Pattern recognition involves building a mental model to interpret incoming inputs. This incoming information is often incomplete and the mental model must extrapolate to complete the patterns, a process that is constrained by the statistical regularities in nature. Examples of pattern completion involve identification of objects presented under unfavorable luminance or interpretation of speech corrupted by acoustic noise.
Recognition of Occluded Objects

Hanlin Tang and Gabriel Kreiman

Pattern recognition involves building a mental model to interpret incoming inputs. This incoming information is often incomplete and the mental model must extrapolate to complete the patterns, a process that is constrained by the statistical regularities in nature. Examples of pattern completion involve identification of objects presented under unfavorable luminance or interpretation of speech corrupted by acoustic noise. Pattern completion is also at the heart of other high-level cognitive phenomena including our ability to discern actions from still images or to predict behavioral patterns from observations.

Pattern completion constitutes a ubiquitous challenge during natural vision. Stimuli are often partially occluded, or degraded by changes in illumination and contrast. While much progress has been made towards understanding the mechanisms underlying recognition of complete objects, the neural computations underlying more challenging recognition problems such as object occlusion remain poorly understood. Here we generically refer to object occlusion to include any transformation where only partial information about an object is accessible (such as the multiple examples illustrated in Fig. 1), and refer to object completion as the ability to infer object identity from partial information (without necessarily implying that subjects perceptually fill in the missing information).

Understanding how the neural representations of visual signals are modified when objects are occluded is critical to developing biologically constrained computational models of occluded object recognition and may also shed light on how to solve manifestations of pattern completion in other domains. The development of feed-forward models for visual recognition of whole objects has been driven by behavioral and physiological experiments establishing the hierarchy of feature tuning and robustness to image transformations. Similarly, by systematically examining when and where neural representations that are robust to occlusion
emerge can help extend our theoretical understanding of vision and develop the
next generation of computational models in vision. Understanding pattern com-
pletion and recognition of occluded objects is a challenging task: performance
depends on the stimulus complexity, the type and amount of occlusion and the task
itself. In this chapter, we summarize recent efforts to examine the mechanisms
underlying recognition of partially occluded objects and discuss several avenues for
future work towards a theory of object completion.

1 Visual System Hierarchy

Object recognition is orchestrated through a semi-hierarchical series of processing
areas along ventral visual cortex (Connor et al. 2007; DiCarlo et al. 2012; Felleman
and Van Essen 1991; Logothetis and Steinberg 1996; Riesenhuber and Poggio
1999; Schmolesky et al. 1998; Tanaka 1996). At each step in this hierarchy, the
feature specificity of the neurons increases in complexity. For example, neurons in primary visual cortex (V1), respond selectively to bars of a particular orientation (Hubel and Wiesel 1959) whereas neurons in inferior temporal cortex respond preferentially to complex shapes including faces and other objects (Desimone et al. 1984; Gross et al. 1969; Perrett et al. 1992; Richmond et al. 1983; Rolls 1991). In addition to this increase in feature complexity, there is a concomitant progression in the degree of tolerance to object transformations such as changes in object position or scale (Hung et al. 2005; Ito et al. 1995; Logothetis et al. 1995). The selective and tolerant physiological responses characterized in the macaque inferior temporal cortex have also been observed in the human inferior temporal cortex (Allison et al. 1999; Liu et al. 2009). The timing of these neural responses provides important constraints on the number of possible computations involved in visual recognition. Multiple lines of evidence from human psychophysical measurements (Potter and Levy 1969; Thorpe et al. 1996), macaque single unit recordings (Hung et al. 2005; Keysers et al. 2001), human EEG (Thorpe et al. 1996) and human intracranial recordings (Allison et al. 1999; Liu et al. 2009) have established that selective responses to whole objects emerge within 100–150 ms of stimulus onset in the highest echelons of the ventral visual stream.

Research over the last several decades characterizing the spatiotemporal dynamics involved in the neural representation of objects in these successive areas has led to the development of a theoretical framework to explain the mechanisms underlying object recognition. An influential theoretical framework suggests that, to a first approximation, processing of visual information traverses through the ventral stream in a feed-forward fashion, without significant contributions from long top-down feedback loops or within-area recurrent computations (Deco and Rolls 2004; Fukushima 1980; LeCun et al. 1998; Mel 1997; Olshausen et al. 1993; Riesenhuber and Poggio 1999; Wallis and Rolls 1997). Consistent with this notion, computational models of object recognition instantiating feed-forward processing provide a parsimonious explanation for the selectivity and tolerances observed experimentally (Serre et al. 2007b). The activity of these computational units at various stages of processing captures the variability in the neural representation from macaque single unit recordings along the visual hierarchy (Cadieu et al. 2014; Yamins et al. 2014). These feed-forward computational models have inspired the development of deep convolutional networks that demonstrate a significant degree of success in a variety of computer vision approaches to object recognition (e.g. Hinton and Salakhutdinov 2006; LeCun et al. 1998; Russakovsky et al. 2015; Sun et al. 2014; Taigman et al. 2014).

