Texture segregation by visual cortex: Perceptual grouping, attention, and learning
Rushi Bhatt 1, Gail A. Carpenter 2, Stephen Grossberg *,3

Department of Cognitive and Neural Systems, Center for Adaptive Systems and Center of Excellence for Learning in Education, Science, and Technology, Boston University, 677 Beacon Street, Boston, MA 02215, USA

Received 30 July 2006; received in revised form 19 June 2007

Abstract

A neural model called dARTEX is proposed of how laminar interactions in the visual cortex may learn and recognize object texture and form boundaries. The model unifies five interacting processes: region-based texture classification, contour-based boundary grouping, surface filling-in, spatial attention, and object attention. The model shows how form boundaries can determine regions in which surface filling-in occurs; how surface filling-in interacts with spatial attention to generate a form-fitting distribution of spatial attention, or attentional shroud; how the strongest shroud can inhibit weaker shrouds; and how the winning shroud regulates learning of texture categories, and thus the allocation of object attention. The model can discriminate abutted textures with blurred boundaries and is sensitive to texture boundary attributes like discontinuities in orientation and texture flow curvature as well as to relative orientations of texture elements. The model quantitatively fits the Ben-Shahar and Zucker [Ben-Shahar, O. & Zucker, S. (2004). Sensitivity to curvatures in orientation-based texture segmentation. Vision Research, 44, 257–277] human psychophysical data on orientation-based textures. Surface-based attentional shrouds improve texture learning and classification: Brodatz texture classification rate varies from 95.1% to 98.6% with correct attention, and from 74.1% to 75.5% without attention. Object boundary output of the model in response to photographic images is compared to computer vision algorithms and human segmentations.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Texture segregation; Object recognition; Image segmentation; Perceptual grouping; Spatial attention; Object attention; Attentional shroud; Visual cortex; Adaptive Resonance Theory (ART)

1. Introduction

1.1. Learning object form and texture

How does the brain effortlessly learn to recognize the global forms of objects while it also learns their surface attributes like local textures? Sometimes objects can be recognized just from their local textures (Biederman, 1981; Gurnsey and Laundry, 1992; Renninger and Malik, 2004). On the other hand, object form, notably texture boundaries, are often an important cue for object recognition (Beck, 1982; Biederman and Ju, 1988; Elder and Zucker, 1998; Grossberg and Mingolla, 1985b; Nothdurft, 1985). This article develops a neural model, called the Distributed ARTEx (dARTEX) model (Fig. 1), which can learn both global object form and local object texture by
operating at different scales of processing. This model also clarifies how spatial and object attention can work together to facilitate object and texture learning and recognition tasks.

The dARTEX model is inspired by two parallel streams of modeling work. The first stream is developing the 3D LAMINART model to explain and predict how the laminar circuits of visual cortex lead to visual percepts; e.g.,

The second stream models how the inferotemporal and prefrontal cortices work together to achieve fast, stable, incremental learning of distributed visual recognition categories in response to complex and changing visual environments. The foundations of this latter modeling stream were laid in articles about Adaptive Resonance Theory, or ART; e.g., Carpenter and Grossberg (1987, 1991), Carpenter, Grossberg, Markuzon, Reynolds, and Rosen (1992), Carpenter, Grossberg, and Reynolds (1991a), Carpenter, Grossberg, and Rosen (1991b), Grossberg (1976, 1978, 1980, 1999b), and Grossberg and Williamson (1999). These foundational ART models used winner-take-all recognition categories. The main ART predictions about brain dynamics have recently received support from neurophysiological and anatomical experiments, in addition to earlier supportive psychological data; see Grossberg (1995, 2003) and Raizada and Grossberg (2003) for reviews.

The ARTEx model of Grossberg and Williamson (1999) joined together visual preprocessing (multiple-scale bottom–up filtering, horizontal grouping, and surface filling-in) as a perceptual front end to an ART classifier to learn and categorize both Brodatz textures and natural textured scenes after they were processed by a synthetic aperture radar (SAR) sensor. The present work extends the Grossberg and Williamson (1999) study. It builds upon a more recent development of ART, called Distributed ART, or dART, which shows how distributed, rather than winner-take-all, learning can be combined with previously established ART properties of fast, stable, incremental learning and recognition (Carpenter, 1997, 2001; Carpenter, Milenova, and Noeske, 1998).

The resulting dARTEX model combines multiple-scale bottom–up filtering, horizontal grouping, top–down spatial and object attention, and a dART classifier, in a laminar cortical circuit model. The dARTEX model is used to quantitatively simulate a large set of challenging human psychophysical data about Orientation-Based Texture Segmentation (OBTS) from the experiments by Ben-Shahar and Zucker (2004). Fig. 2 shows some of the texture stimuli used in these experiments. Fig. 3 describes the data and the dARTEX simulations of these data.

1.2. Explaining OBTS data

OBTS data are important because they illustrate perceptual differences due to variations in orientation gradients and in element configurations at texture boundaries. These configural effects in texture segmentation have long been observed, but a comprehensive, quantitative, and mechanistic explanation of the processes underlying the asymmetries in boundary perception due to the spatial layout of their texture elements and orientations has been lacking (Beck, 1982; Ben-Shahar and Zucker, 2004; Nothdurft, 1985; Olson and Atteave, 1970; Wolfson and Landy, 1995).

Explaining the OBTS data illustrates the biological relevance of dARTEX. It does not, however, show that dARTEX can actually learn complex textures. To demonstrate this, we show that dARTEX achieves favorable benchmarks compared to other texture classifiers in the computer vision literature (Greenspan, Goodman, Chellappa, and Anderson, 1994; Grossberg and Williamson, 1999; Randen and Husoy, 1999). This dARTEX improved texture classification performance is due to top–down spatial attention and autonomous self-supervised learning of novel textures from the visual scene. As will be demonstrated in Section 3, these mechanisms in dARTEX reduce classification errors at texture boundaries by preventing mixing of features from different textures during classification. Prior texture classification approaches do not employ such attentional modulation to prevent the mixing, and therefore are most susceptible to classification errors at the interface of abutted textures.

1.3. Form-fitting spatial attention modulates learning and eye movements

A key new insight of the dARTEX model is to show how spatial attention can significantly enhance texture learning and recognition, including the regulation of this learning by object attention. Thus, the present work proposes how spatial and object attention may be coordinated during object learning and recognition. This insight builds upon work by Fazl, Grossberg, and Mingolla (2005) which shows how such a surface-induced spatial attentional representation, or attentional shroud (Tyler and Kontsevich, 1995), can be used to learn view-invariant object categories while eye movements actively search a scene. The present work shows how multiple scales of learning and recognition can be coordinated by such an attentional shroud to learn both object texture and form.

It should be noted at the outset that our concept of attentional shroud emphasizes different aspects of form-fitting spatial attention than the original concept of Tyler and Kontsevich (1995). Tyler & Kontsevich proposed an attentional shroud that is a self-organizing surface map which morphs to account for momentarily available depth cues. This concept was extended and integrated to include multiple attentional shrouds extending from each attended loca-
Fig. 2. dARTEX boundary grouping outputs denoting texture boundaries in some interesting stimuli from the Ben-Shahar and Zucker (2004) experiments. In each panel, top figure is the input to dARTEX and the bottom figure is the boundary grouping activity in V2 layers 2/3 added over all three scales for 12 orientations. Length of each oriented line segment is proportional to the boundary activity in that orientation at that location. The perceptual saliencies in each case can be explained through grouping boundaries and top–down amplification of elements through boundary-based attention. Here, \( \Delta k_T \) is the tangential curvature discontinuity at the horizontal boundary, \( \Delta k_N \) is the normal curvature discontinuity, \( \Delta h_b \) is the angular discontinuity at the boundary, and \( \Delta h_w \) is the image orientation gradient magnitude. See text for a complete description of these stimuli. (a) \( \Delta k_T = \text{max} \), \( \Delta h_w = 5^\circ \), and \( \Delta h_b = 90^\circ \) (b) \( \Delta k_N = \text{max} \), \( \Delta h_w = 5^\circ \) and \( \Delta h_b = 30^\circ \) (c) \( \Delta k_T = \Delta k_N \), \( \Delta h_w = 5^\circ \) and \( \Delta h_b = 90^\circ \) (d) \( \Delta k_T = \text{max} \), \( \Delta h_w = 15^\circ \), \( \Delta h_b = 90^\circ \) (e) \( \Delta k_N = \text{max} \), \( \Delta h_w = 15^\circ \), \( \Delta h_b = 90^\circ \) (f) \( \Delta k_T = \Delta k_N \), \( \Delta h_w = 15^\circ \), \( \Delta h_b = 90^\circ \).
Fig. 3. Complete simulation results of the Ben-Shahar and Zucker (2004) experiments. Each panel shows detection accuracy curves for a specific image orientation gradient magnitude $\Delta \theta_{\text{within}}$ as the boundary angular discontinuity amount $\Delta \theta_{\text{between}}$ increases from $5^\circ$ to $90^\circ$. Points marked by boxes (blue curves) belong to $\Delta k_T = \max$ configurations, those marked by circles (red curves) are $\Delta k_N = \max$ configurations, and those by crosses (green curves) are $\Delta k_T = \Delta k_N$ configurations. Each of these configurations leads to a different configuration of texture bars. Average performance over all configurations is marked by triangles (black curves). For each panel, the top figure shows experimental data and the bottom panel shows dARTEX salience outputs. See Section 3 for a complete description.
tion to complete the local surface out to the nearest contour boundary. A key aspect of this concept is that there is a unique depth map for whatever scene is before us.

The stimuli that are simulated in the present work are two-dimensional images of texture-defined object shapes, rather than scenes containing three-dimensional objects. These simulations illustrate how a surface-based form-fitting representation of spatial attention may be used to facilitate learning of texture categories, and recognition of surface texture properties. We believe that the two shroud concepts probe the same underlying brain processes. Indeed, related modeling work about 3D shape proposes how perceptual representations of tilted, slanted, and curved three-dimensional object shapes may be formed (Grossberg, Kuhlmann, and Mingolla, 2007; Grossberg and Swaminathan, 2004). Additional modeling proposes how attentional shrouds influence learning of view-invariant representations of two-dimensional object shapes by regulating how multiple view-dependent categories of the shape may be bound together into a view-invariant object representation (Fazl et al., 2005; Fazl, Grossberg, and Mingolla, submitted for publication). We believe that the similarity of the Tyler & Kontsevich perceptual concepts about shrouds with our own concepts that link perception and category learning will become clear when the modeling studies above are generalized to the case of learning view-invariant categories of three-dimensional object shapes. Future experimental tests and modeling projects will clarify if this prediction is correct.

The remainder of this article is organized as follows. Section 2 describes psychophysical and neurophysiological data that are explained by dARTEX. Section 3 introduces the dARTEX model and demonstrates its computational properties with the help of computer simulations. Section 4 describes the orientation-based texture segmentation experiments of Ben-Shahar and Zucker (2004) and compares dARTEX simulations to the experimental data. Section 5 describes dARTEX benchmark texture classification results using a set of images from the Brodatz (1966) album. Section 6 discusses object boundary processing in natural images. Section 7 discusses related prior studies, and Section 8 provides a discussion and conclusions. The dARTEX equations and parameters are given in the Appendices.

2. Psychophysical and neurophysiological data explained by dARTEX

2.1. Processing of orientation based textures: Boundary grouping instead of curvature operators

Effortless, pre-attentive, texture boundary processing in dARTEX requires multi-scale feature filtering and long-range perceptual grouping (Beck, 1982; Bergen and Landy, 1991; Caelli, 1985; Grossberg and Mingolla, 1985b; Sutter, Beck, and Graham, 1989). Differences in texture luminance, color, orientation, density, size, element shape, relative placement, and the statistical distribution of the texture features have all been found to influence texture segmentation (Beck, 1982; Bergen and Julesz, 1983; Grossberg and Mingolla, 1985b; Julesz, 1986; Malik and Perona, 1990; Nothdurft, 1985; Wolfson and Landy, 1998).

Orientation Defined Textures (ODTs) are generally made of tilings of oriented bars. Examples of such textures are shown in Fig. 2. Perceived segregation strength of ODTs depends on both the magnitude of orientation discontinuity at the boundary as well as the image orientation gradient (Ben-Shahar and Zucker, 2004; Nothdurft, 1985, 1992; Olson and Atteave, 1970). An increase in image orientation gradient, or the amount of angular variation, decreases boundary salience, while an increase in angular discontinuity at the boundary increases boundary salience (Nothdurft, 1985). Nothdurft (1992) showed that the ratio of angular discontinuity at the texture boundary and the orientation gradient in the image is a good predictor of perceived boundary salience, with higher ratios resulting in more salient boundaries. Texture element contrast, density, and relative sizes have also been shown to influence segregation strength (Nothdurft, 2000a, 2000b).

The importance of element configuration at the boundaries in ODTs has also been recognized (Beck, 1982; Nothdurft, 1985, 1992; Olson and Atteave, 1970). All other texture parameters being equal, oriented bars that are more parallel to the texture boundary produce stronger segregation than bars that are oblique to the boundaries (Nothdurft, 1985, 1992; Wolfson and Landy, 1995). Consider, for example, the orientation-based textures in Fig. 2. An increase in orientation discontinuity at the boundary results in a corresponding increase in boundary salience. In particular, texture boundaries in Fig. 2c are easier to detect than those in Fig. 2a. As noted earlier, element configuration at the boundary plays an important role in segmentation. Given the same amount of boundary discontinuity and image orientation gradient, elements aligned with the boundaries cause stronger boundary perceptions; compare Fig. 2a and b. As the orientation gradient in the image increases, the detection task gets harder. For example, even with the same amount of boundary discontinuity, texture boundaries in Fig. 2f are harder to detect than those in Fig. 2c. Also, while configurations with elements parallel to the boundary may still yield detectable boundaries (Fig. 2f), the task gets much harder for other element configurations as the orientation gradients increase. For example, the boundaries in Fig. 2d and e are harder to detect than those in Fig. 2f. Fig. 2d–f have the same orientation gradient and boundary discontinuity magnitudes, but differ in element configurations at the boundaries. For each of the images shown in Fig. 2, the accompanying dARTEX boundary grouping simulations below them reproduce these subjective observations on boundary detection.

Using textures like those in Fig. 2, Ben-Shahar and Zucker (2004) systematically investigated how various types of orientation flow discontinuities influence perceived texture
boundary salience. To the best of our knowledge, the Ben-Shahar and Zucker (2004) study is the most comprehensive in terms of the parametric range and the number of different configurations investigated. We therefore chose their Orientation-Based Texture Segmentation (OBTS) data as the main target dataset for dARTEX texture simulations. The stimuli and the parametric manipulations in the Ben-Shahar and Zucker (2004) experiments are described in more detail in Section 4.1. dARTEX simulations of human psychophysical performance with these stimuli are discussed in Sections 4.3 and 4.4.

The sensitivity to element configurations at texture boundaries discussed earlier in this section is difficult to explain using only the Filter-Rectification-Filter (FRF) model of human texture perception that was introduced as part of a more general model of boundary grouping by Grossberg and Mingolla (1985b). The FRF model consists of two Filtering stages with an intermediate Rectifying threshold; see also Bergen and Landy (1991), Bovik, Clark, and Geisler (1990), Graham, Beck, and Sutter (1992), Grossberg (1987), and Sutter et al. (1989). Wolfson and Landy (1995) used a version of the FRF model, with extra weight applied to the oriented channel aligned with the texture boundary, to explain a subset of the configurational effects mentioned above. The extra weight explained the increased salience of boundaries when elements are parallel to them, as compared to when elements are oblique to them.

Ben-Shahar and Zucker (2004, p. 267) noted that: “Unfortunately, such an explanation introduces a chicken-and-egg problem; the outcome (i.e., the orientation edge) must be given as an input to the computational process from which it is supposed to emerge (and after all, the goal of OBTS is to find these edges).” Wolfson and Landy (1995) also recognized this shortcoming of their FRF approach, mentioning that the method of assigning extra weight was “not particularly compelling” (Wolfson and Landy, 1995, p. 2782). In their study, Ben-Shahar and Zucker (2004, Section 4), went on to identify two intrinsic image parameters, named tangential and normal curvatures, and showed them to be intimately linked to their configurational effects. Ben-Shahar and Zucker (2004) further argued that any data fit that simply uses extra weight for certain boundary configurations or orientation gradient values without any regard to the vector nature of the image orientation gradient could not account for all their observations. For example, each panel in Fig. 3 has a fixed image orientation gradient magnitude, denoted by $\Delta \theta_{\text{within}}$, while the tangential and normal curvatures change in a specific way. For certain image orientation gradient magnitudes ($\Delta \theta_{\text{within}} = 15^\circ$, see Fig. 3a, b, and d), there is a crossover between the curves depicted by boxes and crosses (the blue and the green curves), which vanishes when $\Delta \theta_{\text{within}} = 20^\circ$; see Fig. 3e. Ben-Shahar and Zucker argued that selectively enhancing certain configurations (e.g., Wolfson and Landy, 1995) or image orientation gradients does not explain salience crossovers that vanish simply by increasing the image orientation gradient (Fig. 3e).

dARTEX simulates such effects without using explicit curvature computations. In fact, complex visual computations, that are often described using concepts like texture element configuration, curvature and its discontinuity, or other high-level scene descriptions, are explained in dARTEX as emergent properties of local, dynamic, cellular interactions that are consistent with anatomical and neurophysiological data. For example, sensitivity in OBTS to the angular discontinuity magnitude at the boundary emerges from the dynamics of a center-surround competition, wherein higher angular discontinuity results in larger feature contrast and therefore stronger boundaries. However, finer distinctions due to element configuration at the boundary, like those in Fig. 2, cannot be explained simply by the contrast in bottom–up oriented filter activities.

In the article that introduced the FRF concept, Grossberg and Mingolla (1985b) showed how FRF processes form part of a larger perceptual grouping process that is sensitive to configurational effects in OBTS. An earlier example of such sensitivity was demonstrated by Cuthberts, Gove, Grossberg, and Mingolla (1991), who simulated bipartite texture discrimination data of Beck, Prazdny, and Rosenfeld (1983, chap. A theory of textural segmentation). They used FRF followed by grouping to simulate data that FRF alone could not fully explain. dARTEX simulations in Section 4 show that boundary grouping, when combined with object attentional feedback that enhances grouped boundaries, and spatial attentional feedback that enhances features within the surfaces surrounded by these boundaries, together explain the asymmetry in boundary saliencies due to element configuration, and indeed all the main effects in the OBTS data; see Section 4.2.

