Getting into Shape:
The Impact of Early Visual Development on Object Recognition

Zejin Lu* (zekinglu@gmail.com)
Department of Education and Psychology, Freie Universität Berlin
Berlin, BE 14195, Germany
Institute of Cognitive Science, University of Osnabrück
Osnabrück, NI 49090, Germany

Sushrut Thorat* (sthorat@uos.de)
Institute of Cognitive Science, University of Osnabrück
Osnabrück, NI 49090, Germany

Radoslaw M Cichy (rmcichy@zedat.fu-berlin.de)
Department of Education and Psychology, Freie Universität Berlin
Berlin, BE 14195, Germany

Tim C Kietzmann (tim.kietzmann@uni-osnabrueck.de)
Institute of Cognitive Science, University of Osnabrück
Osnabrück, NI 49090, Germany

* shared first author

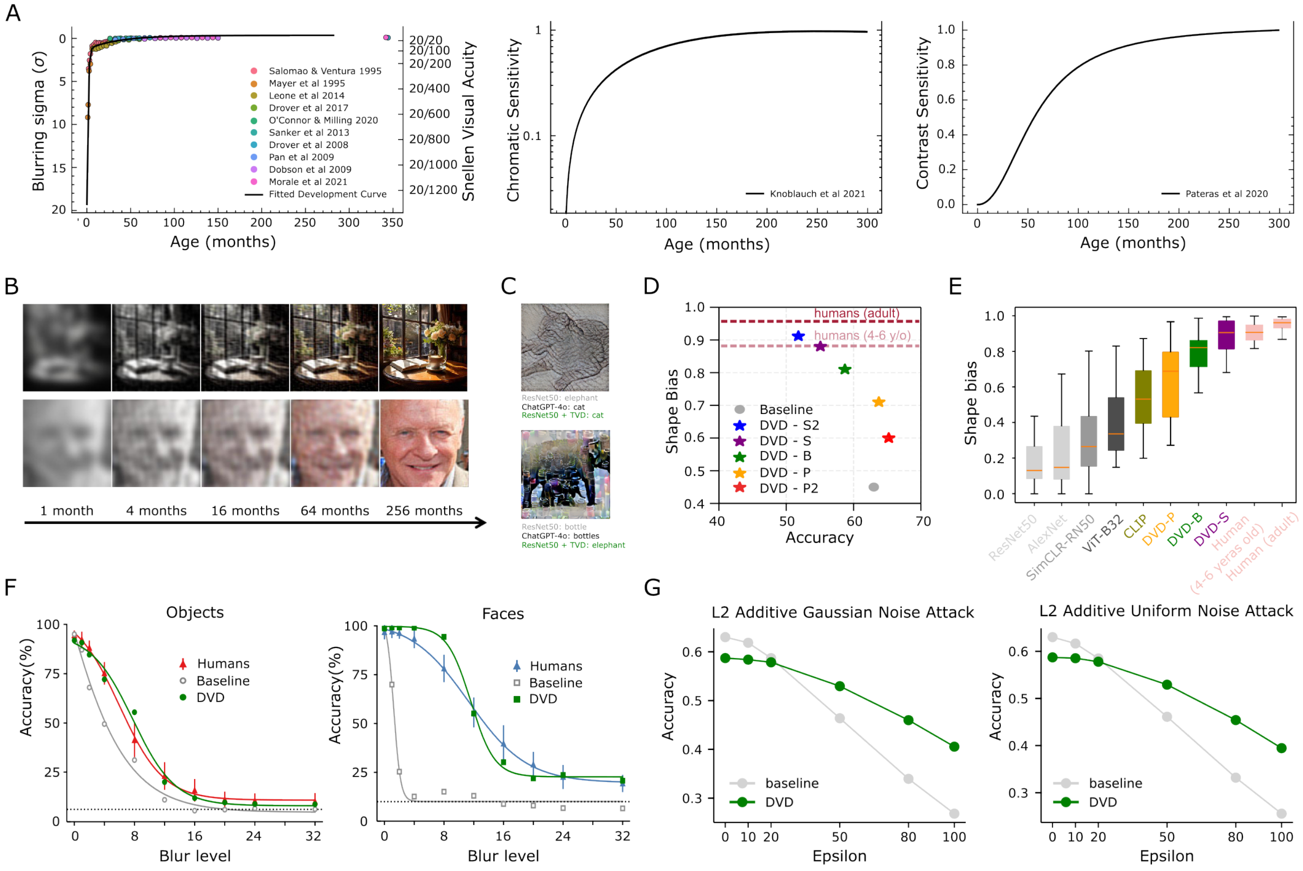

Figure 1: **Developmental visual diet (DVD) promote human-like shape bias and robustness in vision models. A**. Developmental visual diet (DVD) modelling behavioural measurements from infancy to adulthood (0–25 years) for visual acuity (left), chromatic sensitivity (middle), and contrast sensitivity (right). **B**. Example DVD images. **C**. Example cue-conflict images illustrating shape–texture bias. **D**. Accuracy and shape bias in baseline and DVD-trained models. **E**. Comparison of shape bias between DVD and standard vision models - DVD achieves near-human-level shape bias. **F**. DVD-trained models better align with human responses under image degradation. **G**. Enhanced L2-based adversarial robustness of DVD-trained models against additive Gaussian (left) and uniform (right) noise attacks.

Abstract

A prolonged period of immaturity is a key feature distinguishing humans from artificial neural networks (ANNs) and many other animals. For example, various aspects of the visual experience of babies are rather poor and only slowly improve over time. In stark contrast, AI vision models are presented with mature, adult-like input from the start. Here we wondered whether there is a computational advantage to this developmental trajectory, and how far it would impact artificial vision models. Indeed, recent studies indicate that injecting several fixed levels of blur into training can improve robustness and shape bias — longstanding challenges for vision models in object recognition. However, a large margin to human visual robustness remains for all publicly available vision systems. To further explore the possibilities of a human-adjusted visual diet, we introduce