Object-zoomed training of convolutional neural networks inspired by toddler development improves shape bias

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

1. Psychology Research Institute
2. Informatics Institute

University of Amsterdam * **n.muller@uva.nl** # shared senior author

Introduction

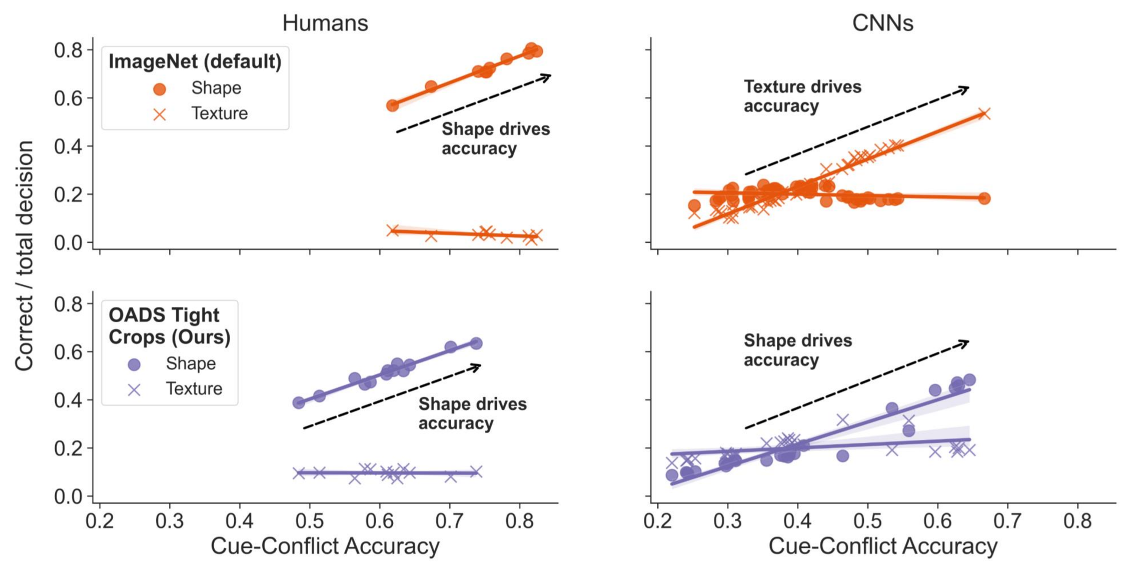
Convolutional Neural Networks (CNNs) surpass human-level performance on visual object recognition [1] and detection [2], but their behavior still differs from human behavior in important ways. One prominent example is that CNNs trained on ImageNet exhibit an image texture bias [3], while humans exhibit a strong bias toward object shape [4].

[3] *Geirhos et al. (2018)*

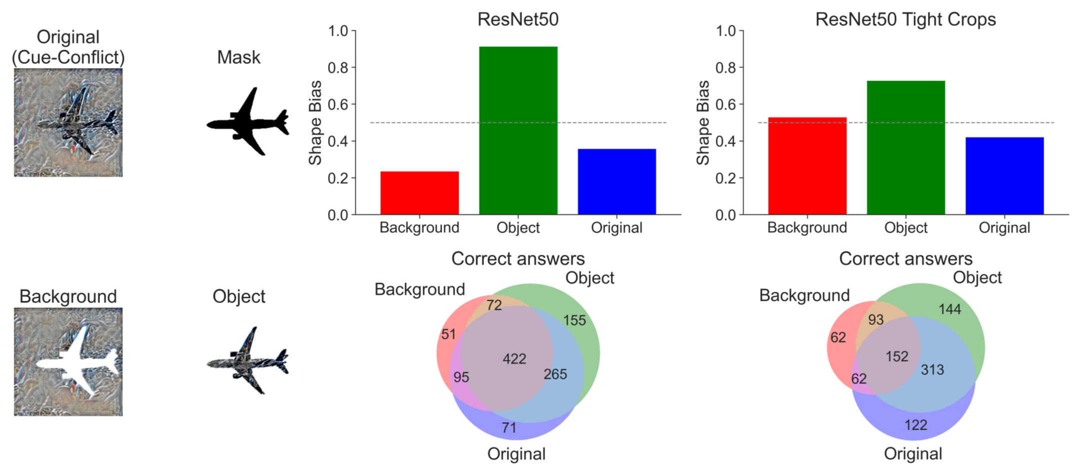
Developmental research suggests that one factor driving human shape bias is that during early childhood, toddlers tend to fill their field-of-view with close-up objects [5].