2.2. Texture classification in texture discrimination

In addition to object boundary detection, region-based and local classification processes also play an important role in texture processing (Bovik et al., 1990; Caelli, 1985, 1988; Greenspan et al., 1994; Grossberg and Williamson, 1999; Jain and Farrokhnia, 1991). Human texture discrimination performance remains robust even when the boundary between juxtaposed textures is blurred, or when there is empty space between the two textures to be discriminated (Gurnsey and Laundry, 1992; Wolfson and Landy, 1998). An abrupt texture gradient, therefore, is not a necessary condition for the discrimination of certain textures. These facts suggest that there is a spatially localized classification process involved in texture discrimination when the textures differ in their overall feature patterns (Grossberg and Williamson, 1999; Gurnsey and Laundry, 1992). Using textures comprising oriented line segments, Wolfson and Landy (1998) observed that better discrimination performance is achieved when two textures with large differences in their mean orientations are abutted and have a sharp transition compared to being spatially separated by blank space. When the same type of textures differ in the standard deviation of line segment orientations, but have the same
mean orientation, their discriminability remains unaffected by spatial separation or abutment. The importance of abutment when textures differ in mean orientation, but not when they differ in the standard deviation with the same mean, furthers the case for local classification-based surface processing in addition to boundary-based processing. In summary, boundary-based processing helps to detect sharp transitions in feature distribution, while classification processes are sensitive to local activity patterns of oriented filters (Grossberg and Williamson, 1999; Gurnsey and Laundy, 1992; Julesz, 1986; Rao and Lohse, 1996; Wolfson and Landy, 1998).

Another important set of data that has guided the development of dARTEX demonstrates contextual modulation of visual cortical cells whose receptive fields lie on or inside figure boundaries. For example, the relative orientation of texture elements outside a V1 cell’s receptive field can suppress or enhance its firing rate (Kapadia, Ito, Gilbert, and Westheimer, 1995; Knierim and van Essen, 1992; Sillito, Grieve, Jones, Cudeiro, and Davis, 1995; Zipser, Lamme, and Schiller, 1996). Some oriented cells in V1 show sensitivity to texture boundaries in their later responses, at around 80–100 ms after stimulus onset (Lee, Mumford, Romero, and Lamme, 1998; Nothdurft, Gallant, and van Essen, 2000). For example, the later responses of V1 cells with vertical preferred orientation are enhanced by a vertical figure boundary even when the figure and ground are made of mutually perpendicular, obliquely oriented texture features (Lee et al., 1998; also see Fig. 4 for a dARTEX simulation). Such modulation due to the presence of texture boundaries has been implicated as a possible mechanism of figure-ground segregation (Lee et al., 1998; Roelfsema, Lamme, Spekreijse, and Bosch, 2002). These contextual modulations in activity occurring around 80–100 ms after stimulus presentation have been linked to top–down modulatory influences (Hupé et al., 1998; Lamme, Supér, and Spekreijse, 1998; Roelfsema et al., 2002).

A combination of figural boundary grouping and boundary-based attentional feedback in dARTEX gives rise to modulations of V1 activities at texture-defined figure boundaries. In dARTEX, boundary-based attention acts through modulatory corticogeniculate feedback (Gove, Grossberg, and Mingolla, 1995) and enhances LGN activities at figure boundaries while suppressing their surround. Fig. 1 shows the model connection from boundary grouping stages in V2 to the LGN via V1. While modulatory corticocortical feedback from V2 to V1 may also enhance activities of oriented V1 cells at grouped image locations (Bullier, Hupé, James, and Girard, 1996; Grossberg, 2003), corticogeniculate feedback is the principal connection in dARTEX that links to the observations by Lee et al. (1998).

Primary visual cortical cell responses are also enhanced when their receptive fields lie in the interior of a textured figure (Lamme, 1995; Zipser et al., 1996). In dARTEX, this is accomplished by form-fitted spatial attentional shrouds which gate V1 simple cell responses.

3. The dARTEX model

dARTEX embodies two processing phases. Initial processing involves fast feed-forward activation of most of the dARTEX stages. In the second phase, surface-based spatial attentional shrouds facilitate dART learning of texture features by selecting surface regions with similar texture features, and thereby enhances the accuracy of texture classification. dARTEX operations are illustrated using the input image in Fig. 4a. Additional computational characteristics of dARTEX will be discussed using other simulations. In particular, simulations of the Ben-Shahar and Zucker (2004) experiment will be discussed in Section 4 after defining relevant dARTEX processes.

Mathematical equations of dARTEX are provided in the Appendix. In order to facilitate cross-referencing, the Appendix equation number pertinent to the text is provided in parentheses. Each dARTEX stage in the model diagram of Fig. 1 is also labeled with the corresponding equation variable in the Appendix.

3.1. Center-surround network, LGN

Retinal and Lateral Geniculate Nucleus (LGN) processing are lumped together for simplicity. As seen in Fig. 1, the model LGN receives a bottom–up retinal input image as well as a top–down, corticogeniculate, boundary-based, modulatory attentional feedback signal. LGN output signals input to simple cells in V1 layers 6 and 4 as well as to a surface filling-in process in V4.

The LGN contains ON cells and OFF cells. ON cells obey membrane, or shunting, equations and interact via on-center, off-surround interactions (Eq. (A.4)). This competition discounts the illuminant in the scene and normalizes local image contrasts. The center-surround operations are carried out at three equally spaced scales of spatial interaction. Cells in the complementary OFF channel interact via an off-center, on-surround network. Both ON and OFF channels provide bottom–up inputs to the simple cells in V1. Grossberg, Mingolla, and Williamson (1995) showed how such preprocessing helps to process complex imagery. See Roska, Molnár, and Werblin (2006) for consistent data about ON and OFF retinal processing.

As seen in Fig. 1, boundary-based corticogeniculate attentional feedback (Section 3.6) acts on LGN cells (Eq. (A.4)). This feedback is derived from boundary grouping cells in V2 layer 2/3. It selectively enhances LGN activities at texture boundaries through an on-center kernel while suppressing features in the boundary neighborhoods through an off-surround kernel. This boundary-based attentional signal enhances LGN activities at the texture boundaries; see Fig. 4d.

3.2. Oriented simple cells, V1 layers 6 and 4

Two sets of oriented simple cells, one each in V1 layers 6 and 4, serve as oriented filters of the image. The simple cells
Fig. 4. dARTEX simulations on an example input image. For panels g-l, darker pixels denote higher activity. Insets in panels b, c, e, and f are magnified versions of the grayed regions. (a) Input. (b) Output of the two competitive stages of V2, without boundary-based attention modulating the LGN activities. (c) Bipole cell outputs without boundary-based attention. Boundaries of the square figure are very weak for the appropriate orientation. (d) LGN stage activities with boundary-based attention amplifying boundary elements through corticogeniculate feedback. (e) V2 competitive stage outputs with boundary-based attention. The boundaries are further amplified compared to those in panel b and horizontal and vertical orientations show improved activity at the figure boundaries. (f) Bipole stage robustly group the vertical and horizontal square figure boundaries. (g) Featural signal to the surface filling-in domain at equilibrium. This signal is a combination of bottom-up arriving LGN activity and top-down volitionally selected surface attention. (h) Surface filling-in activity added over all scales at equilibrium. Features in panel g are contained by boundaries in panel f during filling-in. (i) Attentional shroud at the surface-attention level. The dark spot in the middle of the shroud is the top-down volitional spatial attention signal. Long-range spatial competition inhibits all surface activity except in the attended surface. This signal gates simple cell activities in V1 during feature learning and classification. (j) Featural signal at equilibrium after volitional attention shifts to the background. (k) Corresponding featural filling-in activity at equilibrium. (l) Attentional shroud at the surface-attention level. The dark spot in the background region is the volitional spatial attentional signal.
sample direct bottom–up activity from the LGN ON and OFF channels (Fig. 1). Each set of simple cells represents 24 orientations for each of the three LGN center-surround scales (Eq. (A.6)).

Layer 6 simple cells are identical to the layer 4 simple cells in terms of their bottom–up receptive fields. However, layer 6 simple cells are also gain-modulated by a spatial attentional shroud; see Eq. (A.10). Spatial attention in dARTEX is hypothesized to act at V1 layer 6 through a corticocortical feedback pathway originating at PPC (Fellman and van Essen, 1991). While this may be a multistage pathway in vivo, dARTEX instantiates this as a direct pathway. Formation, volitional selection, and maintenance of spatial attention in dARTEX is described in Section 3.8.

Spatial attentional shrouds in dARTEX serve two purposes: First, much in the same way as the boundary-based attentional modulation of model LGN, spatial attention maintains layer 6 simple cell activities at attended positions while suppressing activity at unattended positions around the locus of attention; see Eq. (A.10). Suppressing unattended activity prevents mixing of surface features across texture boundaries, resulting in improved texture learning and classification performance; see Section 5. Second, spatial attention modulates texture category learning, such that texture features within the boundary of the attended surface may be learned as part of the same texture category. This property is a homolog for texture learning of the demonstration by Fazl et al. (2005) of how spatial attention binds multiple views of an object into a view-invariant object recognition category.

In dARTEX, dART texture category cells are activated bottom–up by V1 layer 6 simple cells (Fig. 1). Bottom–up texture category activation plays two roles: First, it performs a pixel-by-pixel classification of image texture. Second, dART categories activate feature expectations via the layer 6 to 4 network (Callaway, 1998; Grossberg and Raizada, 2000). Layer 4 simple cell activity matched with this feature expectation (Eq. (A.13)) generates bottom–up output to the complex cell filters in layers 2/3 of V1.

In the initial, feedforward, processing phase of dARTEX, spatial attention is inactive, and layer 6 simple cell activity is identical to layer 4 simple cell activity; see Eq. (A.12). Learning at this time converts the matched layer 4 simple cell activity into a noise-suppressed version of bottom–up simple cell activity; see Eq. (A.13). Image noise reduction obtained by this feature match is discussed in Section 3.9. Activation of subsequent stages of dARTEX gives rise to pre-attentively defined figural boundaries and surfaces. Interaction of these surfaces and spatially localized top–down volition gives rise to a form-fitting spatial attentional shroud; see Sections 3.7 and 3.8. Such a shroud acts on layer 6 simple cells and thereby influences dART classification and learning. This dARTEX property is also discussed in Section 3.9. The dARTEX activation phases of feedforward activation, boundary- and surface-based attentional modulation, and texture learning are summarized in Section 3.10.

3.3. Complex cells, V1 layer 2/3

Complex cells in V1 layer 2/3 combine rectified layer 4 simple cell outputs (Eq. (A.10)) of the same scale but opposite contrast polarities (Eq. (A.14)), resulting in 12 orientations of complex cells for each of the three spatial scales (Ferster and Miller, 2000; Hubel and Wiesel, 1959, 1968). Complex cells act as contrast-polarity insensitive, oriented, multi-scale filters.

3.4. Spatial and orientational competition, V2 layers 6 and 4

Contrast enhancement of complex cell output is accomplished by two stages of competitive interactions in layers 6 and 4 of V2 (Grossberg and Mingolla, 1985a, 1985b; Grossberg and Raizada, 2000). As seen in Fig. 1, the first competitive stage in V2 is a spatial center-surround competition between similarly oriented complex cells (Eq. (A.15)). This stage picks out discontinuities in the spatial activity distribution of similarly oriented complex cells, thereby signaling local texture feature contrast and realizing an end stop operation (Hubel and Wiesel, 1977).

The second, orientational, competition stage in V2 is driven by the outcome of the spatial competition and acts at a smaller spatial scale wherein mutually orthogonal orientations inhibit each other the most (Eq. (A.18)). This competition reduces activity at image locations with ambiguous feature orientations and reduces unoriented noise. The orientational competition stage also realizes an end cut operation (Grossberg and Mingolla, 1985b). End cutting produces hyper-acute responses at the endings of oriented bars by disinhibiting orientations that are perpendicular to those that the bar directly activates. End cutting enhances oriented responses at bar endings and thereby facilitates illusory contour grouping between multiple bars (Gove et al., 1995; Grossberg and Mingolla, 1985b; Ross, Grossberg, and Mingolla, 2000).

Taken together, the two competitive stages induce strong featural activities at texture boundaries while suppressing uniformly oriented texture elements within figure interiors and the background; see Fig. 4b.

These processing stages were initially proposed as part of the Boundary Contour System (BCS) of Grossberg and Mingolla (1985a, 1985b) and have since become a standard component in many texture segmentation algorithms and models of human texture processing. In particular, the processing stages of oriented filtering, pointwise rectification, and spatial competition processes in Sections 3.2–3.4 comprise the Filter-Rectification-Filter (FRF) model of texture segregation (Bergen and Landy, 1991; Bovik et al., 1990; Graham et al., 1992; Greenspan et al., 1994; Jain and Farrokhnia, 1991; Sutter et al., 1989).
3.5. Bipole grouping cells, V2 layer 2/3

Long-range interactions among the pyramidal cells in layers 2/3 of V2 (Eq. (A.21)) carry out perceptual grouping by forming and completing oriented figural boundaries in responses to inputs from the orientational competition. Orientationally pooled boundaries, in turn, input to the surface filling-in stage in V4 and also realize boundary-based attentional feedback to the LGN; see Fig. 1.

The recurrent interactions among the layer 2/3 pyramidal cells realize the bipole property, whereby cells that are (approximately) collinear and coaxial with respect to one another across space can excite each other via long-range connections. These long-range connections also activate interneurons that inhibit each other and nearby pyramidal cells via short-range disynaptic inhibition (Eq. (A.29)). The balanced excitation and inhibition at target cells helps to implement the bipole property: (1) A cell can fire when it receives strong bottom–up excitatory input from the orientational competition. (2) When two (almost) collinearly aligned inducing stimuli are present, one on each flank of the cell, a boundary grouping can form even without direct bottom–up input. This is because the inhibitory interneurons, apart from inhibiting the bipole cell, also inhibit each other, thus normalizing the total amount of inhibition emanating from the interneuron pool. Summating excitation from both flanks combined with normalizing inhibition create a case of “two-against-one”, so the target cell is excited above threshold. In contrast, excitation from only one of the horizontal flanks creates a case of “one-against-one” wherein the cell gets commensurate amounts of excitation from the long-range excitatory connections and inhibition from the shared pool of inhibitory interneurons. The bipole cell is then not excited above threshold. As a result, boundaries form inwardly between pairs or greater numbers of active cells, but not outwardly. Bipole cells that are sensitive to different orientations also compete (Eq. (A.28)), thereby sharpening their responses for well-grouped orientations.

Bipole grouping is sensitive to the degree of collinearity and relative distances of features; see Eq. (A.23). This sensitivity is needed to explain OBTS data, as described in Section 4.

Bipole grouping was proposed by Grossberg (1984) and simulated by Grossberg and Mingolla (1985a, 1985b). It has since helped to explain a variety of perceptual grouping percepts (e.g., Gove et al., 1995; Grossberg and Howe, 2003; Grossberg and Swaminathan, 2004; Ross et al., 2000). Psychophysical evidence for perceptual grouping that obeys the bipole property has been reported by several researchers (Field, Hayes, and Hess, 1993; Kellman, 2003; Kellman and Shipley, 1991). Cells with long-range monosynaptic, recurrent, excitatory connections and disynaptic, short-range, inhibitory connections occur in the cortical area V1 (Hirsch and Gilbert, 1991; McGuire, Gilbert, Rivlin, and Wiesel, 1991). The bipole property has been reported in physiological recordings from cells in cortical area V2 (von der Heydt and Peterhans, 1989; von der Heydt, Peterhans, and Baumgartner, 1984). Long-range statistical correlations in co-aligned orientations have been observed in natural scenes, which provides an ecological stimulus for the development of such long-range grouping kernels (Grossberg and Williamson, 2001; Sigman, Cecchi, Gilbert, and Magnasco, 2001).

3.6. Boundary-based attention, from V2 to LGN

In many textured scenes, contrast may not exist at the orientation of a form boundary. For example, although emergent square boundaries in Fig. 4a are perceived to be vertical and horizontal, there is neither luminance nor horizontal and vertical orientation contrast at the boundaries. The end cut mechanism (Section 3.4) disinhibits horizontal and vertical orientations at the ends of the oblique texture bars at the form boundaries (Fig. 4b). Bipole grouping of these vertical and horizontal orientations result in figure boundaries in the appropriate positions and orientations. However, these emergent square figure boundaries are weak compared to the strong oblique orientation contrasts which induced them; see Fig. 4c.

Boundary attention can generate strong vertical and horizontal boundaries. Boundary attention is derived by pooling bipole cell activities over all orientations and scales (Eq. (A.30)). As shown in Fig. 1, it modulates the LGN via V1 (Callaway, 1998; Grossberg, 1976; Guillery, 1967; Murphy and Sillito, 1987; Przybyszewski, Gaska, Foote, and Pollen, 2000; Sillito and Jones, 2002; Sillito, Jones, Gerstein, and West, 1994). Because LGN cells are essentially unoriented, the enhanced LGN feature activity at the form boundaries preferentially activates vertically and horizontally oriented cells, as in Fig. 4c. After orientational competition, the bipole cells group and amplify these signals along the emergent figure boundaries, as seen in Fig. 4f.

Because corticogeniculate feedback in dARTEX is modulatory, it amplifies grouped features without activating new image features (Gove et al., 1995). Such a modulatory feedback process was predicted as a property of ART in Grossberg (1976, 1980). It has been used to explain various perceptual effects (Gove et al., 1995; Grossberg and Grunewald, 2002; Grossberg and Raizada, 2000; Raizada and Grossberg, 2001), including those of Sillito et al. (1994), which ART predicted (Grossberg, 1976).

3.7. Surface filling-in domain, V4

As described in Section 3.1 through Section 3.6, dARTEX computations pre-attentively and automatically define emergent figure boundaries using discontinuities in local feature distributions. A surface filling-in process in V4 (among other cortical areas; see Cao and Grossberg, 2005; Fang and Grossberg, 2005; Grossberg, 1994) is controlled by these pre-attentive boundaries.

Figure boundaries signaled by the bipole cells in V2 layers 2/3 project to surface filling-in domains, where they
gate the filling-in of surface feature signals arriving from the LGN (Cohen and Grossberg, 1984; Grossberg and Mingolla, 1985a; Grossberg and Todorović, 1988); see Fig. 1 and Eq. (A.31). The boundary-gated spread of these surface features tends to generate uniform filled-in activity levels within each boundary compartment. Fig. 4 illustrates this filling-in process, with figure boundaries in Fig. 4f, surface feature inputs in Fig. 4g and j, and filled-in activities in Fig. 4h and k, respectively.

Surface filling-in at each spatial scale contributes in a different way toward the combined filled-in activity. At the smallest scale, boundaries enclose filling-in of individual texture bars. At larger scales, filling-in is contained primarily by object form boundaries, leading to uniformly filled-in activity within each object’s surface. Surface filling-in over all spatial scales therefore supports individual bars as well as object form; see Fig. 4h and k.

