Contour Integration Underlies Human-Like Vision

Ben Lonnqvist[1,2], Elsa Scialom[1], Abdulkadir Gökce[2], Zehra Merchant[1], Michael H. Herzog[1], Martin Schrimpf[2]

1 Laboratory of Psychophysics
2 NeuroAI Laboratory
EPFL, Switzerland

ICML International Conference On Machine Learning

SCAN ME!

EPFL
Swiss National Science Foundation

benlonnqvist.github.io/

Question

Humans group. elements together to perceive objects.[1]

Are current DNNs able to recognize objects that require grouping?

Do models exhibit contour integration like humans?

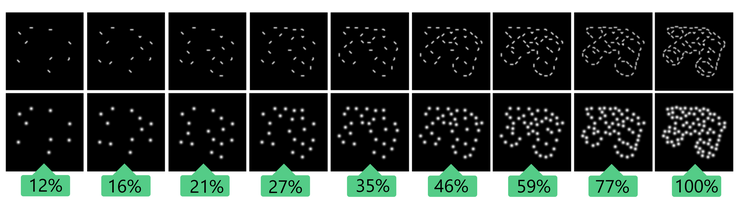
Approach

Contour integration is highly directional in mammals[2] – our dataset captures this effect by using lines and phosphenes.

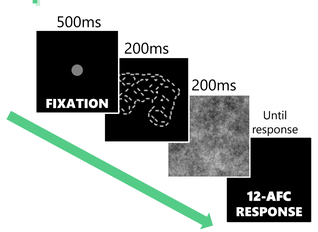
We tested 50 humans and over *1,000* vision DNNs on our dataset.

Dataset

- 4 image types: RGB, Contours, Phosphenes, Segments[3].
- A *wide range* of element densities.

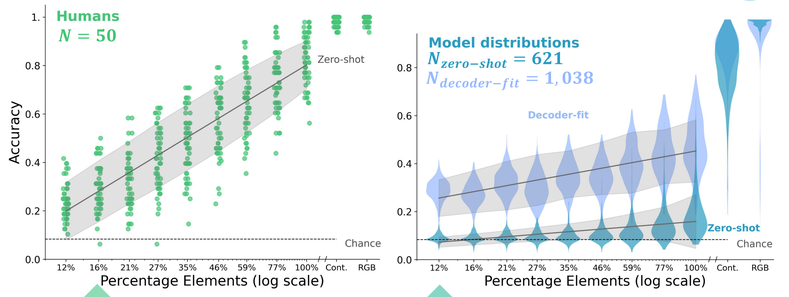


Experimental Methodology



- $N = 50$ subjects total.
- Subjects were split into two groups of 25 subjects, each performing either the phosphene or segment task.
- All subjects saw Contour and RGB images.

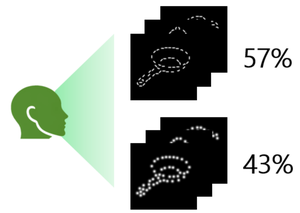
Humans outperform models on average



Human performance scales log-linearly.

Models perform worse for **any** image fragmentation.

Humans and models share integration bias

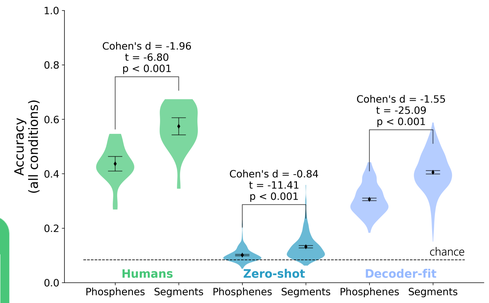


SEGMENT - PHOSPHENE

57% − 43%

HUMAN INTEGRATION BIAS

=14%



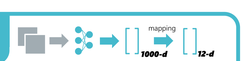
Both humans and models have higher performance on segments than on phosphenes.

Modeling methodology

- A total of 1,038 DNNs.
- Model responses were obtained in *two ways*: zero-shot, as well as by fitting a decoder.

