Quantifying Task-relevant Similarities in Representations Using Decision Variable Correlations

The University of Texas at Austin
Interdisciplinary
Neuroscience Program

center for theoretical and computational neuroscience
ctcn The University of Texas at Austin

Eric (Yu) Qian[1], Wilson Geisler[3] & Xue-Xin Wei[1,2,3]
[1]Institute for Neuroscience. [2]Department of Neuroscience. [3]Department of Psychology, University of Texas at Austin.

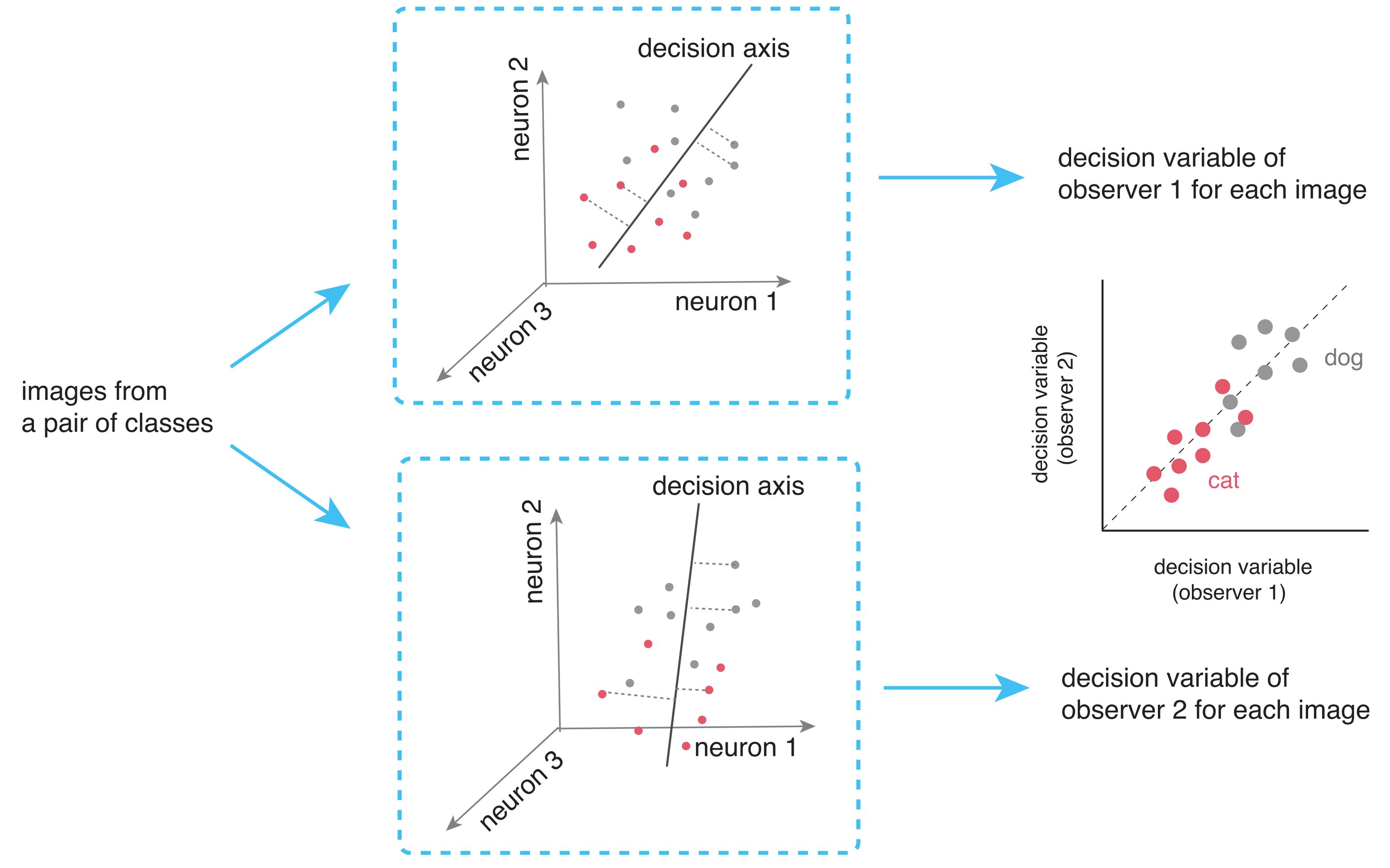
ericqian@utexas.edu
w.geisler@utexas.edu
weixx@utexas.edu

