Configural processing as a fundamental mechanism for robust visual recognition in neural networks across varied viewing conditions

Hojin Jang (hojin4671@korea.ac.kr)
Department of Brain and Cognitive Engineering, Korea University
145, Anam-ro, Seongbuk-gu, Seoul, 02841, South Korea

Pawan Sinha (psinha@mit.edu)
Department of Brain and Cognitive Sciences, MIT
43 Vassar Street, Cambridge, Massachusetts, 02139, The United States

Xavier Boix (xboix@fujitsu.com)
Fujitsu Research
4655 Great America Pkwy, Santa Clara, California, 95054, The United States

Abstract:

A hallmark of face recognition is configural processing; that said, the underlying neurocomputational mechanisms are still elusive despite decades of research. Moreover, despite a few previous studies hinting at the benefits of configural processing under challenging conditions, detailed insights were scarce and mostly empirical. The study posits that recognizing faces through configural cues is more effective than using local features across variations in viewing conditions. This hypothesis was tested using face-like digit stimuli and comparing neural network models trained to recognize them with either local or configural cues. The findings demonstrate that neural networks can indeed discern configural cues, which notably enhance performance against geometric alterations like rotation and scaling. Additionally, when both types of cues are present, models prefer configural over local cues. Our results offer new neurocomputational evidence of the advantages of configural processing in reliably recognizing faces across diverse conditions.

Keywords: configural processing; face recognition; robustness.

Introduction

Humans excel in facial recognition, quickly identifying unique individuals by processing faces holistically instead of separately focusing on individual features (Farah et al., 1995; Farah et al., 1998; Kanwisher, 2000). This involves sensitivity to the spatial configurations and relationships among facial elements like the eyes, nose, and mouth, with even subtle differences detectable by typical individuals (Le Grand et al., 2001). Although configural processing is vital for recognizing faces, the reasons for its importance are not fully understood. Some theories suggest that configural processing originates from early-life blurred vision (Dobson & Teller, 1978; Le Grand et al., 2001; Vogelsang et al., 2018; Jang and Tong, 2021) or from the need to distinguish objects with greater detail and expertise, like faces (Gauthier and Tarr, 2002). This study proposes that it could also be an ecologically optimal strategy for reliable face recognition in varied conditions. The present study introduces a new perspective – that configural processing may be an ecologically optimal strategy for robust face recognition across diverse viewing conditions.

To investigate the computational bases of configural processing, this study employed deep learning models trained on tasks using either local or configural cues with face-like digit stimuli. The results showed that deep neural networks can leverage configural cues and prioritize them over local features, exhibiting greater robustness to geometric transformations like rotation and scaling. These findings offer new computational insights into the advantages of configural processing for reliable face recognition across varying conditions.

Methods

In this study, we created face-like patterns using four alphabetic characters, drawing from the EMNIST database. We then assessed the role of local and configural cues using two tasks (**Figure 1A**): the 'local' task, where categories share configurations but differ in letter combinations, and the 'configural' task, where categories share letters but differ in configurations. We also introduced a combined 'local plus configural' task that merges elements of both tasks (**Figure 2A**).

Various neural network models were then evaluated for each task.

Results

Figure 1B illustrates the accuracy of neural networks on both tasks given two types of transformations, rotation and scaling. It first highlights that local features are effective across various transformation levels. Additionally, it echoes Jang et al. (2023)'s observation that increasing the number of training classes, without altering the volume of training images, enhances generalization. For the configural task, performance improves significantly with the increase in class numbers, indicating the networks' abilities at configural recognition.

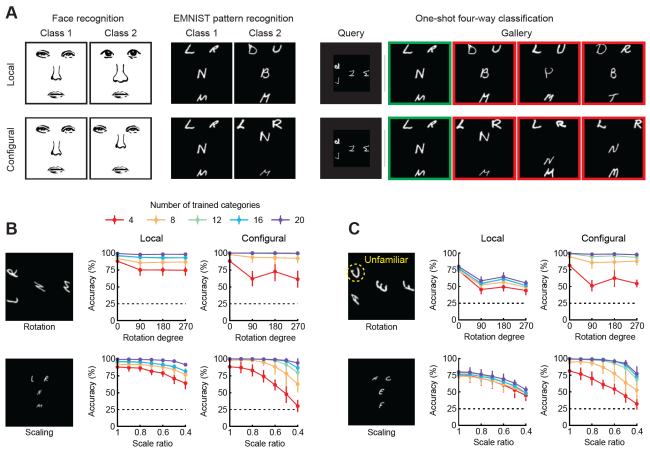

Figure 1: **A** Illustration of the local and configural tasks. Generalization performance on both tasks given rotation and scaling transformations when individual features were familiar (**B**) and unfamiliar (**C**).

