Deep Neural Networks Are Predictive of Neural Data Through Textures

Jessica Loke (j.loke@uva.nl)
Department of Psychology, University of Amsterdam
Postbus 15915, 1001NK Amsterdam, The Netherlands

Lynn K. A. Soerensen (lynnka@mit.edu)
Department of Brain and Cognitive Sciences, McGovern Institute for Brain Research,
Massachusetts Institute of Technology

Iris I. A. Groen (i.i.a.groen@uva.nl)
Institute for Informatics, University of Amsterdam
Science Park 900, 1098 XH, Amsterdam, The Netherlands

H. Steven Scholte (h.s.scholte@uva.nl)
Department of Psychology, University of Amsterdam
Postbus 15915, 1001NK Amsterdam, The Netherlands

Abstract – Human visual processing is well predicted by deep neural networks (DNNs), yet what drives this predictive power is less understood. Interestingly, human visual cortices have recently been reported to represent objects in a texture-like fashion, akin to a texture bias commonly observed in DNNs. We hypothesized that this alignment of DNNs with human neural recordings is driven by DNNs' ability to explain variance related to texture information in images. To test this, we recorded electroencephalography (EEG) signals from human participants (n=57) while they viewed three types of images: natural images, texture-synthesized, and object-only versions. Next, we compared these neural representations with features extracted from five different DNN architectures processing the same images. Our results show that features extracted from texture-synthesized images are just as predictive of EEG responses as features extracted from original images themselves. Moreover, features extracted from texture-synthesized images were most predictive of EEG responses for texture-synthesized images. Our results suggest that DNN's ability to predict neural data derives from a shared bias for textures in the human visual cortex.

Keywords: Visual processing; Object Recognition; EEG; Texture

Introduction

Deep neural networks (DNNs) have demonstrated high predictive performance of human brain responses. Prior studies have revealed a tendency in some DNNs to categorize images based on textures rather than shape (Geirhos et al. 2018; but see Ritter et al. 2017; Hermann et al. 2019). Interestingly, recent studies have also revealed texture-like representations of objects in the human visual cortex (Jagadeesh & Gardner, 2022; Henderson et al. 2019). Based on these findings, we hypothesized that DNNs' ability to predict neural data stems from their ability to encode texture information. This study explores this texture bias by presenting participants with three types of images differentially emphasizing texture and object information: texture-synthesized, object-only and original images. The texture-synthesized condition retained only texture information and removed object cues from the images. The object-only condition retained only the object and removed contextual texture information. The original condition serves as a control preserving the original balance of texture and object information. We posit that DNNs will exhibit the highest predictive performance with texture-synthesized images, as this condition aligns closest to both the DNNs' inherent biases and the texture-like representations in human visual processing.

Methods

We recorded EEG signals from 57 human participants as they passively viewed images. The images were presented in a rapid serial visual presentation (RSVP; Potter & Fox, 2009) design, starting with a fixation cross of 100ms, followed by the stimuli presentation for 100ms – this sequence alternates 120 times before participants were allowed a break. In total, participants were presented with a total of 12,000 stimuli – 200 stimuli x 3 conditions x repeated 20 times.

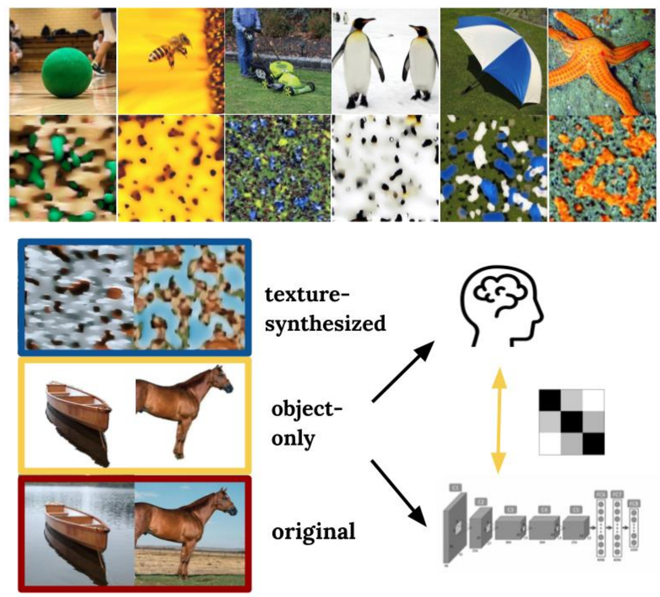

Figure 1: The first row shows original THINGS images and the second row their texture-synthesized counterparts. These images, together with objects-only images, were presented to humans and DNNs.