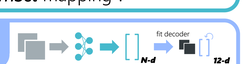
Brain-Score

ZERO-SHOT



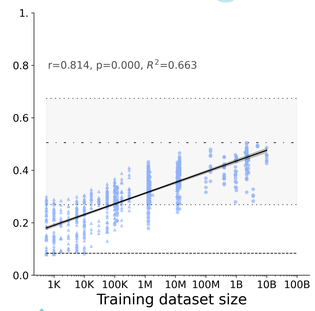
ImageNet labels are mapped to our **12-AFC** task labels using a ***WordNet SynSet*** mapping[4].

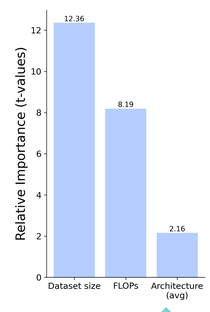
DECODER-FIT



The *penultimate layer activations to* 120 novel samples (10 per class) are used to fit a linear decoder.

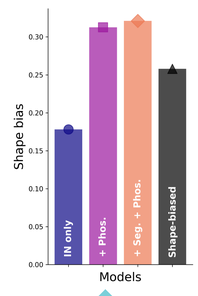
Pre-training dataset size explains performance





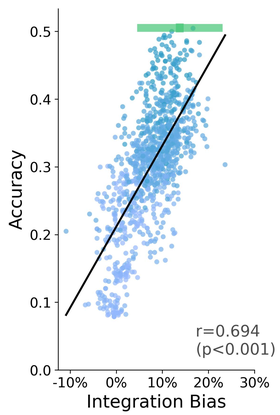
Training dataset size is the strongest predictive variable for performance.

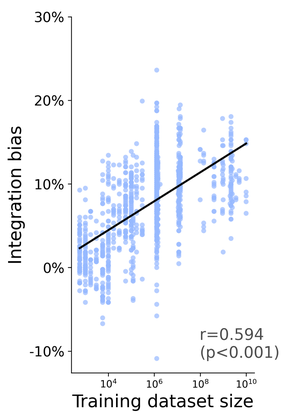
Architecture is largely unimportant for performance.

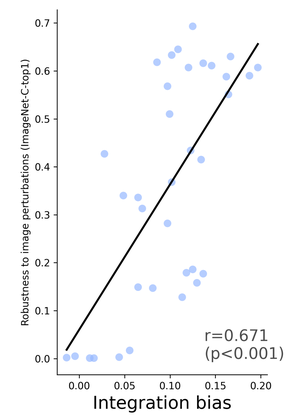


We trained ResNet-18 models on **fragmented versions of ImageNet**. They reach **higher shape bias** than models trained on Stylized ImageNet[4].

Contour integration underlies performance and robustness







Contour integration strongly correlates with performance ($\boldsymbol{r = 0.694}$).

Models learn contour integration to cope with complex and varied images.

The **more contour integration** models have, the **more robust** they are (ImageNet-C).

Financial support

*This work was supported by the Swiss National Science Foundation grant n. 176153 "Basics of visual processing : from elements to figures" and ERA-NET NEURON (Ref Nr: NEURON-051).

References

1. Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A Century of Gestalt Psychology in Visual Perception I. Perceptual Grouping and Figure-Ground Organization. Psychological bulletin.
2. Roelfsema, P. R. Cortical Algorithms for Perceptual Grouping. Annual Review of Neuroscience, 29(1):203–227, 2006. doi: 10.1146/annurev.neuro.29.051605.112939.
3. Rotermund, D., Scialom, E., Reprow, M., Herzog, M., &Ernst, U. (2024). davrot/percept simulator 2023.V1.0.0 (neuroprosthesis). Zenodo. doi: 10.5281/zenodo.10978899
4. Geirhos, R., Narayanappa, K., Mitzkus, B., Thieringer, T.,Bethge, M., Wichmann, F. A., & Brendel, W. (2021). Partial success in closing the gap between human and machine vision. NeurIPS.