Feedforward Neural Networks can capture Human-like Perceptual and Behavioral Signatures of Contour Integration

Fenil R. Doshi (fenil_doshi@fas.harvard.edu)
Department of Psychology, Harvard University
33 Kirkland Street, Cambridge, Massachusetts 02138

Talia Konkle (tkonkle@fas.harvard.edu)
Department of Psychology and Center for Brain Sciences, Harvard University
33 Kirkland Street, Cambridge, Massachusetts 02138

George A. Alvarez (alvarez@wjh.harvard.edu)
Department of Psychology, Harvard University
33 Kirkland Street, Cambridge, Massachusetts 02138

Abstract

Contour integration is the process of linking local edge elements to arrive at a unified perceptual representation of a complete contour, and may thus serve as a critical pre-cursor representation needed to extract global shape information supporting object recognition. Many mechanisms have been proposed for such a feature-linking process (Field, Hayes & Hess, 1993; Kellman & Shipley, 1991), including long-range lateral interactions (Bosking et al., 1997), temporally synchronized cortical oscillations (Engel, Konig & Singer, 1991), and top-down feedback connections (Kim et al., 2019). In this study, we test the alternative possibility that feed-forward, nonlinear convolutional neural networks are able to perform contour integration without lateral connections, recurrence, or top-down feedback. We find that such a feedforward system exhibits sensitivities to global and local contour curvatures comparable to humans, but it requires two critical inductive biases to do so - visual experience of relatively straight-looking smooth contours and an architectural constraint of increasing receptive field size. Through this approach, we provide computational support for the hypothesis that a hierarchical feedforward visual processor can develop and leverage Gestalt-like laws of "good continuation" to detect extended contours in a manner consistent with human perception.

Keywords: Perceptual Grouping; Gestalt vision; Psychophysics; Deep Neural Networks

Methods and Results

The primary question of this work is whether feedforward architectures can show human-like sensitivities to contour integration.

Behavioral Experiment

To first address this question, we conducted a behavioral study with human participants using a similar experimental setup as Field et al. (1993). We recruited 46 participants via Prolific and presented them with a subset of 1000 synthetic images of gabor elements (see **Figure 1a**). Each image was comprised of an array of 256 gabor elements, among which 12 elements form an extended contour, while the remaining 244 elements are positioned randomly without any density cue to segregate the contour elements from the background. The curviness of the path of the hidden contour was parametrically controlled to vary between 15 and 75 degrees. An example image with 15 degrees of curvature is shown in **Figure 1A**.

In the behavioral experiment, each trial consisted of two images, one containing a contour and the other a tightly controlled version with identical contour and background elements but with randomized orientation of the contour elements. The two images were presented for one second each with a 500ms gap, and the participants were asked to identify which of the two images contained a contour.

The results of this behavioral study are shown in **Figure 1B (blue line)**, which show a systematic fall off in detecting contours as the degree of path curvature increases.

Model Experiments

We next investigated whether a standard Alexnet model pretrained for object recognition had the capacity to detect these contours (Krizhevsky, Sutskever & Hinton, 2017). This model is purely feedforward, and thus lacks all the previously hypothesized mechanisms underlying the perceptual representation of curvatures. We reasoned that this model would fail to detect the presence of extended contours in these displays, providing an important baseline architecture over which further mechanistic connections could be added.

To establish this baseline, we attached a linear read-out head at the final stage of the model, and fine-tuned the model end-to-end while training the read-out head on a contour present/absent task. The model was trained on 5000 images and validated on a dataset of 600 images, both containing contours sampled from a broad range of curvatures angles.

However, we found that the fine-tuned models could accurately detect all curvatures--even the curviest contours, where human perceptual capacities were unable (**Figure 1B- gray line**). We use guided backpropagation (Springenberg et al., 2014; **Figure 1C**) to verify that the model is using the contour elements itself to accurately detect the contour, rather than a short cut. Thus, contrary to our assumptions, purely feed forward convolutional architectures are in fact architecturally capable of “path integration” as operationalized by this task.

Inductive Bias: Receptive Field Size

We reasoned that the later stages of the model, which have larger receptive fields encompassing increasingly more of the full display, must be capable of detecting these contours and supporting the contour detection capacity. To test this, we conducted the same read-out