Stimuli

We used a subset of THINGS Image Dataset (Hebart et al. 2019). To vary the contribution of texture information, we created two additional image conditions – texture-synthesized and object-only (Fig. 1). For the texture-synthesized condition, we used a method from Gatys et al. (2015) which utilizes a pretrained VGG-19 to extract and replicate textural patterns from the original images. These textures were spatially confined to the receptive field of VGG-19's first convolutional layer (conv1_1). For the object-only condition, we segmented the objects of interest using Background Remover (Nader, 2024), isolating them from their background to focus on the object rather than texture.

DNNs

We selected five DNN architectures commonly used in the computational modeling of visual processing – AlexNet, VGG-16, ResNet-18, ResNet-50 and ViT-b-16

(Krizhevsky et al. 2012, Simonyan & Zisserman, 2014; He et al. 2016, Dosovitskiy et al. 2021). We initialized five different seeds for every model.

Model comparisons

To characterize visual processing in EEG recordings, we computed representational dissimilarity matrices (RDMs; Diedrichsen & Kriegeskorte, 2017) using cosine similarity of 22 posterior EEG electrodes. For DNN activations, we computed RDMs from activations for all convolutional, pooling and fully-connected layers.

We used Ridge Regression to regress the EEG RDMs onto DNN RDMs. Using these estimated linear mappings, we predicted representations on a held-out dataset of 15 subjects and 100 stimuli. We then computed the correlations between the predicted representations with EEG RDMs. The prediction performance of different models was compared using pairwise comparisons (Kruskal-Wallis test) of the area under the curve (AUC) of the correlation time courses.

Results

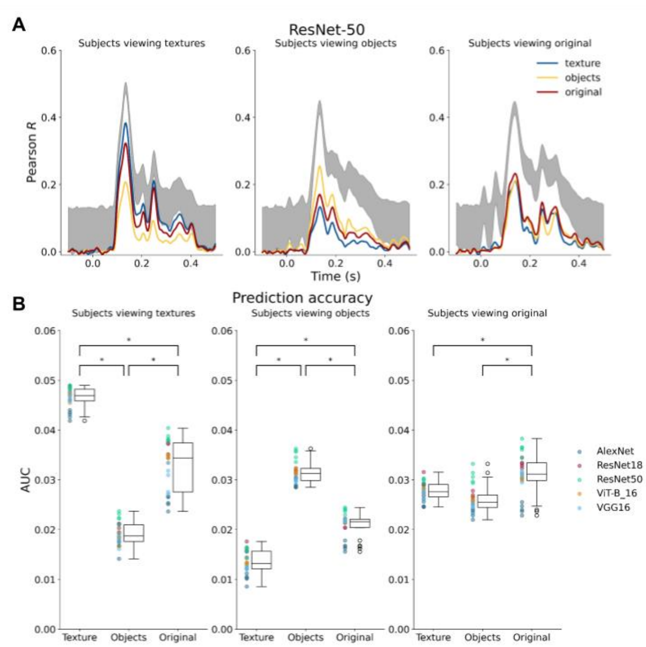

Figure 2: (A) Prediction performance of ResNet-50 from -0.1s to 0.5s relative to stimuli onset. The model shows the highest representational correlation for EEG responses during the presentation of texture-synthesized images. (B) AUC of prediction performance for different DNNs across all time points.

Enhanced representational correlation between DNN and humans with texture-synthesized images

By isolating texture information and removing object cues from images, we increased the representational correlation between DNNs and humans (Fig. 2A^1, 2B^1). This result is robust held for all DNNs, suggesting that DNN prediction of neural signals was indeed sensitive to the presence of texture information in neural signals. When human subjects viewed original images, features from original images performed only marginally better than features from texture-synthesized images (Fig. 2A^3, 2B^3). This is remarkable because the texture-synthesized images are devoid of semantic information (Fig. 1). We also performed a partial correlation to assess the unique variance attributed to each feature. Similarly, we saw an increased contribution of unique variance from texture features in EEG responses towards texture-synthesized images (Fig. 3A^1, 3B^1).

