Increasing behavioural alignment of DNNs and humans using high resolution images

Niklas Müller[1*], Iris I.A. Groen[2#], H. Steven Scholte[1#]

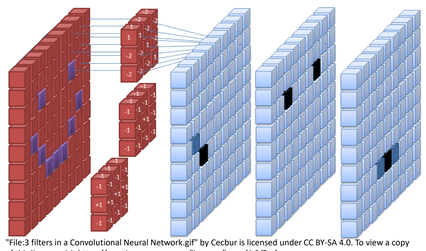
1) Psychology Research Institute, University of Amsterdam 2) Informatics Institute, University of Amsterdam

* **n.muller@uva.nl** **# shared senior author**

Deep Neural Networks

Computational models have long been used for explaining and understanding processes of human behaviour or human physiology. However, models that could explain both at the same time for real-world sensory input were so far out of reach.

With increased computational resources, **Deep Neural Networks (DNNs)** emerged as candidate models to study neural mechanisms underlying human behaviour, e.g. object recognition in real-world scenes. However, most DNNs for vision are trained on small, low–quality images.



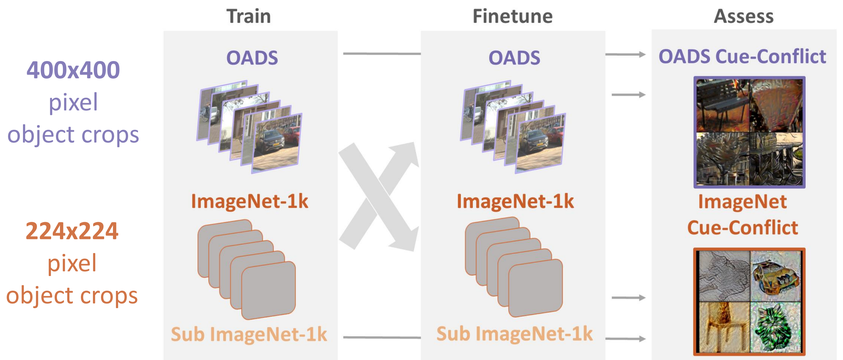
We investigate if matching visual input of DNNs (images) to that of humans (retinal output) improves their behavioural alignment.

Training and fine-tuning DNNs

Using images that exhibit conflicting cues for object shape and object texture, we evaluate the texture-shape-bias of DNNs on a high-resolution and a low-resolution dataset, respectively.

Open Amsterdam Data Set (OADS)

6130 Scene images 5496x3672 pixels 71.870 object annotations

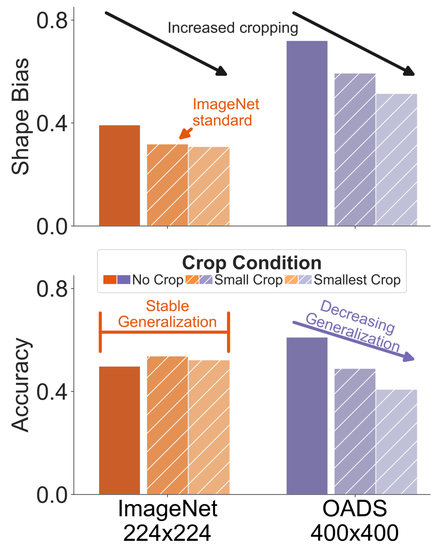


We evaluate DNN texture-bias and assess the specific factors that lead to human and DNN behavioural (mis-)alignment.

Access to global object shape

Cropping images is commonly done when training DNNs. We vary the amount of cropping used during training and find that **stronger cropping increases the texture-bias of DNNs**.

This effect is independent from the training resolution or dataset.