[5] *Smith et al. (2011)*

Aligning classification strategies



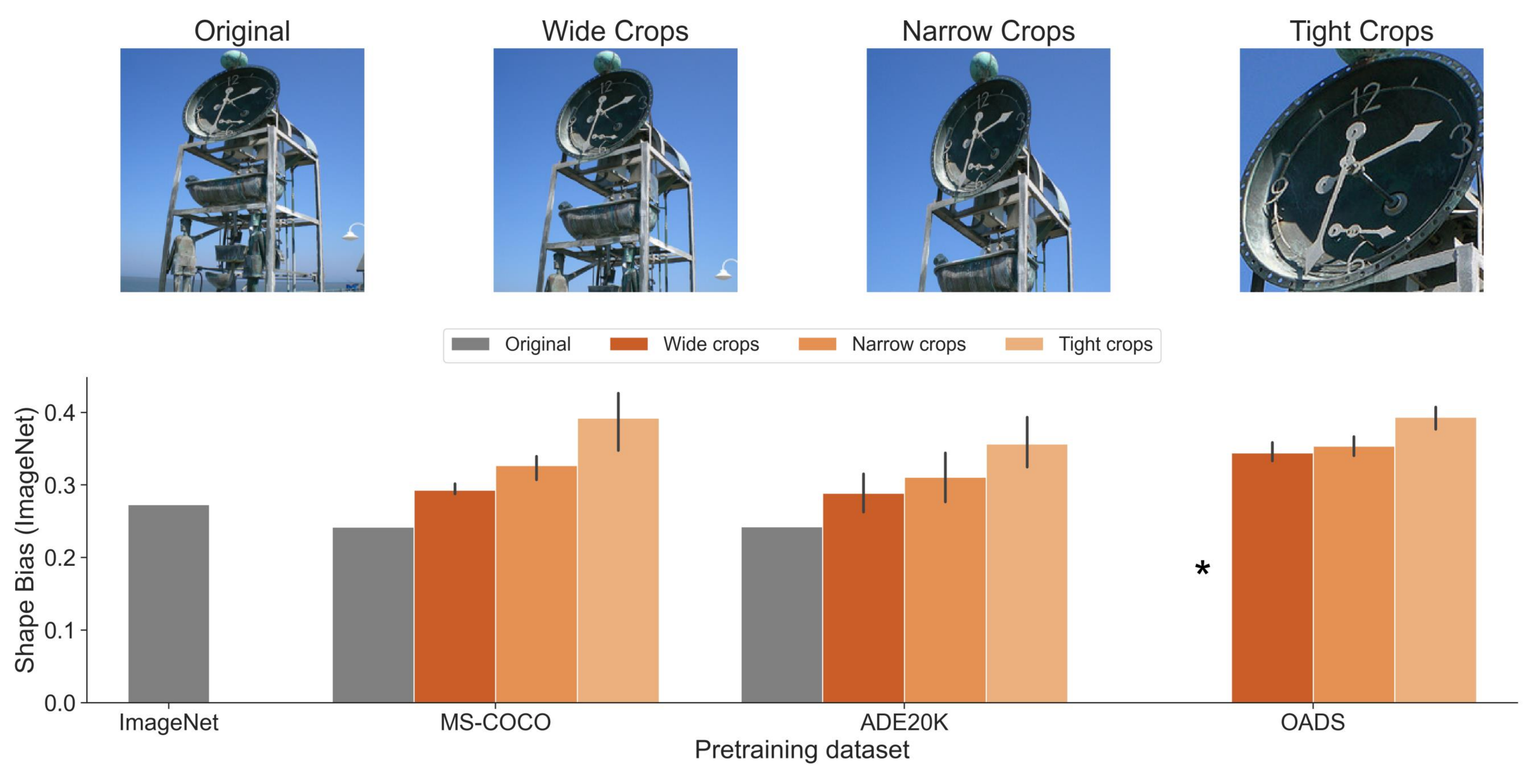
Networks trained on tightly cropped objects are largely **using object shape for classification**, just like for humans.



Further, networks **rely less on image background** for their classification compared to the default ImageNet-trained networks.

Zoomed-in object training increases ImageNet shape bias

Here, we operationalize this close-up as a zoom-in on objects during CNN training which we show increases shape bias without any additional training or data augmentation.