3.8. Spatial attention shrouds, PPC

Attentional shrouds arise within dARTEX from feedback between the surface filling-in process in V4 and the spatial attention process in posterior parietal cortex, or PPC (Colby and Goldberg, 1999; Deubel and Schneider, 1996); see Fig. 1. Spatial attention (Eq. (A.33)) receives filled-in surface activities from V4 filling-in and projects back to V4. At the spatial attention stage, pre-attentively filled-in surfaces in a scene bid for attention. Depending on task demands, volitional top-down spatial attention (a “spotlight” of attention; Posner, 1980) may also select a location of interest by locally enhancing its activity (see term $I_{mo}$ in Eq. (A.33)). Due to the back projection from spatial attention to surface filling-in (Eq. (A.31), activity of the volitionally selected surface gets enhanced (Reynolds & Desimone, 2003; Reynolds, Pasternak, & Desimone, 2000; Tse, 2005) through boundary-gated spread of the volitional enhancement signal throughout the bounded surface. Enhanced filled-in surface activity feeds forward to the spatial attention stage, further enhancing an entire region that fits the attended surface form (cf., Cavanagh, Labianca, & Thornton, 2001; Pylyshyn, 1989; Tyler & Kontsevich, 1995). This form-fitting distribution of spatial attention is an attentional shroud. At the same time, other regions at the surface attention level are suppressed by the winning attentional shroud by long-range spatial competition (Carpenter & Grossberg, 1991; Desimone, 1998; Grossberg, 1994; Reynolds, Chelazzi, & Desimone, 1999).

Simulations in Fig. 4 illustrate shroud formation. Fig. 4g shows the total input to the filling-in domain at equilibrium, which combines bottom-up LGN input (Eq. (A.4)) and top-down spatial attention input (Eq. (A.31)). The bipole activities (Fig. 4f) are boundaries that gate the filling-in process. Boundary-gated filling-in leads to the square surface shown in Fig. 4h. Spatial attention is strongly activated at the winning surface and inhibited elsewhere (Fig. 4i). The dark spot close to the middle of the square surface in Fig. 4g and i is the top-down volitional attention signal (Eq. (A.33)) that biases the competition in favor of the attended surface.

Fig. 4j–l shows how a shift in the volitional signal triggers a shift of form-fitting spatial attention to the newly selected surface. In the surface filling-in domain, boundary-gated diffusion of the shifted focal volitional attention increases filled-in activity in the figure background. Through long-range spatial competition, activity in the previously attended form is eliminated and the surviving spatial attentional signal fits the entire background. A spatially localized volitional attention can hereby determine which form-fitting locus becomes the spatial attentional shroud. In the absence of a volitional attentional spotlight, a number of other factors, including surface size, boundary strength, and feature intensity automatically determine the winning form-fitting shroud.

A number of investigators have reported that spatial attention can mold itself to an object’s shape. Our simulations illustrate how this can happen through feedback between surface filling-in and spatial attention. Our simulation of how an attentional shroud (Tyler & Kontsevich, 1995) forms adapts the process described in Fazl et al. (2005), who showed how such shrouds can be used to regulate the learning of view-invariant object categories from multiple object views. Other investigators have reported data showing how the allocation of spatial attention can depend upon the spatial location of objects (Connor, Predie, Gallant, & van Essen, 1997; Connor, Gallant, Predie, & van Essen, 1996; Johnson & Yantis, 1995; Posner, 1980; Pylyshyn & Storm, 1988; Sagi & Julesz, 1986; Yeshurun & Carrasco, 2000), specific stimulus features (Cavanagh, 1992; Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1990; Wolfe, Cave, & Franzel, 1989), and the salience of the visual object (Blaser, Pylyshyn, & Holcombe, 2000; Duncan, 1984; Mitchell, Stoner, Fallah, & Reynolds, 2003; O’Craven, Downing, & Kanwisher, 1999; Treisman & Paterson, 1984). Both Fazl et al. (2005) and the present work illustrate how spatial attention can regulate category learning and, during that learning process, the allocation of object attention (see Section 3.9).

3.9. dART-based texture learning, classification, and top-down attentive matching

The dART network in the model is driven by spatially pooled V1 layer 6 simple cell activities (Fig. 1, Eq. (A.34)). The dART network learns to assign a texture category to each image pixel. Carpenter (1997) introduced the dART model to explain how fast, incremental, and stable category learning could be achieved while allowing for distributed category activation. Distributed category representation has desirable category generalization and compression characteristics (Carpenter et al., 1998).

The spatial distribution of feature activity in natural images is typically variable, even for a seemingly homogeneous texture. During bottom-up texture category learning by the dART network, top-down modulatory expectations
(Eq. (A.35)) are also learned within adaptive connections from layer 6 to 4 (Callaway, 1998; Grossberg & Raizada, 2000; Salin & Bullier, 1995; Wiser & Callaway, 1996) and matched against layer 4 simple cells (Fig. 1, Eq. (A.13)). This match operation is an interface between pre-attentive texture processing and attentive local texture classification (Grossberg, 1999a, 2003).

According to the distributed ART match rule (Carpenter, 1997, 2001; Carpenter et al., 1998), the matched feature is the smaller of the layer 4 simple cell activity and the top–down expectation (Eq. (A.13)). This match operation suppresses noisy pixels that deviate significantly from prototypical texture features, leading to feature activity due to texture features rather than noisy deviations of them. By eliminating such noise, texture category matching facilitates completion of a boundary grouping around the texture-induced form, and thus the formation of a form-fitting attentional shroud. The shroud, in turn, further facilitates texture category learning by eliminating texture exemplars that belong to different forms; cf., Eq. (A.10).

Fig. 5a depicts a noisy textured image. Due to the high variability in bar orientation, the competitive stages of V2 have a hard time in detecting feature discontinuities (Fig. 5b). Using the matched simple cells, on the other hand, the competitive stages can suppress the interior features and pick out the boundaries successfully (Fig. 5c). The bipole grouping stage can now signal the figure boundaries correctly (Fig. 5d).

Fig. 1 shows the top–down spatial attention pathway that gates inputs to the dART network in V1 layer 6. During dART learning, the attentional shroud enables learning of attended surface features. During supervised dART learning, input feature vectors (Eq. (A.34)) at attended image pixels are associated with a class name, or label. This class label may be externally supplied, or automatically generated based on task context. Benchmark simulations in Section 5 generate class labels using image presentation sequence number. The dART learning algorithm is described in Section B.1.

3.10. Summary of dARTEX operation modes

During initial bottom–up activation, layer 4 match activity is the noise-suppressed simple cell activity over the whole scene (Sections 3.2 and 3.9). Bottom–up activation of subsequent dARTEX stages through these matched simple cells gives rise to a form-fitting spatial attentional shroud (Section 3.8). Activation of the shroud initiates attentive texture classification, whereby spatial attention suppresses layer 6 simple cell activity at unattended positions surrounding the attended form. This suppression prevents feature mixing across texture boundaries during classification and improves model performance. Persistent activation of the attentional shroud is maintained in a (surface)–(spatial attention) resonance during which features within the resonant surface are learned as the same texture (Section 3.9). During supervised learning, a texture class label is also supplied at attended spatial locations during this resonance.

A large mismatch between bottom–up layer 4 feature and its top–down expectation may suppress layer 4 match activity at such positions. A big enough mismatch may trigger the activation of another shroud. Volitional top–down surface attention signals, in the absence of bottom–up inputs, may sensitize, or prime dART categories, but cannot create suprathreshold activity in layer 4. However, such top–down priming may prepare a category cell to react more quickly and vigorously to subsequent bottom–up input that matches the top–down prime, resulting in rapid recognition of the primed feature.

4. Simulation of orientation-based texture segmentation experiments

The Orientation-Based Texture Segmentation (OBTS) experiments of Ben-Shahar and Zucker (2004) and their simulation by dARTEX are described in this section.
4.1. OBTS experiments of Ben-Shahar and Zucker (2004)

Ben-Shahar and Zucker (2004) used a two alternative, forced choice experiment (2AFC) to investigate the differences in OBTS due to texture element configuration at the boundaries and the role of image orientation gradients. Illustrative experimental stimuli are shown in Fig. 2. The stimulus images were made of a regularly spaced tiling of $21 \times 21$ bright elongated bars on a dark background. Orientation of the texture bars varied at a constant rate in the vertical direction, except at the wedge-shaped form boundaries. That is, from top to bottom of the image, each left-to-right raster of oriented bars had the same orientation and each such raster differed from the next by a constant angle. The image orientation gradient, which defines the rate of change of orientation at each image location and denoted by $\nabla \theta$, was thus constant and vertically oriented. An abrupt, larger than average change in orientation, or equivalently a discontinuity in the image orientation gradient, led to an emergent figure boundary. Wedge shaped emergent figures were generated through such discontinuities in image orientation gradient. While the horizontal and diagonal limbs of the wedge shape were due to orientation discontinuity, the vertical limb of the wedge form was due to the image boundary itself. Fig. 2 shows examples of left-pointing wedge shapes. Stimulus images spanned 10° of visual angle.

Subjects had to respond, using a button-press, whether the wedge-shaped figures were left-pointing or right-pointing. The only difference between these two alternatives was the location of horizontal discontinuity line which appeared either 2.5° above or below the center of the image. Since the diagonal discontinuity line was the same, either a left-pointing or a right-pointing wedge emerged, depending on the location of the horizontal boundary. Stimulus images were presented for 200 ms, preceded and followed by masks of randomly oriented bars. Performance accuracy of 75% or more was considered statistically significant.

Ben-Shahar and Zucker (2004) hypothesized that the diagonal discontinuity line in the image did not affect the subjects’ decision process because it maintained the same bar configuration regardless of the location of the horizontal boundary. The two possible responses were therefore due to the location of the horizontal boundary, being either above (for a right-pointing wedge) or below (for a left-pointing wedge) the fixation point. All subjects were pretrained, and were aware that the discontinuity line determining the response was always horizontally oriented and could occur at exactly one of the two possible locations.

Ben-Shahar and Zucker (2004) varied two parameters to produce the stimulus set. The first parameter was based on the orientation gradient. Since direction of the gradient was always vertical, the only attribute available for variation was its magnitude. The orientation gradient magnitude, denoted by $\Delta \theta_{\text{within}}$, was parametrically varied from 5° to 30° per degree of visual angle, in steps of 5°. Task difficulty increased with an increase in the angular gradient. The wedge shape in Fig. 2c (with $\Delta \theta_{\text{within}} = 5^\circ$) is easier to detect than in Fig. 2f (with $\Delta \theta_{\text{within}} = 15^\circ$).

The second parameter was the amount of angular discontinuity at the boundaries, denoted by $\Delta \theta_{\text{between}}$. The parameter $\Delta \theta_{\text{between}}$ varied from 5° to 90° in steps of 5°. Task difficulty decreased with an increase in this parameter; the wedge shape in Fig. 2a is harder to detect (with $\Delta \theta_{\text{between}} = 30^\circ$) than in Fig. 2c (with $\Delta \theta_{\text{between}} = 90^\circ$).

The relation between local bar orientation $\theta$ and image orientation gradient vector $\nabla \theta$ gives rise to two scalar fields, or curvatures, defined over the entire image. These curvatures measure the initial rates of change of orientation in directions tangential (the tangential curvature $k_T$) and normal (the normal curvature $k_N$) to bar orientation at each image location. Specifically, the two curvatures are (Ben-Shahar & Zucker, 2004):

$$k_T = \nabla \theta \cdot (\cos(\theta), \sin(\theta)),$$

$$k_N = \nabla \theta \cdot (-\sin(\theta), \cos(\theta)),$$

where $\nabla \theta$ is the image orientation gradient vector, and $\theta$ is the bar orientation. Operator $\cdot \mathbb{E}$ denotes a vector inner product.

As seen in Eq. (1), texture bar orientation and the two curvatures are intimately linked through the image orientation gradient vector. Furthermore, the two curvatures are co-variant: the value of one completely specifies the other. Specifically, $\Delta \theta_{\text{within}} = \| \nabla \theta \| = \sqrt{k_T^2 + k_N^2}$ (Ben-Shahar & Zucker, 2004). With the image orientation gradient in the vertical direction, the normal curvature $k_N$ is maximal and $k_T$ is zero wherever texture bars are horizontal; see Eq. (1). Similarly, the tangential curvature $k_T$ is maximal and $k_N$ is zero wherever texture bars are vertical. The orientation gradient discontinuity in the image causes a corresponding discontinuity in the tangential and normal curvatures, denoted by $\Delta k_T$ and $\Delta k_N$, respectively. Furthermore, different relative values of $\Delta k_T$ and $\Delta k_N$ give rise to different bar configurations at the texture boundary. The main goal of the Ben-Shahar and Zucker (2004) study was to relate the curvature discontinuities to OBTS. A complete description of the differential geometric interpretation of orientation defined textures can be found in Ben-Shahar and Zucker (2004). In what follows, a qualitative description is provided of the two curvature discontinuities and resultant bar configurations at the horizontal texture boundary in the experimental stimuli of Ben-Shahar and Zucker (2004).

Three different curvature discontinuity conditions at the horizontal texture boundary were explored by Ben-Shahar and Zucker (2004). The first combination was defined by $\Delta k_N = 0$, the second by $\Delta k_T = 0$, and the third by $\Delta k_T = \Delta k_N$. In a given trial, the quantities $\nabla \theta$, $\Delta \theta_{\text{between}}$, $\Delta k_N$, and $\Delta k_T$ completely specified the stimulus image, and were kept constant throughout the stimulus presenta-
tion. The following discussion assumes a vertical orientation gradient vector $\nabla \theta$.

In the first curvature discontinuity combination, termed the $\Delta k_T = \max$ condition, $\Delta k_T$ is maximal and $\Delta k_N$ is zero at the horizontal figure boundary. As seen in Eq. (1), the two curvatures are continuous functions of the orientation gradient over the whole image, except at the form boundaries defined by discontinuity in orientation. The only way $\Delta k_N = 0$ condition can be achieved at the horizontal boundary is if $k_N$ remains the same at points equidistant in the vertical direction from the horizontal boundary on its either side. This requirement, along with the continuity of curvatures on either side of the horizontal boundary, requires that $k_N$ pass through a maximum and $k_T$ simultaneously pass through zero. As described earlier, these transitions in $k_N$ and $k_T$ result in horizontal bar orientation at the horizontal figure boundary. Fig. 2a shows an example of this condition, where bars on either side of the horizontal boundary deviate from the horizontal orientation by equal angular amounts $\Delta \theta_{\text{between}}/2$ but in opposite directions. In other words, except for the angular discontinuity, the orientation gradient would have given rise to horizontal bar orientations at the horizontal form boundary. For the $\Delta k_T = \max$ condition, as $\Delta \theta_{\text{between}}$ increases, the bars become increasingly oblique on either side of the boundary.

The second curvature discontinuity condition is defined as $\Delta k_T = \Delta k_N$. For this condition, by a similar argument as above and referring to Eq. (1), the orientation at the horizontal boundary passes through an oblique, or 45$^\circ$, orientation. Fig. 2c shows an example of this condition with $\Delta \theta_{\text{between}} = 90^\circ$. Bars on each side of the horizontal boundary deviate by 45$^\circ$ from the oblique orientation, resulting in bars parallel to the boundary on one side and perpendicular on the other side.

The third curvature discontinuity combination, termed the $\Delta k_N = \max$ condition, is similar to the first except $\Delta k_N$ is maximal and $\Delta k_T = 0$. Bar orientation on either side of the horizontal figure boundary is equidistant in angle from the vertical direction; see the horizontal boundary in Fig. 2b.

Using the curvature discontinuity conditions described above, Ben-Shahar and Zucker (2004) found that the same amount of $\Delta \theta_{\text{between}}$ resulted in different perceived boundary saliencies when the flow of oriented bars was interrupted in different ways. This asymmetry was most pronounced at small within-region orientation gradient magnitudes ($\Delta \theta_{\text{within}} 10^\circ$). For these low gradient amounts, the $\Delta k_T = \max$ condition was salient even for low values of $\Delta \theta_{\text{between}}$; for example, in Fig. 2a the wedge form can be clearly seen with $\Delta \theta_{\text{within}} = 5^\circ$ and $\Delta \theta_{\text{between}} = 30^\circ$. Despite the same amount of boundary discontinuity, however, the wedge form in $\Delta k_N = \max$ condition is much harder to see: Horizontal boundaries in Fig. 2a appear clearer than in Fig. 2b. For small $\Delta \theta_{\text{within}}$ and $\Delta \theta_{\text{between}}$, the $\Delta k_T = \max$ conditions were even more salient than the $\Delta k_T = \Delta k_N$ conditions, but the $\Delta k_T = \Delta k_N$ configurations eventually became stronger with increasing $\Delta \theta_{\text{between}}$.

This can be seen in the data plots of Fig. 3a, where the detection accuracies for the $\Delta k_T = \max$ (the blue curve with boxed points) are higher than any other condition at low values of $\Delta \theta_{\text{between}}$. However, as $\Delta \theta_{\text{between}}$ becomes large, the $\Delta k_T = \Delta k_N$ condition becomes more salient; see the green curve with crossed points in Fig. 3a. As noted earlier in this section, an increase in orientation gradient magnitude $\Delta \theta_{\text{within}}$ is detrimental to successful boundary detection: Note the drop in accuracies as $\Delta \theta_{\text{within}}$ progressively increases in Fig. 3a through 3f. Specifically, salience of the $\Delta k_T = \max$ configurations diminishes for $\Delta \theta_{\text{within}} = 10^\circ$ compared to $\Delta \theta_{\text{within}} = 5^\circ$, and ultimately disappears with increasing orientation gradient magnitude $\Delta \theta_{\text{within}}$; see the blue curves with boxed points in Fig. 3a through 3f. For $\Delta \theta_{\text{within}} = \{15, 20\}$, only the $\Delta k_T = \Delta k_N$ conditions were detectable even for large $\Delta \theta_{\text{between}}$ values (Fig. 3c and d), which also eventually disappears (Fig. 3e and f).

Parallel configurations that arose with $\Delta k_T = \Delta k_N$ and $\Delta \theta_{\text{between}} = 90^\circ$, where bars on one side of the horizontal boundary were horizontal, were found to be more salient than other configurations: Compare horizontal boundary strengths of Fig. 2d–f, where the boundary in Fig. 2f is the strongest. As a second experiment, Ben-Shahar and Zucker (2004) further investigated the relative salience of the $\Delta k_T = \Delta k_N$ conditions to parallel configurations. In this second experiment, $\Delta \theta_{\text{between}}$ was set at 45$^\circ$. The purpose of this experiment was to investigate whether parallel configurations were always salient regardless of the specific values of $\Delta \theta_{\text{within}}$. Ben-Shahar and Zucker (2004) found that, at low angular gradient magnitudes ($\Delta \theta_{\text{within}} 10^\circ$), parallel configurations had higher detection accuracy. However, with higher angular gradients (15$^\circ$ $\Delta \theta_{\text{within}} 20^\circ$), the $\Delta k_T = \Delta k_N$ configurations became more salient, suggesting that the abrupt termination of the "flow" was more salient.

Possible reasons for the salience asymmetries due to boundary configurations and changes in $\Delta \theta_{\text{within}}$ are discussed in Section 3.3 using dARTEX simulations.

