

Hand-Engineered Image-Computable Models Can Still Outperform DNNs in V1 Similarity

Tanish Mendki (tmendki@ucsb.edu)
Dept. of Psychological and Brain Sciences
UC Santa Barbara

Sudhanshu Srivastava
Dept. of Cognitive Science
UC San Diego

Ansh Soni
Dept. of Psychology
University of Pennsylvania



1. Background and Introduction

- Deep neural networks (DNNs) are now a standard computational tool for studying the visual stream. Early work showed a strong relationship between a model's object recognition accuracy and its alignment with neural responses in higher visual cortex, especially inferotemporal (IT) cortex.
- Task-optimized DNNs were found to capture key properties of IT representations, with higher-performing networks exhibiting greater IT predictivity (Yamins & DiCarlo 2014; Cadieu 2014; Schrimpf 2020). This established the idea that improving recognition accuracy naturally leads to better models of biological vision.
- However, recent evaluations challenge this narrative: Linsley et al. (2023) found that newer, higher-performing networks are worse at predicting IT responses than earlier architectures. Whether this decline in neural alignment also occurs earlier in the visual hierarchy, specifically in primary visual cortex (V1) remains an open question.
- V1 is a well-characterized stage of processing that constructs orientation- and frequency-tuned representations forming the basis for later invariant and category-relevant features, therefore hand-crafted models should also be considered.

2. Methods

2a) Models

Biologically inspired mathematical models:

- HMAX
- Gabor Wavelet Pyramid
- Steerable Pyramid
- Neocognitron

Standard Convolutional Neural Networks (CNNs)

- AlexNet
- ResNet-50

Transformers

- ViT
- CvT-13
- ConvNeXt-Large

Self-supervised encoders

- CLIP
- SLIP

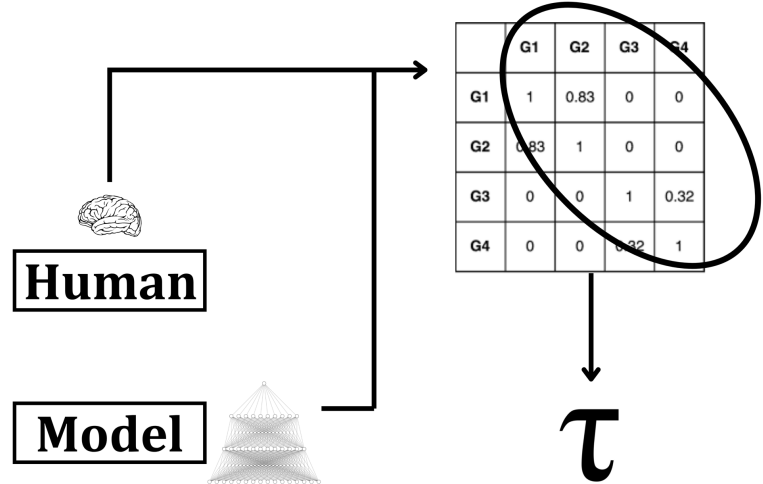
VOneNet (V1 front-end module)

- AlexNet
- ResNet-50
- CORnet-S

2b) Metrics

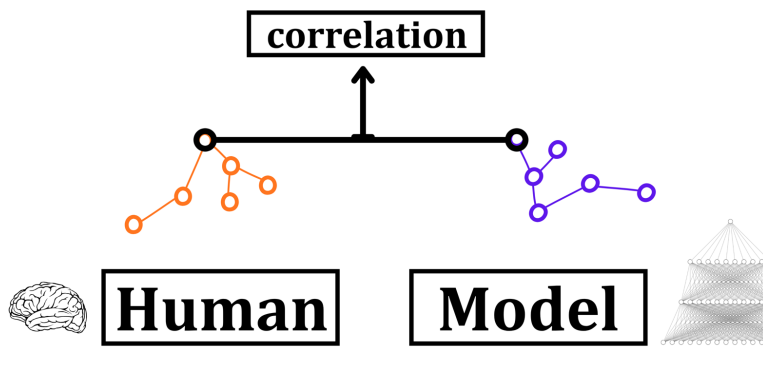
Representational Similarity Analysis (RSA)

- $N \times N$ representational dissimilarity metric (RDM).
- Compare between images by computing Kendall's tau.



