Investigating links between artificial neural networks and human visual perception

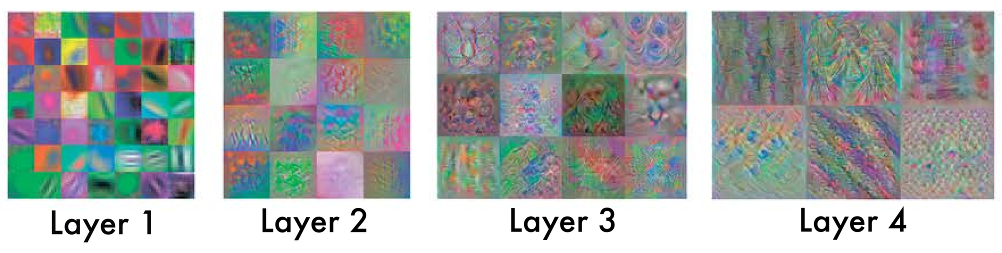
Kirill Korotaev, Georgiy Zhulikov and Joseph W. MacInnes

HIGHER SCHOOL OF ECONOMICS
NATIONAL RESEARCH UNIVERSITY

Motivation

Advances in machine learning using Deep Neural Networks (DNN) rival human-level performance in visual recognition tasks (He, Zhang, Ren, & Sun, 2015; LeCun, Bengio, & Hinton, 2015).

Although most studies have engineering goals and don't account for biological plausibility it has been shown that features learned in the hidden layers of neural network resemble hierarchical processing stages in the human visual cortex(Cadieu et al., 2014; Güçlü & van Gerven, 2014; Kriegeskorte, 2015)



Visualization of Convolutional Neural Network based on AlexNet (Krizhevsky, Sutskever, & Hinton, 2012). Feature detectors gradually increase in size and complexity, as in the primate visual cortex. Visualization technique by Zeiler & Fergus, 2014

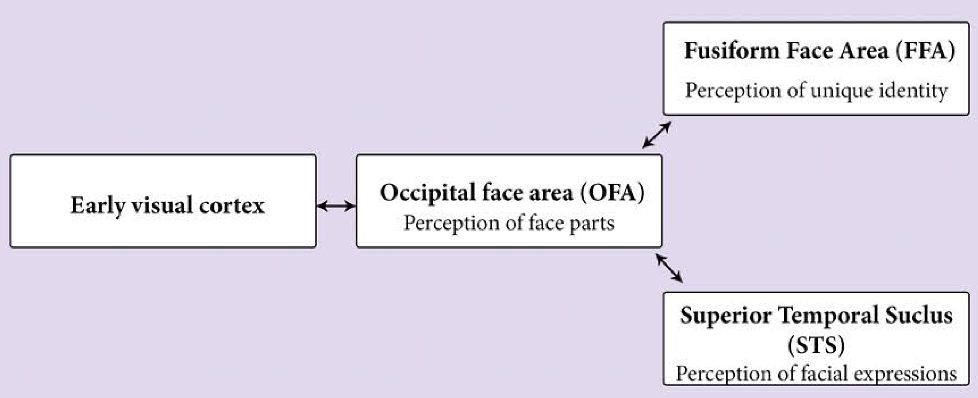
Background

Face Inversion Effect

Converging evidence from psychology and neuropsychology indicated the presence of Face Inversion Effect (FIE) in humans.

Although it's harder to recognize any object from upside-down, face recognition is **disproportionately** impaired by inversion (Epstein, Higgins, Parker, Aguirre, & Cooperman, 2006; Moscovitch, Winocur, & Behrmann, 1997; Yin, 1969).

Face Perception Network



The effect is partly explained by hierarchical Face Perception Network - according to neuroscience data humans possess brain regions that activate selectively to faces, but less or not at all to other objects (Haxby et al., 2000; Kanwisher et al., 1997; Yovel & Kanwisher, 2005).

Research question

Do layers of artificial deep neural networks process visual information in a functionally similar way with areas of human visual cortex?