### 4.2. Texture boundary contrast measure for the OBTS task simulation

In order to compare model performance to perceptual data, the following boundary contrast measure was used:

$$S = \mu \frac{E_{\text{Lower}} - E_{\text{Upper}}}{v + E_{\text{Lower}} + E_{\text{Upper}} + F} + 50,$$

where $\mu$ and $v$ are constant numbers, and $E_{\text{Lower}}$ and $E_{\text{Upper}}$ are bipole cell activities pooled over horizontal bands centered at each possible boundary location. Factor $F$ in the above equation is average boundary activity over the whole image. The contrast measure above is derived from the equilibrium solution of a shunting on-center, off-surround
network (Grossberg, 1973, 1980). The contrast measure hypothesizes that subjects divided their spatial attention to locations around the two possible horizontal discontinuity locations that determine the responses, and chose the location with greater activity. The probability of a correct response is thus hypothesized to be proportional to the contrast between boundary activities at the two attended spatial locations. Section A.10 mathematically defines the terms in this salience measure.

Two methods of computing the boundary activities $E_{\text{Lower}}$ and $E_{\text{Upper}}$ in the boundary contrast measure are compared. The two methods differ in the manner in which bipole activities are pooled across orientations within the upper and lower possible boundary regions. The first pooling method uses a spatial attentional gain (Johnson and Yantis, 1995; Posner, 1980; Pylyshyn and Storm, 1988; Shaw, 1982) where total boundary grouping activity, summed over all orientations, is compared between two possible boundary regions (Eq. (2)). This pooling concept mathematizes the fact that the subjects knew that the correct response could be determined completely by the location of the horizontal discontinuity line, and that the line could appear at exactly one of the two possible locations in the scene.

The second method differentially weighs boundary orientations in addition to spatial locations (Cavanagh, 1992; Corbetta et al., 1990; Wolfe et al., 1989). The horizontal orientation is given the most weight, with a Gaussian fall-off for nearby orientations (Eq. (A.39)). This boundary contrast measure incorporates the fact that the response-determining boundaries are always horizontal, and assumes that subjects paid attention to the two possible regions and looked for a horizontal line within them. Wolfson and Landy (1995) used a similar scheme of assigning more weight to a specific orientation of their second-order filters to explain the increased salience of parallel-perpendicular bar configurations at texture boundaries.

4.3. Simulation of orientation-based texture segmentation data

dARTEX simulations clarify the role of boundary grouping and attentional amplification of well-grouped boundaries in Orientation-Based Texture Segmentation (OBTS). For the OBTS simulations, the volitional spatial attention ($I_{pq}$ in Eq. (A.33)) was kept fixed at the image center, where the experimental subjects were instructed to fixate. In the interest of conserving computer simulation time, the dART network was kept untrained. We have verified that an appropriately trained dART network does not alter the simulation outputs for $\Delta \theta_{\text{within}} = 5^\circ$.

Fig. 2 shows simulation outputs of some experimental stimuli. Fig. 2a shows a stimulus with $\Delta k_T = \text{max}$, $\Delta \theta_{\text{within}} = 5^\circ$, and $\Delta \theta_{\text{between}} = 30^\circ$. In this case, texture elements are roughly aligned with the horizontal texture boundary and the orientation discontinuity detected by the competitive stages in V2 is readily grouped by the horizontally oriented bipole cells. The automatic boundary attention to the LGN (Section 3.6) further amplifies these boundaries while suppressing their neighboring elements by acting on the surround kernels in the LGN. As a result of amplifying boundary positions at the LGN, the activities of simple cell filters at the boundaries are higher compared to their neighborhood. This enhanced contrast in the simple and complex cell activities supplements the orientation contrast, resulting in further amplification of the boundary. Thus, grouping and feedback produce a strong boundary even for low values of $\Delta \theta_{\text{between}}$. As seen in Fig. 6b and c, regardless of the method of boundary activity pooling, strong boundary contrast develops for the $\Delta k_T = \text{max}$ configuration, which is plotted with boxed points and blue curves.

Fig. 2b shows a stimulus with the same orientation parameters ($\Delta \theta_{\text{within}} = 5^\circ$, $\Delta \theta_{\text{between}} = 30^\circ$) as in the foregoing discussion, but with $\Delta k_N = \text{max}$. For this case, the texture bars are almost orthogonal to the boundary on both sides. This configuration results in very weak grouping in the horizontal bipole orientation. Furthermore, bipole grouping in the vertical orientation spreads the activity orthogonal to and away from the texture boundary, resulting in a thick band of increased activity around the texture boundary at the LGN. Neither the center-surround filters in the LGN, nor the spatial competitive stages in V2, can sharpen this diffuse boundary activity. As a result, the total bipole activity at the horizontal figure boundary is not very different from the average activity over the whole image. Extra weight to the horizontal bipole orientations in the boundary contrast measure only decreases the weighted bipole activity at the boundary, resulting in a further deterioration of boundary salience: Compare data points in Fig. 6b and c for $\Delta \theta_{\text{between}} = 30^\circ$ and $\Delta k_N = \text{max}$ condition, plotted with circled points and red curves.

Fig. 2c shows a stimulus with very strong segmentation. In this case, with $\Delta k_T = \Delta k_N$, the orientation gradient is low ($\Delta \theta_{\text{within}} = 5^\circ$), and the boundary discontinuity amount is maximal ($\Delta \theta_{\text{between}} = 90^\circ$). Here, orientation contrast is readily detected by the competitive stages in V2, and the co-aligned boundary features are strongly grouped by the horizontally oriented bipole cells. Furthermore, feedback modulation of the LGN increases the activities at the horizontal boundary. In fact, of all the examples simulated with $\Delta \theta_{\text{within}} = 5^\circ$, this configuration yielded the strongest boundaries: Compare dARTEX simulation outputs of the three conditions for $\Delta \theta_{\text{between}} = 90^\circ$ in Fig. 6b.

The grouping and boundary-based attentional processes described above also help to explain the rest of the outputs in Fig. 6b. As described in Section 3.1, the $\Delta k_T = \text{max}$ condition bars start out parallel to the horizontal boundary and increasingly become oblique on both sides of the boundary as $\Delta \theta_{\text{between}}$ increases. On the other hand, the $\Delta k_N = \text{max}$ configurations start out roughly oblique to the boundaries and gradually become parallel to the boundary on one side as $\Delta \theta_{\text{between}}$ increases. Lastly, the $\Delta k_N = \text{max}$ configurations start out with bars perpendicular-
lar to the horizontal boundary on both sides and gradually assume the same oblique orientation on both sides of the boundary as the $\Delta k_T = \text{max}$ case.

The $\Delta k_T = \text{max}$ configurations, therefore, initially receive a boost in the boundary strength through grouping, but this advantage diminishes with increasing $\Delta \theta_{\text{between}}$ as the horizontal filter activity gradually decreases. However, with increasing $\Delta \theta_{\text{between}}$, boundary contrast, and therefore the boundary-based attentional feedback to the LGN, increase in magnitude, which amplifies the end cuts (see Section 3.4), thus retaining the horizontal boundary grouping. For the $\Delta k_T = \Delta k_N$ configurations, boundary elements become increasingly parallel to the horizontal boundary with increasing $\Delta \theta_{\text{between}}$, greatly increasing the bipole grouping and feedback enhancement of LGN features. Saliency curves therefore cross in Fig. 6b, where $\Delta k_T = \Delta k_N$ configuration saliencies surpass $\Delta k_T = \text{max}$ at higher $\Delta \theta_{\text{between}}$ values. Lastly, while saliencies for $\Delta k_N = \text{max}$ configurations do increase with increasing $\Delta \theta_{\text{between}}$, they are never strong enough to surpass either of the other configurations due to the lack of grouping and feedback processes working in concert as described earlier.

The complete set of OBTS simulations is shown in Fig. 3. As image orientation gradient magnitude $\Delta \theta_{\text{within}}$ increases, activities of the oriented filters of a given orientation become increasingly non-uniform over the whole image. As a result, the spatial competitive stages in V2 begin to fail to suppress activities in the region interiors, and the boundary contrast diminishes rapidly. With increasing $\Delta \theta_{\text{within}}$, boundary activity contrast for all configurations decreases regardless of the amount of $\Delta \theta_{\text{between}}$, and the overall image “clutter” increases (see Fig. 2d–f), where background activity is higher with $\Delta \theta_{\text{within}} = 15^\circ$ than in Fig. 2a–c with $\Delta \theta_{\text{within}} = 5^\circ$. The boundary contrast measure in Eq. (2) is penalized by this image clutter through a long-range inhibitory term $F$, denoting average boundary activity in the image. Nevertheless, grouping and feedback modulation continue to play a role, and the $\Delta k_T = \Delta k_N$ configurations for higher values of $\Delta \theta_{\text{between}}$ where bars are roughly parallel to the horizontal boundary,
do eventually surpass detection thresholds for $\Delta \theta_{\text{within}}$ $20^\circ$.

To control for model parameter overfitting, dARTEX parameters and the salience measure were tuned using stimuli with $\Delta \theta_{\text{within}} = \{5,15,25\}$. The goodness of data fit can be inferred in the $\Delta \theta_{\text{within}} = \{10,20,30\}$ simulations.

Table 1 enumerates dARTEX parameters that required tuning for a good data fit and qualitatively describes their effect on model behavior. Computational cost is prohibitively high for an exhaustive mechanical search for optimal parameter values. We therefore started with parameter values that resulted in a reasonable model output on the input image in Fig. 4a. Subsequent data fit optimization involved iteratively tuning cell output thresholds, spatial scale and passive decay rate of V1 simple cells, intensity and range of V2 spatial and orientational competition, and spatial scale of long-range bipole grouping. Achieving a balance between local competition and long-range co-operation among bipole cells was one of the most important optimizations. Intensity of the boundary feedback to LGN also plays an important role in the data fit. The salience measure parameters were also iteratively optimized to achieve the quantitative data fit. Model parameter values obtained through the iterative process are generally in line with other models from which dARTEX draws inspiration, including those of Gove et al. (1995), Grossberg and Raizada (2000), and Ross et al. (2000).

Certain aspects of the Ben-Shahar and Zucker (2004) data are not accurately captured by the model. For example, the salience measure of Eq. (2) tends to overestimate the $\Delta k_N = \max$ accuracies at small values of $\Delta \theta_{\text{between}}$, while underestimating the $\Delta k_T = \max$ accuracies for $\Delta \theta_{\text{between}} = 5^\circ$. This is seen in model outputs for $\Delta \theta_{\text{within}} = 10^\circ$ (Fig. 3b). This discrepancy can be attributed to the salience measure. For $\Delta k_N = \max$ cases, co-aligned texture bars occur at a horizontal boundary location opposite the true location. Fig. 7a and b shows examples of such co-alignment where the “true” horizontal discontinuity line is at the bottom location and the horizontal alignment occurs close to the top boundary location. This horizontal arrangement results in a strong percept of a “line” at the top location which may have biased the subjects’ judgment in favor of the top boundary location.

The salience measure underestimates human performance for two reasons. The first reason is attributed to a lack of spatial scale selection in the salience measure. The smallest spatial scale in the model signals bar co-alignment. However, spatial competition at medium and large spatial scales is sensitive to feature discontinuity and suppresses uniform feature activity. Since the salience measure assigns the same weight to boundaries at all scales, the resulting boundary salience is lower than assigning maximal weight to the smallest spatial scale. A possible solution for future study involves a flexible scale selection mechanism through competition across scales. See Grossberg et al. (2007) for an example of such inter-scale competition in generating percepts of 3D shape-from-texture.

A second reason for the discrepancy in data fit may be traced to how boundary activities are pooled at the two possible horizontal discontinuity locations. Two pooling kernels are used, each being a horizontal band centered at a possible boundary location with a Gaussian drop-off in the vertical direction. As seen in Fig. 7a and b, bars are co-aligned slightly below the center of the top pooling kernel. The top pooling activity is therefore lower than what it could be, were the pooling kernel centered at the co-aligned bars. A more complex salience measure that slightly shifted the center of the pooling kernel could provide a better data fit. Estimating such a kernel would involve an additional process that automatically shifts spatial attention to figure boundaries of interest. Computation of such a spatial attention shift may involve visual areas V4, parietal cortex, and beyond, modeling of which is beyond the focus of dARTEX. Fazl et al. (2005, submitted for publication) present a model of how this may be accomplished. Future work will synthesize both models.

Fig. 8 shows a scatter plot of model fit using Eq. (2) and the experimental data. Values for Ben-Shahar and Zucker (2004) data were found from published data plots. As seen in Fig. 8a and b, dARTEX fits observed data fairly well when $\Delta \theta_{\text{within}} = 5^\circ$ and $10^\circ$; correlation coefficients for $\Delta \theta_{\text{within}} = 5^\circ$ and $10^\circ$ are 0.95 and 0.91, respectively. However, increasing $\Delta \theta_{\text{within}}$ to $15^\circ$ deteriorates the data fit; the correlation coefficient for $\Delta \theta_{\text{within}} = 15^\circ$ is 0.77. As seen in Fig. 8c, the model performs at chance for $\Delta \theta_{\text{within}} = 15^\circ$, while human subjects’ performance varies significantly around chance performance level. This deteriorates the correlation between model fit and observed data. Sources of such decision bias remain to be investigated.

As in the second experiment of Ben-Shahar and Zucker (2004), the $\Delta k_T = \Delta k_N$ simulations were compared to parallel configurations where bars on the inside of the horizontal boundary were parallel to the boundary. For this comparison, boundary contrasts were computed for $\Delta \theta_{h} = 45^\circ$ and $\Delta \theta_{\text{within}} = \{5,10,15,20\}$. Ben-Shahar and Zucker (2004) observed that the parallel configurations were more salient than the configurations with $\Delta \theta_{\text{within}} = \{5,10\}$ but became less salient with $\Delta \theta_{\text{within}} = \{15,20\}$. The percentage detection accuracy difference between the two conditions in the experiments was $12,12,12,12$ for $\Delta \theta_{\text{within}} = \{5,10,15,20\}$, respectively, with stronger parallel configuration assuming positive values. dARTEX output differences, on the other hand, were $[3.1,1.1,2.1,4.5]$ percent for $\Delta \theta_{\text{within}} = \{5,10,15,20\}$, respectively, matching the observed relative saliencies of the two configurations.

For textures with $\Delta \theta_{\text{within}} = \{15,20\}$, the $\Delta k_T = \Delta k_N$ bar configurations at the boundaries look like Ts. For the same parameter values, the parallel configuration looks like horizontally oriented Vs, with one arm of the V parallel to the boundary. The discrepancies in the exact percentage values between dARTEX output and observed data may be due to the inherent importance of junctions for figure-ground segregation. We suspect additional mechanisms, such as
the angle cells simulated by Grossberg and Swaminathan (2004), may be contributing towards a heightened salience of the T configurations. For the smaller differences in the simulated values between the two configurations for \( \Delta \theta_{\text{within}} = \{5, 10\} \), additional amplification of boundaries due to a V2–V1 modulatory feedback (Grossberg and Raizada, 2000), not implemented in dARTEX, may be responsible.

### 4.4. The role of boundary-based attention and horizontal bipole grouping

The consequences of inactivating either boundary grouping or boundary-based attention are shown in Fig. 6d and e. With the boundary-based attentional amplification of the LGN turned off (Fig. 6d), the obliquely oriented elements at the boundary for high values of \( \Delta \theta_{\text{between}} \) are not contrast-enhanced. This results in much diminished salience at high values of \( \Delta \theta_{\text{between}} \) in the \( \Delta \kappa_T = \max \) configurations.

With the bipole grouping switched off, all curves are generally lower, as seen in Fig. 6e. Furthermore, contrary to the experimental data, the contrast measure in Eq. (2) does not increase for \( \Delta \kappa_T = \max \) or \( \Delta \kappa_N = \max \), as \( \Delta \theta_{\text{between}} \) increases. This is due to the lack of amplification of horizontal boundary orientations by bipole grouping. In the plots of Fig. 6d and e, extra weight was applied to the horizontal orientations in the salience measure, as described in Eq. (A.36). Since grouping was inactive in Fig. 6e, the V2 competition stage outputs, and the vertical axis was rescaled to match the other plots.

With bipole grouping disabled and equal weight given to all orientations while pooling V2 competition stage outputs, the salience symmetries due to boundary configurations disappear, as seen in Fig. 6f. Bipole grouping is a prime driver for producing the salience asymmetries: With

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Functional importance</th>
<th>Too big</th>
<th>Too small</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_a = 0.25 ) (A.4), ( \tau_p = 0.25 ) (A.6), ( \tau_a = 0.05 ) (A.18)</td>
<td>Contrast normalization factors</td>
<td>No normalization</td>
<td>Noise amplification</td>
</tr>
<tr>
<td>( \delta = 2 ) (A.6)</td>
<td>Simple cell elongation</td>
<td>No response to curves</td>
<td>No orientation</td>
</tr>
<tr>
<td>( \theta = 0.01 ) (A.15)</td>
<td>Tonic activity in spatial competition</td>
<td>Deteriorates orientational selectivity</td>
<td>Deteriorates end-cuts (Section 3.4)</td>
</tr>
<tr>
<td>( \alpha_b (s) = [4.5, 9, 13.5] ) (A.16)</td>
<td>Surround extent in spatial competition</td>
<td>Misses small surfaces</td>
<td>Misses boundaries</td>
</tr>
<tr>
<td>( \omega_a = 0.43 ) (A.16)</td>
<td>Extent of similar orientations competing</td>
<td>Lack of orientational contrast</td>
<td>Cannot suppress small, systematic orientational gradients like in ( \Delta \theta_{\text{within}} = 5^\circ ) case</td>
</tr>
<tr>
<td>( \omega_b = 0.84 ) (A.19)</td>
<td>Extent of dissimilar orientations competing</td>
<td>Lack of orientational contrast</td>
<td>Small magnitude orientational discontinuities lost, like in ( \Delta \theta_{\text{between}} = 15^\circ )</td>
</tr>
<tr>
<td>( \beta_l = 7.5 ) (A.21)</td>
<td>Long-range cooperative grouping</td>
<td>No grouping</td>
<td>Bipole cells saturate and lose graded response</td>
</tr>
<tr>
<td>( \gamma_l = 0.005 ) (A.13)</td>
<td>Cell output threshold</td>
<td>No grouping</td>
<td>Spurious grouping of noise</td>
</tr>
<tr>
<td>( \sigma_l = L_B ) (A.29)</td>
<td>Length of bipole long axis</td>
<td>Misses local curvature</td>
<td>Cannot group the bars</td>
</tr>
<tr>
<td>( \omega_l = 0.001 ) (A.25)</td>
<td>Width of bipole short axis</td>
<td>Boundaries thick and diffuse</td>
<td>Grouping fails unless elements perfectly aligned</td>
</tr>
<tr>
<td>( \psi_l = 0.0001 ) (A.27)</td>
<td>Extent of nearby orientational contribution in grouping</td>
<td>Lack of orientational selectivity</td>
<td>Can group only straight lines</td>
</tr>
<tr>
<td>( \omega_m = 2.5 ) (A.28)</td>
<td>Extent of dissimilarly oriented bipoles competing</td>
<td>Deteriorates grouping of oblique bar ends at horizontal boundary. E.g., ( \Delta \kappa_T = \max ) and ( \Delta \theta_{\text{between}} = 90^\circ )</td>
<td>Deteriorates orientational contrast of boundaries when orientation gradient is present</td>
</tr>
<tr>
<td>( C_y = 7.5 ) (A.29)</td>
<td>Interneuron mutual inhibition</td>
<td>Bipole cells saturate</td>
<td>No grouping</td>
</tr>
<tr>
<td>( \lambda = 25 ) (A.4)</td>
<td>Boundary feedback strength</td>
<td>Noise amplification</td>
<td>No bar-end amplification</td>
</tr>
<tr>
<td>( \mu = 67 ) (A.36)</td>
<td>Amplification factor</td>
<td>Incorrect salience range</td>
<td>Incorrect salience range</td>
</tr>
<tr>
<td>( \nu = 0.13 ) (A.36)</td>
<td>Boundary contrast normalization</td>
<td>No normalization</td>
<td>Salience saturates</td>
</tr>
<tr>
<td>( \sigma_D = 15 ) (A.38)</td>
<td>Width of spatial attention band</td>
<td>No contrast between upper and lower regions</td>
<td>High variability of salience measure due to lack of smoothing</td>
</tr>
<tr>
<td>( A_F = 2.3 \times 10^{-6} )</td>
<td>Average boundary activity pooling coefficient</td>
<td>Rapid performance drop with increasing ( \Delta \theta_{\text{between}} )</td>
<td>No effect of increasing ( \Delta \theta_{\text{between}} )</td>
</tr>
</tbody>
</table>

Column 1: Parameter symbols and the equation numbers in which they appear. Column 2: Functional importance of each parameter in the model. Columns 3 and 4: Model output consequences when parameters and chosen too big or too small. These parameters were systematically varied to improve data fit after an initial tuning based on the input in Fig. 4a.
the same weight applied to all orientations, with bipole grouping enabled, the relative boundary contrasts are preserved in Fig. 6c but not in Fig. 6f, where the bipole grouping is disabled.

