Deep neural networks optimized for both face detection and face discrimination most accurately predict face-selective neurons in macaque inferior temporal cortex

Kohitij Kar (k0h1t1j@yorku.ca)
Department of Biology, York University
Toronto, Ontario, M3J1P3, Canada

Nancy Kanwisher (ngk@mit.edu)
Brain and Cognitive Sciences, Massachusetts Institute of Technology
Cambridge, Massachusetts, 02139, USA

Katharina Dobs (katharina.dobs@psychol.uni-giessen.de)
Justus-Liebig University Giessen
Giessen, 35394, Germany

Abstract

Artificial neural network (ANN) models trained on face identification (e.g., VGG-face) outperform models trained on object categorization (e.g., VGG-16, CORnet-Z) in predicting human face recognition behavior. Why then does the opposite hold for prediction of the responses of face-selective neurons in both humans and macaques? Here we test the hypothesis that face-specific neural machinery is optimized for both detecting and discriminating faces. Consistent with this hypothesis, we find that face-selective neural responses in macaque inferior temporal (IT) cortex are best fit by ANNs trained on both face identification and face detection (faces vs. objects).

Keywords: Face recognition; Neural encoding; Face neurons; Deep Convolutional Neural Networks; Inferior temporal cortex

Introduction

Face processing holds considerable significance in neuroscience, as a multifaceted cognitive process at the core of human social interactions and communication. A rich understanding of this system would include an image-computable computational model that can sufficiently explain face-related behaviors and the underlying brain circuitry that support them.

Previous research presents a puzzling contrast. On the one hand, networks trained on object recognition (VGG-obj) perform poorly compared to those trained on faces (VGG-face) on face recognition tasks (Dobs, Martinez, Kell, & Kanwisher, 2022). On the other hand, studies suggest that training on face identification is not a prerequisite for the emergence of brain-like neural face representations (Chang, Egger, Vetter, & Tsao, 2021; Vinken, Prince, Konkle, & Livingstone, 2022) This indicates that generic object features may contribute to constructing a representational space similar to the neural face space. Moreover, models trained on broad stimulus categories (e.g., ImageNet, Places) predict neural face-specific responses similarly well or even better than those trained on faces only (Grossman et al., 2019; Ratan Murty, Bashivan, Abate, DiCarlo, & Kanwisher, 2021).

Here we test the hypothesis that face-specific neural populations are engaged both in face detection (i.e., discriminating faces from other objects) and face identification (i.e. discriminating different faces from each other). Thus, ANN models trained on both faces and objects will outperform models trained on only faces or only objects in predicting face-selective neurons.

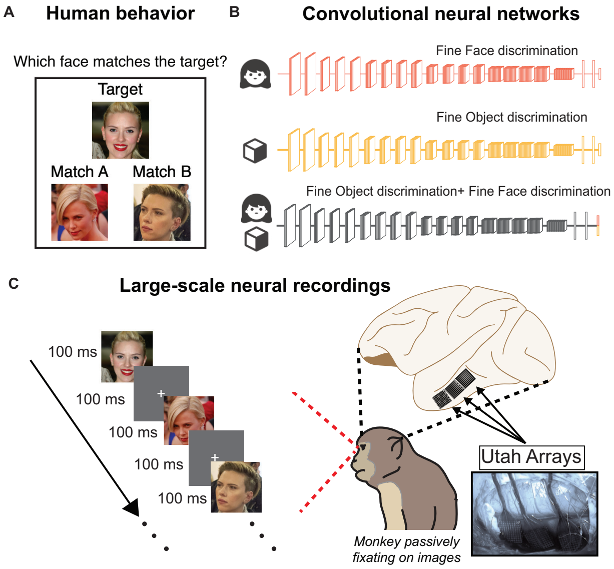
A
Human behavior
B
Convolutional neural networks
Which face matches the target?
Target
Match A
Match B
Fine Face discrimination
Fine Object discrimination
Fine Object discrimination+ Fine Face discrimination
C
Large-scale neural recordings
100 ms
100 ms
100 ms
100 ms
100 ms
Utah Arrays
Monkey passively fixating on images

Figure 1: **A.** Behavioral task to test facial identity discrimination performance in humans, **B.** development of ANNs with varying training diet and task objectives, and **C.** large-scale neural recordings across macaques IT to measure the representations of the face images (n=200) used in **A.**