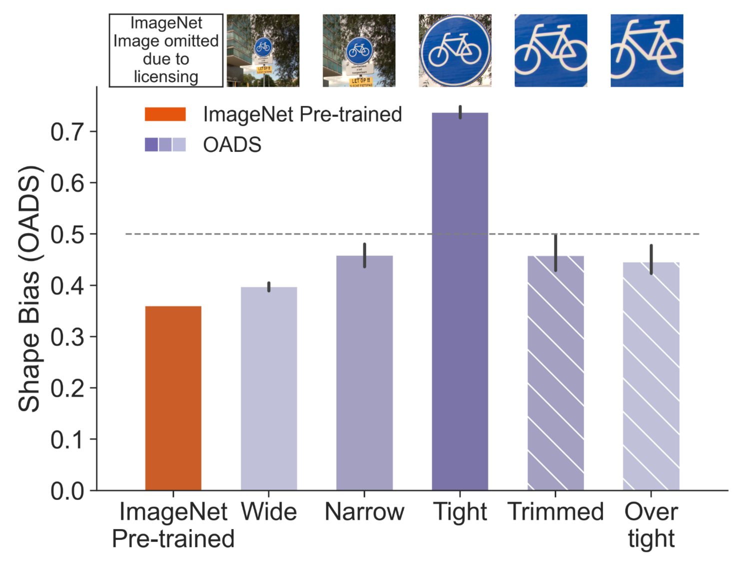


Shape bias increases when reducing the amount of background included in training crops across all four datasets.

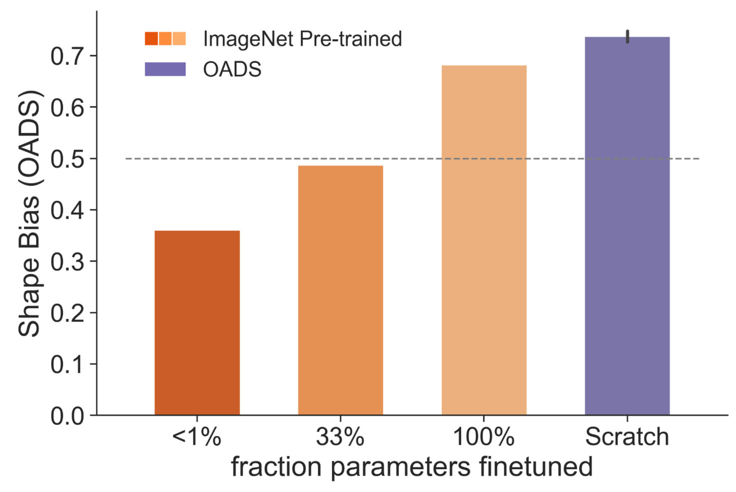
Highest shape bias on tightly-cropped images

A network trained and tested on tightly cropped OADS images exhibits a **shape bias that is twice as large** as that of a standard ImageNet-trained CNN.

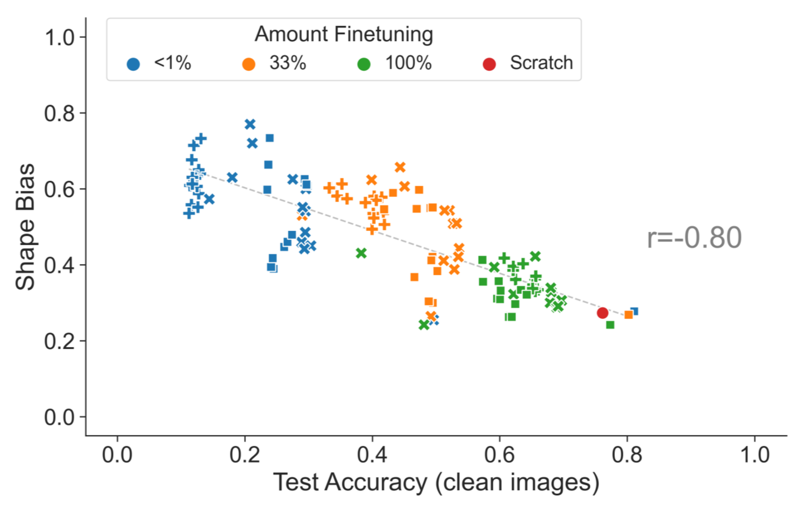
Interestingly, however, continuing to crop images result in lower shape bias, indicating that networks might **rely more on textures when the global object shape is removed.**



