Check for updates

Editorial: Perceptual organization in computer and biological vision

James H. Elder[1,2†], Mary A. Peterson[3,4†] and Dirk B. Walther[5*†]

[1]Department of Electrical Engineering and Computer Science, York University, Toronto, ON, Canada, [2]Department of Psychology, York University, Toronto, ON, Canada, [3]Cognitive Science Program, University of Arizona, Tucson, AZ, United States, [4]Department of Psychology, University of Arizona, Tucson, AZ, United States, [5]Department of Psychology, University of Toronto, Toronto, ON, Canada

Editorial on the Research Topic
Perceptual organization in computer and biological vision

A principal challenge for both biological and machine vision systems is to integrate and organize the diversity of cues received from the environment into the coherent global representations we experience and use to make good decisions and take effective actions. Early psychological investigations date back more than 100 years to the seminal work of the Gestalt school (Koffka, 1935; Wertheimer, 1938). Yet in the last 50 years, neuroscientific and computational approaches to understanding perceptual organization have become equally important, and a full understanding requires integration of all three approaches (Wagemans et al., 2012; Elder, 2018). Perceptual organization can be defined as the process of establishing meaningful relational structures over raw visual data, where the extracted relations correspond to the physical structure and semantics of the scene. The relational structure may be simple, such as set membership for image segmentation, or more complex, such as sequence representations of contours, hierarchical representations of surfaces, or layered representations of scenes. These structures support 3D scene understanding, object detection, object recognition, and activity recognition, among other tasks.

This Frontiers Research Topic brings together 13 contributions to Frontiers in Psychology and Frontiers in Computer Science, with the aim of presenting a single, unified collection that will encourage integration and cross-fertilization across disciplines. Together, these contributions explore how the brain forms representations of contours, surfaces, and objects over 3D space and time, and the degree to which representations formed by recent deep learning models may be similar or different. Here we briefly introduce these 13 contributions and highlight how they interrelate.

"Shape from dots: a window into abstraction processes in visual perception": What constraints and rules does the visual system use to organize simple visual elements into meaningful contours? Displays of dots provide an interesting way of exploring these grouping rules. Unlike Gabor patches (Field et al., 1993; Kovacs and Julesz, 1993, 1994) and line elements (Pettet, 1999; Drewes et al., 2016; Elder et al., 2018; Baker et al., 2021), dots do not provide orientation information. Nevertheless, observers group dots into contours. Baker and Kellman explore under which geometric conditions people perceive a spatial sequence of dots as executing a smooth vs. abrupt change in orientation. They find that a triplet of dots forming an obtuse angle (more than 90 degrees) is perceived as a smooth contour, whereas a triplet forming an acute angle (<90 degrees) is perceived as an abrupt vertex. Dot displays that describe curvilinear contours as opposed to sharp-angled

vertices allowed for clearer perception, better mental rotation, and more accurate detection of shapes. These results may reflect the underlying statistics of smooth contour curvatures and abrupt orientation changes we encounter in the visual world.

"Combining contour and region for closed boundary extraction of a shape": Ultimately, neural mechanisms must organize local spatial features coded in early stages of the visual system into the coherent object representations we perceive. The grouping cues that support this computation include geometric regularities of the object's bounding contour (e.g., good continuation) as well as photometric regularities within the object (e.g., color similarity; Elder and Zucker, 1996, 1998). In this contribution, Hii and Pizlo propose a foveated shortest-path model of contour grouping to explore the potential fusion of geometric contour and color cues in recovering complete object boundaries. Psychophysical results demonstrate that the human visual system can synergistically combine geometric and color grouping cues, in qualitative agreement with their computational model.

"Specific Gestalt principles cannot explain (un)crowding": The contribution from Hii and Pizlo concerned how local elements on the retina are organized into a representation of a coherent figure or object. But perceptual organization extends beyond a single figure to determine how we perceive collections of figures or objects in the scene. One window into this process is provided by the study of *crowding*. Crowding is the phenomenon wherein fine spatial judgements can be made more difficult if extraneous "distractor" elements are brought near to the stimulus being judged. *Uncrowding* refers to the remarkable fact that adding a regular pattern of multiple distractors can release this effect. This phenomenon has generally been attributed to the perceptual organization of these extraneous elements into a perceptual group apart from the stimulus being judged. However, in their psychophysical study, Choung et al. find that, while the degree of uncrowding is strongly correlated with perceived grouping, simple models of perceptual grouping fail to account for this relationship. This suggests that the formation of perceptual groups may depend upon subtle interplays and higher-level perceptual interpretations of the visual stimulus that are not easily captured by a simple combination of Gestalt laws.