5. Texture classification benchmark simulations

A benchmark image dataset derived from the Brodatz (1966) album illustrates the ability of dARTEX to process natural textures, while elucidating the importance of spatial attention in texture learning and classification. The benchmark comparisons show that detrimental effects on texture learning and classification due to feature mixing at texture boundaries can be circumvented by surface-induced attentional shrouds. Spatial attention hereby acts as a supervisory signal, delineating textured regions that can be learned as part of the same texture category.

Each image in the dataset consists of a square textured figure on a textured background. Ten different texture materials were used from the Brodatz (1966) album for generating the dataset. The Brodatz images chosen are Grass, Herringbone Weave, Wool, French Canvas, Paper, Wood, Cotton Canvas, Oriental Cloth, Jeans, and Raffia. Fig. 9a shows an example image.

The effect of spatial attention was quantified by training and testing with or without attention on images with either two textures (e.g., Fig. 9a) or on single texture images. Classification results for all combinations of training and testing with or without attention, and on single-texture or two-texture images is summarized in Table 2. A study with nine images was also undertaken and summarized in Table 2, both to illustrate model robustness and to facilitate its comparison with other studies that used nine textures. The dataset generation and benchmark procedure is described next.

For training, a sequence of 40 images was generated, with each of the 10 textures appearing four times in succession as the central square figure with another texture from the remaining nine textures for the background. The dART network was initially untrained. The image sequence was
presented to dARTEX in blocks of four images. The volitonal local attentional signal $I_{ij}$, or spotlight, in Eq. (A.33) was a 3-by-3 pixel spot of activity located at a randomly chosen position in the central square figure. When this signal leads to a surface-(spatial attention) resonance, a form-fitting spatial attentional shroud emerges. The shroud determines the pixels to be associated with the supplied class label for each image presentation. A total of 1300 pixels, being 2% of the available training data, were randomly selected from those falling within the surface attention in the block of presentations. In order to compare classification performance with other supervised algorithms, the block sequence number was used as a class label for supervised training. Spatially pooled simple cell activities at image pixels selected through the above procedure and their assigned class labels constituted the dART training dataset (Eq. (A.34)). The dART training algorithm is described in Section B.1. All dART network parameters were optimized using this training dataset by 10-fold cross-validation and five training epochs. Cross-validation error did not decrease with more than 1300 training samples per class.

A second dataset was generated by presenting each of the 10 textures by itself as the input image, and using feature vectors from 1300 randomly selected pixels along with their true texture names as class labels. This dataset gave an estimate of the best achievable classification performance.

To investigate the role of surface attention for preventing feature mixing at the boundaries, a third dataset was constructed using the same 40 image set used for the first dataset. For this dataset, surface attention was prevented from modulating the layer 6 simple cells (Eq. (A.10)). The lack of surface attention resulted in mixing of features at texture boundaries, thereby reducing the quality of training samples. The training set was generated by randomly selecting 325 image pixels from the central square figure of each image, with true texture names of the square figure as class labels. With four presentations of each texture, this amounted to 1300 samples per class, the same as for other datasets. When compared to classification performance on the first dataset, this dataset gave an estimate of classification error due to feature mixing at the boundaries during training.

Test classification accuracy was estimated using two datasets. The first dataset was generated with each of the 10 textures presented by itself as the input image. The second dataset used every possible figure and background

Table 2

<table>
<thead>
<tr>
<th>Testing</th>
<th>Training</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2 Textures/image with attention</td>
</tr>
<tr>
<td>10-Texture library</td>
<td></td>
</tr>
<tr>
<td>One texture/image with attention</td>
<td>97.2%</td>
</tr>
<tr>
<td>2 Textures/image with attention</td>
<td>95.1%</td>
</tr>
<tr>
<td>2 Textures/image no attention</td>
<td>90.6%</td>
</tr>
<tr>
<td>9-Texture library</td>
<td></td>
</tr>
<tr>
<td>One Texture/image with attention</td>
<td>97.6%</td>
</tr>
<tr>
<td>2 Textures/image with attention</td>
<td>95.4%</td>
</tr>
<tr>
<td>2 Textures/image no attention</td>
<td>90.9%</td>
</tr>
</tbody>
</table>

Performance was evaluated using two image libraries containing 10 and 9 textures each. The different training and test conditions were based on the number of textures in input images and the presence or absence of top-down surface attention.
combination of the 10 Brodatz textures, resulting in a total of 45 test images. Image patches used in training were not used for testing. The benefit of surface attention in eliminating classification errors at the figure boundary was estimated by computing classification accuracies with and without surface attention for the second test dataset mentioned above. Test classification using surface attention was performed in two phases. In the first phase, volitional attention was set at the horizontal center of the input image and offset a little towards the top, similar to the image in Fig. 4i. In this phase, surface attention covered the central square figure, and dART inputs on the outer side of the form boundary were suppressed; cf., Eq. (A.12). All pixels within the attended surface were classified. In the second phase, volitional attention was set at a location 96 pixels to the left of its previous location, similar to the image in Fig. 4l. The shift in volitional attention resulted in a background-fitting attentional shroud, activating features surrounding the central square figure and suppressing those inside the square. These newly attended pixels were classified, and class labels from the two phases were combined to obtain a final classification; see Fig. 9b for an example output. Classification without surface attention was performed pixel-by-pixel over the whole input image in a single phase of processing.

The top half of Table 2 contains classification results for every combination of the training and testing conditions described above on the 10-texture library described earlier. As noted above, the training and testing procedures above were also performed on a 9-texture library obtained by removing one texture material (the Jeans texture image) from the image set. The bottom half of Table 2 contains classification results for this 9-texture library.

This benchmark study leads to two main conclusions: First, training on two texture images with top–down attention gives classification rates similar to training on single texture images; compare the first two columns of Table 2. Second, test performance significantly deteriorates when trained without surface attention on the two textures per input image dataset; compare the last column of Table 2 to the first two. These two observations suggest that most of the classification performance decrement is due to mixing texture features during training. Furthermore, surface attention usually significantly improves test performance: Compare rows 2 and 3 of Table 2 for the 10-texture library and rows 5 and 6 for the reduced 9-texture library.

Fig. 9 shows an example of dARTEX texture classification output. In particular, Fig. 9a shows the best case input image on which 97.6% classification accuracy was achieved by using surface attention. Fig. 9b shows the corresponding texture label output. Without surface attention, accuracy on the same image was 93.2%. In comparison, Grossberg and Williamson (1999, p. 1396) reported a 79.5% correct classification rate on a Brodatz texture mosaic. In their ARTEX model, Grossberg and Williamson (1999) used single-texture images for training and did not utilize surface attention. As seen in Fig. 5 of Grossberg and William-
images. Fig. 9c shows a gray-level map of the total number of times each location in the images was misclassified without surface attention but was correctly classified with attention. Darker pixels in Fig. 9c denote higher counts of improvement in classification. In the same color-scale, Fig. 9d shows the reverse case, where a previously correct classification was misclassified due to surface attention. This analysis shows that surface attention yields classification improvement along the figure boundary without incurring noticeable deterioration.

6. Object boundary processing in natural images

This section illustrates that the dARTEX boundary grouping process is competent to segment natural scenes. The texture classification study in Section 5 illustrated that dARTEX can complete form boundaries even if no local edges have the orientation of the emergent form boundary, and can support the learned classification of natural textures. The present section notes that correct form boundaries in natural scenes can be detected, over multiple spatial scales, even if image contrasts are low.

dARTEX performance on two images from a human-segmented database of images (Martin, Fowlkes, Tal, and Malik, 2001) is shown in Fig. 10. dARTEX boundary detection, grouping, and attention produce image boundaries that are in fair correspondence to human segmented boundaries. For the two images used, the first two rows in Fig. 10 show object boundaries assigned by human subjects (Fig. 10b and f), the second moment matrix (2MM) edge detection algorithm (Fig. 10c and g, Konishi, Yuille, Coughlan, and Zhu, 1999), and dARTEX (Fig. 10d, h). With dARTEX, co-operative boundary grouping and the attentional modulation of these grouped boundaries results in relatively strong boundary activity even where the image contrast is low. For instance, consider the vertical boundary of the head of the elephant in the foreground in Fig. 10e. The dARTEX output in Fig. 10h successfully detects the boundary, while the 2MM boundary detection algorithm in Fig. 10g is unable to do so.

Multiple-scale processing of dARTEX also has its use in signaling varying amounts of detail about the scene. For example, for the input image in Fig. 10e, equilibrium activities of bipole grouping cells at the three successively larger scales signal different aspects of the scene. At the lowest spatial scale, individual surface features are preserved (Fig. 10i). Visually, however, the medium scale is the most informative for identifying the elephant (Fig. 10j). The largest scale bipole groupings isolate the two animals from the image background (Fig. 10k).

While dARTEX yields favorable boundary processing results compared to the feature contrast-based edge detection approach, it still does not encompass many aspects of human vision. For example, dARTEX does not have figure-ground segregation abilities and the considerable amount of domain knowledge used by humans when dealing with natural scenes. An example of the utilization of such domain knowledge can be seen in human segmentations of Fig. 10b, where human subjects disregarded all of the windows in the building because their task was to segment the major components of the scene and not the finer details. Neither dARTEX nor computer algorithms mentioned here have access to such strategies. For an extension of boundary and surface processing that can explain various figure-ground data, see articles about FACADE theory, e.g., Grossberg (1994, 1997); Grossberg and Swaminathan (2004); Grossberg and Yazdanbakhsh (2005); and Kelly and Grossberg (2000).

7. Related work

Many models of texture segregation follow the general Filter–Rectification–Filter (FRF) scheme of two filter stages with an intermediate point-wise rectifying nonlinearity (Bergen and Landy, 1991; Grossberg and Mingolla, 1985b; Grossberg and Williamson, 1999; Malik and Perona, 1990; Sutter et al., 1989; Wilkinson and Wilson, 1998). A related class of models utilizes oriented filter outputs, similar to the first stage of filtering in FRF, for local texture classification (Fogel and Sagi, 1989; Grossberg and Williamson, 1999; Jain and Farrokhnia, 1991; Rubner and Tomasi, 1999). A survey of filter energy-based methods, especially the FRF, and their relation to human texture segmentation may be found in Bergen and Landy (1991). Some of the more recent texture-based image segmentation attempts also follow a similar filter-based approach (e.g., Arivazhagan and Ganesan, 2003; Krumm and Shafer, 1994).

Malik and Perona (1990) used a boundary-based approach for computing texture gradients and were able to fit psychometric functions of texture discriminability. Instead of using orientational competition as in Grossberg and Mingolla (1985b) and dARTEX. Malik and Perona (1990) used a winner-take-all competition in orientation at every image pixel. The dARTEX simulations of Orientation-Based Texture Segmentation (OBTS) data presented in this paper show that the graded responses from all oriented channels are required for a quantitative data fit. In addition, dARTEX uses mechanisms beyond FRF, including bipole grouping, boundary attention, and surface attention, that enable quantitative simulations of OBTS data.

Another texture classification approach has utilized Markov random fields to probabilistically estimate and cluster texture features (Chellappa and Chatterjee, 1985; Cross and Jain, 1983; Manjunath and Chellappa, 1991; Mao and Jain, 1992; Zhu, Wu, and Mumford, 1998). Yet another approach utilizes probabilistic or data-driven methods for nonparametric estimation of filters, or for combining filter responses (Malik, Belongie, Leung, and Shi, 2001; Martin et al., 2001; Puzicha, Hoffmann, and Buhmann, 1997; Varma and Zisserman, 2003). While all these approaches result in good texture classification performance for isolated texture patches, they do not have the explicit means to utilize top-down attention to guide
texture learning or to prevent misclassifications at the boundaries due to feature mixing.

The ARTEX model of Grossberg and Williamson (1999) utilized boundary processing to drive local texture classification. In the ARTEX model, a Boundary Contour System (Grossberg and Mingolla, 1985a, 1985b) acts as a front-end to a Gaussian ARTMAP (Williamson, 1996) classifier for pixel-by-pixel texture classification. The dARTEX model differs from ARTEX in three ways: First, the dART network used here is more tightly integrated with boundary processing, so that learned texture feature expectations from the dART network directly influence the subsequent dynamics and the final output of the boundary system. Second, as we saw in Section 5, top-down surface-mediated attentional shrouds guide attentive classification and texture learning. This method substantially improves classification performance in the benchmark studies and also helps boundary processing. Third, dART network in the current model accomplishes preliminary local texture categorization and is sensitive to image scale and orientation. In contrast, ARTEX models texture categorization in the IT cortex and addresses scale-and-orientation independent texture classification using long-range image attributes like surface brightness.

Image processing approaches that combine region (cf., surface) and edge (cf., boundary) based techniques are becoming increasingly popular. For a review, see Munoz, Freixenet, Cufi, and Marti (2003). For example, Paragios and Deriche (2002) used a filter-based technique to recognize texture and used the recognition information to repel figure boundary contours from uniformly textured regions. Mirmehdi and Petrou (2000) used color and texture infor-
mation to develop initial estimates of figure boundaries, which were refined using iterative re-classification at progressively smaller spatial scales. Martin et al. (2001) used a combination of texture, color, and luminance contrast to partition image pixels into regions of uniform texture. All of these approaches are similar to dARTEX in that they begin with a hypothesis about the texture boundary locations which are then refined to improve performance through a combination of region and boundary information. In dARTEX, the hypothesis about figure boundaries leads to surface attention that prevents feature mixing at the boundaries. Texture feature expectations from dART, in turn, result in image noise suppression that improves form boundary processing. dARTEX thus makes explicit the role of top–down spatial and object attention and the ways in which boundary and region-based processing may interact in a laminar cortical framework. The other models mentioned above do not explicitly address these issues.

Thielscher and Neumann (2003) have proposed a texture boundary processing model that is more similar to dARTEX boundary processing. In their model, top–down, modulatory, boundary feedback from V4 to prior cortical areas was shown to be necessary for orientation-defined texture boundary processing. The goal of their model was to provide a qualitative explanation of the fact that textures with higher image orientation gradient require a correspondingly higher orientation contrast at the boundary for successful segregation (Nothdurft, 1985, 1992). Since the Thielscher and Neumann (2003) model predates the observations of Ben-Shahar and Zucker (2004), it is not known whether it can quantitatively fit the complete OBTS experiments. The dARTEX model, although similar in spirit, differs from that of Thielscher and Neumann (2003) in two major ways. First, dARTEX uses multiple scales of interaction at each processing stage and the scale channels interact via boundary and surface feedback. The Thielscher and Neumann (2003) model, on the other hand, uses a single scale of processing which gets larger with each successive stage. Spatial scales of interaction in dARTEX also increase with each successive processing stage. In dARTEX the smallest scale groups the texture elements by their proximity in space and similarity in orientation, while the larger scales compute figure boundaries and corners using the same network interactions. Similar information is available in the Thielscher and Neumann (2003) model, but at different stages of processing. Second, dARTEX utilizes spatial attentional shrouds to provide a self-supervising signal for the dART module to bind local feature views into texture categories. Spatial attention in dARTEX thus closes the loop between the bottom-up, pre-attentive, boundary processing and the top–down, modulatory, attentive texture learning (Grossberg, 2003). The Thielscher and Neumann (2003) model does not include texture learning or classification and has not been used to address the role of spatial attention for texture classification. On the other hand, the two models share many common properties and address similar perceptual data using feature contrast-based boundary finding, modulatory feedback for figure boundary enhancement, and long-range boundary grouping using bipole cells. The two models have been inspired by the same prior work on how the laminar circuits of the visual cortex lead to visual percepts, in particular by BCS/FCS model (Grossberg and Mingolla, 1985a, 1985b) and the 3D LAMINART model (Grossberg, 2003; Grossberg and Raizada, 2000; Raizada and Grossberg, 2003).

8. Discussion and conclusions

8.1. Local texture and global form

The dARTEX model combines object boundary and local texture feature computations in a unified, biologically plausible framework. The model uses emergent object boundaries to control surface filling-in which, in turn, generates form-fitting spatial attentional shrouds that direct attention to, and thereby enable learning of, textures that belong to the corresponding objects. Such attentional processing improves both texture classification and object boundary processing itself. When combined with other ART studies that show how object boundaries and surface properties may be used to categorize properties of object form (e.g., Bradski and Grossberg, 1995; Carpenter and Ross, 1995; Carpenter and Grossberg, 1987; Carpenter et al., 1992, 1991a), a unified ART system for classifying both local and global object properties at multiple levels of the visual system can be discerned.

