Project Report
Human Face Recognition

Radhika Garg (2023201030)
Rutuja Deshpande (2023201032)
Atharva Pande (2023201065)

***Abstract*—The paper titled "Qualitative similarities and differences in visual object representations between brains and deep networks" explores the comparison between object representations in deep neural networks and visual cortical areas in the brain. The study recasts well-known perceptual and neural phenomena in terms of distance comparisons to investigate whether these phenomena are present in feedforward deep neural networks trained for object recognition. The research highlights both qualitative similarities and differences in visual object representations between brains and deep networks, providing insights into the properties that could be incorporated to improve deep networks. The findings offer valuable implications for understanding the relationship between human vision and deep learning systems.**

I. Qualitative similarities and differences in visual object representations between brains and deep networks

We have replicated 9 experiments on **VGG-16 pretrained and VGG-16 Random** Models conducted by IISc Institute. The experiments that we have replicated are:

- Thatcher effect
- Weber's Law
- 3D Effect
- Mirror Confusion
- Relative Size
- Surface Invariance
- Selectivity along Multiple Dimensions
- Occlusion
- Global Shape Advantage

Feature Extraction: For feature extraction, we fed each image into the network and recorded the activations of every layer as a separate column vector. Thus, for a single image, we obtained 37 feature vectors, each representing a layer's activations.To measure the dissimilarity between images A and B, we computed the Euclidean distance between their respective activation vectors.

II. Experiments Replicated

A. *Thatcher Effect*

Thatcher effect is an elegant demonstration of how upright faces are processed differently from inverted faces, presumably because we encounter mostly upright faces.All faces were grayscale, upright and front-facing. To thatcherize a face, we inverted the eyes and mouth while keeping rest of the face intact. We implemented inversion by first registering facial landmarks on frontal faces using an Active appearance model-based algorithm.

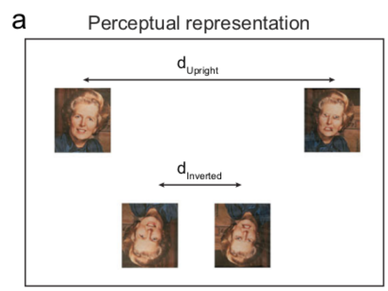


Fig. 1. Perceptual representation of Thatcher Effect

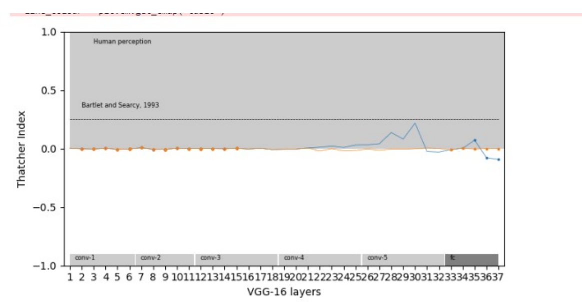


Fig. 2. Results of Thatcher(Orange represents VGG-16 rand, blue represents VGG-16)

B. *Weber's Law*

The law states that the just-noticeable difference (JND) between two stimuli is proportional to the magnitude of the stimuli and can be expressed as a constant ratio. If the correlation between pairwise distances and relative changes in length is larger than the correlation with absolute changes in length, we deemed that layer to exhibit Weber's law. The correlation difference was initially negative in the early layers of the network, meaning that the early layers were more sensitive to absolute changes in length.

C. *3D Effect*

The study aimed to determine if deep networks exhibit similar processing patterns to humans when presented with stimuli that vary in 3D shape features. The findings revealed that deep networks did not show sensitivity to 3D shape

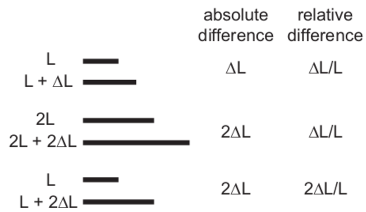


Fig. 3. Weber's law

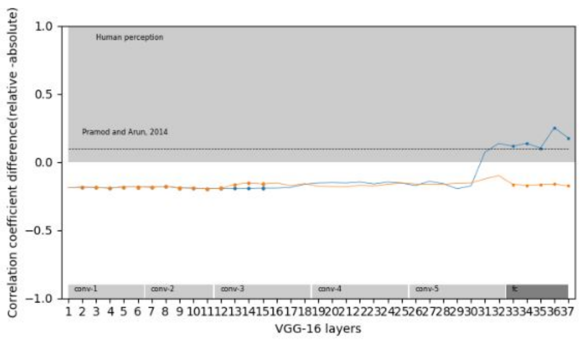


Fig. 4. Results Weber's Law(Orange represents VGG-16 rand, blue represents VGG-16)

