A Benchmark Dataset for Perceiving Object Physical Relationships in Human and Machine Vision

Georgina Woo, Nancy Kanwisher, RT Pramod
Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

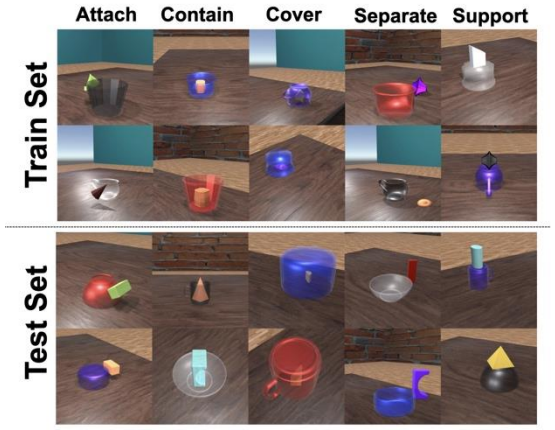
Introduction

Humans perceive physical relationships between objects automatically and accurately. **What computational mechanisms underlie this ability?**

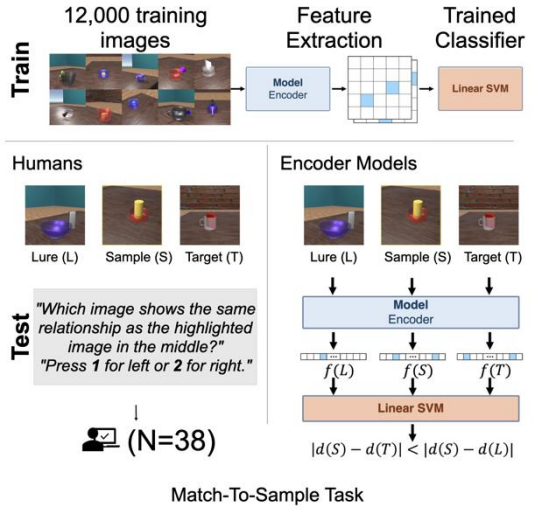
We introduce **ObjectRel**: 14,400 synthetic images of five physical relationships, used to test **deep learning models** as hypotheses of human object-relation understanding.

ObjectRel

Train and test objects **never overlap**, so models must generalize to novel objects.

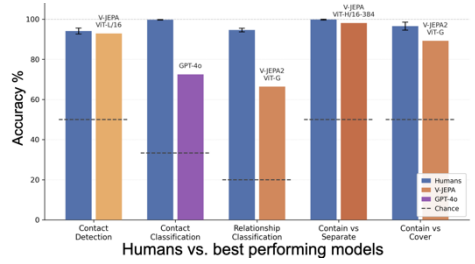


Methods



Match-To-Sample Task

Other classification tasks show similar gaps

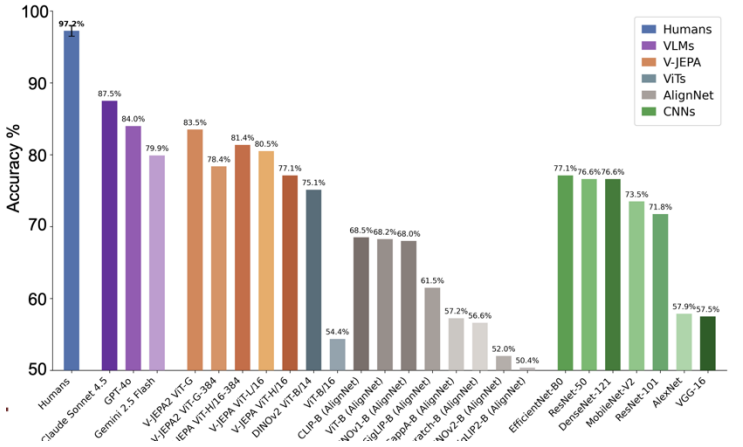


Humans vs. best performing models

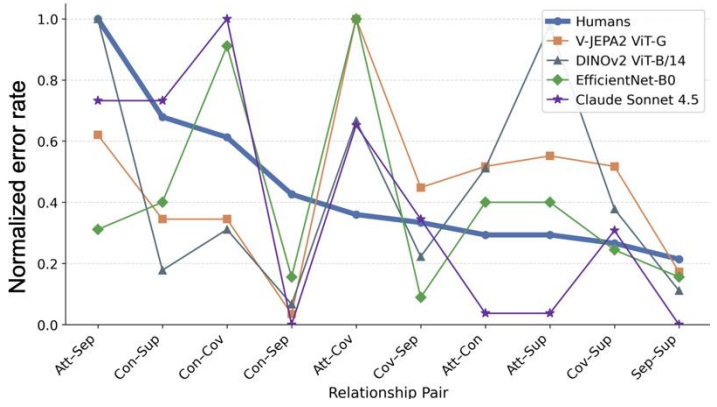
Robustness checks

- Cosine & MLP never beat SVM readout
- Accuracy-matched errors: r ≈ 0

Match-to-sample: all models fall short of human accuracy

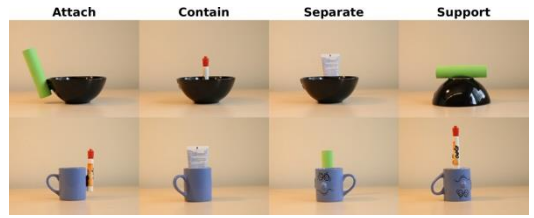


Model and human error patterns are uncorrelated



Per-pair error rates, normalized to each model's maximum and sorted by human errors.

The gap holds for real photographs



Humans 96.5% vs. best model 81.3%. Synthetic accuracy predicts real-photo accuracy (r = 0.63).

Next steps

Video set: each relationship is created by a hand, then tested by a collision.

