Different generalization capabilities of face and object processing in a developmental sequence of blurry to clear images

Hojin Jang and Frank Tong

Department of Psychology, Vanderbilt University

VANDERBILT
Vision Research Center

TONGLAB
VISION, NEUROIMAGING, AND COMPUTATIONAL NEUROSCIENCE

Introduction

❑ Motivation

- Visual acuity is poor at birth and gradually improves over the first 6 months[1]
- Clinical research suggests that initial blurry vision may be important for the development of integrative visual processing, and perhaps especially so for faces[2, 3]
- A recent convolutional neural network (CNN) study found that CNNs trained with a sequence of blurry to clear face images could maintain robustness to blur, consistent with more integrative spatial processing[4]
- Given that face and object processing involve distinct neural mechanisms[5, 6], would initial experience with blurry images necessarily enable robust object recognition?

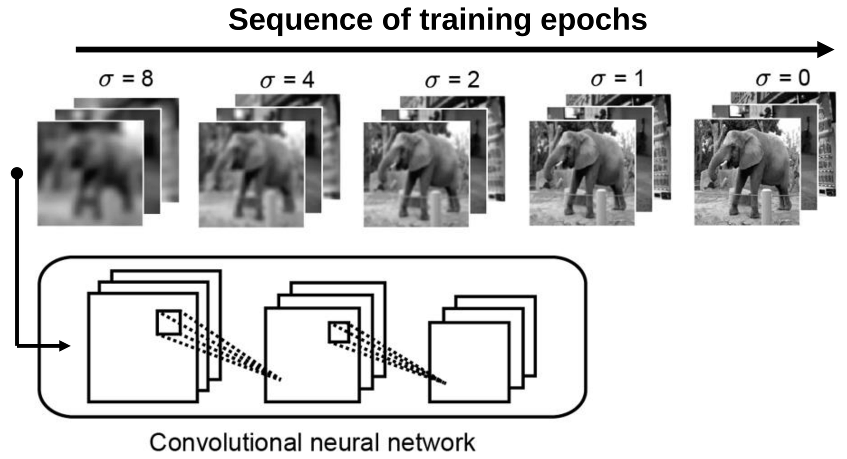
❑ Aim of the study

- To investigate whether training CNNs with a sequence of blurry to clear object images can account for the blur-robust nature of human object recognition

Methods

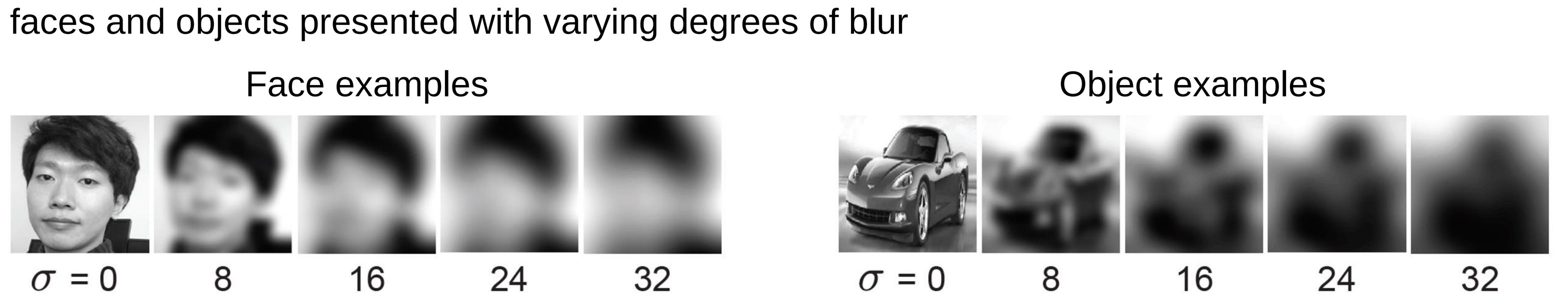
❑ CNN training procedures

- We trained 4 separate CNNs (AlexNet architecture[7]) to classify either faces (FaceScrub[8]) or objects (ImageNet[9]) using the following protocol:
  1. Blurry to clear face-trained CNN
  2. Clear face-trained CNN
  3. Blurry to clear object-trained CNN
  4. Clear object-trained CNN