Results

We used three VGG-16 networks (**Figure 1B**) trained from scratch on object categorization (VGG-obj), face identification (VGG-face) and both object and face discrimination (VGG-dual) (Dobs, Martinez, et al., 2022). By extracting activation patterns from the networks' penultimate layers for each of 200 face images used in a behavioral target-matching task (**Figure 1A;** Dobs, Yuan, Martinez, & Kanwisher, 2022), we compared the ANNs to human face discrimination accuracy. Consistent with previous findings, we observed that indeed VGG-face significantly outperformed VGG-obj (**Figure 2A**) and achieved human-like behavioral accuracy.

Using previously established methods (Kar, Kubilius, Schmidt, Issa, & DiCarlo, 2019), we recorded large-scale neural activity across macaque IT cortex (monkey A: 53 sites; monkey B: 67 sites) while the monkey passively fixated the same face images, presented for 100 ms on their central field of view (8 deg; **Figure 1C**). In line with previous results (Ratan Murty et al., 2021; Grossman et al., 2019; Chang et al., 2021), we found that VGG-obj better predicted macaque IT neurons' responses to the face images (*see* **Figure 2B**).

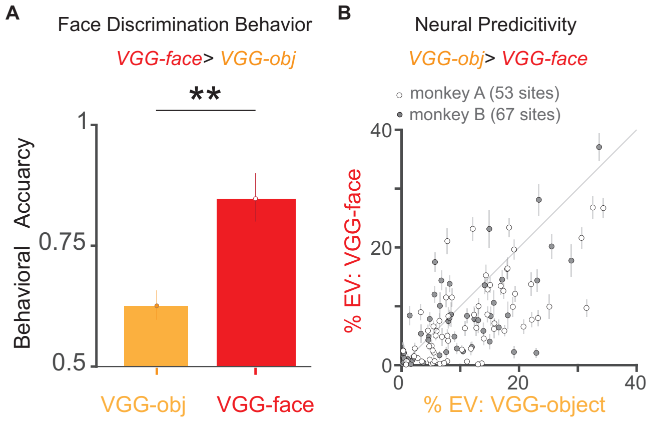

Figure 2: Behavior vs. Neural predictivity of ANNs. **A**. VGG-face significantly outperforms VGG-obj in the task shown in Figure 1A. **B.** VGG-obj better predicts macaque IT neurons' responses to the face images used in the behavioral tasks. Each dot refers to a neural site (white dots: monkey A, 53 sites; black dots: monkey B, 67 sites); paired t-test; p<0.0001; t(119) =5.6813. EV = explained variance

Next we tested the IT predictivity of a network trained on both face and object discrimination (VGG-dual; Dobs, Martinez, et al., 2022). Interestingly, we found that it predicted neural responses in macaque IT better than VGG-obj (**Figure 3A**). Furthermore, this improved performance increased as a function of the face selectivity of the recorded neural site (**Figure 3B**). But what is it about VGG-dual that enables it to outperform VGG-obj? If networks must be optimized both for face detection and face discrimination to account for neural responses, then a network that maintains separate output categories for each face but assigns all objects to one category (VGG-faceObj) should outperform a network trained on the same stimuli that assigns all faces to one category but maintains separate output categories for each object type (VGG-objFace; **Figure 4A**). Indeed, preliminary evidence supports this hypothesis, with improvement in neural predictivity of VGG-faceObj over VGG-objFace increasing with the face selectivity of the neuron (**Figure 4B**).

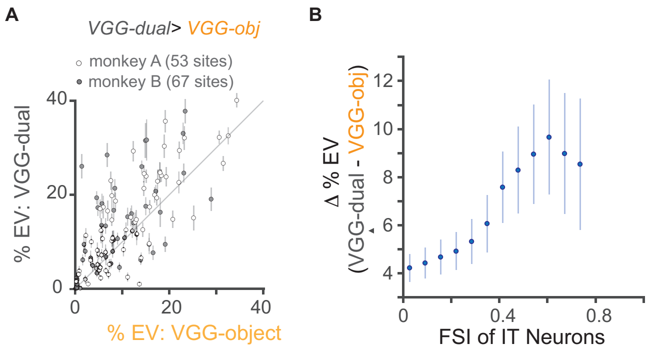

Figure 3: **A.** ANNs optimized for both tasks (VGG-dual) outperform VGG-obj, in predicting macaque IT neurons' responses to the face images used in the behavioral tasks. Each dot refers to a neural site (white dots: monkey A, 53 sites; black dots: monkey B, 67 sites). p<0.0001; t(119) =5.35. **B.** The increase in % Explained Variance is significantly correlated with the face-selectivity of the neural sites. R = 0.34 (neuron-by-neuron; 120 neurons), p <0.0001. Neurons are grouped together based on their face selectivity index (FSI: x-axis).