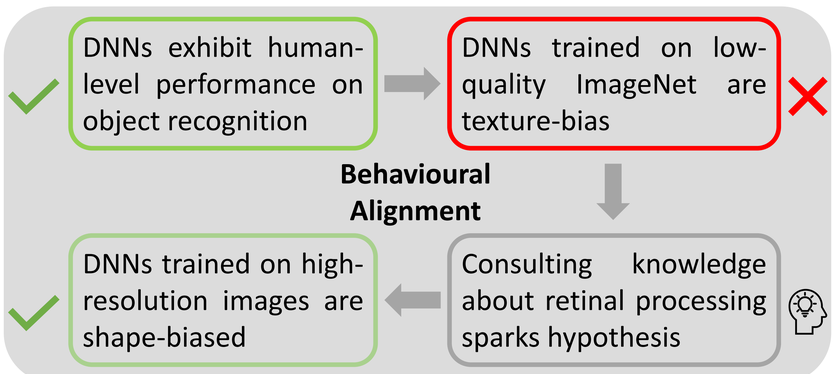


Cropping images removes a substantial amount of the **global object shape** and thus forces DNNs to use object texture instead

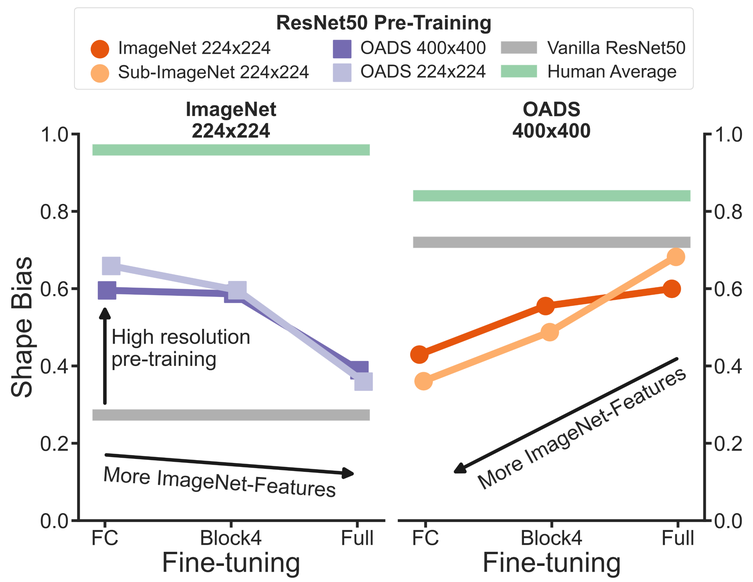
Texture-bias in Deep Neural Networks

DNNs have been found to behave differently than humans when classifying objects. While humans recognize and classify objects according to their shape, DNNs strongly rely on texture[1].

Many methods have been proposed to tackle this issue, mostly targeting model architecture or data augmentation (e.g., [2,3]). Instead, we rely on **knowledge about the physiology of the human visual system** for increasing shape-bias in DNNs and thus for improved behavioural alignment between humans and DNNs.



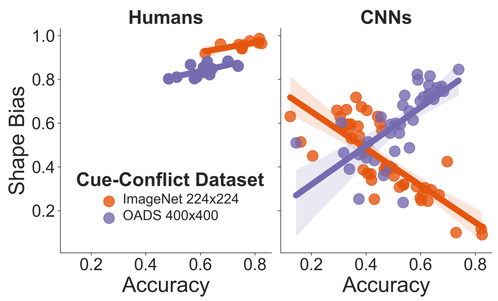
High shape-bias in DNNs



(Pre-)Training on **high-resolution images** yields **human-level shape-bias**. ImageNet-tuned models are biased towards texture.

Conclusions: Alignment with Humans

Accuracy and shape-bias are positively correlated for both humans and DNNs when trained on high-resolution images, but not when trained on low-resolution images.



Deep learning models display **human-like behaviour** when trained on high-resolution images.

High-resolution images resemble **ecologically valid** input data.

This showcases the effect and importance of **aligning DNN input to human "input"** when attempting to align their behaviour.

Acknowledgements | This work was funded by the UvA Data Science Center Interdisciplinary PhD Program **References** | [1] Geirhos et al. *arXiv*, 2018; [2] Tuli et al. *CogSci*, 2021; [3] Hermann et al. *NeurIPS*, 2020.