NEURAL INFORMATION PROCESSING SYSTEMS

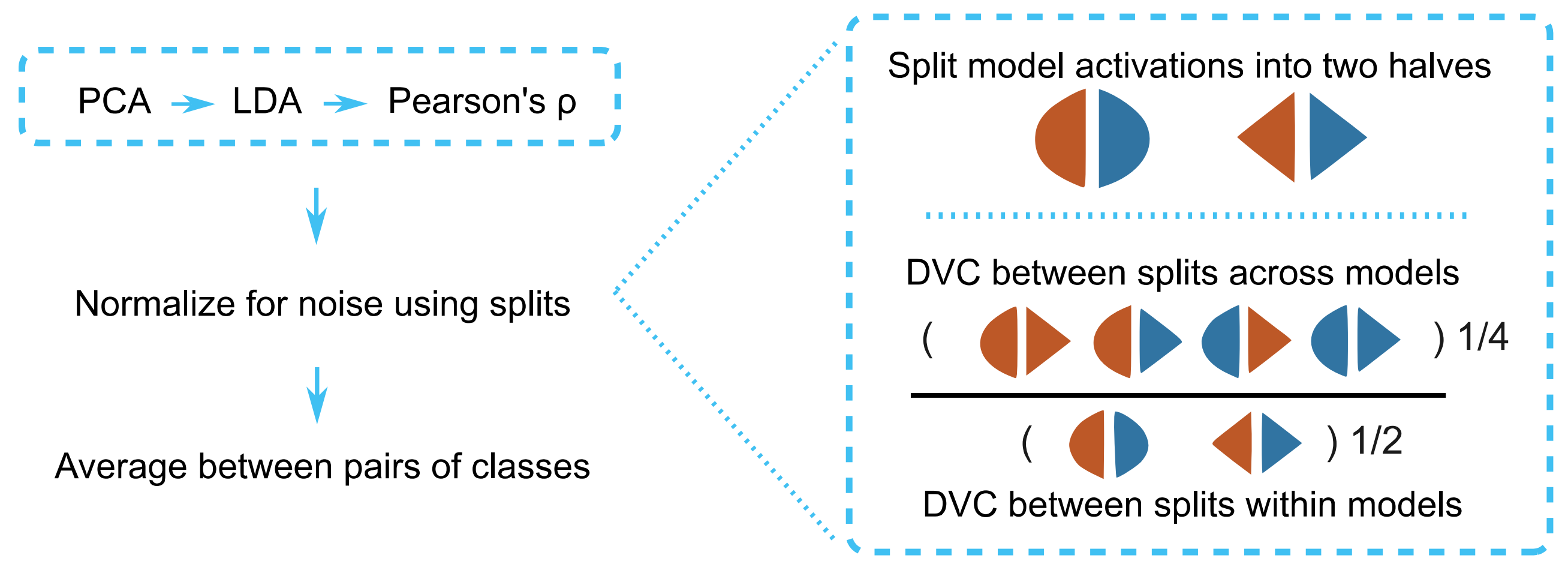
Decision Variable Correlation (DVC)

We propose using the decoded decision variable (DV) as a gateway to studying the strategies an observer uses to solve a task. The DVC between two observers performing the same task reflects the task-relevant similarity of their internal representations. Compared to existing similarity measures for network models, DVC is principled and uniquely task-focused, providing a new lens for analyzing brain and network representations.