Hypothesis:

If DNN perceive visual information in the same way as humans they would :

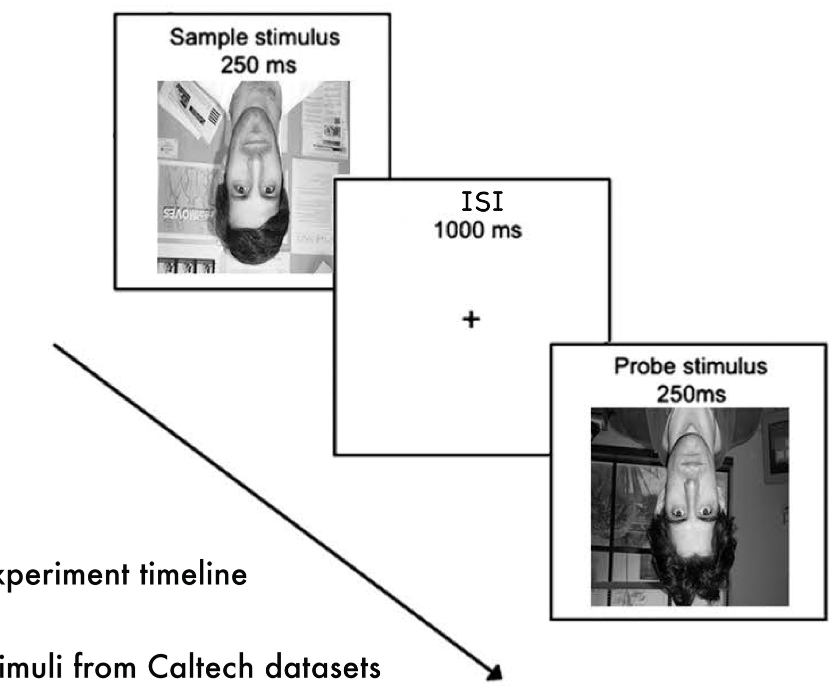
1) recognize inverted objects worse than upright

2) recognize inverted faces worse than inverted houses

To test the hypothesis we acquired human data and compared it with neural network performance, using stimuli from Caltech Faces and Caltech Houses datasets

