Humans can rapidly and accurately recognize objects in spite of variation in viewing parameters (e.g. position, pose, and size) and background conditions. To understand this invariant object recognition ability, one can construct and test models that aim to accurately capture human behavior, including making the same patterns of errors over all images. Here, we applied this straightforward visual “Turing test” to the leading feed-forward computational models of human vision (hierarchical convolutional neural networks, HCNNs) and to a leading animal model (rhesus macaques). Our results reveal a failure of current HCNNs to fully replicate the image-level behavioral patterns of primates.

### Methods

**Purpose:** To enforce invariant object recognition performance, have proven to be a dramatic advance in the state-of-the-art of artificial vision systems.

- **HCNN models** match primates in their core object recognition behavior with respect to object-level difficulty and object-level confusion patterns.
- **HCNN models** fail to match primates in behavior/image-level difficulty and fail to match image-level confusion patterns.

**Behavioral paradigm:** We tested object recognition performance on a two alternative forced choice (2AFC) match-to-sample task using brief (100ms) presentations of naturalistic synthetic images of basic-level objects.

- **High-throughput psychophysics:** Over a million behavioral trials across hundreds of humans and five monkeys were aggregated to characterize each object on each image. To collect such large behavioral datasets, we used Amazon Mechanical Turk for humans and a novel home-cage touchscreen system for monkeys.

**Stimuli:** To enforce invariant object recognition behavior, we generated several thousand naturalistic synthetic images (2400 images, 24 objects, 100 images/subject) on a random natural image background. Each object can produce myriad images under variations in viewing parameters, with some images being more challenging than others.

**Results:**

- Consistent with previous work, macaque monkeys and all tested HCNN models are highly consistent with humans with respect to object-level metrics (O1, O2), while a lower level representation (V1) is less consistent.

**Behavioral metrics:**

- To measure the behavioral consistency between two systems with respect to a given behavioral metric, we computed a noise-adjusted correlation, which accounts for the reliability (i.e. internal consistency) of behavioral measurements.

**Conclusion:**

- In recent years, specific HCNN models, optimized to match human-level object recognition performance, have proven to be a dramatic advance in the state-of-the-art of artificial vision systems.
- HCN models match primates in their core object recognition behavior with respect to object-level difficulty and object-level confusion patterns.
- HCN models fail to match primates in behavior/image-level difficulty and fail to match image-level confusion patterns.

- Attempts to rescue the image-level gap in consistency, including adding retinal sampling on the model front-end, adding various decoders on the model back-end, and supplementing model training with similar images, did not reduce the gap between models and primates.

- Furthermore, no particular property of the image (luminance, spatial frequency) or of the object in the image (size, pose) was strongly correlated with model failure to capture primates’ image-level performance.

Taken together, these results suggest that high-resolution behaviour could serve as a strong constraint for discovering models that more precisely capture the neural mechanisms underlying primate object recognition.