In addition, **Figure 1C** demonstrates the impact of using unfamiliar alphabet letters on task performance. The local task suffers in accuracy but remains above chance, while the configural task maintains consistent performance, underscoring configural processing's robustness and independence from local features.

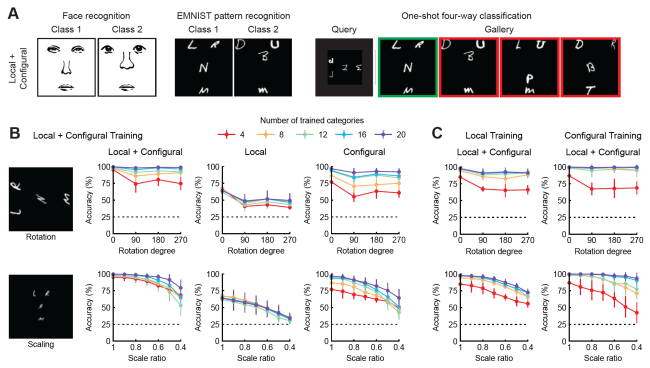

Figure 2: **A** Illustration of the local plus configural task. **B** Generalization performance of neural networks trained on the local plus configural task and tested on each of the three tasks. **C** Generalization performance of neural networks trained on either the local or the configural task, and tested on the local plus configural task.

The study also investigates which type of cue, local or configural, is more dominant in scenarios involving both. From the local plus configural task, networks showed a clear preference for configural information when tasked to use both cues (**Figure 2B**). This preference suggests configural cues' superior role in ensuring reliable recognition across different transformations. When networks trained on either local or configural tasks were evaluated on the same test, specifically the local plus configural task, it was observed that the networks trained on the configural task demonstrated marginally better robustness in performance compared to those trained on the local task, as illustrated in **Figure 2C**.

Conclusions

In our study, we used deep neural networks to examine how configural cues aid in recognizing alphabet patterns undergoing geometric transformations such as rotation and scaling. We found that configural cues are crucial for accurate recognition amidst varying local features, supporting the idea that configural processing is key for robust face recognition under varied viewing environments.

Acknowledgments

This research was supported partly by the following grants from Korea University, K2404751, and also by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via [2022-21102100009]. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

References

Dobson, V., & Teller, D. Y. (1978). Visual acuity in human infants: a review and comparison of behavioral and electrophysiological studies. Vision Research, 18(11), 1469–1483.

Farah, M. J., Wilson, K. D., Drain, H. M., & Tanaka, J. R. (1995). The inverted face inversion effect in prosopagnosia: Evidence for mandatory, face-specific perceptual mechanisms. Vision research, 35(14), 2089-2093.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is"" special"" about face perception?. Psychological review, 105(3), 482.

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: bridging brain activity and behavior. Journal of Experimental Psychology: Human Perception and Performance, 28(2), 431.

Jang, H., & Tong, F. (2021). Convolutional neural networks trained with a developmental sequence of blurry to clear images reveal core differences between face and object processing. Journal of vision, 21(12), 6-6.

Jang, H., Zaidi, S. S. A., Boix, X., Prasad, N., Gilad-Gutnick, S., Ben-Ami, S., & Sinha, P. (2023). Robustness to Transformations Across Categories: Is Robustness Driven by Invariant Neural Representations?. Neural Computation, 35(12), 1910-1937.

Kanwisher, N. (2000). Domain specificity in face perception. Nature neuroscience, 3(8), 759-763.

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2001). Early visual experience and face processing. Nature, 410(6831), 890–890.

Vogelsang, L., Gilad-Gutnick, S., Ehrenberg, E., Yonas, A., Diamond, S., Held, R., et al. (2018). Potential downside of high initial visual acuity. Proceedings of the National Academy of Sciences, 115(44), 11333–11338.