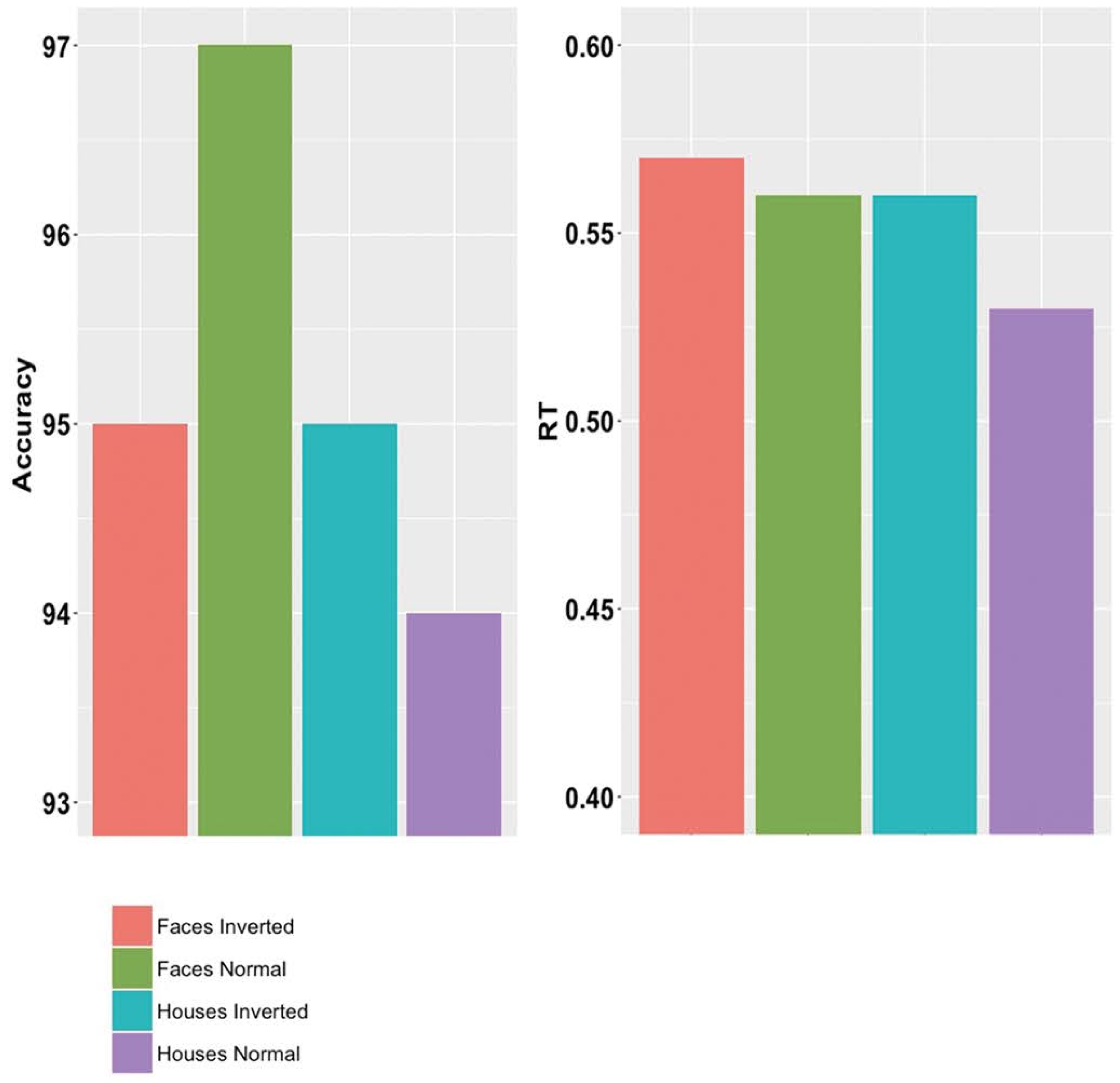
Pilot experiment

In our experiment subjects participated in n-1 matching task, that consisted of four blocks of images with fixed orientation : upright faces, upright houses, inverted faces and inverted houses with counterbalancing across blocks. Each stimuli was presented for 250 ms



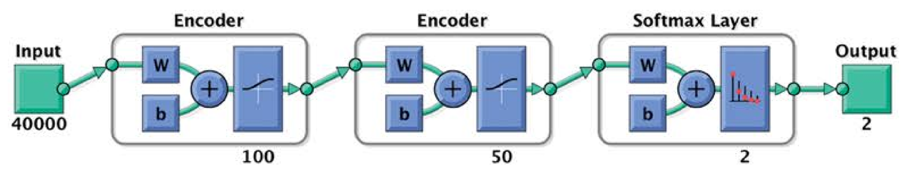
Experiment timeline

Stimuli from Caltech datasets



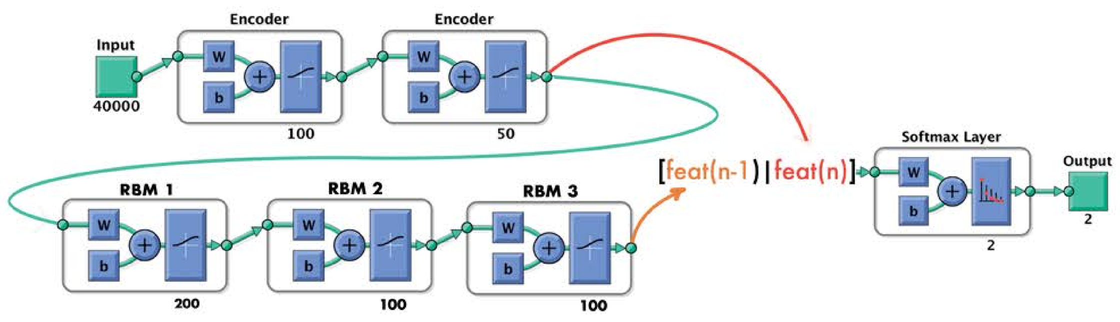
Computational Modeling

Attempt 1: Deep Autoencoder



Our initial results with Deep Neural Network stacked of 2 Autoencoders and a Softmax layer are inconsistent with behavioral data. Although neural network achieved reasonable accuracy when it was tested on upright images of faces and houses, it didn't perform rotation well.