"Good continuation in 3D: the neurogeometry of stereo vision": The studies discussed so far provide intriguing insight into perceptual organization in the 2D image plane. But how does this relate to the structure of our 3D visual world? Bolelli et al. note that the back-projected boundaries of solid objects are generally not planar curves (Koenderink, 1984), and their 3D structure can be important to perceptual organization and object understanding. Fortunately, this 3D structure can potentially be recovered via the geometry of the binocular projection. Bolelli et al. introduce a mathematical framework relating the projected geometry of these 3D curves to binocular neural selectivity. Based on tools from sub-Riemannian geometry, their model makes predictions about how interactions between neurons in early visual cortex should depend upon the ocularity and joint position-orientation tuning of the neurons. This model provides a framework for understanding the stereo correspondence problem as well as torsional eye movements.

"The coherent organization of dynamic visual images": The challenge of perceptual organization extends not only over the

three dimensions of space but also the dimension of time. The review article by Lappin and Bell details how the brain uses spatiotemporal regularities in moving images to perceptually organize the visual stream into continuous surface representations that support the discrimination of fine spatiotemporal judgements with hyperacuity precision.

"Visual cortical processing—From image to object representation": The foveated shortest-path object grouping model of Hii and Pizlo entails an incremental construction of progressively more global, complex, and complete representations. While Hii and Pizlo do not suggest a specific mapping of their model to brain regions, it is common to assume that such computations proceed hierarchically from early to later visual areas. However, a body of work from Zhou et al. (2000), Craft et al. (2007), von der Heydt (2015), Williford and von der Heydt (2016), and others, provides an alternative account. These findings include neural sensitivity in earlier areas of visual cortex to illusory contours and figure/ground assignment that could only emerge from more global computations, challenging the conventional view. In particular, the identification of *border ownership cells* in cortical area V2 that respond selectively to a contour depending upon the figure/ground sign is strong evidence against a feedforward, hierarchical view of object perception. What is the alternative? von der Heydt reviews computational and neurophysiological research supporting the existence of *grouping cells* (*G* cells) that pre-attentively link neurons in early visual areas that are selective for contours to form representations of global "proto-objects" via recurrent processing. von der Heydt conjectures that these G cells might be located outside of the object pathway in the ventral stream, since recordings in areas V1, V2, and V4 have failed to confirm their existence.

"Backward masking implicates cortico-cortical recurrent processes in convex figure context effects and cortico-thalamic recurrent processes in resolving figure-ground ambiguity": Peterson and Campbell also present evidence against a feedforward account of visual perception. They show that recurrent processing plays an essential role in the perception of classic figure-ground displays that were long taken as evidence that convexity is an important prior in building objects in a bottom-up fashion. Previously, Peterson and Salvagio (2008) and Goldreich and Peterson (2012) found that convexity is a weak figural prior unless it is supplemented by a background prior. The background prior requires homogeneous fill-in concave regions alternating with convex regions. Peterson and Campbell show that the convexity prior and the background prior conflict in traditional displays where both convex and concave regions are homogeneously colored and that recurrent processing resolves this conflict before conscious perception. Furthermore, they identify both cortico-cortical and cortical-thalamic recurrent processes in the perceptual organization of the classic displays. Their experiments show that dynamical recurrent interactions are involved in some of the foundational experiments taken as evidence for a feed-forward model of figure-ground perception.