8.2. Explaining orientation-based texture segmentation

The study by Ben-Shahar and Zucker (2004) showed how tangential and normal curvatures and their discontinuities are intimately linked to Orientation-Based Texture Segmentation (OBTS). Their model attempted to explain OBTS using these curvatures and their discontinuities at the texture boundaries as intrinsic image parameters. Specifically, they showed that the textures segregated most readily when discontinuities in both the curvatures were simultaneously maximized. They also showed that the two curvatures and their discontinuities are an efficient way of systematically describing and studying a large class of Orientation Defined Textures (ODTs). In the process of describing ODTs with the help of curvatures, Ben-Shahar and Zucker (2004) also put forward a strong case against models of OBTS that detect discontinuities in filter energy and rely on amplification of filter activities based on element orientation relative to the boundaries (e.g., Wolfson and Landy, 1995). In their paper, Ben-Shahar and Zucker (2004) pointed out two objections to using the boundary for selecting the features for amplification or assigning extra weight during salience computation. The first objection was the “chicken-and-egg problem”, where the out-
come of the processing – namely, the orientation of the texture boundary – was needed as an input to select which features were assigned the most weight. The second, and more serious, problem was that, even if such selection were allowed, the boundary features may get amplified simply based on their absolute orientation, without any selectivity for their relation to other features away from the boundary. These other features are known to affect segmentation, regardless of whether they are characterized by curvature (Ben-Shahar and Zucker, 2004), orientation variability (Wolffson and Landy, 1998), or arrangement (Beck, 1982; Grossberg and Mingolla, 1985b). Ben-Shahar and Zucker (2004) further argued that such modifications to the energy-based model do not suffice to explain their whole suite of results, especially the crossover of curves for certain image orientation gradients but not for others; see Fig. 3.

By fitting the OBTS data of Ben-Shahar and Zucker (2004), dARTEX alleviates most of the concerns about modifying the Filter–Rectifier–Filter (FRF) model raised earlier. While encompassing a FRF computation, dARTEX also includes long-range, recurrent boundary and surface processes, top–down spatial attention that fits the shape of an object surface, and automatic top–down attention that selectively enhances only the features that lie on a texture boundary by an amount proportional to the boundary grouping strength. Through boundary-based attention derived from bipole cells, dARTEX amplifies only the image features that group well and, at an appropriate spatial scale, lie on the figure boundary. See Fig. 2 for examples of how this automatic attention leads to boundaries that mimic the percepts.

Sensitivity to image orientation gradient arises in dARTEX through two factors: First, at high orientation gradients, the feature contrast detection stage in V2 is unsuccessful in suppressing featural activity in the figure interior. As a result, the activity difference between texture boundary grouping and individual bar grouping decreases, resulting in no net LGN feature amplification through boundary feedback. The second factor is the formulation of the boundary contrast measure, which supplements the first one: At high orientation gradients, the contrast measure is low because the boundaries due to feature contrast grouping are roughly the same as boundaries linking individual texture bars.

In the spirit of Wolffson and Landy (1995), the OBTS boundary contrast measure used here applies more weight to the horizontal boundary orientation. Such an explicit weighting applied to spatial locations and orientations of boundaries may seem at first to be subject to the same “chicken-and-egg” problem. However, additional mechanisms are hypothesized to play a role in this task-specific modulation of orientations. For example, object attentional modulation that weights boundaries at certain locations or orientations can be computed using center-surround computations much in the spirit of the two competitive stages of Grossberg and Mingolla (1985b), wherein the attended spatial locations are winners of a spatial competition over multiple trials due to the presence of strong boundaries at those locations. Similarly, the horizontal orientation would always win in an orientational competition at these spatial locations, since the form boundaries are always horizontal. As seen in the V2 layer 2/3 bipole cell activities plotted in Fig. 2, even without the task-specific spatial and orientational attentional modulation, the pre-attentive grouping and automatic, boundary-based attentional feedback produce strong boundaries at the correct locations and orientations. Large-scale spatial and orientational competition activated by bipole inputs over multiple trials may thus lead to top–down spatial attention to boundaries during segmentation, further amplifying the differences due to element configuration, and producing a quantitative data fit. Also compare dARTEX simulations in Fig. 6b (with higher weight in the horizontal orientation) and c (with equal weight to all orientations) that preserve the crossovers in the experimental data in Fig. 6a.

While fitting the OBTS data of Ben-Shahar and Zucker (2004), the dARTEX model also outperforms related texture classification approaches (Grossberg and Williamson, 1999; Randen and Husoy, 1999). Improved performance in texture classification is due to the ability of spatial attention to eliminate feature mixing at form boundaries. Emergent form boundary completion in response to noisy images is also facilitated by the feature noise rejection afforded by dART-based feature match during attentive texture classification.

8.3. Unifying boundary and surface attention

dARTEX attempts to clarify the parallel roles of boundary-mediated object attention and surface-mediated spatial attention. Boundary-mediated object attention is critical for object form processing, while surface-mediated spatial attention is critical in local surface feature learning and recognition. In the process of clarifying the roles of these processes, dARTEX elaborates the claim that boundaries and surfaces are the units of visual attention (Grossberg, 1999a, 2003). Boundary-mediated attention in dARTEX is shown to proceed in two parts: In the automatic part, pre-attentively computed boundaries amplify features at boundary selected positions. The task-specific part of the boundary-mediated attention uses feedback to LGN to further amplify boundaries at attended spatial positions.

Surface-mediated attention in dARTEX proceeds in a similar fashion, but subserves a complementary role: Surfaces identified by pre-attentive boundary-gated filling-in bid for spatial attention. Surfaces with the highest filled-in activity win spatial attention automatically. In task-specific spatial attention, a volitional focus of attention, or attentional “spotlight”, can bias the competition in favor of the attended surface. A form-fitting surface attentional shroud from the spatial attention stage leads to a surface-
(spatial attention) resonance, that binds multiple texture views from the attended surface into local texture categories learned by the dART network. Experiments on perceptual learning have shown how visual task performance can alter the detection thresholds for elementary image features like orientation and local motion (Hochstein and Ahissar, 2002; Watanabe, Nanez, and Sasaki, 2001). Some of these improvements are positionally specific, thereby suggesting that these improvements occur at early stages of the visual hierarchy (see Ahissar and Hochstein, 2004; for a survey), and perhaps at the very locations of feature representation (Grossberg, 1999a, 2003). In a similar vein, the ARTSCAN and perhaps at the very locations of feature representation hierarchy (see Ahissar and Hochstein, 2004; for a survey), that these improvements occur at early stages of the visual improvements are positionally specific, thereby suggesting that these improvements occur at early stages of the visual hierarchy. Therefore, once these stages reached equilibrium, their activities were sampled and incorporated into the two stages that directly receive them: Boundary attention derived from the bipole cells in Eq. (A.21) influences the LGN (Eq. (A.4)), and surface-based attention from the PPC (Eq. (A.33)) influences the V1 layer 6 simple cells (Eq. (A.10)). After incorporating feedback signals into the early stages, the entire dARTEX hierarchy was re-calculated and the numerical integration of the three stages mentioned above was performed again, leading to the next equilibrium state. Model outputs did not change perceivably after three such iterations of equilibration and re-calculation. All results shown are after three iterations, except the images in Fig. 4b and c that denote activity before the feedback kicks in. These figures were sampled at equilibrium just before the first re-calculation of dARTEX stages.

In what follows, each dARTEX processing stage is described in order of its activation in response to bottom–up retinal input.

### A.1. Center-surround field, LGN

The LGN stage consists of a spatial competitive stage with on-center, off-surface filters (ON channel) as well as off-center, on-surround filters (OFF channel). This stage discounts the illuminant and normalizes image contrast.

### Appendix A. System equations

This section mathematically describes the dARTEX model. Parameter values used in the simulations immediately follow the equation descriptions.

Each model neuron is typically modeled as a single voltage compartment in which the membrane potential, \( v \), is given by

\[
C_m v = - (v - E_{\text{leak}}) g_{\text{leak}} - (v - E_{\text{excit}}) g_{\text{excit}} - (v - E_{\text{inhib}}) g_{\text{inhib}},
\]

where \( C_m \) is the membrane capacitance, the \( E \) terms represent reversal potentials, \( g_{\text{leak}} \) is a constant leakage conductance, and the time-varying conductances \( g_{\text{excit}} \) and \( g_{\text{inhib}} \) represent the total inputs to the cell (Grossberg, 1973; Hodgkin, 1964). Most of the following network equations are instances of this general membrane equation, where, for simplicity, the capacitance term \( C_m \) was set equal to 1, the leakage conductance is relabeled as \( g_{\text{leak}} = A \), the excitatory and inhibitory reversal potentials are relabeled as: \( E_{\text{excit}} = B \) and \( E_{\text{inhib}} = -C \), and the leakage reversal potential is set to \( E_{\text{leak}} = 0 \). Then Eq. (A.1) can be rewritten as a membrane, or shunting equation

\[
v = -A v + (B - v) g_{\text{excit}} - (v + C) g_{\text{inhib}},
\]

where \( A \) is a constant decay rate, \( B \) is an excitatory saturation potential, \( g_{\text{excit}} \) is the total excitatory input, \( C \) is a hyperpolarization parameter, and \( g_{\text{inhib}} \) is the total inhibitory input.

Throughout the dARTEX description, subscripts \( ij \) or \( pq \) denote two-dimensional Cartesian image co-ordinates \( (i,j) \) and \( (p,q) \), respectively. Subscript \( s \) is reserved for spatial scale, and subscripts \( k \) and \( r \) are reserved for orientations.

Unless otherwise noted, two-dimensional isotropic Gaussian spatial kernel is denoted by \( G_{pq(\sigma)} \). In particular, Gaussian spatial kernel \( G_{pq(\sigma)} \) for position \((p,q)\) and scale \( s \) is defined as:

\[
G_{pq(\sigma)} = \frac{1}{2\pi\sigma^2(s)} \exp \left( -\frac{(p-i)^2 + (q-j)^2}{2\sigma^2(s)} \right),
\]

where the kernel \( G_{pq(\sigma)} \) is centered at position \((i,j)\) and has a scale-dependent spatial variance \( \sigma(s) \). Unless otherwise noted, all kernels are normalized to add to 1. All kernel parameters are reported in units of image pixels, and all input images used for simulations are \( 256 \times 256 \) pixels wide.

The operator \( |E| \) denotes half-wave rectification.

All simulations were carried out using the Matlab simulation package (version 6.5, mathworks.com). All but three model processing stages were explicitly solved for their equilibrium values in order of their activation in the model hierarchy. The remaining three stages, namely, the bipole, surface filling-in, and surface attention cells (Eqs. (A.21), (A.31), and (A.33)) were numerically integrated until equilibrium using a second-and-third order Runge-Kutta formula with adaptive integration step size (Matlab function ode23). Equilibrium for the latter stages was defined as the sum-of-squares activity not changing by more than 0.25% for three successive integration steps for each model stage. The boundary and surface attention stages constitute feedback signals that influence the initial stages of the model and thus the entire processing hierarchy. Therefore, once these stages reached equilibrium, their activities were sampled and incorporated into the two stages that directly receive them: Boundary attention derived from the bipole cells in Eq. (A.21) influences the LGN (Eq. (A.4)), and surface-based attention from the PPC (Eq. (A.33)) influences the V1 layer 6 simple cells (Eq. (A.10)). After incorporating feedback signals into the early stages, the entire dARTEX hierarchy was re-calculated and the numerical integration of the three stages mentioned above was performed again, leading to the next equilibrium state. Model outputs did not change perceivably after three such iterations of equilibration and re-calculation. All results shown are after three iterations, except the images in Fig. 4b and c that denote activity before the feedback kicks in. These figures were sampled at equilibrium just before the first re-calculation of dARTEX stages.

In what follows, each dARTEX processing stage is described in order of its activation in response to bottom–up retinal input.
As seen in Fig. 1, in addition to bottom–up retinal input, the LGN stage also receives top–down modulatory boundary-based attention.

The activity \( x_{ij}^+ \) of the LGN ON channel neuron at spatial position \((i,j)\) and scale \(s\) obeys a shunting on-center, off-surround equation:

\[
x_{ij}^+ = -\alpha_s x_{ij}^+ + (1 - \lambda_s f_{ij}) \left[I_{ij}(1 + \lambda_s f_{ij})\right] - (1 + x_{ij}^+) \left[\sum_p I_{pq} \left(1 + \lambda_s f_{pq}\right) G_{pqij(s)}\right].
\]

In Eq. (A.4), the term \( x_{ij}^+ \) on the right hand side defines a passive decay at rate \( \alpha_s = 0.25 \). In the excitatory term \( (1 - x_{ij}^+) [I_{ij}(1 + \lambda_s f_{ij})] \), retinal input \( I_{ij} \) at position \((i,j)\) is gain-modulated by a boundary-based attentional signal \( f_{ij} \) defined in Eq. (A.30). The signal \( f_{ij} \) amplifies LGN activity at figure boundaries. The inhibitory term \( (1 + x_{ij}^+) \left[\sum_p I_{pq} \left(1 + \lambda_s f_{pq}\right) G_{pqij(s)}\right] \) suppresses the neighborhood around the boundaries through an inhibitory surround. In Eq. (A.4), the top–down gain factor \( \lambda_s = 25 \), and the Gaussian surround kernel \( G_{pqij(s)} \) in the inhibitory term is defined in Eq. (A.3), with scale-dependent variances \( \sigma_s(s) = [4.5, 9.0, 13.5] \) determining the extent of surround inhibition for each of the three successively increasing spatial scales.

The equilibrium solution for Eq. (A.4) is:

\[
x_{ij}^+ = \frac{(1 + \lambda_s f_{ij}) I_{ij} - \sum_p (1 + \lambda_s f_{pq}) I_{pq} G_{pqij(s)}}{\alpha_s + (1 + \lambda_s f_{ij}) I_{ij} + \sum_p (1 + \lambda_s f_{pq}) I_{pq} G_{pqij(s)}}.
\]

The LGN OFF channel activities, denoted by \( x_{ij}^- \), are defined as \( x_{ij}^- = -x_{ij}^+ \). They are the activities of an analogous shunting off-center, on-surround equation. This formulation for the model’s simplified LGN stage is adapted from Gove et al. (1995). In some treatments, OFF cells are tonically active in the absence of inputs (e.g., Grossberg et al., 1995). This property was not needed in the present simulations.

### A.2. Oriented simple cells, V1 layers 6 and 4

Two sets of oriented multiple-scale simple cells are simulated in layers 4 and 6 of V1, respectively. As seen in Fig. 1, both sets of simple cells are driven by bottom–up activation from the LGN. The simple cells act as polarity-of-contrast sensitive filters that detect oriented features in the image along the filter’s preferred orientation. In model simulations, 24 different simple cell orientations are used for each of the 3 spatial scales. The simple cells in layer 6 are identical to layer 4 simple cells in terms of their bottom–up inputs, but they are also modulated by top–down spatial attention. The spatial attention stage is described in Section A.8.

V1 layer 4 simple cells are bottom–up activated by LGN ON and OFF activities sampled through spatially elongated and offset Gaussian kernels. In particular, layer 4 simple cell \( p_{ijk} \) at position \((i,j)\), orientation \(k\), and scale \(s\) obeys the equation:

\[
p_{ijk} = -\beta_p p_{ijk} + (1 - p_{ijk}) \times \sum_p \left(\left[x_{pq}^+\right]^+ R_{pqijk(s)}^+ + \left[x_{pq}^-\right]^+ R_{pqijk(s)}^+\right) - (1 - p_{ijk}) \sum_p \left(\left[x_{pq}^+\right]^+ R_{pqijk(s)}^- + \left[x_{pq}^-\right]^+ R_{pqijk(s)}^-\right).
\]

In Eq. (A.6), the passive decay rate \( \alpha_p = 0.25 \). In the (A.6) excitatory term, LGN ON cell output signals \( x_{pq}^+ \) are filtered by oriented spatially elongated Gaussian kernel \( R_{pqijk(s)}^- \), while the LGN OFF cell output signals \( x_{pq}^- \) are filtered by a similar kernel \( R_{pqijk(s)}^+ \). The centers of the kernels \( R_{pqijk(s)}^- \) and \( R_{pqijk(s)}^+ \) are offset in mutually opposite directions from each simple cell’s centroid along an axis perpendicular to the simple cell’s direction of elongated sampling. In the inhibitory term of Eq. (A.6), the same kernels sample an LGN channel complimentary to the one in the excitatory term. The net activity of simple cells is thus a measure of image feature contrast in its preferred orientation. In mathematical terms, the vertically oriented kernels \( R_{pqijk(s)}^- \) and \( R_{pqijk(s)}^+ \) in (A.6) are:

\[
R_{pqijk(s)}^- = A_p(s) \exp\left(-\left(i - p \pm \sigma_p(s) / 2, j - q \pm \delta / 2\right)^2 / 2\sigma_p^2(s)\right),
\]

where superscripts + and − denote positive and negative lobes of the kernel that are centered at \((i \pm \sigma_p(s) / 2, j)\), respectively. Each kernel has a scale-dependent spatial variance \( \sigma_p(s) = [1, 2, 3] \) along its narrow axis that determines the spatial region over which average contrast is measured, and an elongation factor \( \delta = 2 \) for the long axis. Other lobe orientations are generated by coordinate system rotation. The normalization factor \( A_p(s) = (2\pi\sigma_p^2(s)\delta)^{-1} \) in Eq. (A.7) ensures that kernel lobes individually add to 1.

The equilibrium solution of simple cell activity of Eq. (A.6) equals:

\[
p_{ijk} = \frac{A - B}{\alpha_p + A + B},
\]

where the terms \( A \) and \( B \) are:

\[
A = \sum_p \left(\left|x_{pq}^+\right|^+ R_{pqijk(s)}^+ + \left|x_{pq}^-\right|^+ R_{pqijk(s)}^-\right),
\]

\[
B = \sum_p \left(\left|x_{pq}^+\right|^+ R_{pqijk(s)}^- + \left|x_{pq}^-\right|^+ R_{pqijk(s)}^+\right).
\]

By (A.8), the simple cell equilibrium activity detects luminance contrast in the preferred orientation, indeed contrast-normalizes its responses according to a Weber law (Grossberg, 1980). The magnitude of the Weber law parameter, or, equivalently, the passive decay rate \( \alpha_p \), determines how quickly contrast normalization sets in as
input activities increase. In dARTEX simulations, $\gamma_p$ is chosen to achieve a balance of high contrast normalization and sensitivity to top–down attentional modulation.

Symmetric sampling of LGN ON and OFF channel activities by simple cells in Eq. (A.6) ensures that a simple cell encoding a light-to-dark edge in a given orientation has the same activity as another simple cell of complementary contrast polarity preference when it responds identically dark-to-light edge. This balance in activation is important for processing form boundaries independent of their luminance contrast polarity. The simple cell Eqs. (A.6) and (A.8) are adapted from Ross et al. (2000).