Attempt 2 : Deep Autoencoder + DBN

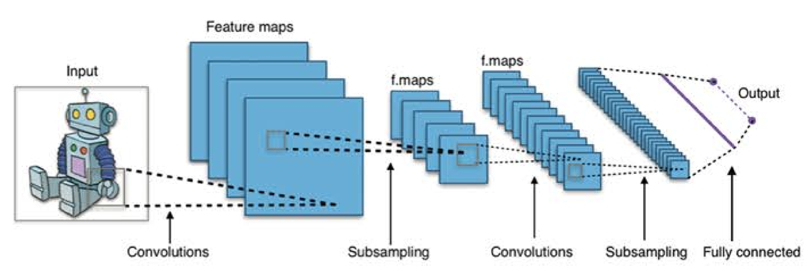


Because matching and classification are different, we attempted to add stacked RBMs to our network for modeling temporal memory storage.

Autoencoder learns unsupervised, and DBN memorizes its learned features. Softmax layer receives simultaneous inputs from Autoencoder and DBN and decides whether features were generated from same image or not.

However, it always concludes that images are different

Attempt 3: ConvNet



Considering that Convolutional Neural Networks (CNN) currently dominate computer vision and remain state-of-the art at object recognition, we decided to test how they would respond to inversion.

As in attempt 1, we trained it on upright images of faces and houses while tested on both upright and inverted. ConvNet achieved 100% accuracy on upright images and it performance dropped only by 0,6% when tested on inverted

References:

Cadieu, C. F., Hong, H., Yamins, D. L. K., Pinto, N., Ardila, D., Solomon, E. A., ... DiCarlo, J. J. (2014). Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition. PLoS Computational Biology, 10(12). https://doi.org/10.1371/journal.pcbi.1003963

Epstein, R. A., Higgins, J. S., Parker, W., Aguirre, G. K., & Cooperman, S. (2006). Cortical correlates of face and scene inversion: A comparison. Neuropsychologia, 44(7), 1145–1158. https://doi.org/10.1016/j.neuropsychologia.2005.10.009

Güçlü, U., & van Gerven, M. A. J. (2014). Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Brain's Ventral Visual Pathway, 35(27), 10005–10014. https://doi.org/10.1523/JNEUROSCI.5023-14.2015

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. Trends in Cognitive Sciences, 4(6), 223–233. https://doi.org/10.1016/S1364-6613(00)01482-0

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. CoRR, abs/1502.0. https://doi.org/10.1109/ICCV.2015.123

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 17(11), 4302–11. https://doi.org/10.1098/Rstb.2006.1934

Kriegeskorte, N. (2015). Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. Annual Review of Vision Science, 1(1), 417–446. https://doi.org/10.1146/annurev-vision-082114-035447

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Advances In Neural Information Processing Systems, 1–9. https://doi.org/http://dx.doi.org/10.1016/j.protcy.2014.09.007

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Moscovitch, M., Winocur, G., & Behrmann, M. A. R. L. E. N. E. (1997). What Is Special about Face Recognition? Nineteen Experiments on a Person with Visual Object Agnosia and Dyslexia but Normal Face Recognition. Journal of Cognitive Neuroscience, 9(5), 555–604. https://doi.org/10.1162/jocn.1997.9.5.555

Yin, R. K. (1969). Looking at upside-down faces. Journal of experimental psychology, 81(1), 141.

Yovel, G., & Kanwisher, N. (2005). The neural basis of the behavioral face-inversion effect. Current Biology, 15(24), 2256–2262. https://doi.org/10.1016/j.cub.2005.10.072

Zeiler, M. D., & Fergus, R. (2014). Visualizing and Understanding Convolutional Networks arXiv:1311.2901v3 [cs.CV] 28 Nov 2013. Computer Vision–ECCV 2014, 8689, 818–833. https://doi.org/10.1007/978-3-319-10590-1_53