Leveraging Artificial Neural Networks to Enhance Diagnostic Efficiency in Autism Spectrum Disorder: A Study on Facial Emotion Recognition

Kushin Mukherjee*(kmukherjee2@wisc.edu)
Department of Psychology, University of Wisconsin-Madison
Madison, WI-53715, USA

Na Yeon Kim* (nayeon@caltech.edu)
Division of the Humanities and Social Sciences, California Institute of Technology
Pasadena, CA-91125, USA

Shirin Taghian Alamooti (staghian@yorku.ca)
Department of Biology, York University
Toronto, Ontario, M3J1P3, Canada

Ralph Adolphs (radolphs@caltech.edu)
Division of the Humanities and Social Sciences, Division of Biology and Biological Engineering, California Institute of Technology
Pasadena, CA-91125, USA

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

∗ indicates equal contribution

Abstract

The ability to recognize emotions and intent in facial expressions varies significantly between neurotypical (NT) individuals and those with autism spectrum disorder (ASD). Traditional inferential models often utilize high-level categorical descriptors of stimuli, neglecting the variance introduced by image-level sensory representations. This study investigates whether accounting for image-level differences can improve the development of diagnostic image-sets for emotion recognition tasks. We employ image-computable artificial neural network (ANN) models of primate vision, fine-tuning them on an existing dataset to predict the behavior of NT and ASD adults. Using these ANNs, we select a new set of images that are predicted to yield the largest differences in performance between NT and ASD subjects. Subsequently, we conduct facial emotion discrimination tasks and find that the ANN-selected images produce significantly larger behavioral gaps between the groups compared to a random selection of images. Notably, the diagnostic efficiency of the selected images can be predicted by the ANNs' ability to predict NT subject behavior. Our findings suggest that ANN models of vision could offer valuable clinical translation benefits for autism research, opening up new avenues of exploration.

Keywords: autism; artificial neural networks; facial emotion discrimination; diagnostic efficiency

Introduction

Autistic individuals exhibit differences in social behavior due to distinct processing of emotions from faces compared to NT individuals. A better understanding of the computations that underlie these differences in emotion processing could help in efficiently identifying the behavioral markers of autism. However, the absence of image-computable models has historically limited the mapping of the image-level properties in photographs of faces to behavioral patterns. In recent years, several artificial neural networks (ANNs) have been developed that can match human behavior on object recognition tasks and whose internal representations can explain patterns of neural activity in the primate ventral visual pathway (Yamins et al., 2014; Rajalingham, Schmidt, & DiCarlo, 2015; Nayebi et al., 2018; Kar, Kubilius, Schmidt, Issa, & DiCarlo, 2019). Kar (2022) examined how one can leverage the image-level predictions of such brain-mapped ANNs to more efficiently probe the behavioral differences observed between the NT population and individuals with ASD (Wang & Adolphs, 2017). Kar (2022) observed that ANN models of primate vision trained on varied objectives can perform human-like facial emotion judgments. Interestingly, the ANNs' image-level behavioral patterns better matched the NT subjects' behavior than those measured in adults with ASD. In this study, we tested whether these ANNs can help guide experimental design by accurately predicting the differences in emotional judgements on a set of images across NT and ASD populations. These predictions can therefore be leveraged to generate more diagnostic stimuli for investigating the differences between these two populations.

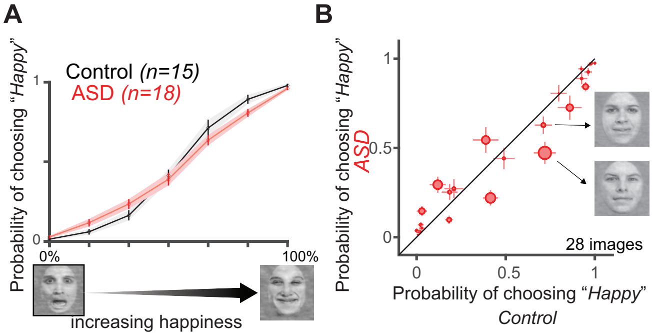

Figure 1: **A.** Wang and Adolphs (2017) tested NT (Control; n=15) and ASD (n=18) subjects on a facial emotion discrimination task. ASD (red curve), on average, showed lower specificity (slope of the psychometric curve) compared to the Controls (black curve). The shaded area and errorbars denotes SEM across participants. **B.** Image-level differences in NT behavior vs. ASD in the Wang and Adolphs (2017) study. Each red dot corresponds to an image. The size of the dot is scaled by the difference in behavior between the Controls and ASD. Errorbars denote SEM across subjects. Two example images are highlighted that show similar emotional ("happiness") judgments by the Controls but drive significantly different behaviors in individuals with ASD — demonstrating the importance of investigating behavior at the image-level.