"Perceptual organization and visual awareness: the case of amodal completion": It has long been debated whether the process of amodal completion of partially occluded objects demands attention and awareness or whether it can occur autonomously. Here, Kimchi et al. report four experiments investigating this

question, using a variant of a color-opponent flicker technique in which a priming stimulus can be presented for a duration necessary for perceptual completion while remaining outside perceptual awareness. Kimchi et al. used this technique to create priming stimuli that cued either a local, global, or ambiguous interpretation of a subsequent target stimulus. They found that when the prime indicated a local completion, local targets were classified faster than global targets, suggesting that local completion can take place without visual awareness. However, when the prime cued a global or ambiguous interpretation, target responses were unaffected by the prime, which they take as evidence that awareness is necessary to resolve ambiguity and to generate a global completion.

"Visual and haptic cues in processing occlusion": Vision is only one of the human senses, and fusion with haptic sensing could be particularly important to inform the perceptual organization of partially occluded objects that are only partially visible to the eye. Prior work has shown that partially occluded faces are more easily recognized when the occluders are stereoscopically rendered to appear in front, rather than behind, the faces. Here, Takeichi et al. use virtual reality to investigate how both visual and haptic information about the relative depth of the occluder affects recognition of katakana characters. While the haptic cue was found to increase the confidence of observer judgements of the relative depth of the occluder, there was no effect on character recognition. Also, counter to prior work with faces, character recognition was better when the "occluder" was rendered to be behind, rather than in front, of the character, suggesting that 3D processing may be different for specialized 2D stimuli like textual characters than for faces.

"The mid-level vision toolbox for computing structural properties of real-world images": The research reviewed above largely follows in the tradition of Gestalt psychology in using highly simplified stimuli to isolate specific perceptual factors and test hypotheses. However, the maturation of computer vision technologies provides opportunity to explore whether principles of perceptual organization generalize to real-world scenes in all of their complexity. Walther et al. provide a useful resource for this endeavor with their Mid-Level Vision (MLV) Toolbox. The toolbox offers algorithms for extracting contours from photographs and for computing a variety of contour properties: orientations, curvature, length, and contour junctions. Relying on the medial axis transform as a dual representation of scene contours, the toolbox provides code to compute measures of local parallelism, local mirror symmetry, and contour separation. The toolbox also contains code for visualizing these properties and for manipulating contour drawings based on them.

"Does training with blurred images bring convolutional neural networks closer to humans with respect to robust object recognition and internal representations?": The success of deep learning models in solving computer vision problems has led to their adoption as potential models for predicting neural and behavioral response to visual stimuli. While these models do capture many aspects of neural and behavioral response, there are intriguing divergences in how networks handle out-of-distribution perturbations such as image blur. Here, Yoshihara et al. find that training convolutional networks with a mix of blurry and sharp images makes them more

human-like in their robustness to blur and weighting of shape vs. texture in making classification decisions (Geirhos et al., 2018).

"Shape-selective processing in deep networks: integrating the evidence on perceptual integration": Training with blurred stimuli likely knocks out fine-scale texture cues that networks tend to rely on by default, upweighting the use of shape cues. But what is the nature of the shape cues that these networks can use? While humans make profound use of configural shape information, recent research suggests that deep networks struggle to organize these global shape cues, relying more on local shape features (Baker et al., 2018; Baker and Elder, 2022). In their contribution, Jarvers and Neumann perform a new analysis of deep neural network shape sensitivity that suggests that the addition of recurrent or residual connections can enhance sensitivity to non-local shape, although not to the extent seen in humans. These results suggest future directions for neural network design that may lead to models that are better able to capture the human ability to organize local features into representations of global object shape.

"Self-attention in vision transformers performs perceptual grouping, not attention": Deep learning models have made substantial gains in performance through mechanisms of "self-attention" and "cross-attention" that allow for multiplicative interactions between data inputs and are the basis for more recent state-of-the-art transformer architectures. Here, Mehrani and Tsotsos argue that the effect of self-attention is in fact more appropriately described as *perceptual organization* based on feature similarity. In a series of computational experiments, they demonstrate that vision transformers learn to group stimuli based on features such as hue, lightness, saturation, shape, size, or orientation and suggest that this can be thought of as a form of horizontal relaxation labeling. This novel view provides insight into how transformer architectures may solve difficult perceptual organization problems that challenge convolutional architectures.