V1 layer 6 simple cells have receptive fields identical to those of layer 4 simple cells. Additionally, layer 6 simple cells are modulated by top–down spatial attention; see Eq. (A.33). In the initial feedforward phase of input activation, when spatial attention is not yet active, simple cells in layers 4 and 6 exhibit identical responses. However, as spatial attention activates, it suppresses unattended layer 6 simple cells through surround inhibition from nearby attended locations. Specifically, layer 6 simple cell activity $a_{ijks}$ at position $(i,j)$, orientation $k$, and scale sobes the equation:

$$a_{ijks} = -a_0 a_{ijks} + (1 - a_{ijks}) (1 + \Theta(r_{ij})) \times \sum_{pq} \left[ \left( [X]_{pq}^+ R_{pq(i)k(s)} + [X]_{pq}^- R_{pq(i)k(s)} \right) \right]$$

$$- (1 + a_{ijks}) \left[ 1 + \sum_{mn} \Theta(r_{mn}) G_{mn(i)j(s)} \right] \times \left( \sum_{pq} \left[ [X]_{pq}^+ R_{pq(i)k(s)} + [X]_{pq}^- R_{pq(i)k(s)} \right) \right].$$  \( \text{(A.10)} \)

In (A.10), the passive decay rate $a_0$ is 0.25. The on-center and off-surround terms in Eq. (A.10) that sample LGN activity $[X]_{pq}^\pm$ using simple cell kernel lobes $R_{pq(i)k(s)}$ are the same as Eq. (A.6). The on-center is top–down attenuatively amplified through the modulatory term $(1 + \Theta(r_{ij}))$. The signal $\Theta(r_{ij})$ in this term is input from spatial attention $r_{ij}$ in Eq. (A.33). Through a similar modulatory term in the off-surround, $(1 + \sum_{mn} \Theta(r_{mn}) G_{mn(i)j(s)}$, unattended cells around the locus of spatial attention are strongly inhibited.

The attentional signal function $\Theta(r_{ij})$ in Eq. (A.10) normalizes all above-threshold spatial attention activities $r_{ij}$ to approximately the same signal size. Specifically,

$$\Theta(r_{ij}) = \frac{\lambda_a [r_{ij} - \gamma_a]^+}{\varepsilon_a + [r_{ij} - \gamma_a]^+},$$  \( \text{(A.11)} \)

where $\lambda_a = 2$, $\gamma_a = 0.05$, and the Weber law parameter $\varepsilon_a = 0.05$. A large gain $\lambda_a$ and a small Weber law parameter $\varepsilon_a$ ensure that unattended simple cells surrounding the attended ones are strongly inhibited through the term $(1 + \sum_{mn} \Theta(r_{mn}) G_{mn(i)j(s)})$ in Eq. (A.10), while simple cell activity at attended locations is uniformly amplified. This attentional modulation in Eq. (A.10) prevents mixing of dART input features across texture boundaries during local texture classification and learning, improving overall model performance. This property of dARTEX is described in Section 3.9. The Gaussian kernel $G_{pq(i)j(s)}$ in Eq. (A.16) is defined in Eq. (A.3) with scale-dependent variance $\sigma_{pq}^2(s) = [4.5, 9, 13.5]$, and determines the spatial extent of attentive inhibition of unattended features.

The equilibrium solution of V1 layer 6 simple cell activity in Eq. (A.10) obeys

$$a_{ijks} = \frac{A - B + \left( A \Theta(r_{ij}) - B \sum_{mn} \Theta(r_{mn}) G_{mn(i)j(s)} \right)}{\varepsilon_a + A + B + \left( A \Theta(r_{ij}) + B \sum_{mn} \Theta(r_{mn}) G_{pq(i)j(s)} \right)},$$  \( \text{(A.12)} \)

where the terms $A$ and $B$ are defined in Eq. (A.9). The equilibrium solution in Eq. (A.12) shows that layer 6 activity is identical to layer 4 activity in absence of attention (when $\Theta(r_{pq}) = 0$, for all $p$ and $q$); see Eq. (A.8). When spatial attention activates, the term $B \sum_{mn} \Theta(r_{mn}) G_{mn(i)j(s)}$ in the numerator of Eq. (A.12) ensures that unattended simple cells in the surround (where $\Theta(r_{pq}) = 0$) of the attended locations (where $\Theta(r_{pq})$ is close to 1) are inhibited.

As seen in Fig. 1, V1 layer 6 simple cells activate a distributed learned category representation in the dART network (Carpenter, 1997). This category representation reads-out learned expected simple cell activities at each image location; see Eq. (A.35). The layer 4 simple cell activity is matched with this learned expectation via the layer 6–4 modulatory network (Callaway, 1998; Grossberg and Raizada, 2000). Specifically, match activity $m_{ijks}$ for position $(i,j)$, orientation $k$, and spatial scale $s$ is (Carpenter, 1997):

$$m_{ijks} = 0.5 \left[ \min(\sigma_{ijks}, [p_{ijks}]^+) + 1 - \min(\sigma_{ijks}, 1 - [p_{ijks}]^+) \right],$$  \( \text{(A.13)} \)

where $\sigma_{ijks}$ is the dART expected ON simple cell activity, and $\sigma_{ijks}$ is the expected OFF simple cell activity. This complementarity of ON and OFF features will be described as part of the definition of the dART input feature vector in Eq. (A.34). The feature expectation term $\sigma_{ijks}$ is defined in Eq. (A.35). Simple cell signals $[p_{ijks}]^+$ from V1 layer 4 are defined in Eq. (A.6). By Eq. (A.13), the match signal is an average of matched activities of the ON and OFF simple cell activities. This matching operation facilitates dARTEX processing by reducing feature noise, and thereby facilitates the completion of object form boundaries, as described in Section 3.9. The matched features in Eq. (A.13) are the bottom–up inputs to complex cells of V1 layer 2/3.

### A.3. Complex cells, V1 layer 2/3

Each complex cell receives activity from pairs of simple cells that are selective to opposite contrast polarities, and thereby acts as an oriented polarity-of-contrast insensitive filter. The complex cell activity $q_{ijks}$ for position $(i,j)$, orientation $k$, and scale $s$ obeys the equation:

$$q_{ijks} = [m_{ijks}]^+ + [m_{ijks}]^+,$$  \( \text{(A.14)} \)
where \( m_{ijkz} \) is the matched simple cell activity described in Eq. (A.13). The orientation index \( K \) in Eq. (A.14) denotes the matched simple cell of contrast polarity opposite to the one indexed by \( k \). Since there are 24 simple cell orientations, the above operation produces 12 unique orientations of complex cells for each spatial scale. The complex cell activities are adapted from Gove et al. (1995). However, there was no dART matching of simple cell activities in that work.

### A.4. Spatial and orientational competition, V2 layers 6 and 4

As seen in Fig. 1, the two competitive stages in V2 are driven by complex cell activities arriving from V1 layers 2/3. The spatial competition stage activity \( u_{ijks} \) at position \((i,j)\), orientation \( k \), and scale \( s \) obeys the equation:

\[
u_{ijks} = -\alpha_u u_{ijks} + (1 - u_{ijks}) (J + q_{ijks}) - (1 + u_{ijks}) \left( \sum_{pq} q_{pqrs} K_{pqrijk}(s)u_{pqrs} \right),
\]

where the passive decay rate \( \alpha_u = 0.25 \), \( J = 0.01 \) is a constant tonic activity, and \( q_{ijks} \) is the complex cell activity arriving from V1 layers 2/3 as defined in Eq. (A.14). The spatial competition kernel \( K_{pqrijk}(s)u_{pqrs} \) in Eq. (A.15) is a product of two Gaussian kernels, one that depends on orientation, and the other that depends on spatial distance. In particular, kernel \( K_{pqrijk}(s)u_{pqrs} \) at position \((i,j)\) and orientation \( k \) is:

\[
K_{pqrijk}(s)u_{pqrs} = A_v(s) \exp \left( \frac{-(\min(r-k, O-r+k))^2}{2\sigma^2_v} \right) G_{pqrs}(u_{pqrs}),
\]

The first Gaussian kernel in (A.16) is the orientation kernel: it is maximal for the orientation \( k \) with a Gaussian fall-off for more dissimilar orientations, \( O = 12 \) is the total number of complex cell orientations, and the orientation-dependent Gaussian drop-off variance \( \sigma_o = 0.43 \). The second spatial Gaussian kernel in (A.16) is defined in Eq. (A.3) with a scale-dependent variance \( \sigma_s(s) = [4.5, 9, 13.5] \). Thus, in Eq. (A.15), similar orientations in a local image neighborhood inhibit each other the most. Coefficient \( A_v(s) = ((2\pi)^{1/2} \sigma_o)^{-1} \) in Eq. (A.16) is a scale-dependent normalization factor such that the kernel adds to 1.

As a result of spatial competition between like-oriented complex cell outputs in Eq. (A.15), uniform activity in luminance and orientation is suppressed, while discontinuities in their spatial distribution are enhanced. The equilibrium solution of the spatial competition stage in Eq. (A.15) is:

\[
u_{ijks} = J + q_{pqrs} - \sum_{pq} q_{pqrs} K_{pqrijk}(s)u_{pqrs} \]

As seen in Fig. 1, the orientational competition stage in V2 is driven by the outcome of the spatial competition. At this stage, mutually orthogonal orientations inhibit each other the most in a small spatial neighborhood. Specifically, orientational competition activity \( v_{ijks} \) at position \((i,j)\), orientation \( k \), and scale \( s \) obeys the equation:

\[
v_{ijks} = -\alpha_v v_{ijks} + (1 - v_{ijks}) \left( [u_{ijks} - \gamma_v^i]^+ \right) - (1 + v_{ijks}) \left( \sum_{pq} [u_{pqrs} - \gamma_v^i]^+] L_{pqijk}(s) \right),
\]

where the passive decay rate \( \alpha_v = 0.05 \). Signal \([u_{ijks} - \gamma_v^i]^+\) is the input from the spatial competition stage defined in Eq. (A.15), with output threshold \( \gamma_v = 0.005 \). The orientational competition kernel is a product of two Gaussian kernels, one that depends on orientation, and the other that depends on spatial distance. In particular, kernel \( L_{pqijk}(s) \) centered at position \((i,j)\) and orientation \( k \) is:

\[
L_{pqijk}(s) = A_v(s) \exp \left( \frac{-(O/2 - \min(r-k, O-r+k))^2}{2\sigma^2_v} \right) G_{pqrs}(u_{pqrs}),
\]

where the first, orientation Gaussian kernel is maximal for an orientation perpendicular to the orientation \( k \) with a Gaussian fall-off for more similar orientations, \( O = 12 \) is the total number of complex cell orientations, and the orientation-dependent Gaussian drop-off variance \( \sigma_o = 0.84 \). The second spatial Gaussian kernel in (A.19) is defined in Eq. (A.3) with a scale-dependent variance \( \sigma_s(s) = [1, 2, 3] \). Thus, in Eq. (A.18), dissimilar orientations in a local image neighborhood inhibit each other the most. Factor \( A_v(s) = ((2\pi)^{1/2} \sigma_o)^{-1} \) in Eq. (A.19) is a scale-dependent normalization factor such that the kernel adds to 1. The equilibrium solution for the orientational competitive stage defined in Eq. (A.18) is:

\[
v_{ijks} = [u_{ijks} - \gamma_v^i] - \sum_{pq} [u_{pqrs} - \gamma_v^i] L_{pqrijk}(s)u_{pqrs} + \alpha_v v_{ijks} + [u_{ijks} - \gamma_v^i] + \sum_{pq} [u_{pqrs} - \gamma_v^i] \]

The two competitive stages defined in Eqs. (A.15) and (A.18) were introduced into the neural modeling of boundary grouping in Grossberg (1984) and are adapted from Grossberg and Mingolla (1985a, 1985b).

### A.5. Bipole grouping cells, V2 layer 2/3

The bipole grouping stage in layer 2/3 of V2 is driven by competitive stage output from layer 4 of V2, as seen in Fig. 1. Bipole cells receive long-range monosynaptic excitatory inputs from other bipole cells and short-range disynaptic inhibitory connections from a shared pool of interneurons. Specifically, bipole cell activity \( z_{ijkz} \) at position \((i,j)\), orientation \( k \), and scale \( s \) obeys the equation:
\[ z_{ijks} = -z_{ijks} + (1 - z_{ijks}) (\lambda_z[v_{ijks} - \gamma_z])^+ + \beta_z[h_{ijks}^L - \gamma_z]^+ + \beta_z[h_{ijks}^R - \gamma_z]^+ \]
\[ - (1 + z_{ijks}) (m_{ijks} + s_{ijks} + b_{ijks}), \]  
(A.21)

where \( \lambda_z[v_{ijks} - \gamma_z]^+ \) is the bottom-up excitatory input signal from the orientational competition stage defined in Eq. (A.18) with gain \( \lambda_z = 1 \) and output threshold \( \gamma_z = 0.005 \). Long-range horizontal excitatory signals \( \beta_z[h_{ijks} - \gamma_z]^+ \) are received from bipole cells on the two coaxial flanks (denoted by L and R) of each bipole cell, with gain \( \beta_z = 7.5 \) and output threshold \( \gamma_z = 0.005 \). The signals \( [h_{ijks} - \gamma_z]^+ \) are defined in Eq. (A.22). The spatial and orientational competition term \( m_{ijks} \) in Eq. (A.21) receives signals from other bipole cells as defined in Eq. (A.28). The di-synaptic inhibition terms \( s_{ijks} \) in Eq. (A.21) realize the bipole property as defined in Eq. (A.29).

In what follows, the horizontal bipole orientation is assumed for the sake of notational simplicity. Other bipole orientations are derived by co-ordinate system rotation.

The long-range horizontal excitatory kernels \( h_{ijks}^L/R \) in Eq. (A.21) obey:
\[ h_{ijks}^{L/R} = \sum_{pqr} [z_{pqr} - \gamma_z]^+ h_{pqrkij}^{L/R} (s)^m_{ijks}, \]  
(A.22)

which sum output signals \( [z_{ijks} - \gamma_z]^+ \) from bipole cells defined in Eq. (A.21) with output threshold \( \gamma_z = 0.005 \). The left bipole half-kernel \( h_{pqrkij}^{L} \) is a product of three Gaussian kernels which determines the amount of sampling from other bipole cells based on their distance, lateral offset from the horizontal axis, and deviation from the preferred orientation at various locations in the receptive field. As a result, proximal, coaxially located, and co-aligned bipole cells are sampled with higher weight. Specifically, the left half-kernel \( h_{pqrkij}^{L} \) at position \((i,j)\), orientation \(k\), and scale \(s\) is:
\[ h_{pqrkij}^{L}(s) = A_H(s) \exp\left(-\frac{(p-i)^2 + (q-j)^2}{2\sigma^2(s)}\right) \exp\left(-\frac{K^2_{pqr|kij(s)}^L}{2\gamma_w} \right) \]  
if \( i < 0 \)
\[ 0 \]  
otherwise
(A.23)

The first term
\[ \exp\left(-\frac{(p-i)^2 + (q-j)^2}{2\sigma^2(s)}\right), \]  
(A.24)
in (A.23) is an isotropic spatial Gaussian with scale-dependent variance \( \sigma^2(s) = [100, 200, 300] \). This term controls the spatial extent of the bipole half-kernel, with larger scales performing boundary groupings over a longer spatial range. The second term in Eq. (A.23),
\[ \exp\left(-\frac{K^2_{pqr|kij(s)}^L}{2\gamma_w} \right), \]  
(A.25)
determines contributions from locations lateral to the horizontal axis of the kernel: Term \( K_{pqr|kij} \) ensures that farther and more laterally offset locations contribute less. Drop-off of the kernel with distance and thickness is governed by parameter \( \gamma_w = 0.001 \). A larger value of \( \gamma_w \) results in a kernel with a thicker band around the principal axis. The factor \( K_{pqr|kij} \) in Eq. (A.25) for a bipole cell at position \((i,j)\) is:
\[ K_{pqr|kij} = \arctan\left(\frac{(p-i)(q-j)}{(p-i)^2 + (q-j)^2}\right). \]  
(A.26)

Factor \( K_{pqr|kij} \) above is sensitive to both the distance between locations \((i,j)\) and \((p,q)\) and the spatial offset of position \((p,q)\) from the horizontal axis. For example, along the principal axis \( q = j, K_{pqr|kij} = 0 \), and the term in Eq. (A.25) is at its maximum of 1. As position \((p,q)\) departs the principal axis, and moves farther from position \((i,j)\), \( K_{pqr|kij} \) diverges to infinity. In this case, Eq. (A.23) vanishes in the limit, disabling the sampling from such locations. Eq. (A.25) thus controls thickness of the bipole half-kernel in Eq. (A.23).

Finally, the third term in Eq. (A.23),
\[ \exp\left(-\frac{(z_{pqr} - K_{pqr|kij})^2}{2\gamma^2}\right), \]  
(A.27)
determines the contribution from other bipole cells based on their orientation \(r\) and position \((p,q)\). This term is sensitive to three attributes: Orientation of the contributing bipole cell \((r)\), its offset from the horizontal axis, \((q-j)\), and its distance from the position \((i,j)\) of target bipole cell receiving the input. Through this term, contributions from the same bipole orientation are maximal along the principal axis. Contributions from dissimilar bipole orientations increase while those from similar bipole orientations decrease as offset from the principal axis increases. By controlling the amount of pooling from dissimilar orientations, the term in Eq. (A.27) limits the maximum curvature of the boundaries that can be completed. In particular, parameter \( \psi_w = 0.0001 \) in Eq. (A.27) controls the highest curvature that can be grouped. A larger value of \( \psi_w \) enables grouping of more dissimilar orientations, and thereby grouping of boundaries with higher curvatures. The constant \( O = 12 \) in the above equation denotes the total number of orientations. Finally, coefficient \( A_H(s) = 0.5(2m^2\sigma^2(s))^{-1} \) in Eq. (A.23) normalizes the kernel to add to 0.5.

The spatial and orientational competition term \( m_{ijks} \) in Eq. (A.21) sharpens boundaries and reduces orientational ambiguity. For position \((i,j)\), orientation \(k\), and scale \(s\) term \( m_{ijks} \) is:
\[ m_{ijks} = A_m \sum_{pqr} [z_{pqr} - \gamma_m]^+ L_{pqrkij}(s)\sigma_m, \]  
(A.28)
where the constant \( A_m = 2.5 \) is a gain factor, terms \( [z_{pqr} - \gamma_m]^+ \) in Eq. (A.28) are output signals from bipole cells defined in Eq. (A.21) with \( \gamma_m = 0.005 \), and the orientational competition kernel \( L_{pqrkij}(s)\sigma_m \) is defined in Eq. (A.19) with \( \sigma_m = 2.5 \) and scale-dependent spatial variance \( \sigma_m(s) = [1, 2, 3] \).
The left (L) and right (R) interneuron activity $s_{i,j,k,s}^{L/R}$ at position $(i,j)$, orientation $k$, and scale $s$ in Eq. (A.21) is:

$$
s_{i,j,k,s}^{L/R} = D_x \left( -s_{i,j,k,s}^{L/R} + \lambda_x \left[ x_{i,j,k,s}^{L/R} - \gamma_x \right]^+ - C_x s_{i,j,k,s}^{L/R} \left( s_{i,j,k,s}^{L/R} \right)^+ \right),
$$

(A.29)

where the gain $D_x = 50$ makes the inhibitory interneuron settling time more rapid than that of the bipole cell. Excitatory signal $x_{i,j,k,s}^{L/R} = [x_{i,j,k,s}^{L/R} - \gamma_x]^+$ comes from bipolar cells defined in Eq. (A.22), with gain $\lambda_x = 7.5$ and output threshold $\gamma_x = 0.005$. The left and the right interneurons mutually inhibit each other with inhibition gain $C_x = 7.5$.