a visual training trajectory that simulates the progressive developmental visual diet (DVD) of humans, spanning from newborns to 25-year-old adults. Three aspects of vision are considered: visual acuity, chromatic sensitivity, and contrast sensitivity. We demonstrate that training on vision tasks with DVD yields models with near-human-level shape bias, better alignment with robust human perception under signal deterioration, and much enhanced robustness to a variety of adversarial attacks. Importantly, the DVD improvements are observed on regular vision models, such as ResNet, trained on regular vision tasks, such as ecoset or ILSVRC, thus enabling robust visual inference outside of large-scale, large-data, multimodal models. DVD thereby offers a promising approach to bridging the gap between artificial neural networks and human visual systems.

Keywords: artifical neural networks; early visual development; shape bias; degradation robustness; adversarial robustness

Introduction

Whereas numerous species rapidly develop near-adult vision soon after birth, mammals - especially humans - undergo a prolonged developmental period, transitioning from low visual acuity, limited chromatic sensitivity and contrast sensitivity, to fully mature perception (Teller, 1983). Whether this prolonged developmental window is merely a biological limitation or a crucial mechanism for fostering robust, human-like visual intelligence is a longstanding question. If the latter were the case, would a comparably slow developmental visual diet (DVD) aid artificial vision systems, too?

Compared to artificial vision systems, one aspect where human vision shines is the so-called shape bias (Fig. 1C). Previous studies show that human visual behaviour is 96% reliant on shape when facing a shape-texture cue-conflicting condition(Geirhos et al., 2018). In contrast, modern artificial vision systems, ranging from convolutional neural networks (CNNs) to Vision Transformers (ViTs), predominantly rely heavily on texture-based cues (Geirhos et al., 2018; Baker et al., 2018). Recent studies show that training vision models with several fixed levels of blurry input — one feature of early visual experience — applied in a coarse blurry-to-clear or clear-to-blurry fashion, yields improved robustness (Vogelsang et al., 2018) and shape bias (Jang et al., 2024), but these advances, while encouraging, remain far below human-level performance.