Author contributions

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Baker, N., and Elder, J. H. (2022). Deep learning models fail to capture the configural nature of human shape perception. *iScience* 25:104913. doi: 10.1016/j.isci.2022.104913

Baker, N., Garrigan, P., and Kellman, P. J. (2021). Constant curvature segments as building blocks of {2D} shape representation. *J. Exp. Psychol. Gen.* 150:1556. doi: 10.1037/xge0001007

Baker, N., Lu, H., Erlikhman, G., and Kellman, P. J. (2018). Deep convolutional networks do not classify based on global object shape. *PLOS Comput. Biol.* 14:e1006613. doi: 10.1371/journal.pcbi.1006613

Craft, E., Schutze, H., Niebur, E., and von der Heydt, R. (2007). A neural model of figure-ground organization. *J. Neurophysiol.* 97, 4310–4326. doi: 10.1152/jn.00203.2007

Drewes, J., Goren, G., Zhu, W., and Elder, J. H. (2016). Recurrent processing in the formation of shape percepts. *J. Neurosci.* 36, 185–192. doi: 10.1523/JNEUROSCI.2347-15.2016

Elder, J. H. (2018). Shape from contour: computation and representation. *Ann. Rev. Vision Sci.* 4, 423–450. doi: 10.1146/annurev-vision-091517-034110

Elder, J. H., Oleskiw, T., and Fründ, I. (2018). The role of global cues in the perceptual grouping of natural shapes. *J. Vision* 18, 1–21. doi: 10.1167/18.12.14

Elder, J. H., and Zucker, S. W. (1996). "Computing contour closure," in *Proceedings of the 4th European Conference on Computer Vision*, eds. B. F. Buxton and R. Cipolla (Cham: Springer Verlag), 399–412.

Elder, J. H., and Zucker, S. W. (1998). Evidence for boundary-specific grouping. *Vision Res.* 38, 143–152. doi: 10.1016/S0042-6989(97)00138-7

Field, D. J., Hayes, A., and Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local "association field". *Vision Res.* 33, 173–193. doi: 10.1016/0042-6989(93)90156-Q

Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., Brendel, W., et al. (2018). ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv preprint*. arXiv:1811.12231.

Goldreich, D., and Peterson, M. A. (2012). A Bayesian observer replicates convexity context effects. *Seeing Perc.* 25, 365–395. doi: 10.1163/187847612X634445

Koenderink, J. J. (1984). What does the occluding contour tell us about solid shape? *Perception* 13, 321–330. doi: 10.1068/p130321

Koffka, K. (1935). *Principles of Gestalt Psychology*. Harcourt: Brace and World.

Kovacs, I., and Julesz, B. (1993). A closed curve is much more than an incomplete one: Effect of closure in figure-ground discrimination. *Proc. Natl. Acad. Sci. USA* 90, 7495–7497. doi: 10.1073/pnas.90.16.7495

Kovacs, I., and Julesz, B. (1994). Perceptual sensitivity maps within globally defined visual shapes. *Nature* 370, 644–646. doi: 10.1038/370644a0

Peterson, M. A., and Salvagio, E. (2008). Inhibitory competition in figure-ground perception: context and convexity. *J. Vision* 8, 1–13. doi: 10.1167/8.16.4

Pettet, M. W. (1999). Shape and contour detection. *Vision Res.* 39, 551–557. doi: 10.1016/S0042-6989(98)00130-8

von der Heydt, R. (2015). Figure–ground organization and the emergence of proto-objects in the visual cortex. *Front. Psychol.* 6:1695. doi: 10.3389/fpsyg.2015.01695

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., et al. von der. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychol. Bullet.* 138:1172. doi: 10.1037/a0029333

Wertheimer, M. (1938). "Laws of organization in perceptual forms," in *A Sourcebook of Gestalt Psychology*, ed. W. D. Ellis (London: Routledge and Kegan Paul), 71–88.

Williford, J. R., and von der Heydt, R. (2016). Figure-ground organization in visual cortex for natural scenes. *eNeuro* 3:6. doi: 10.1523/ENEURO.0127-16.2016

Zhou, H., Friedman, H., and von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *J. Neurosci.* 20, 6594–6611. doi: 10.1523/JNEUROSCI.20-17-06594.2000