As seen in the inhibitory term of Eq. (A.29), the left and the right interneurons provide mutual shunting inhibition. Therefore, excitation of the two interneurons from their associated bipolar flanks results in shunting normalization of the interneuron activities and thus normalization of the total inhibitory interneuron input to the bipole cell; see Eq. (A.21). This normalization of total inhibition from the interneurons allows the bipole cell to fire strongly even in the absence of direct bottom–up input. On the other hand, in the absence of bottom–up input to the bipole cell, and when only one of the two interneurons receives excitation from its associated bipolar flank, equal amounts of excitation through direct horizontal input to the bipole and inhibition from the interneuron prevents the bipole cell from firing. This mechanism prevents an outward spread of activity while enabling inward oriented completion of boundaries. The present formulation of the bipole cell-network is adapted from Ross et al. (2000) and from Grossberg and Swaminathan (2004).

A.6. Boundary-based attention, from V2 to LGN

As seen in Fig. 1, boundary-based attention is derived by summing the bipolar activities in V2 layers 2/3 over all orientations and scales. This modulatory boundary-based attention signal (Gove et al., 1995) amplifies LGN activity at grouped boundaries while suppressing their neighborhood by acting on the LGN surround kernels. The pooling across all scales ensures that features at the boundaries are eventually amplified in every spatial scale even when some scales are initially unsuccessful at detecting object boundaries. The LGN activity is defined in Eq. (A.4). The boundary-based attention signal $f_{ij}$ at position $(i,j)$ and scale $s$ is defined as:

$$
f_{ij} = -\alpha_{x_i,j} \beta_{x_i,j} + \sum_{k_s} [z_{i,j,k_s} - \gamma_y]^+ ,
$$

(A.30)

where passive decay rate $\alpha_{x_i,j} = 1$, and $[z_{i,j,k_s} - \gamma_y]^+$ is the input from the bipolar cells defined in Eq. (A.21), with $\gamma_y = 0.005$.

A.7. Surface filling-in domain, V4

Bipole signals added over all orientations and scales define a long-range figural boundary contour. If the boundary is a closed contour, the boundary-gated featural filling-in process can define a figural surface (Grossberg, 1994; Grossberg and Howe, 2003; Grossberg and Todorović, 1988). Surface filling-in activity $h_{ij}$ at position $(i,j)$ and scale $s$ obeys a boundary-gated diffusion equation:

$$
h_{ij} = -h_{ij} + \alpha_{x_{ij}}^+ + \beta_y \Phi (r_{ij}^+ )
$$

$$
+ \sum_{p \in N(i)} (h_{ip} - h_{pq} ) P_{pq_{ij}},
$$

(A.31)

where $x_{ij}^+$ is the LGN ON channel activity defined in Eq. (A.4), gain factor $\beta_y = 3$, $r_{ij}$ is the spatial attentional signal defined in Eq. (A.33), signal function $\Phi (x) = \frac{1}{1 + x}$, and the diffusion permeability $P_{pq_{ij}}$ is defined in Eq. (A.32). $N(i)$ in Eq. (A.31) is the set of eight nearest-neighbor cells adjacent to the cell at position $(i,j)$ with which the $(i,j)$ cell directly interacts during filling-in. The permeability $P_{pq_{ij}}$ that gates the diffusion process in Eq. (A.31) is defined by the equation:

$$
P_{pq_{ij}} = \frac{\lambda_p}{\alpha_p + \sum_k (z_{pq_{k}}^+ + z_{i,q_{k}}^+ )},
$$

(A.32)

with diffusion gain factor $\lambda_p = 21 \times 10^5$. Weber law parameter $\alpha_p = 3.3 \times 10^{-5}$, and $z_{i,q_{k}}$ is the bipole activity defined in Eq. (A.21).

A.8. Spatial attention, PPC

Fig. 1 shows connectivity of the spatial attention stage, which generates a form-fitting attentional shroud in response to feedback with the surface filling-in domain. In particular, this stage receives as bottom–up activation the surface filling-in domain output from V4, as well as a spatially localized top–down volitional attention signal. Specifically, spatial attention activity $r_{ij}$ at position $(i,j)$ is defined by the equation:

$$
r_{ij} = -\alpha_{x_{ij}} \beta_{x_{ij}} + ( 1 - r_{ij} ) \left( \sum_{q} \lambda_{x_q} \Phi \left( \sum_{p} h_{pq} \right) + \beta_{y} I_{pq} \right) G_{pq_{ij}},
$$

$$
- ( 1 - r_{ij} ) \left[ \sum_{q} \lambda_{x_q} \Phi \left( \sum_{p} h_{pq} \right) G_{pq_{ij}},
$$

(A.33)

where passive decay rate $\alpha_{x_i,j} = 1$. Bottom–up input signal $\Phi (h_{pq})$ is from the surface filling-in domain in Eq. (A.31) with signal function $\Phi (x) = x/(1 + x)$ and gain factor $\lambda_p = 100$. Term $I_{pq}$ is the volitional top–down attentional signal, modeled by a $3 \times 3$ pixel wide spot of activity assigned to a location inside the attended surface, with gain $\beta_{y} = 2$. The volitional attentional signal $I_{pq}$ locally enhances the spatial attentional activity. Through spatial attentional signal to the filling-in domain, this signal diffuses within the attended surface; see Eq. (A.31). The spread of volitional attention in the surface filling-in domain in turn enhances the activity of the entire attended form in Eq. (A.33) through bottom–up inputs. As the attended shroud grows due to volitional attention, other weaker shrouds are suppressed in Eq. (A.33) by long range spatial competition. A volitionally selected, form-filling spatial attentional shroud is thus formed in this stage. Gaussian kernels
and $G_{pqij}$ in Eq. (A.33) are defined in Eq. (A.3), with variances $\sigma_{r1} = 3$ and $\sigma_{r2} = 40$, respectively.

**A.9. dART-based texture learning, classification, and top-down attentive matching**

Input to the dART network is a vector that combines spatially pooled layer 6 ON simple cell signals $a_{ijks}$ defined in Eq. (A.10), and OFF simple cell signals $1 - a_{ijks}$. These ON and OFF cell signals normalize to pairwise add to 1. In ART, this preprocessing stage is called complement coding (Carpenter et al., 1991b). Complement coding ensures that the input feature vector of length $2M$ adds to a constant $M$. Specifically, the complement-coded dART input feature vector $A_{ij}$ at position $(i,j)$ is:

$$A_{ij} = \left[ \sum_{pq} a_{pqij}^+ G_{pqij}(s_i) \cdot \sum_{pq} [a_{pqij}]^+ G_{pqij}(s_i) \right],$$

$$+ \sum_{pq} [a_{pqij}]^+ G_{pqij}(s_i) \cdot \sum_{pq} a_{pqij}^+ G_{pqij}(s_i),$$

$$1 - \sum_{pq} a_{pqij}^+ G_{pqij}(s_i),$$

$$1 - \sum_{pq} [a_{pqij}]^+ G_{pqij}(s_i) \right],$$

(A.34)

where $[a_{ijks}]^+$ are output signals from V1 layer 6 simple cells defined in Eq. (A.10) and the Gaussian kernel $G_{pqij}(s_i)$ is defined in Eq. (A.3), with spatial scale dependent variance $\sigma(s_i) = \{1, 2, 3\}$. In Eq. (A.34), $K = 24$ is the number of simple cell orientations, and $S = 3$ is the number of scales. The dART input is thus a 144-dimensional complement-coded feature vector.

The importance of spatial pooling of simple cells in texture classification was investigated by Grossberg and Williamson (1999), where they observed that larger spatial pooling extents resulted in better texture classification, but also caused classification errors at texture boundaries due to feature mixing. In dARTEX, the detrimental effect of feature mixing at texture boundaries is circumvented through spatial attentional modulation of dART inputs, as described in Eq. (A.10).

Section B.2 describes how the dART network responds to complement coded input vectors. In response to this bottom-up input, the dART network reads out top-down learned expectations of simple cell activities, which are described next.

The dART top-down simple cell expectation $\sigma_{ijks}$ at position $(i,j)$, orientation $k$, and scale $s$ is (Carpenter, 1997):

$$\sigma_{ijks} = \sum_{n=1,N} [y_{nij} - \tau_{nks}]^+, $$

(A.35)

where the summation is over all $N$ categories in the dART network. In the dART network, $y_{nij}$ is the $n$th category cell activity at position $(i,j)$ defined in Eqs. (B.10) and (B.11). Learned feature threshold $\tau_{nks} = 1 - w_{nks}$ in Eq. (A.35) is derived from the $n$th dART category weight $w_{nks}$ for orientation $k$ and scale $s$; also see Section B.1, Step 2.

**A.10. Texture boundary contrast measure for the OBTS task simulation**

The texture boundary salience measure $S$ for comparing simulation results to the psychophysical data of Ben-Shahar and Zucker (2004) is defined by the equation:

$$S = \mu \frac{E_{Lower} - E_{Upper}}{v + E_{Lower} + E_{Upper} + F} + 50,$$

(A.36)

where $\mu = 67$ and $v = 0.13$. The pooled boundary activities at the two possible locations of the horizontal discontinuity figure boundaries $E_{Lower/Upper}$ are defined in Eq. (A.37). The factor $F$ is the average bipole activity over the whole image, defined in Eq. (A.40). The above equation arises from a shunting on-center, off-surround network (Grossberg, 1973, 1980) that detects pooled boundary contrast at two possible figure boundary locations. The term $F$ in Eq. (A.36) normalizes the measure and requires that, in order to be salient, the attended boundary contrast must be stronger than the average boundary activity in the scene. For dARTEX simulations, the horizontal boundary was always located at the lower location. Therefore, any increase in the measure in Eq. (A.36) denotes an increased detection accuracy.

The boundary activity $E_{Lower/Upper}$ in Eq. (A.36) is derived by pooling the bipole activities using a product of two kernels. The first kernel determines the spatial region of pooling, while the second determines the relative weights applied to each boundary orientation during pooling. Specifically, the pooled boundary activity at the $E_{Lower/Upper}$ is defined as:

$$E_{Lower/Upper} = \sum_{ijks} [z_{ijks} - \gamma E] D_{ij}^{Lower/Upper} O_{jk},$$

(A.37)

where $[z_{ijks} - \gamma E]^+$ is the output signal from the bipole cell in Eq. (A.21) with $\gamma = 0.005$. The spatial pooling kernel $D_{ij}^{Lower/Upper}$ is a horizontal band centered at the either the lower or the upper possible boundary location, respectively, and defined in Eq. (A.38). The second term, $O_{jk}$, defined in Eq. (A.39), is the orientational pooling kernel that determines the weights of different orientations during pooling.

The spatial pooling kernel $D_{ij}^{Lower/Upper}$ is horizontally elongated over the whole image and is centered at a possible boundary location (either Lower or Upper) with a Gaussian fall-off in the vertical direction. Specifically, the spatial pooling kernel $D_{ij}^{Lower/Upper}$ at vertical image position $j$ is:

$$D_{ij}^{Lower/Upper} = \frac{1}{98 \sqrt{2 \pi} \sigma D} \exp \left( -\frac{(j - q_{Lower/Upper})^2}{2 \sigma D^2} \right),$$

(A.38)

where variance $\sigma D = 15$. The vertical center of this kernel is either at the lower boundary location (mean = $q_{Lower}$) or at the upper boundary location (mean = $q_{Upper}$). The orienta-
tion pooling kernel \(O_{kr}\) in Eq. (A.39) is either a uniform normalized kernel for all \(k\), or a Gaussian centered at orientation \(R\):

\[
O_{kr} = \frac{1}{\sqrt{2\pi}\omega_0} \exp \left(-\frac{(k-R)^2}{2\omega_0^2}\right),
\]

(A.39)

where variance \(\omega_0 = 3.175\). Parameter \(R\) is chosen to be the horizontal orientation, being the maximally active orientation at the attended boundary locations. Term \(F\) in Eq. (A.36) adds all bipolar output signals:

\[
F = A_F \sum_{ijk} [\gamma_E^i - \gamma_E^j]^+, \tag{A.40}
\]

where coefficient \(A_F = 2.3 \times 10^{-6}\), \(\gamma_E\) is the bipolar activity described in Eq. (A.21), and the output threshold \(\gamma_E = 0.005\). The term \(F\) penalizes the salience measure in Eq. (A.36) whenever average boundary activity in the scene is high.

Appendix B. Default ARTMAP implementation

The default ARTMAP algorithm (Carpenter, 2003) implements the dART network in the dARTX model presented here. It is a special case of the distributed ART algorithm described in Carpenter (1997). In the default ARTMAP implementation, supervised training occurs in a Winner-Take-All (WTA) mode where the maximally active category node grows with learning, and its size \(C\) is determined by learning task demands. For a complete description of how this occurs in the distributed ART framework, see Carpenter (1997). Description of the default ARTMAP in this section is limited to essential implementation details. For further implementation information, see Carpenter (1997) and Carpenter et al. (1998).

B.1. Default ARTMAP training

During training, default ARTMAP learns to associate an \(M\)-dimensional feature vector \(a\) to a supplied class label. The training algorithm is described by the following steps:

1. For all positions \((i,j)\) to be learned, generate complement coded input feature vectors \(A_{ij}\) defined in Eq. (A.34).

2. Set initial values: Set the initial feature-to-category weight \(w_{kn} = 1\) from input feature of orientation \(k\) and scale \(s\) to category \(n\). Also set the initial category-to-feature weight \(w_{nks} = 1\) from category \(n\) to input cell of orientation \(k\) and scale \(s\). Set initial class weights \(W_{mn} = 1\) from each category \(n\) to each class label \(m\). Set number of committed category nodes \(C = 1\). For the Winner-Take-All (WTA) mode of learning utilized here, the bottom-up feature-to-category weights \(w_{kn}\) and the top-down category-to-feature weights \(w_{nks}\) remain identical (Carpenter, 1997). Both these weight vectors for category n are therefore denoted by \(w_n\).

3. Select the first input vector in the dataset, \(A_{ij}\), at position \((i,j)\), with associated output class \(K\).

4. Set initial weights for the newly committed category node \(n = C\), where \(C = 1\) is the latest committed category node index:

\[
w_C = A_{ij}, \tag{B.1}
\]

and set the class weight from the category node \(C\) to output class \(K\):

\[
W_{ck} = \begin{cases} 1 & \text{if } k = K, \\ 0 & \text{if } k \neq K. \end{cases} \tag{B.2}
\]

5. Set vigilance \(\rho\) to its baseline value \(\hat{\rho} = 0.6\):

\[
\rho = \hat{\rho}, \tag{B.3}
\]

and reset the code at position \((i,j)\):

\[
y_{ij} = 0. \tag{B.4}
\]

6. Select input vector \(A_{ij}\) at the next position \((i,j)\), with associated actual output class \(K\) (until the last input of the last training epoch).

7. At position \((i,j)\), calculate signals \(T_{nij}\) to committed category nodes \(n = 1, \ldots, C\) using the choice-by-difference signal function (Carpenter and Gjaja, 1994):

\[
T_{nij} = |A_{ij} \wedge w_n| + (1 - x)(M - |w_n|), \tag{B.5}
\]

where \(w_n\) is the weight vector for category \(n\). Note that for an uncommitted node \(|w_n| = 2M\), so \(T_{nij} = xM\). Small values of the signal rule parameter \(x\) cause the system to select category nodes that would minimize learned changes in \(w_n\) during learning. For present simulations, \(x = 0.075\).

8. Search order: Search the committed categories with \(T_{nij} > xM\) in order of \(T_{nij}\) values (max to min) for position \((i,j)\).

9. Search for a category \(R\) that meets the matching criterion and predicts the correct output class \(K\), as follows:

a. Code: For the next category node \(R\) that meets the matching criterion

\[
\left(\frac{|A_{ij} \wedge w_R|}{M} > \rho\right), \text{ set } y_{Rij} = 1 \text{ (winner-take-all).} \tag{B.6}
\]
B.2. Default ARTMAP testing (distributed code)

1. For all positions \((i,j)\) to be classified, generate complement coded input feature vectors \(A_{ij}\) defined in Eq. (A.34).
2. Select input vector \(A_{ij}\) at the next position \((i,j)\).
3. Reset the category code at position \((i,j)\): \(y_{ij} = 0\).
4. Calculate input signals \(T_{mij}\) to all categories \(n\) at position \((i,j)\) according to the choice-by-difference signal function (Carpenter and Giaja, 1994):

\[
T_{mij} = |A_{ij} \wedge w_n| + (1 - \beta)|M - |w_n|\|.
\]  

(8.9)

During testing, the signal rule parameter \(\beta\) is set to 0.075, same as during training. Recall that the signal \(T_{mij}\) to each uncommitted node \(n\) is the constant \(\beta M\).
5. \(A\) is the set of indices of categories firing above threshold, and \(A'\) is the set of indices of categories exactly matching the input. In particular, \(A = \{\lambda = 1, \ldots, C: T_{mij} > \beta M\}\) and \(A' = \{\lambda = 1, \ldots, C: T_{mij} = M\} = \{\lambda = 1, \ldots, C: w_n = A_{ij}\}\).
6. Increased Gradient (IG) CAM Rule: A Content Addressable Memory (CAM) rule specifies a function that characterizes the steady-state activity of the short term memory response to a given vector of inputs converging on a field of neurons. The Increased Gradient (IG) CAM rule used here contrast-enhances the input differences as represented in the distributed category code (Carpenter, 1997; Carpenter et al., 1998).

a. Point box case occurs when at least one category exactly encodes the input and the set \(A'\) is not empty. In this case, only such categories are activated: If \(A'\) is not empty (i.e., \(w_n = A_{ij}\) for some \(n\)), set

\[
y_{mij} = \frac{1}{|A'|} \quad \text{for each } n \in A'.
\]

(B.10)

b. In cases other than a point box case, when \(A\) is not empty, a distributed category activation of coding neurons is realized with the best matching categories activating the most:

\[
y_{mij} = \frac{1}{\sum_{n \in A} \frac{1}{|M - T_{mij}|^p}} \quad \text{for each } n \in A,
\]

(B.11)

where \(T_{mij}\) is the input signal to category \(n\) at position \((i,j)\), defined in Eq. (B.9). In Eq. (B.11), power law parameter \(p = 5\) determines the amount of code contrast enhancement. As \(p\) increases, the category activation increasingly resembles a winner-take-all code in which only the category with highest bottom-up signal survives (Carpenter, 1997). Constant \(M\) in Eq. (B.11) is the number of input features to dART.
7. Calculate distributed output class predictions for position \((i,j)\):

\[
\psi_{mij} = \sum_{n=1}^{C} W_{nm}y_{mij}.
\]

(B.12)

8. Predict output classes from \(\psi_{mij}\) values for position \((i,j)\):

\[
K_{ij} = \arg \max_{n} \psi_{mij}.
\]

(B.13)

9. While there are more test inputs, return to Step 2.

References


