Decoding Contact Information in Pre-Trained Deep Learning Models

Georgina Woo[1,2]*, RT Pramod[2]*, Nancy Kanwisher[2]
[1]Hunter College, City University of New York
[2]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

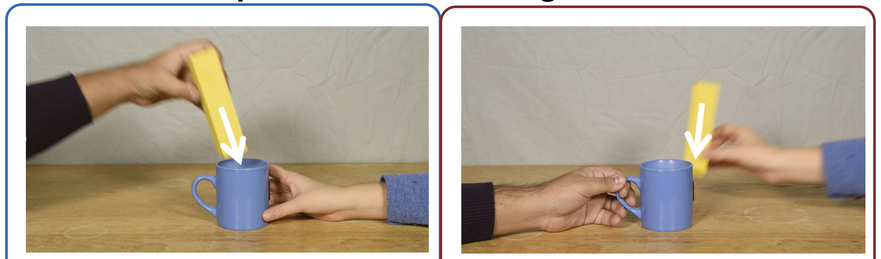
The brain's **Physics Network** detects and predicts contact. What computations underlie this ability?



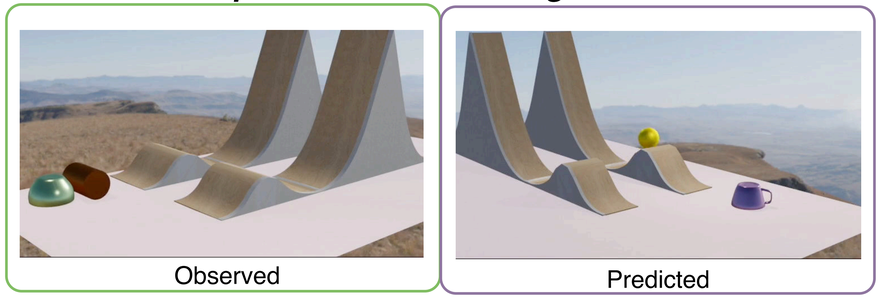
Do deep learning models capture and represent information for contact detection and prediction? If so, how do they compare to the brain?

Methods

Experiment 1: Detecting Contact



Experiment 2: Predicting Contact

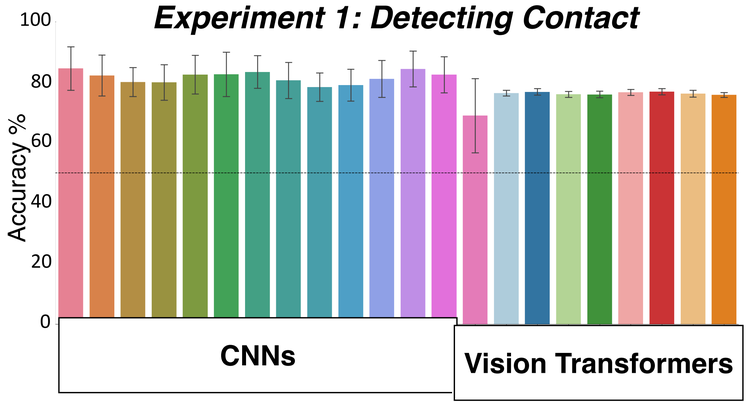


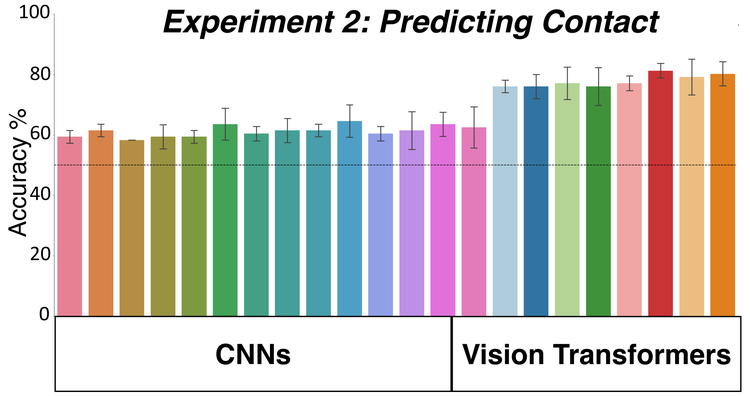
Models

- 13 CNNs and 1 Vision Transformer trained on Kinetics-400
- 8 Vision Transformers trained on Kinetics-700 and Physion

We trained linear SVM classifiers on the model features for contact detection and prediction.

Results

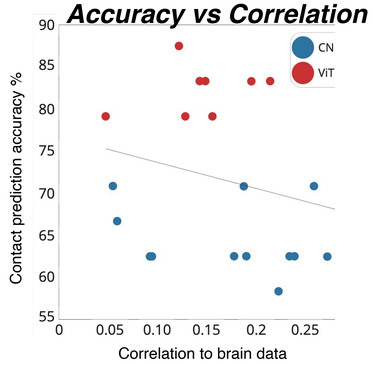


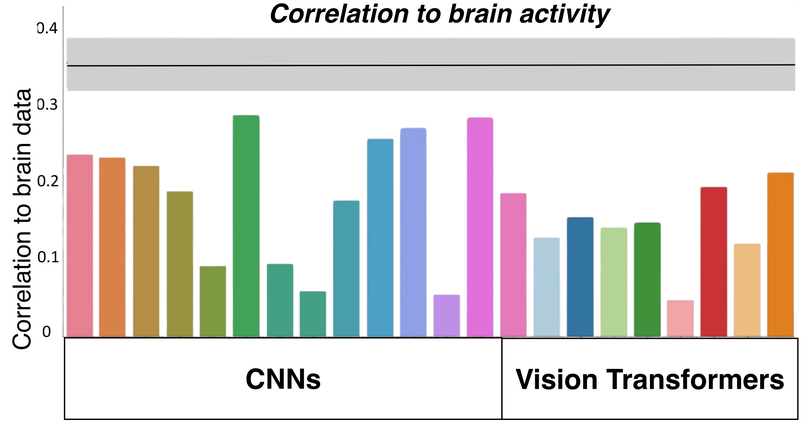


Fits to neural data

We compared the model features to the Physics Network *(see below)* using Representational Similarity Analysis.

There is no correlation between model task performance and fits to neural data *(see right)*.





Conclusion and further study

- The models show above chance accuracy in contact detection and prediction.
- We do not find a strong correlation between the models' internal representation and brain data.

Further study:

- Evaluate other models i.e. language or generative models for contact information and fits to brain data.
- Test if finetuning models with relevant stimuli leads to improvements in task performance and fits to brain data.