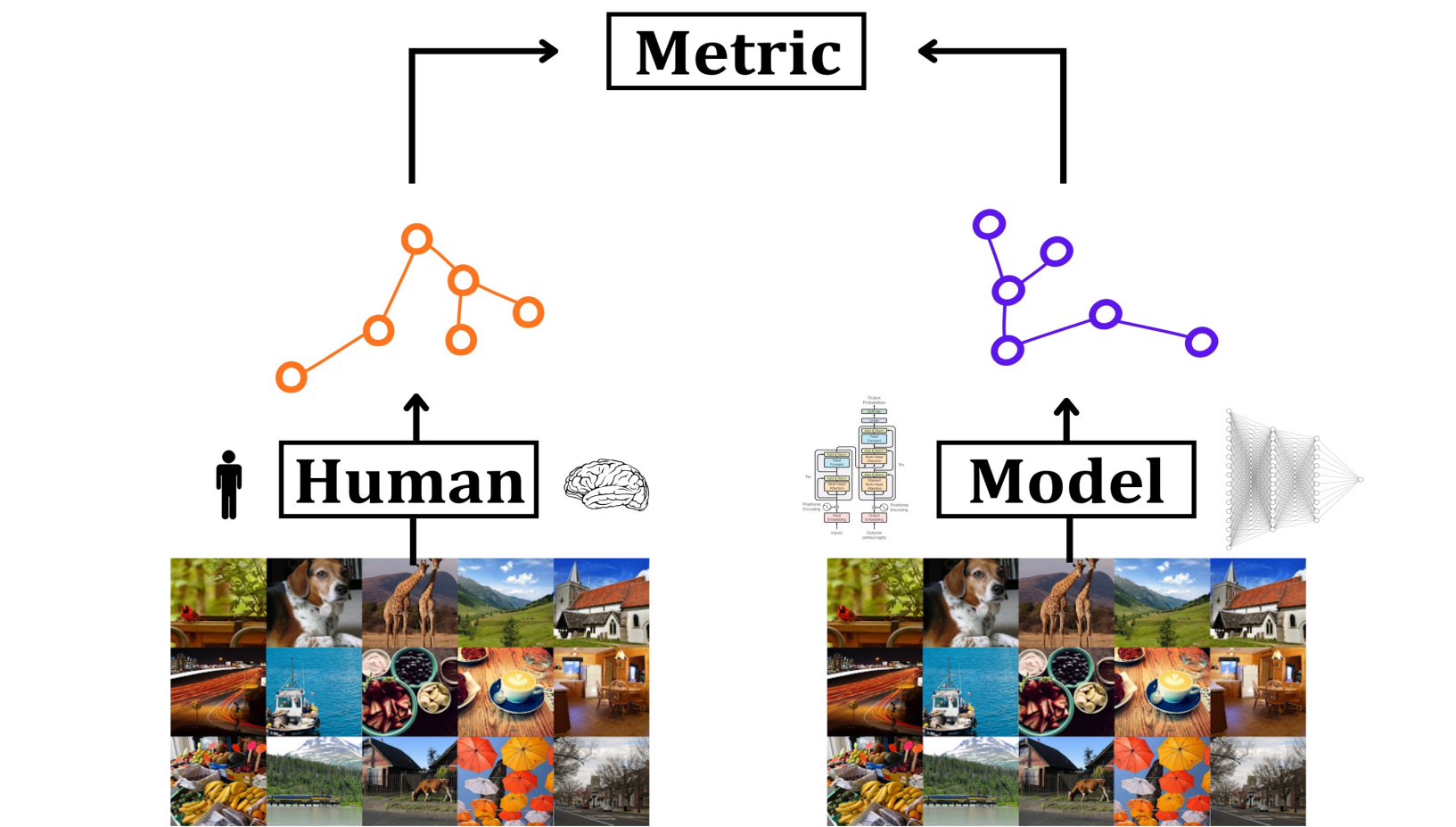
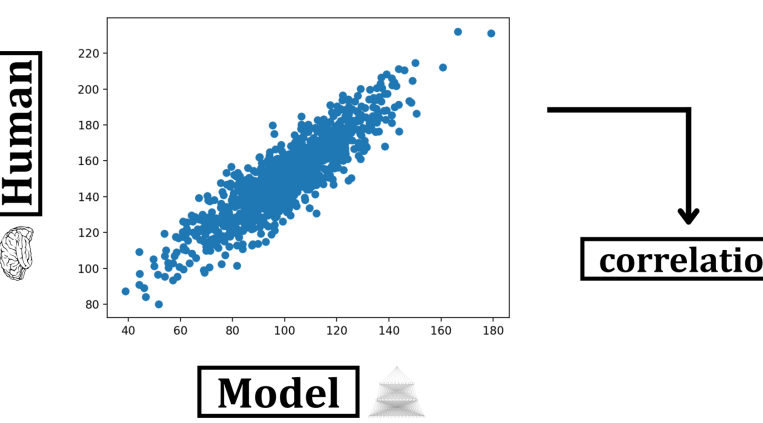
Pairwise matching

- Match model unit to one in the brain with highest correlation.
- Evaluate correlations on a held-out test set (Khosla and Williams, 2024).