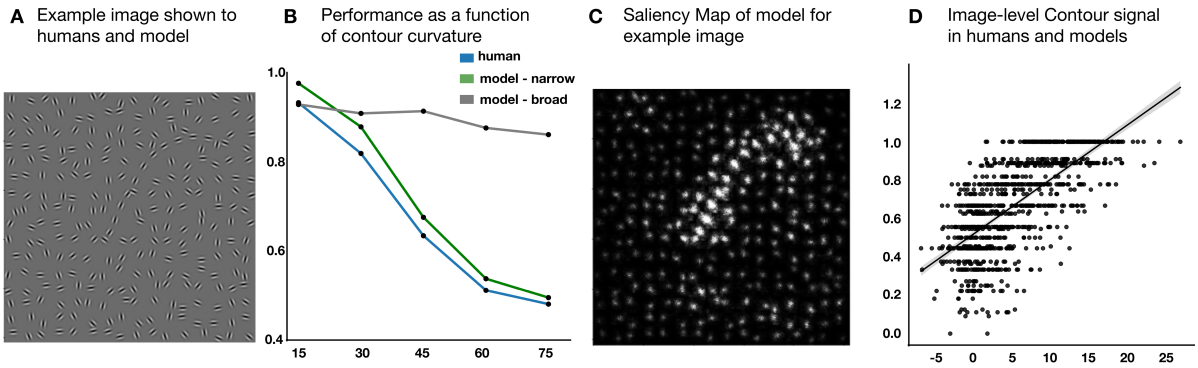

Figure 1: (A) Example image containing gabor elements, subset of which fall along an extended contour with 15° curvature (B) Human and Model (finetuned on narrow and broad curvatures) performance as a function of curvature (C) Saliency map to detect contour from the finetuned model (D) Image-level comparison between model signal strength (x-axis) and human percent correct (y-axis)

and fine-tuning analyses across different layers. Indeed, contour detection accuracy increased with model depth, peaking around the final convolutional layer. Further, we additionally tested bag-net models (Brendel & Bethge, 2019) which do not show hierarchically increasing receptive field properties, and these models were unable to detect the contours across all training regimes we tested. Thus, these results demonstrate that later representational stages of these convolutional feed-forward models are capable of capturing long-range contour information, through units with larger receptive field sizes.

Inductive Bias: Trained Orientation

Given the models were actually too good at detecting paths with extreme curvature, we next examined whether visual training on a narrower range of curvature values would lead to a more human-like degradation of contour detection. We focused on models fine-tuned from the avg-pool layer of Alexnet, based on the first set of results. We then conducted a series of parametric experiments to train models on a variety of different curvature ranges, probing their generalization curves across the full range of curvatures.

We found that when models are trained on relatively straight smooth contours, with a peak at 18 degrees, they show a gradual fall-off with curvature similar to human participants (**Figure 1B – green line**)

We quantified this more rigorously by developing a trial-level correlation measure based on a signal detection framework. For the human data, percent correct on each trial of the 2-AFC task can be taken as a direct measure of the trial-level contour signal strength. For the models, which are noiseless, we directly used the distance from the decision boundary as a measure of contour signal strength for each trial.

Across models, we find a peak correlation between trial-level accuracy between humans and models of 0.695 (**Figure 1D**), for models trained at 18 degrees (generalizing across models trained with some jitter around this peak). Thus, we find that training contour detection in this narrower window of relative straight curvatures naturally leads to systematic and human-like fall-off of contour detection as curvature increases.

Conclusions

Traditional frameworks of contour perception focus on mechanisms operating on early representational layers (e.g. lateral connections of V1), or within local recurrence (e.g. V2-V1 coordinated processing). While our initial aim was to introduce these mechanisms into feed forward deep neural network models, we instead discovered computational support for an alternative mechanism. That is, purely feedforward hierarchical processing can develop and leverage Gestalt-like laws of "good continuation" to detect extended contours in a manner consistent with human perception. We link this capacity to the role of larger receptive fields. And, further, we offer an alternative explanation of human-like contour detection as one that arises from mechanisms aimed at relatively straight (off-meridian) contours. Less accurate detection of increasingly curvy contours naturally follows. It will be important for future work to discover the generality of these claims—e.g. is 18 degrees somehow a critical curvature level in natural image statistics, or is this number more a result of the particular configuration of receptive field sizes present in the Alexnet hierarchical architecture. Broadly, these results raise new empirical avenues to explore contour sensitivity in higher level regions and support the hypothesis that a narrow edge-detection mechanism may provide a simpler but productive explanation of a broader range of contour detection perceptual abilities.

Acknowledgments

This work was supported by NSF CAREER BCS-1942438 to TK and NSF PAC COMP-COG 1946308 to GAA.

References

Bosking, W. H., Zhang, Y., Schofield, B., & Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *Journal of neuroscience*, *17*(6), 2112-2127.

Brendel, W., & Bethge, M. (2019). Approximating cnns with bag-of-local-features models works surprisingly well on imagenet. *arXiv preprint arXiv:1904.00760*.

Engel, A. K., König, P., & Singer, W. (1991). Direct physiological evidence for scene segmentation by temporal coding. *Proceedings of the National Academy of Sciences*, *88*(20), 9136-9140.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: evidence for a local “association field”. *Vision research*, *33*(2), 173-193.

Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive psychology*, *23*(2), 141-221.

Kim, J., Linsley, D., Thakkar, K., & Serre, T. (2019). Disentangling neural mechanisms for perceptual grouping. *arXiv preprint arXiv:1906.01558*.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84-90.

Springenberg, J. T., Dosovitskiy, A., Brox, T., & Riedmiller, M. (2014). Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806*.