Inferring DVC from neural representations

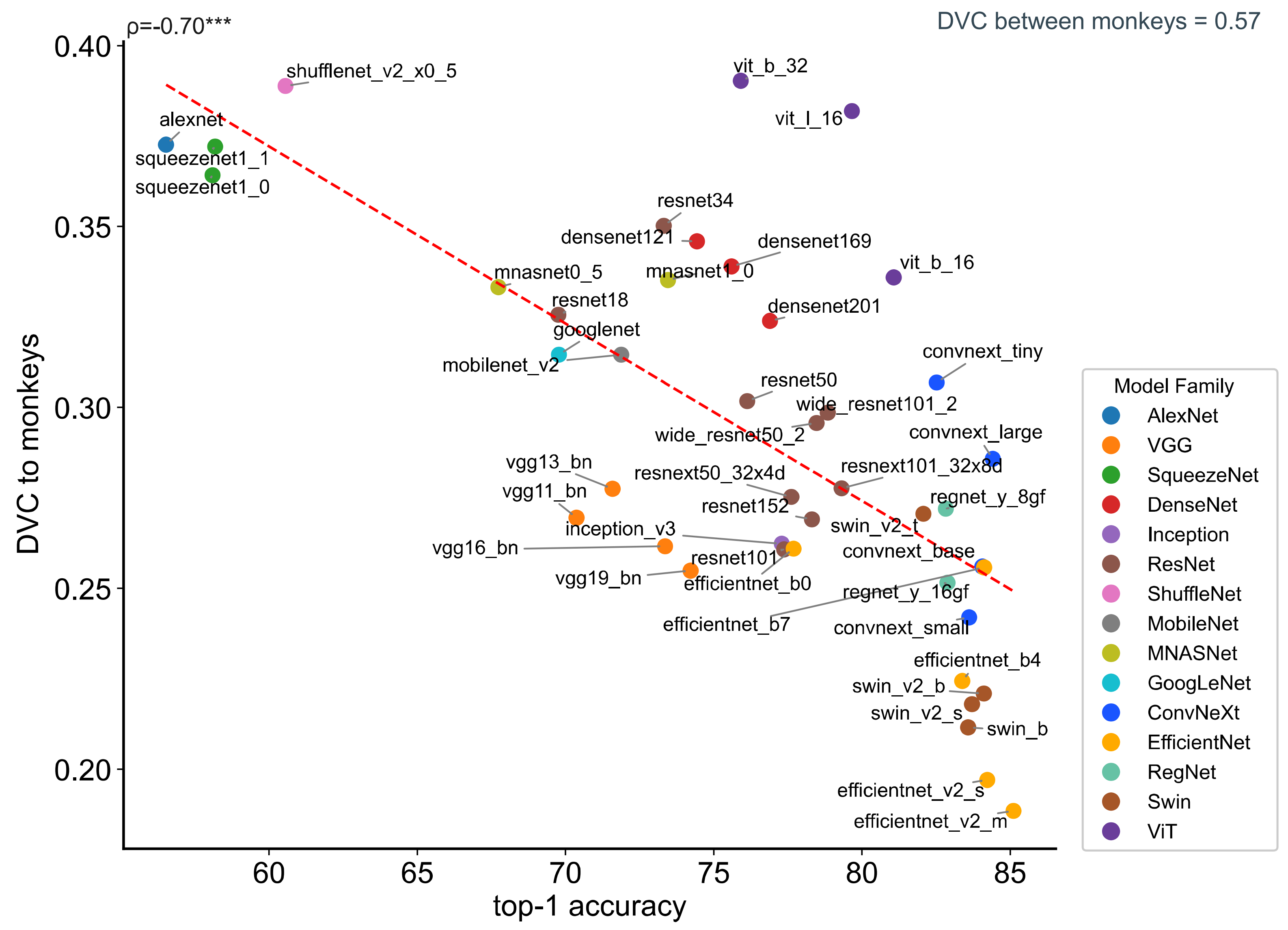


We use optimal linear classifier to infer the DVs of individual observers and then quantify the consistency of the DVs.