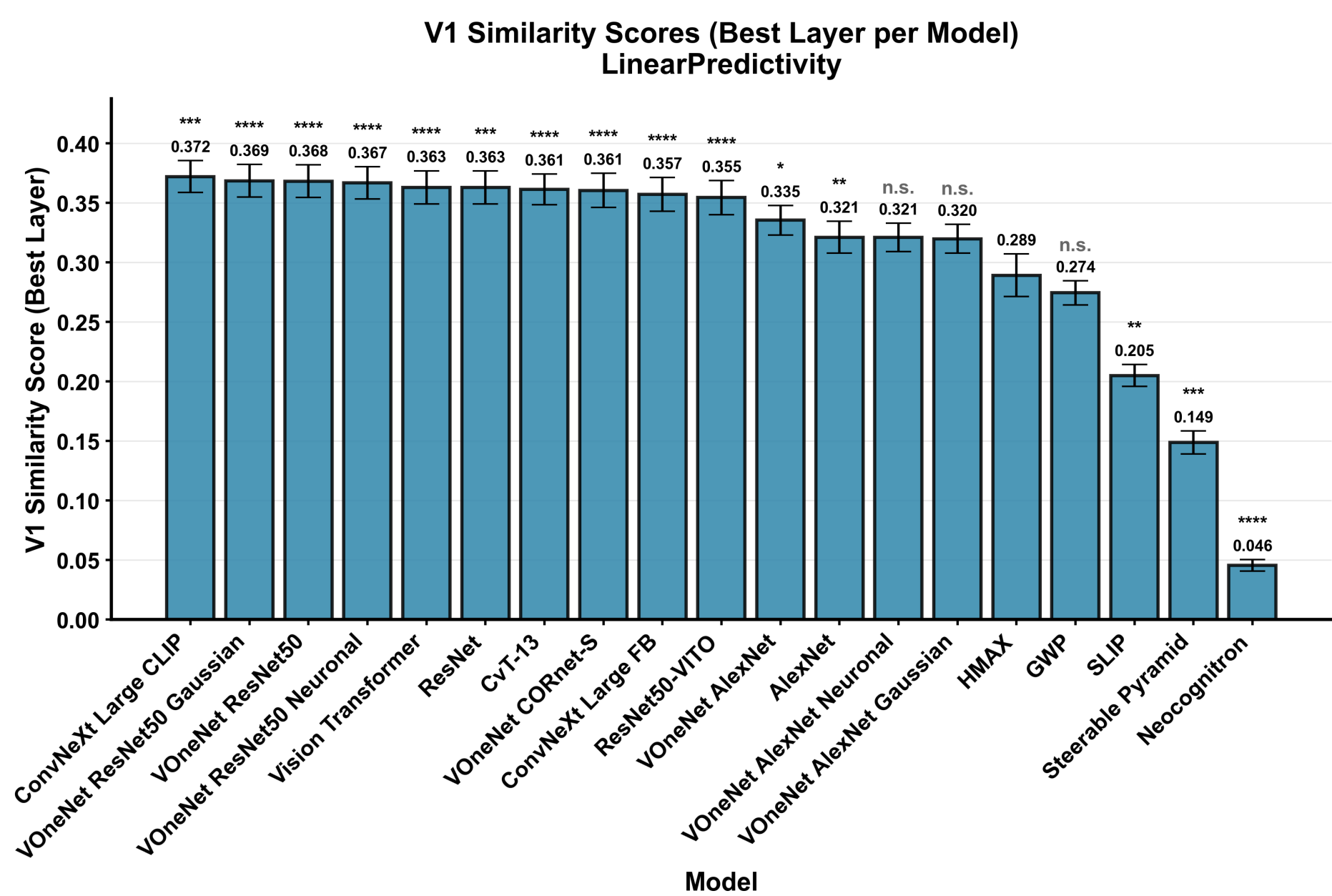
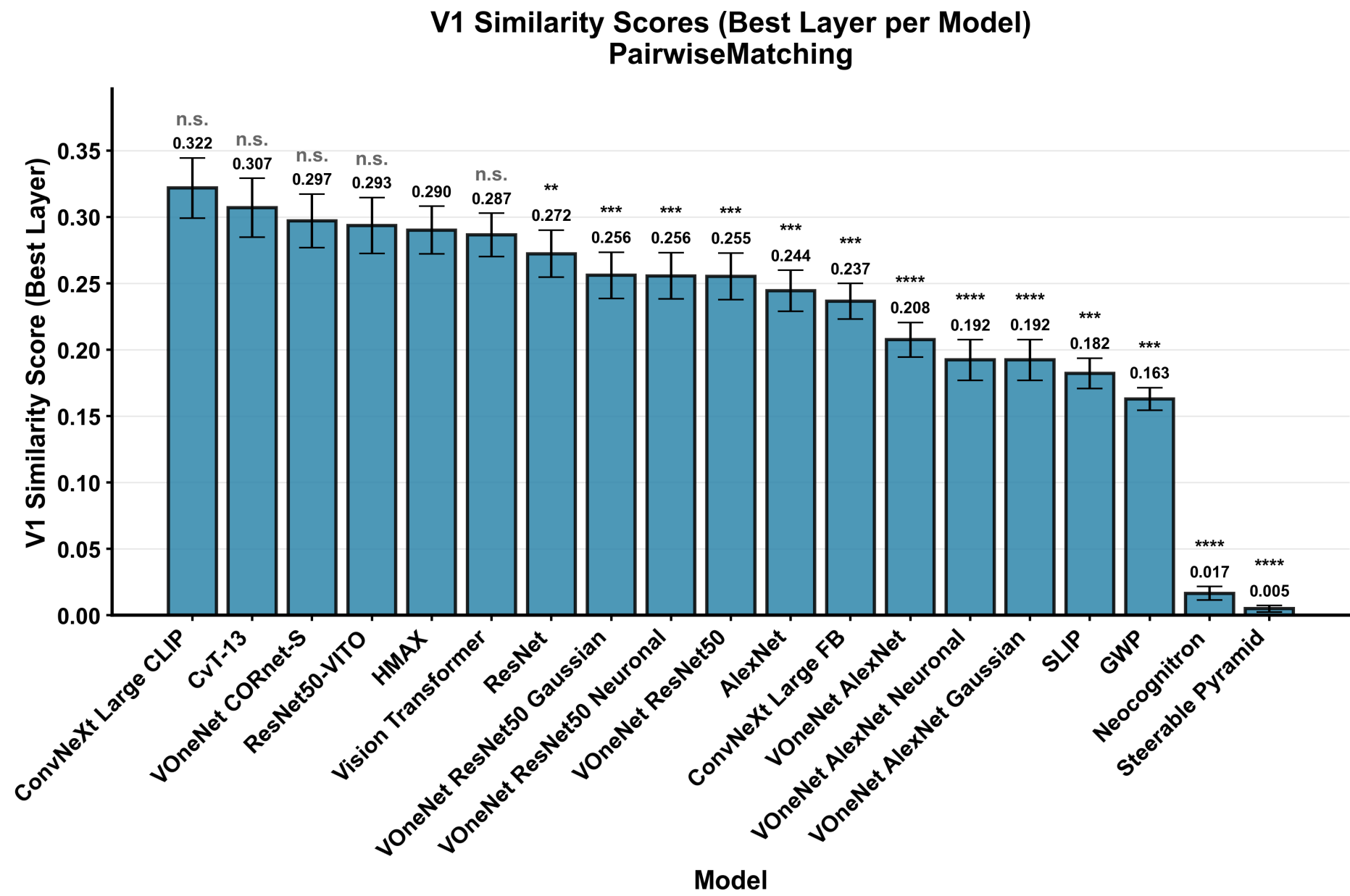