❑ Behavioral experiments

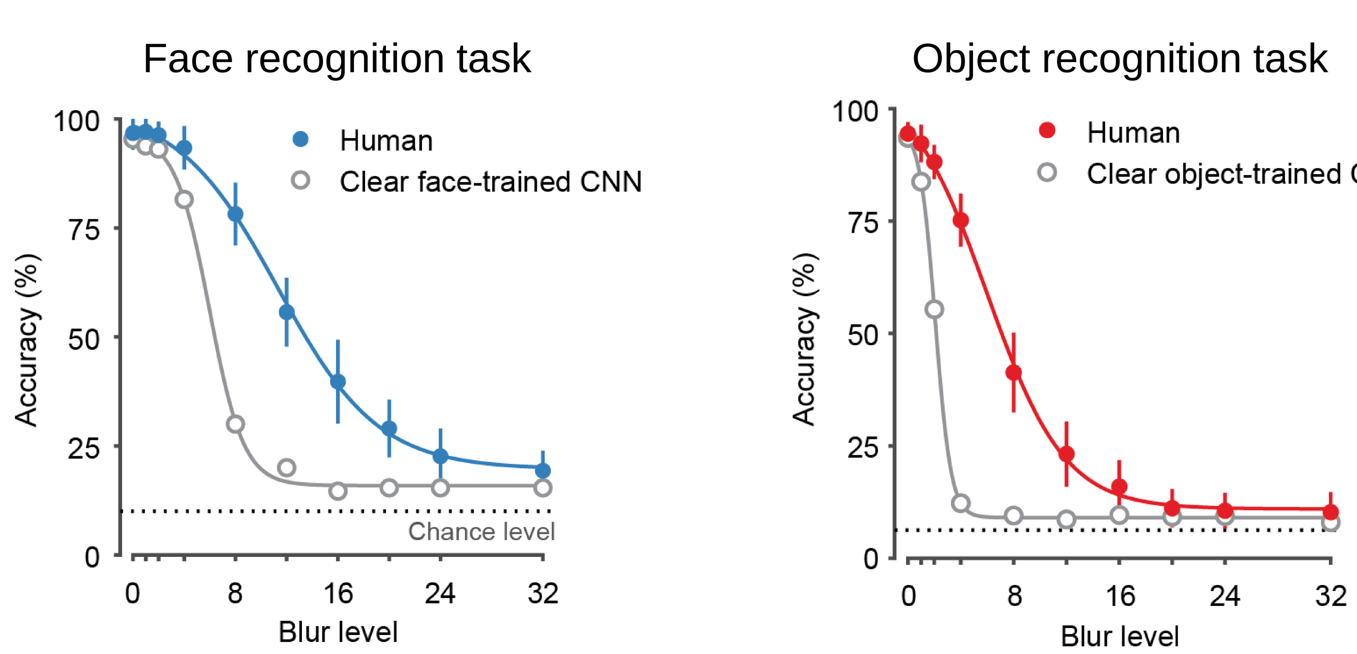
- We compared the performance of human observers (*n*=20 for each task) and CNNs at recognizing faces and objects presented with varying degrees of blur



(σ, standard deviation of a Gaussian kernel; Images of one of the authors are shown (with permission) instead of celebrities to avoid photographic copyright issues)