processing, unlike humans. The researchers observed that the 3D processing index, which measures the difference in distances between stimuli in 3D shape conditions and equivalent 2D conditions, was consistently near zero or negative across all layers of the VGG-16 network. This suggests that deep networks do not encode 3D shape information in a manner similar to human perception.

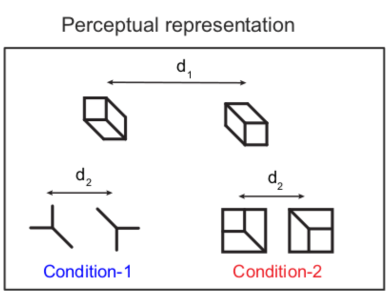


Fig. 5. Perceptual Representation of 3D effect

D. *Mirror Confusion*

Mirror confusion, where vertical mirror reflections are perceived as more similar than horizontal ones. The effect is observed in both human behavior and monkey visual cortex. To assess this phenomenon in deep neural networks, a mirror confusion index is calculated based on distances between horizontal and vertical mirror image pairs. In the VGG-16 network, the index increases across layers, indicating stronger mirror confusion for vertical mirrors. This trend is absent in a randomly initialized network, suggesting deep networks share human-like mirror confusion.

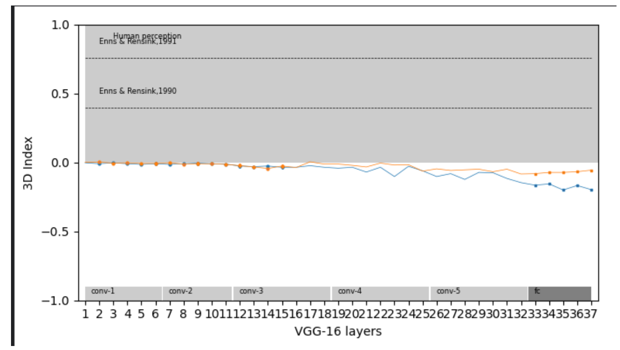


Fig. 6. Results of 3D effect (Orange represents VGG-16 rand, blue represents VGG-16)

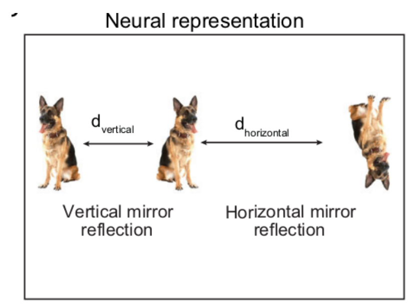


Fig. 7. Neural Representation of mirror confusion

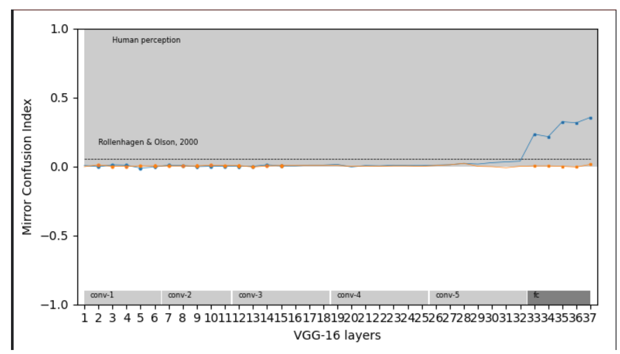


Fig. 8. Results of Mirror Confusion(Orange represents VGG-16 rand, blue represents VGG-16)

E. *Relative Size*

The sensitivity of neurons in high-level visual areas to the relative size of items in a display. In a study, it was found that neural responses were more similar when two items underwent proportional changes in size. This effect was observed in a small fraction of neurons. To explore if this effect is present in the VGG-16 network, a similar analysis was performed. The relative size index showed a modest increase across layers, indicating representation of relative size in the network. This effect was weaker than in actual neurons but present nonetheless, suggesting its presence due to network architecture and training.