Results

Similar to methods proposed by Kar (2022), we fine-tuned representations from IT-analogous layers of several ANNs (**Figure 2A**) to predict ASD and NT facial emotion judgements using the data collected by Wang and Adolphs (2017). Specifically, we extracted ANN representations from the pre-classification layer for a set of 7 ANN vision models — AlexNet (Krizhevsky, Sutskever, & Hinton, 2012), ResNet-50 (He, Zhang, Ren, & Sun, 2016), VGG-19 (Simonyan & Zisserman, 2014), ConvNext-base (Liu et al., 2022), ViT-base (Dosovitskiy et al., 2020) , CLIP-ViT (Radford et al., 2021), and CORNet-S (Kubilius et al., 2019). For CLIP we extracted the feature activation from the 'visual' layer of the visual encoder. We then used these ANN models to select images from a new imageset (Montreal Set of facial displays of emotion (MSFDE): consisting of emotional facial expressions by men and women of European, Asian, and African descent) that were predicted to yield large differences in judgement performance between NT and ASD subjects. Finally, we conducted a new in-lab behavioral experiment with NT and ASD subjects to test whether these hypothesized diagnostic images were more predictive of behavioral differences than images chosen at random from the same dataset. In this

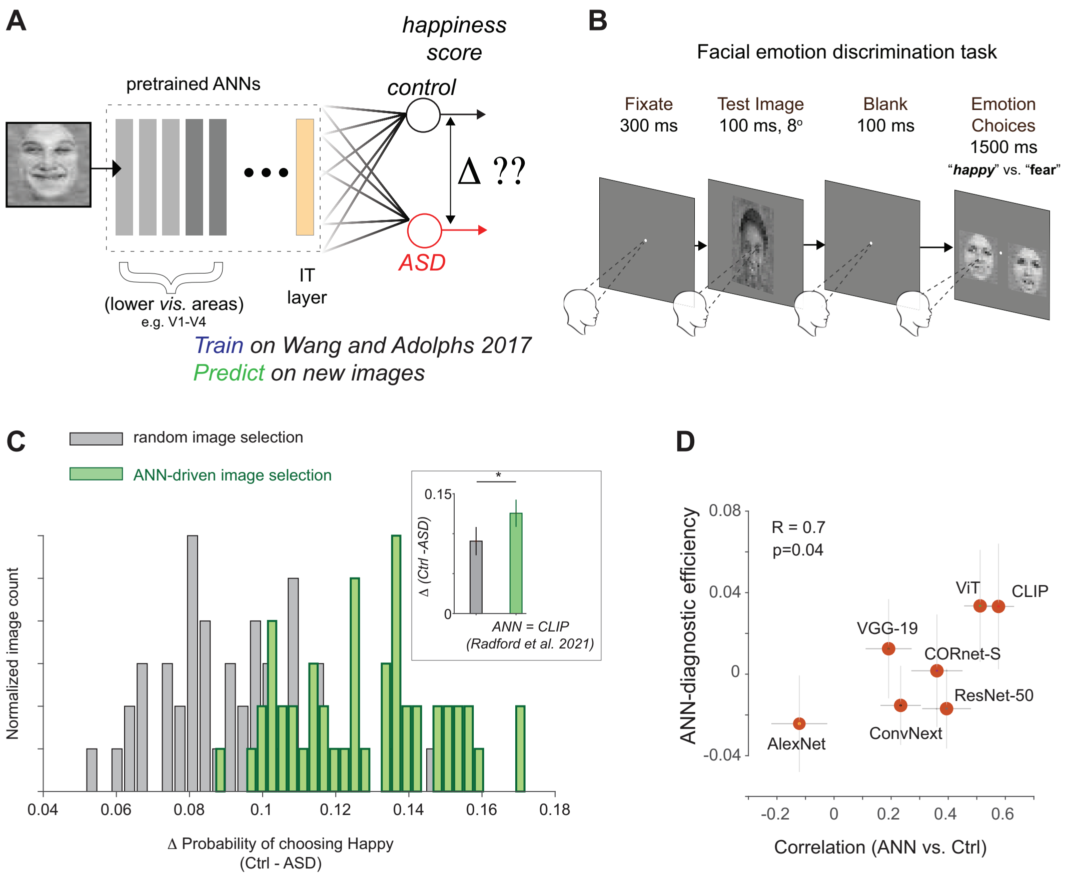

Figure 2: **A.** Our approach: We fine-tune the ANNs on the data (both ASD and NT) from the Wang and Adolphs (2017), and make predictions about the Δ behavior for a new set of images. **B.** On the new set of images (n=80), human participants, both NT (Control; n=13) population and autistic individuals (ASD; n=12) viewed a Test image (face) for 100 ms in their central ∼8 deg, followed by a choice screen with two images of an extreme happy and an extreme fearful face. Subjects had to press a key to indicate which emotion was present in the Test image (fearful or happy?). **C.** Distribution of the difference Δ in scores from NT (Ctrl) and autistic (ASD) adults for two sets of images. The green bars refer to 24 images that were selected based on ANN-based efficiency, and the gray bars are 24 random selected images. The inset shows that the average difference is statistically significant. **D.** ANN diagnostic efficiency (the difference between the NT vs. ASD gap for the diagnostic (ANN-predicted) image-set and the randomly selected image-set) shown as function of the ANN models' original predictivity of the NT behavior.

experiment, we recruited 12 adults with ASD (2 females, mean age = 32.1, age range = 23-42 years) who met the DSM-5 diagnostic criteria for ASD and the Autism Diagnostic Observation Schedule-2 (ADOS-2; module 4) criteria, and 13 NT subjects (2 females, mean age = 32.2, age range = 24-39 years) with no history of psychiatric or neurological disease and no family history of autism. All subjects had normal-range IQ and corrected-to-normal visual acuity. In this task, subjects were shown one image at a time and asked to indicate whether the emotion expressed by the face in the image was 'fearful' or 'happy'. The images from MSFDE had 5 morph levels between fearful and happy.