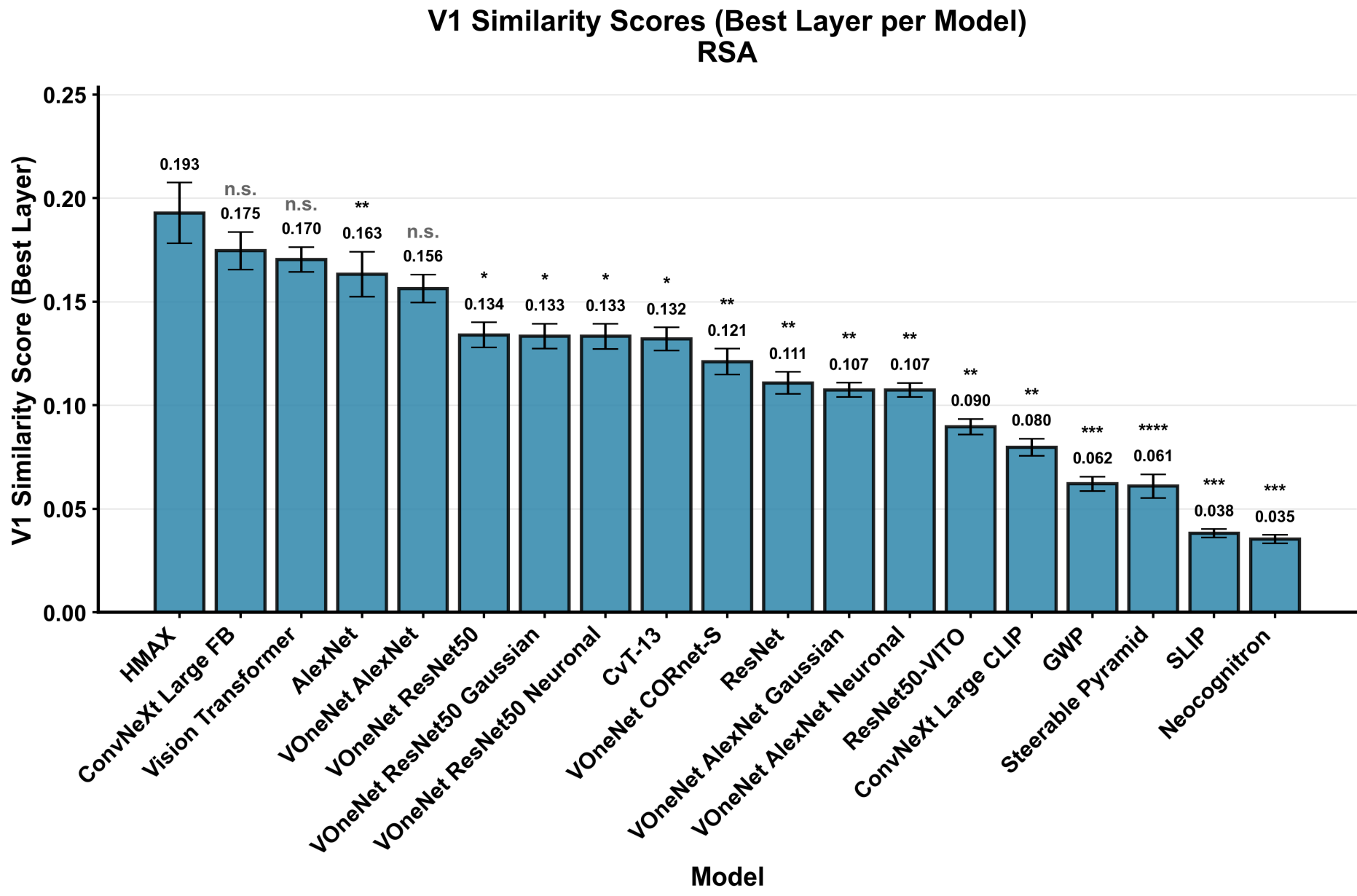