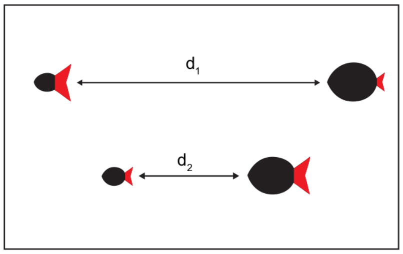


Fig. 9. Neural representation

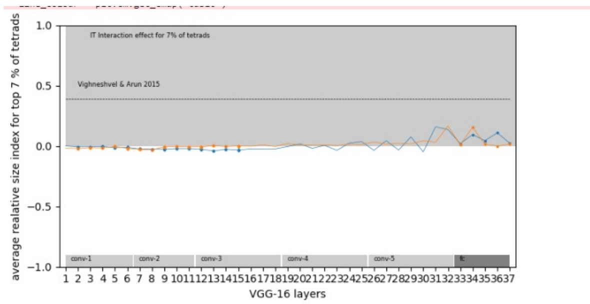


Fig. 10. Results of Relative Size(Orange represents VGG-16 rand, blue represents VGG-16

F. Surface Invariance

IT neurons respond more similarly when a pattern and a surface undergo congruent changes in curvature or tilt. To assess if the VGG-16 network exhibits this property, units with increased interaction between surface and pattern features were identified. The surface invariance index was consistently below zero across layers for both the VGG-16 network and the randomly initialized VGG-16, indicating that deep neural networks trained for object classification do not show surface invariance.

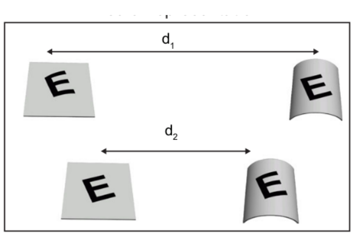


Fig. 11. Neural Representation

G. Selectivity along Multiple Dimensions

IT neurons respond more similarly when a pattern and a surface undergo congruent changes in curvature or tilt. To assess if the VGG-16 network exhibits this property, units with increased interaction between surface and pattern features were identified. The surface invariance index was consistently below zero across layers for both the VGG-16 network and the randomly initialized VGG-16, indicating that deep neural networks trained for object classification do not show surface invariance.

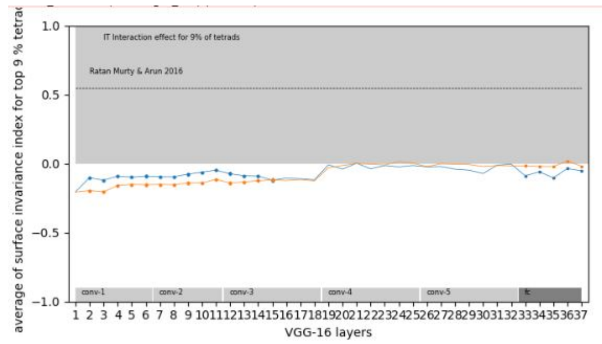
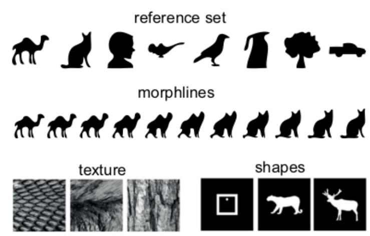


Fig. 12. Results of Surface Invariance(Orange represents VGG-16 rand, blue represents VGG-16)



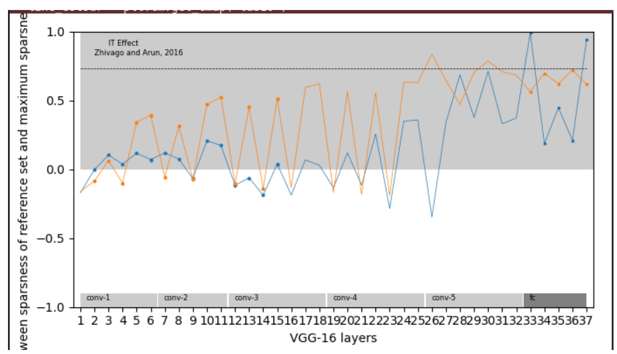


Fig. 13. Results of Corelated Sparsness(Orange represents VGG-16 rand, blue represents VGG-16

H. Occulusion

The automatic processing of occlusion relations between objects in human perception. It highlights how searching for targets with occluded objects is harder than searching for the same objects unoccluded. The study aimed to determine if similar effects are present in the VGG-16 network by calculating an occlusion index. The results showed that the occlusion index remained consistently negative but approached zero across layers, indicating that deep networks do not represent occlusions and depth ordering in a human-like manner.