ANN-driven images produces larger behavioral gap across NT and ASD subjects We found that these selected images from several ANN predictions (specifically CLIP and ViT) did indeed lead to significantly larger behavioral differences (estimated by permutation tests) between ASD and NT subjects relative to a random selection of images from the same set. The degree to which the selected images were diagnostic of behavioral differences was predicted by the how well a given ANN could predict NT responses (Pearson $R = 0.7,\ p = 0.04$; **Figure 2D**). This highlights that higher performing ANNs also show greater utility as diagnostic tools for guiding experimental design.

Conclusion

Our finding suggests that higher performing ANNs can serve as effective diagnostic tools for guiding experimental design in autism research.

Acknowledgement

This work was supported by the Canada Research Chair Program and the Simons Foundation Autism Research Initiative (SFARI)-Pilot award to K.K. This work was funded in part by a SFARI Human Cognitive and Behavioral Science grant to R.A.

References

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., . . . others (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Kar, K. (2022). A computational probe into the behavioral and neural markers of atypical facial emotion processing in autism. *Journal of Neuroscience*, *42*(25), 5115–5126.

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.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*, 1097–1105.

Kubilius, J., Schrimpf, M., Hong, H., Majaj, N. J., Rajalingham, R., Issa, E. B., . . . DiCarlo, J. J. (2019). Brain-Like Object Recognition with High-Performing Shallow Recurrent ANNs. In H. Wallach, H. Larochelle, A. Beygelzimer, F. D'Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in neural information processing systems* (pp. 12785—-12796). Curran Associates, Inc.

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A convnet for the 2020s. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 11976–11986).

Nayebi, A., Bear, D., Kubilius, J., Kar, K., Ganguli, S., Sussillo, D., . . . Yamins, D. L. (2018). Task-driven convolutional recurrent models of the visual system. *Advances in neural information processing systems*, *31*.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., . . . Sutskever, I. (2021). *Learning transferable visual models from natural language supervision.*

Rajalingham, R., Schmidt, K., & DiCarlo, J. J. (2015). Comparison of object recognition behavior in human and monkey. *Journal of Neuroscience*, *35*(35), 12127–12136.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Wang, S., & Adolphs, R. (2017). Reduced specificity in emotion judgment in people with autism spectrum disorder. *Neuropsychologia*, *99*, 286–295.

Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the national academy of sciences*, *111*(23), 8619–8624.