Linear predictivity

- Ridge regression from model features to neural responses.
- Five-fold CV and log spaced ridge grid.



3. Results



4. Discussion

- Early work in using DNNs for studying alignment suggested that task optimization led to increased alignment in IT, but this claim has been refuted in recent work that showed how IT alignment has decreased over the years while model complexity has grown exponentially.
- Performance optimization does not seem to lead to emergent unified representations, and we show this gap here in V1. The HMAX benchmark remains unbeaten when alignment is defined by representational geometry or by strict pairwise structure. The persistence of this benchmark suggests that key biases of early vision remain underrepresented in current computational vision and models.
- Localized features and simple pooling (as in HMAX) continues to capture V1 statistics very well, so preserving early-stage fidelity might be necessary in future work.
- More work with more data modalities, along with testing to see why this difference in metrics arises would be highly beneficial to our understanding of this field.

5. References

Yamins, D. L. K., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *PNAS*.
Cadieu, C. F., et al. (2014). Deep neural networks rival the representation of primate IT cortex for core object recognition. *PLoS Comput Bio*.
Schrimpf, M., et al. (2020). Integrative benchmarking to advance neurally mechanistic models of human intelligence. *Neuron*.
Linsley, D., et al. (2023). Performance-optimized deep neural networks are evolving into worse models of inferotemporal visual cortex. *NeurIPS*.
Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*.
Serre, T., et al. (2007). Robust object recognition with cortex-like mechanisms. *TPAMI*.
Khosla, M., & Williams, A. H. (2024). Soft matching distance: A metric on neural representations that captures single-neuron tuning. *UniReps Workshop / PMLR*.
Soni, A., et al. (2024). Conclusions about neural network-brain alignment are profoundly impacted by the similarity measure. *bioRxiv*.