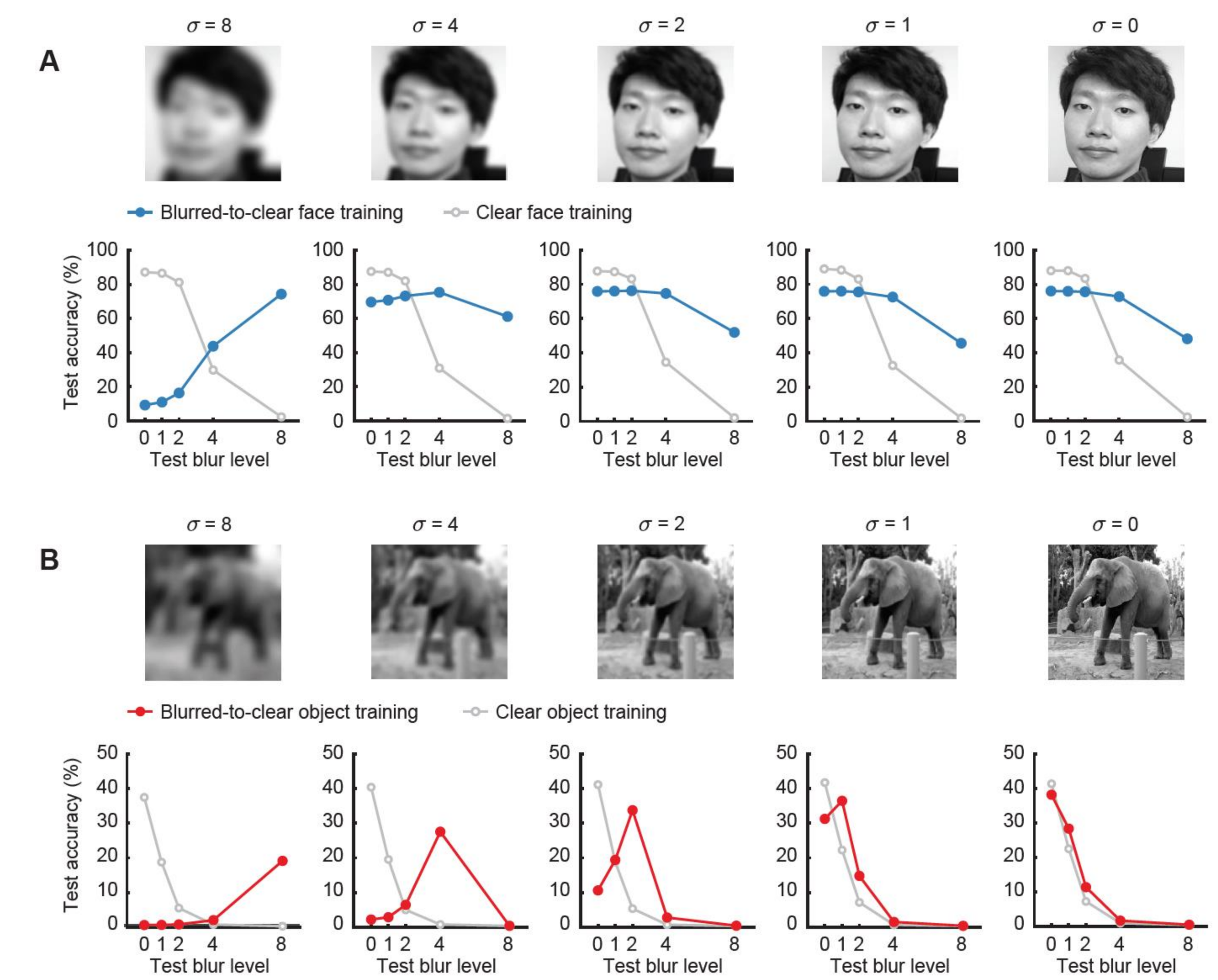
- A subset of face and object examples were used for behavioral experiments (10 celebrities, 5 females and 5 males; 16 object categories, 8 animate and 8 inanimate)