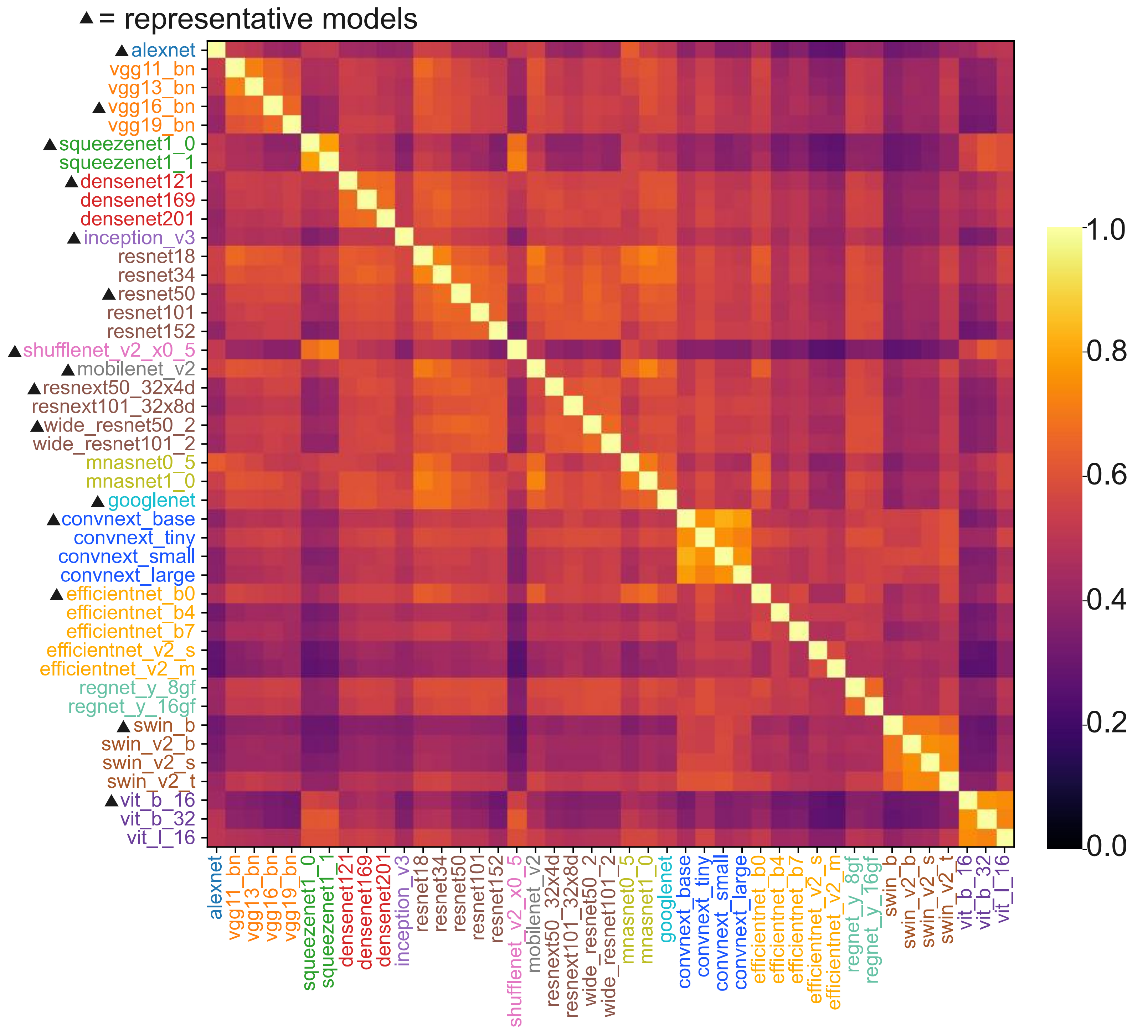

PCA → LDA → Pearson's ρ
Normalize for noise using splits
Average between pairs of classes
Split model activations into two halves
DVC between splits across models
( ) 1/4
( ) 1/2
DVC between splits within models

