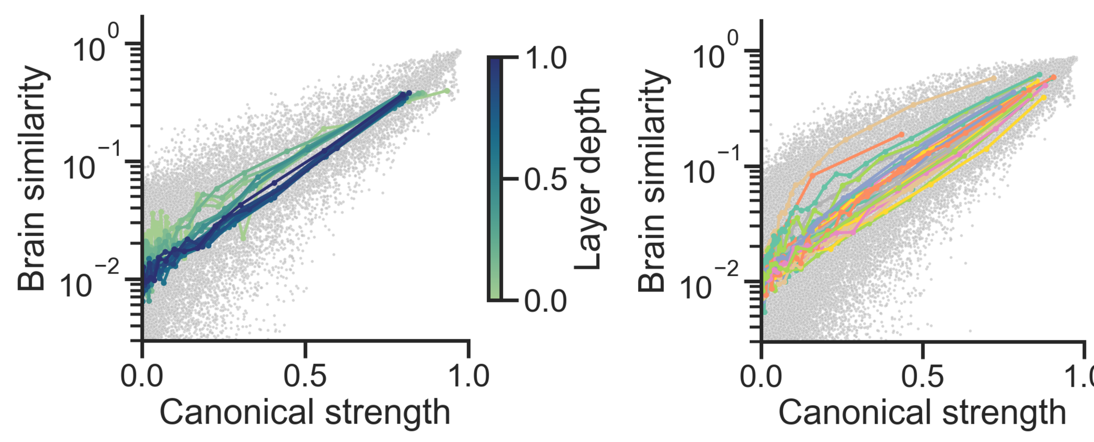
Convolutional neural network (N=11)

MLP-mixer (N=2)

Vision transformer (N=6)

DNNs learn canonical dimensions independent of architectures.

3 Are these effects driven by specific layers or models?



Consistent across layers and models

4 What images strongly activate the canonical dimensions?

Takeaway

- Biologically relevant visual features are generically learnable and are largely independent of constraints on task or architecture.
- Suggests that core statistical principles across biological and artificial vision give rise to canonical representational 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