Results: Human vs. CNN performance



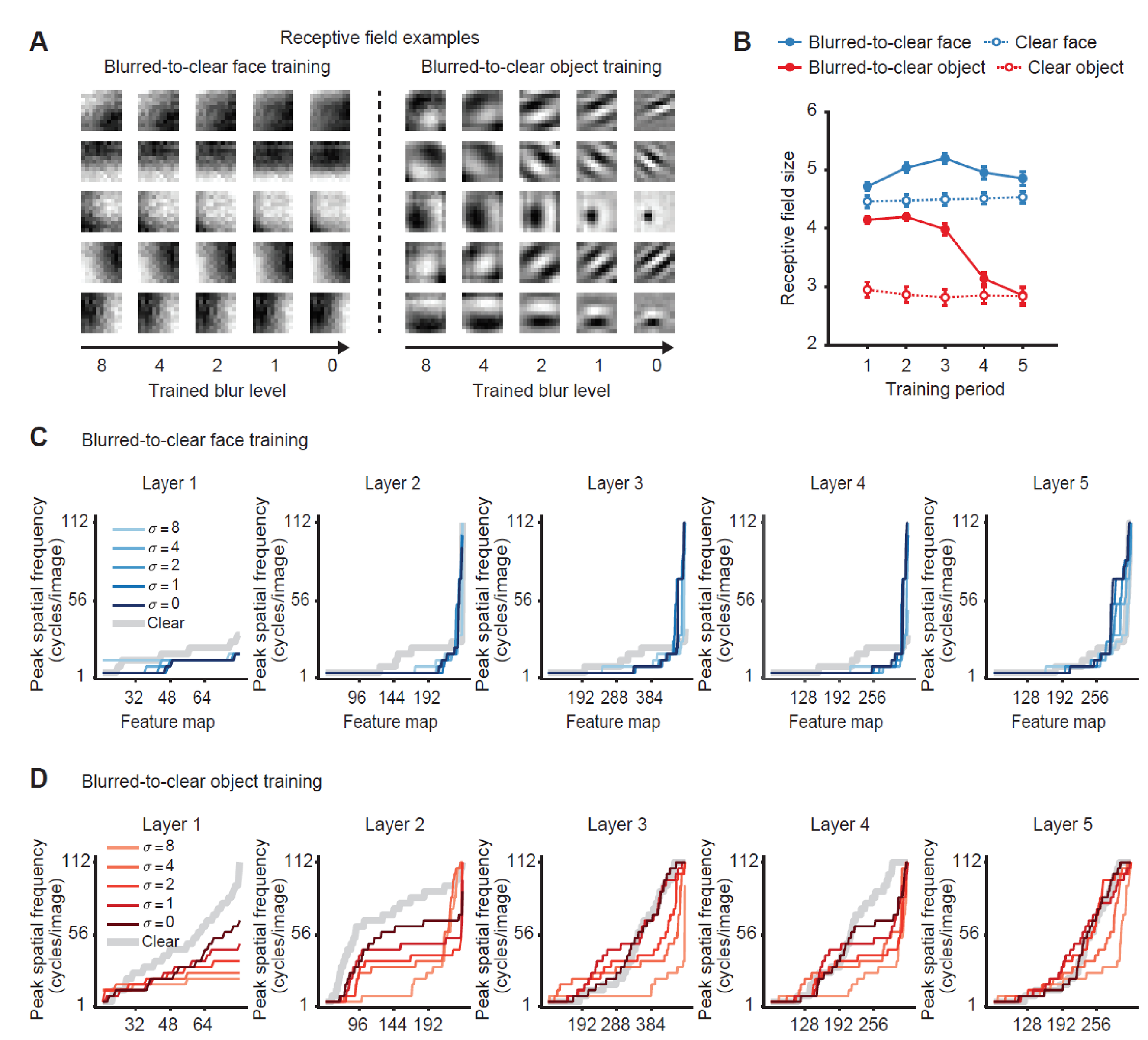
- Human observers outperformed CNNs on both face and object recognition tasks
- Blurry objects were more challenging than blurry faces for both humans and CNNs

Results: CNN performance across training epochs



- The CNN trained on blurry to clear faces shows sustained robustness to blur (**A**)
- By contrast, the CNN trained on blurry to clear objects quickly loses its initial robustness to blur (**B**)
- These results indicate a critical difference between face and object recognition tasks when dealing with blurred images

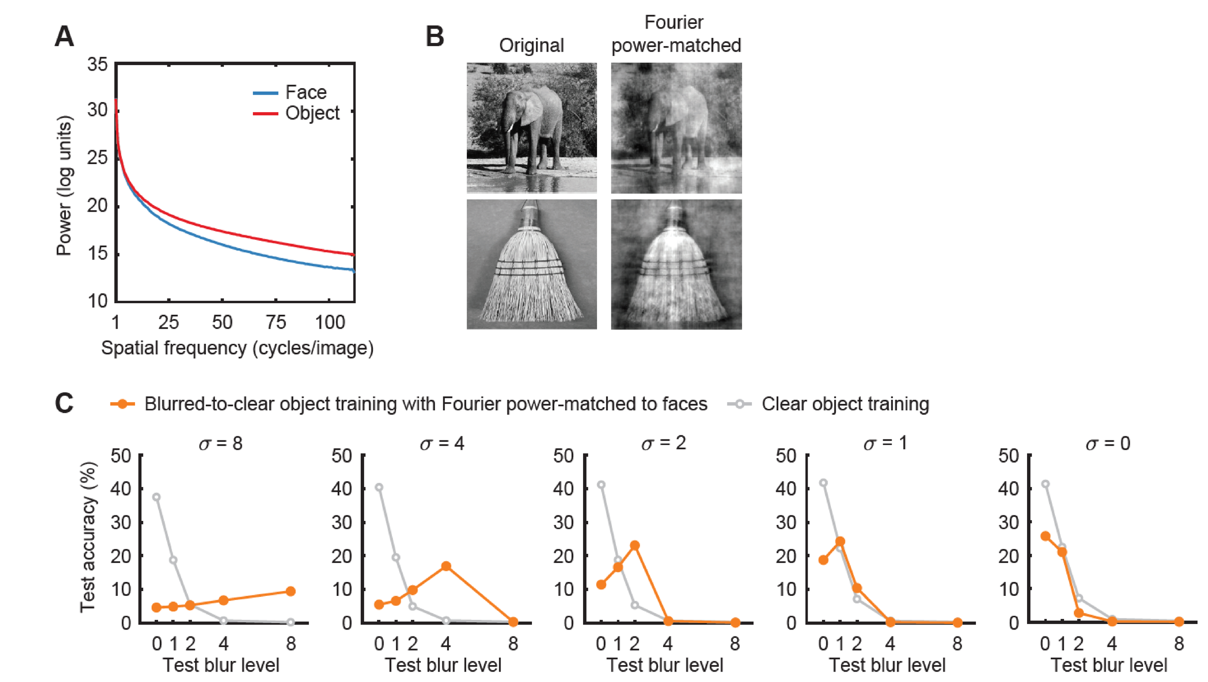
❑ Receptive field and spatial frequency tuning analyses