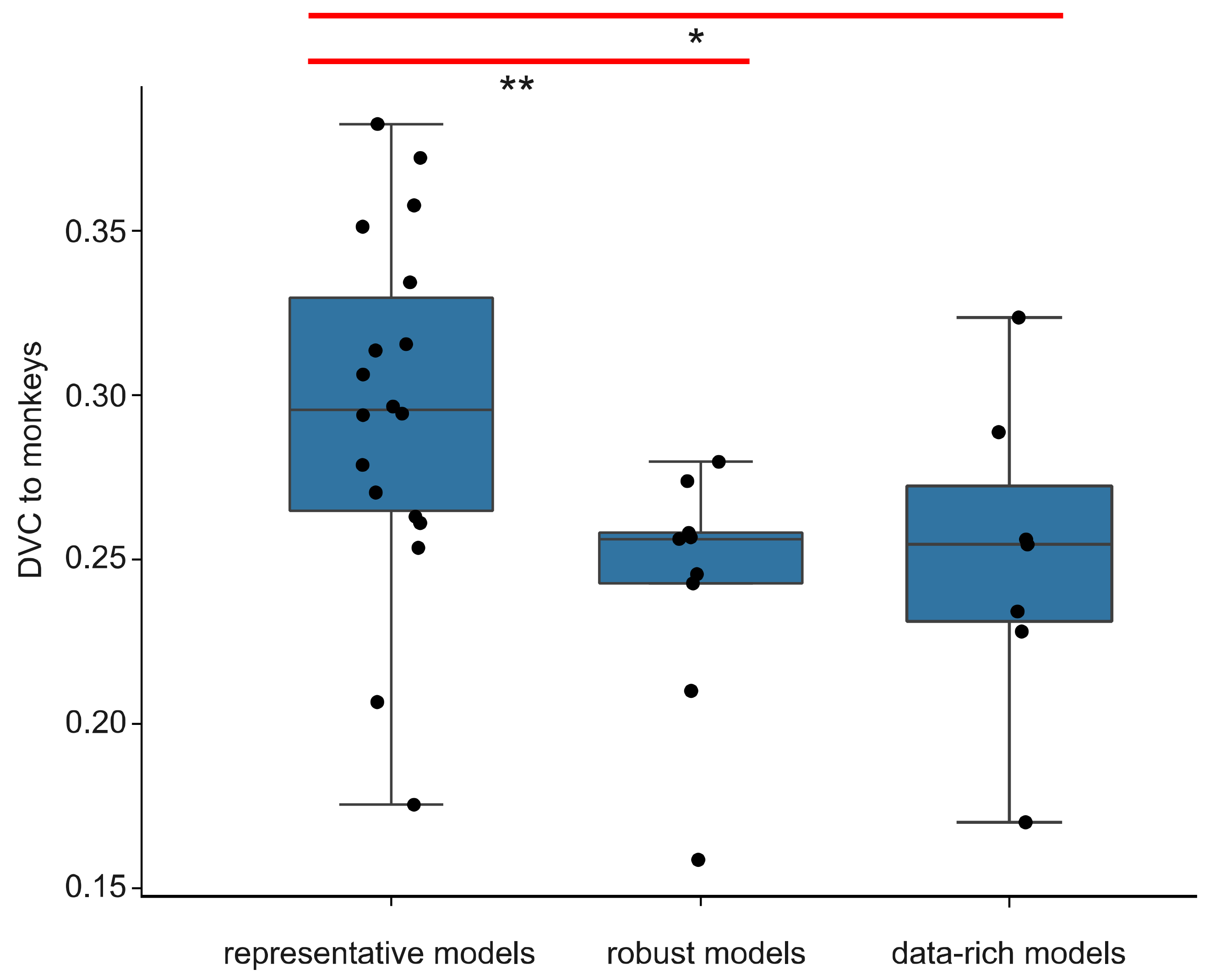
In practice, the representations are reduced to the same dimension, and a linear classifier is obtained for each pair of image categories. After the representations are projected onto the decision axis to obtain the decoded DV, DVC between two observers are further normalized to account for the effect of noise in the representations.