I. Global Shape Advantage

The study created 49 hierarchical stimuli by combining seven shapes at global and local scales. For human perception, distances were calculated as the reciprocal of the average search time for each image pair. For CNNs, features were extracted from each layer, and the Euclidean distance between all pairs of stimuli was calculated. Global distance was deter-

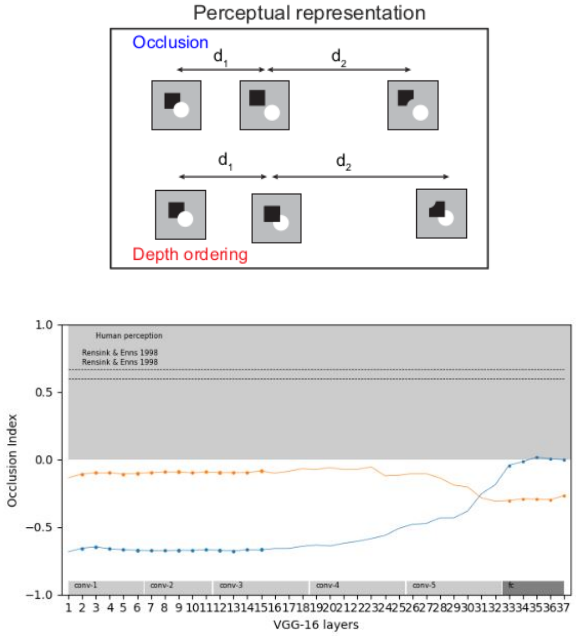


Fig. 14. Results of Occulusion(Orange represents VGG-16 rand, blue represents VGG-16

mined as the mean distance between image pairs with only global changes, while local distance was calculated as the mean distance between image pairs with only local changes.

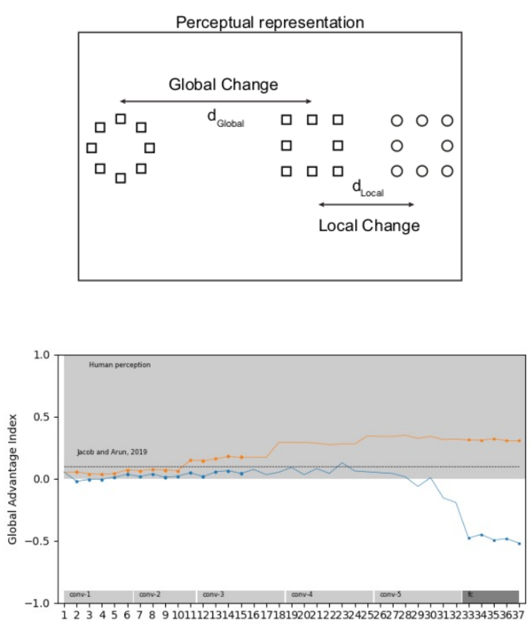


Fig. 15. Results of Global Shape Advantage(Orange represents VGG-16 rand, blue represents VGG-16

References

[1] Georgin Jacob, R. T. Pramod, Harish Katti, & S. P. Arun. (2021). Qualitative similarities and differences in visual object representations between brains and deep networks. Nature Communications, 12(1), 1872.

Perceptual effect	**VGG-16**	**VGG16 Random**
Thatcher effect	No	No
Mirror confusion	Yes	No
Weber's law	Yes	No
Relative size	Yes	Yes
Surface invariance	No	No
3D processing	No	No
Occlusion	No	No
Global advantage	No	Yes
Correlated sparseness	Yes	Yes

Fig. 16. Presence/absence of each effect across deep networks tested.

[2] Serre, T. Deep learning: the good, the bad, and the ugly. Annu. Rev. Vis. Sci. 5, 399–426 (2019).

[3] Sinz, F. H., Pitkow, X., Reimer, J., Bethge, M., & Tolias, A. S. (2019). Engineering a less artificial intelligence. Neuron, 103, 967–979.

[4] Belongie, S., Malik, J., & Puzicha, J. (2002). Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(4), 509–522.