- Layer 1 receptive fields remained stable across training epochs for blurry to clear faces (**A**). However, they became progressively smaller and shifted towards preferring finer detail during blurry to clear training with objects.
- We presented multiple grating patterns to estimate the preferred spatial frequencies of the network across layers. While the blurry to clear face-trained CNN showed stable spatial frequency preferences across epochs (**C**), the object-trained CNN underwent substantial changes towards preferring higher spatial frequencies (**D**).

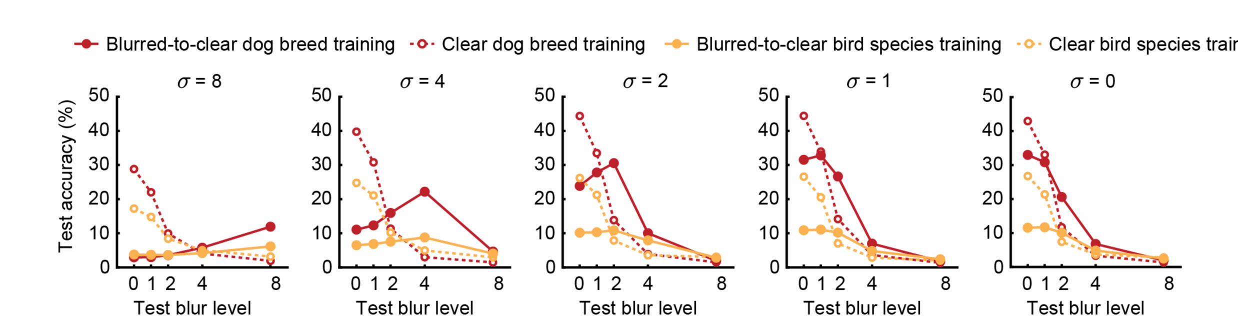
Results: Control analyses

❑ Fourier power spectrum control analysis



- Faces and objects differ in their spatial frequency content (**A**)
- If object images are matched to the Fourier power spectrum of faces (**B**), is this enough to enable robust learning?
- We matched the power spectrum of the object training images to that of the face images (**B**), but still failed to see any beneficial effects (**C**)

❑ Subordinate-level object recognition



- Because all faces share a common configuration, whereas objects do not, would restricting the object classification task to a single basic level category enable successful blurry to clear image training?
- We tested CNNs trained on subordinate-level categorization (i.e., different dog breeds or bird species) but found negligible benefit of blurry to clear image training

Conclusions

- We observed that only the face-trained CNN benefited from the early training with blurry images. This result may reveal a fundamental difference in face and object processing, such that faces can be processed in a more holistic manner than objects.
- We failed to find that early experiences with blurry objects accounts for the robust nature of human vision to blur. That said, our findings do not rule out the possibility that poor initial acuity in infants might have developmental benefits for adult facial recognition.

References

1. 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.
2. Geldart, S., Mondloch, C. J., Maurer, D., De Schonen, S., & Brent, H. P. (2002). The effect of early visual deprivation on the development of face processing. Developmental Science, 5(4), 490-501.
3. Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2001). Early visual experience and face processing. Nature, 410(6831), 890-890.
4. 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.
5. Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. Journal of Neuroscience, 17(11), 4302-4311.
6. Tsao, D. Y., Freiwald, W. A., Tootell, R. B., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. Science, 311(5761), 670-674.
7. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 1097-1105.
8. Ng, H. W., & Winkler, S. (2014). A data-driven approach to cleaning large face datasets. Paper presented at the 2014 IEEE International Conference on Image Processing.
9. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). Imagenet large scale visual recognition challenge. International Journal of Computer Vision, 115(3), 211-252.

This study was supported by a grant from the National Eye Institute (R01EY029278)