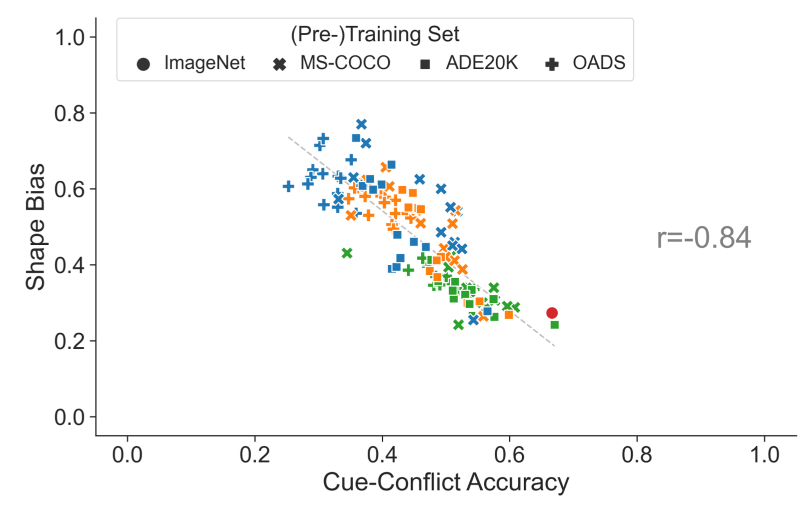
Shape bias increases with the number of parameters finetuned on OADS: **high texture bias is not an inherent property of CNNs** but can be mitigated using images with a more optimal background-object ratio.



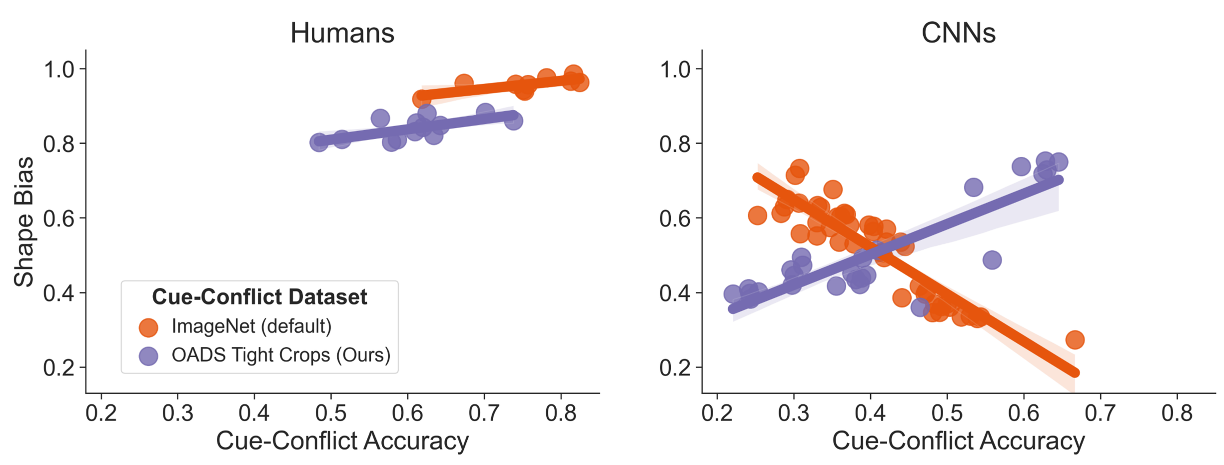
Accuracy-Shape bias tradeoff



Human shape bias increases with increasing accuracy.



CNN shape bias decreases with increasing accuracy on ImageNet.



Training CNNs on zoomed-in object images **resolves this trade-off** and makes high-performing CNNs shape-biased and thus human-aligned.

Conclusion

- Increasing object zoom-in during training increases ImageNet shape bias of convolutional neural networks
- CNN texture bias decreases by **reducing usage of background information** through training on close-up objects
- Removal of background information increases alignment with humans on both shape bias and accuracy
- Developmentally-inspired background-object ratio results in highest CNN shape bias

Preprint!

bioRχiv

Overall, our results suggest that CNN shape bias is largely driven by the training data distribution and that it is not an inherent property of convolutional architectures.

References

[1] Krizhevsky et al., *NeurIPS* (2012); [2] He et al., *CVPR* (2016);
[3] Geirhos et al., *ICLR* (2018); [4] Landau et al., *Develop. Sc.* (2011);
[5] Smith et al., *Develop. Sc.* (2011).

Acknowledgements

This work was funded by the Interdisciplinary PhD Programme of the Data Science Centre of the University of Amsterdam.