In this work, we demonstrated that adhering closely to the human developmental visual trajectory, reflecting gradual improvements in visual acuity, chromatic sensitivity, and contrast sensitivity, is an effective means towards achieving robust artificial vision with nearly human-level shape bias, heightened resilience to image degradations and adversarial attacks.

Results

To approximate visual experience changes in human vision from infancy through adulthood, we made use of available quantitative behavioural data on visual acuity, chromatic sensitivity, and contrast sensitivity obtained via psychophysics experiments from newborns to 25-year-old adults.

Data from each of the three aspects of vision was fit with a smooth, monotonic exponential curve, capturing how sensitivity in different visual aspects evolves over the first 25 years of life (Fig. 1A). These curves underpin our DVD framework: during training, images undergo progressive transformations, initially newborn-level visual acuity, color and contrast sensitivity, then gradually shifting to more adult-like vision. Importantly, model testing is performed on regular input images.

DVD training leads to near-human-level shape bias

Shape bias is a hallmark of human visual perception, with observers consistently prioritizing shape over texture (96% in adults and 88% in four-year-old children)(Huber et al., 2023).

We find that ANN models trained with DVD exhibited substantial improvements in shape-based decisions, while retaining object categorization performance (see Figure 1C). Compared to a baseline (ResNet50, 45% shape bias), DVD shape favored variants (DVD-S and DVD-S2) achieved between 87%-91% shape bias, approaching human-level preference (88%-96%). Also, performance-favored variants (DVD-P and DVD-P2) attain both improved recognition accuracy (from 63% to 65% in ecoset dataset) and elevated shape bias (from 45% to 71%). Under a balanced regime, DVD-B achieved a substantial increase in shape bias (from 45% to 81%) with only a modest accuracy trade-off (4.3%). The various DVD configurations reflect differences in hyperparameters, modulating the quantity and strength of early visual experience introduced during training. Importantly, DVD-trained models (Figure 1D) exceed the shape bias of all conventional vision models trained on naturalistic inputs.

DVD training brings enhanced degradation and adversarial robustness

After training with DVD, models exhibit improved performance under image degradations. As blur intensifies, baseline models experience a sharp decline in both object and face recognition accuracy, whereas DVD-trained models maintain much higher accuracy, aligning more closely with human performance (Fig. 1E, human data from (Jang et al., 2021)). Furthermore, DVD-trained models demonstrate superior adversarial robustness against L2-based Gaussian and uniform noise perturbations (Fig. 1F).

In summary, training with developmental visual trajectories enables models to gradually develop visual acuity, chromatic sensitivity, and contrast sensitivity from low to high levels, yielding near-human-level shape bias, and improving robustness to both image degradations and adversarial perturbations in object recognition.

Methods

Neural network training: We adopted a standard ResNet50 as our baseline architecture, training it on the ecoset dataset (Mehrer et al., 2021) for object classification.

To simulate the trajectory of visual development(DVD), we integrated psychophysical data on visual acuity, chromatic sensitivity, and contrast sensitivity from newborns to 25-year-old adult, sourced from previous psychophysics studies (Braddick et al., 2011; Caltrider et al., 2024; Dobson et al., 2009; Drover et al., 2008; El-Gohary et al., 2017; Katzhendler et al., 2019b; Inal et al., 2018; Katzhendler et al., 2019a; Kugelberg, 1992; Leone et al., 2014).

Based on these developmental trajectories, we established a “months-per-epoch” (mpe) training schedule (e.g., 2mpe over 150 epochs to represent 25 years). This schedule mapped each epoch’s data augmentation parameters to corresponding age-based visual acuity, chromatic and contrast sensitivity. By adjusting this schedule and the contrast-threshold hyper-parameters, we derived DVD variants emphasizing shape bias, performance, or a balanced trade-off.

Shape bias and robustness analysis: Shape bias is measured using cue-conflict stimuli, where texture and shape cues are in conflict, and quantified as the proportion of images classified based on shape over texture (Geirhos et al., 2018).