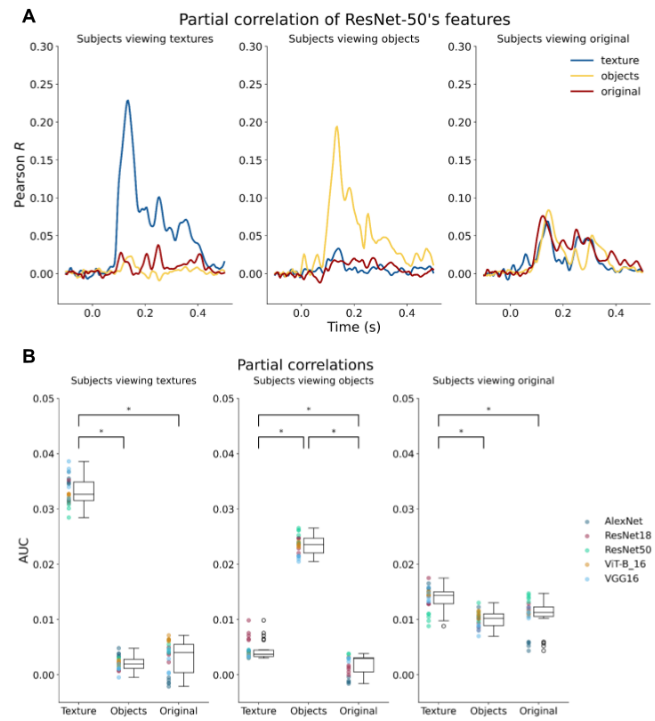

Figure 3: (A) Partial correlations between EEG responses and ResNet-50's features. When texture information is isolated (i.e., subjects viewing textures), texture features show an increase in unique variance. (B) AUC of partial correlations for different DNNs across all time points. All models showed a consistent increase in unique variance captured when texture information was isolated.

Conclusion

Our results demonstrate an increased alignment between neural processing and DNN features when isolating texture information from images. This suggests that the predictive power of DNN features for neural signals – as captured by EEG – is largely derived from a shared sensitivity for image texture.

References

Diedrichsen, J., Kriegeskorte, N. (2017). Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLoS Computational Biology, 13*(4), e1005508. https://doi.org/10.1371/journal.pcbi.1005508

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv.* https://arxiv.org/abs/2010.11929

Geirhos, R., Rubisch, P., Michaelis, C., & Bethge, M. (2018). ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv Preprint arXiv.* https://arxiv.org/abs/1811.12231

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE.* http://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html

Hebart, M. N., Dickter, A. H., Kidder, A., Kwok, W. Y., Corriveau, A., Van Wicklin, C., & Baker, C. I. (2019). THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PloS One, 14*(10), e0223792.

Hermann, K. L., Chen, T., & Kornblith, S. (2019). The origins and prevalence of texture bias in convolutional neural networks. *arXiv [cs.CV].* https://proceedings.neurips.cc/paper/2020/file/db5f9f42a7157abe65bb145000b5871a-Paper.pdf

Jagadeesh, A. V., & Gardner, J. L. (2002). Texture-like representation of objects in human visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 119*(17), e2115302119.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In F. Pereira, C. J. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems, 25.* Curran Associates, Inc. https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

Nader. (2024). backgroundremover: Remove the background of any image with the state of the art models. GitHub. Retrieved from https://github.com/nadermx/backgroundremover

Potter, M. C., & Fox, L. F. (2009). Detecting and remembering simultaneous pictures in a rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance, 35*(1), 28-38. https://doi.org/10.1037/a0013624

Ritter, S., Barrett, D., Santoro, A., & Botvinick, M. (2017). Cognitive psychology for deep neural networks: A shape bias case study. *International Conference on Machine Learning.* https://www.semanticscholar.org/paper/39fb9fa2615620f043084a2ecbbdb1a1f8c707c9

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