Model similarity to brain is decreasing with acc



According to DVC, accuracy on Imagenet-1k is negatively correlated with similarity to brains. Model architecture also influences similarity, as models in the same family tend to be more similar in their representations.

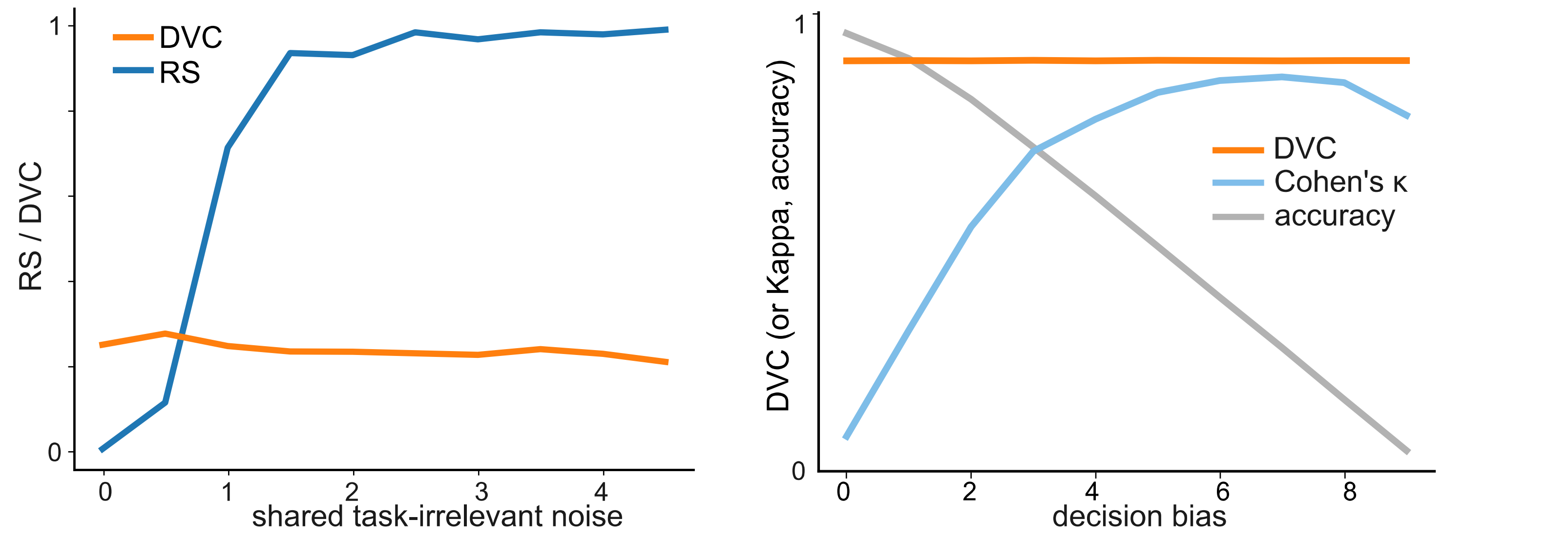




Similarity to brains is not improved by adversarial training (robust models) or scaling data size (data-rich models). Brain-like feature engineering and data diets can still potentially bridge this gap.

DVC is invariant to task-irrelevant correlations (vs. RSA) or decoder bias (vs. Cohen's κ)

Whereas shape metrics like RSA focus on alignment of overall geometry, DVC is invariant to correlations on dimensions that are not task-relevant. Behavioral similarity measures like Cohen's κ is conflated with bias in the criterion, but DVC is not. These properties enable DVC to supersede previous methods in many applications.



References

Data from 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". Journal of Neuroscience.

Sebastian, S., & Geisler, W. S. (2018). "Decision-variable correlation". Journal of Vision.