Degradation robustness was evaluated based on object/face recognition accuracy under progressively increasing image blur (Jang et al., 2021). Adversarial robustness was quantified by the model’s top-1 accuracy under L2-based Gaussian and uniform noise attacks of escalating intensity.

Acknowledgments

The authors acknowledge support by the European Research Council (ERC) StG grant 101039524 TIME (Kietzmann, Thorat), ERC Consolidator grant ERC-CoG-2024101123101(Cichy), CSC grant 202106120015 (Lu), and acknowledge that all computations were funded by the DFG – 456666331.

References

Baker, N., et al. (2018). Deep convolutional networks do not classify based on global object shape. *PLoS computational biology*, *14*(12), e1006613.

Braddick, O., et al. (2011). Development of human visual function. *Vision Research*, *51*(13), 1588–1609. doi: 10.1016/j.visres.2011.02.018

Caltrider, D., et al. (2024). Evaluation of visual acuity. In *Statpearls.* StatPearls Publishing.

Dobson, V., et al. (2009). Normative monocular visual acuity for early treatment diabetic retinopathy study charts in emmetropic children 5 to 12 years of age. *Ophthalmology*, *116*(7), 1397–1401. doi: 10.1016/j.ophtha.2009.01.019

Drover, J. R., et al. (2008). Normative pediatric visual acuity using single surrounded hotv optotypes on the electronic visual acuity tester following the amblyopia treatment study protocol. *Journal of AAPOS*, *12*(2), 145–149. doi: 10.1016/j.jaapos.2007.08.014

El-Gohary, A., et al. (2017). Age norms for grating acuity and contrast sensitivity measured by lea tests in the first three years of life. *International Journal of Ophthalmology*, *10*, 1150–1153. doi: 10.18240/ijo.2017.07.20

Geirhos, R., et al. (2018). Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. In *International conference on learning representations.*

Huber, L. S., et al. (2023). The developmental trajectory of object recognition robustness: children are like small adults but unlike big deep neural networks. *Journal of vision*, *23*(7), 4–4.

Inal, A., et al. (2018). Comparison of visual acuity measurements via three different methods in preschool children: Lea symbols, crowded lea symbols, snellen e chart. *International Ophthalmology*, *38*(4), 1385–1391. doi: 10.1007/s10792-017-0596-1

Jang, H., et al. (2021). Convolutional neural networks trained with a developmental sequence of blurry to clear images reveal core differences between face and object processing. *Journal of vision*, *21*(12), 6–6.

Jang, H., et al. (2024). Improved modeling of human vision by incorporating robustness to blur in convolutional neural networks. *Nature Communications*, *15*(1), 1989.

Katzhendler, G., et al. (2019a). *Blurred images lead to bad local minima.*

Katzhendler, G., et al. (2019b). Potential upside of high initial visual acuity? *Proceedings of the National Academy of Sciences*, *116*(38), 18765–18766. doi: 10.1073/pnas.1906400116

Kugelberg, U. (1992). Visual acuity following treatment of bilateral congenital cataracts. *Documenta Ophthalmologica*, *82*(3), 211–215. doi: 10.1007/BF00160767

Leone, J. F., et al. (2014). Normative visual acuity in infants and preschool-aged children in sydney. *Acta Ophthalmologica*, *92*(7), e521–e529. doi: 10.1111/aos.12366

Mehrer, J., et al. (2021). An ecologically motivated image dataset for deep learning yields better models of human vision. *Proceedings of the National Academy of Sciences*, *118*(8), e2011417118.

Teller, D. Y. (1983). Measurement of visual acuity in human and monkey infants: the interface between laboratory and clinic. *Behavioural Brain Research*, *10*(1), 15–23.

Vogelsang, L., et al. (2018). Potential downside of high initial visual acuity. *Proceedings of the National Academy of Sciences*, *115*(44), 11333–11338. doi: 10.1073/pnas.1800901115