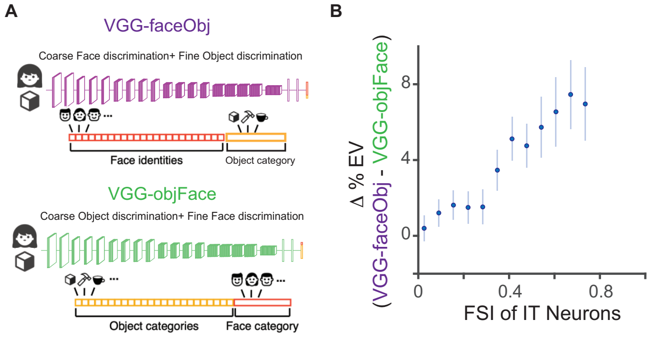

Figure 4: **A.** Two novel ANNs: VGG-faceObj trained on face-detection (face vs. object) and facial identity discrimination, VGG-objFace trained on face-detection (face vs. object) and object categorization. **B.** VGG-faceObj better predicts IT neurons. The difference in % Explained Variance is significantly correlated (R = 0.27, p = 0.0029) with the face-selectivity of the neural sites.Neurons are grouped together based on their face selectivity index (FSI: x-axis).

Discussion

This study supports the hypothesis that face-specific neural machinery is optimized for both face detection and face discrimination, as evidenced by the superior performance of VGG-dual and VGG-faceObj in predicting face-selective neural responses in macaque IT cortex. Moreover, these results provide an answer to why models trained on faces only - thereby being optimized for face discrimination but not face detection - do not predict face-selective neural responses as well as models trained on faces and objects.

Acknowledgments

We thank Sarah Goulding for support during data collection. This work was supported by the Canada Research Chair Program and the Simons Foundation Autism Research Initiative (SFARI)-Pilot award to K.K., by the DFG Collaborative Research Center SFB/TRR 135, and Research Cluster 'The Adaptive Mind' funded by the Excellence Program of the Hessian Ministry of Higher Education, Science, Research and Art to K.D., and by the National Institutes of Health grant DP1HD091947 to N.K. and National Science Foundation Science and Technology Center for Brains, Minds, and Machines.

References

Chang, L., Egger, B., Vetter, T., & Tsao, D. Y. (2021). Explaining face representation in the primate brain using different computational models. *Current Biology*, *31*(13), 2785–2795.

Dobs, K., Martinez, J., Kell, A. J., & Kanwisher, N. (2022). Brain-like functional specialization emerges spontaneously in deep neural networks. *Science advances*, *8*(11), eabl8913.

Dobs, K., Yuan, J., Martinez, J., & Kanwisher, N. (2022). Using deep convolutional neural networks to test why human face recognition works the way it does. *bioRxiv*, 2022–11.

Grossman, S., Gaziv, G., Yeagle, E. M., Harel, M., Mégevand, P., Groppe, D. M., ... others (2019). Convergent evolution of face spaces across human face-selective neuronal groups and deep convolutional networks. *Nature communications*, *10*(1), 4934.

Kar, K., Kubilius, J., Schmidt, K., Issa, E. B., & DiCarlo, J. J. (2019). Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nature neuroscience*, *22*(6), 974–983.

Ratan Murty, N. A., Bashivan, P., Abate, A., DiCarlo, J. J., & Kanwisher, N. (2021). Computational models of category-selective brain regions enable high-throughput tests of selectivity. *Nature communications*, *12*(1), 5540.

Vinken, K., Prince, J. S., Konkle, T., & Livingstone, M. (2022). The neural code for 'face cells' is not face specific. *bioRxiv*, 2022–03.