These purely feedforward architectures do not incorporate any feedback or recurrent connections. However, at the anatomical level, feedback and recurrent connections figure prominently throughout the visual system (Felleman and Van Essen 1991). In fact, quantitative anatomical studies have suggested that feedback and recurrent connections significantly outnumber feedforward ones (Callaway 2004; Douglas and Martin 2004). The computational contributions of these feedback and recurrent projections are largely underexplored in existing computational models of visual recognition because their underlying roles remain unclear. Several
investigators have suggested that these feedback and recurrent projections could play an important role during object recognition under conditions where the visual cues are impoverished (e.g. poor illumination, low contrast) or even partially missing (e.g. visual occlusion) (Carpenter and Grossberg 2002; Hopfield 1982; Mumford 1992; Tang et al. 2014; Wyatte et al. 2012a).

Following the approach suggested by Marr in his classic book on vision (Marr 1982), we subdivide our discussion of recognition of occluded objects into three parts: (i) definition of the computational problem by describing behavioral performance during recognition of occluded objects, (ii) characterization of the implementation at the physical level describing the neural responses to occluded objects and (iii) initial sketches of theoretical ideas instantiated into computational models that aim to recognize occluded objects.

2 The Computational Problem of Object Completion

Figure 1 shows examples of several images that induce object completion. In the natural world, objects can be partially occluded in multiple different ways due to the presence of explicit occluders, shadows, camouflage and differential illumination. Object completion is an ill-posed problem: in general, there are infinite ways of completing contours from partial information. The visual system must be able to infer what the object is despite the existence of all of these possible solutions consistent with the visual input.

2.1 Amodal Completion

Occluded shapes can be perceived as whole (Fig. 1a, compare left and right panels). Object completion can be amodal when there is an explicit occluder and the subject cannot see the contours behind the occluder despite being aware of the overall shape (Singh 2004). In contrast, in the famous illusory triangle example (Fig. 1b), Kanizsa describes the phenomenon known as modal completion whereby the object is completed by inducing illusory contours that are perceived by the observer (Kanizsa 1979). Because these illusory inducers are rare in natural vision, in this chapter we focus on amodal completion. Even though occluded or partial objects such as the ones shown in Fig. 1c, d are segmented, observers view the object as a single percept, not as disjointed segments. Amodal completion is also important for achieving this single `gestalt’. Investigators have used a variety of different stimuli to probe the workings of object completion, ranging from simple lines and geometric shapes to naturalistic objects such as the ones shown in Fig. 1e–g.

Psychophysical studies of amodal completion have provided many clues to the underlying computations (Kellman et al. 2001; Sekuler and Murray 2001). Amodal completion relies on an inferred depth between the occluder shape and the occluded
object, which in turns generates a surface-based representation of the scene (Nakayama et al. 1995). In fact, presence of the occluder aids in identifying the occluded object, as powerfully illustrated by the Bregman’s occluded B letters (Bregman 1981) Grouping of different parts into a complete whole, and the ‘completion’ of missing lines and contours, represent an important component of object recognition. There are infinite possible ways of completing occluded objects. The ambiguities arise from the many combinations with which occluded edges, called ‘inducers’ can be paired together, as well as the infinite possible contours between two pairs of inducers (Kellman et al. 2001; Nakayama et al. 1995; Ullman 1976). Despite the many possible solutions, the visual system typically arrives at a single (and correct) interpretation of the image.

The temporal dynamics of shape completion can constrain the computational steps involved in processing occluded images. Psychophysics experiments have measured the time course of amodal completion with a diverse array of experimental paradigms. The most common method is a contrast of an occluded shape against its mosaic parts (e.g. Figure 1a). For example, in the prime matching paradigm, subjects are first primed with a stimulus, and then asked to judge whether a pair of test stimuli represent the same or a different shape. Subjects are faster to correctly respond ‘same’ when the primed shape is the same as the test stimuli. When partly occluded objects are used as the prime, this priming effect depends on the exposure time (Sekuler and Palmer 1992). At short durations (50 ms), occluded objects primed subject’s responses towards mosaic shapes, suggesting that 50 ms is not enough time for amodal completion of the prime stimulus. At longer durations (100 ms or more), the priming effect switched to favor whole shapes. Therefore, the authors estimate that amodal completion for simple geometric shapes takes between 100 and 200 ms, depending on the amount of occlusion (Sekuler et al. 1994).

A different set of behavioral experiments suggests approximately the same time scales for amodal completion: in several studies, subjects are asked to discriminate shapes in a timed forced-choice task. Response times to occluded shapes lagged those to whole shapes by about 75–150 ms (Murray et al. 2001; Shore and Enns 1997). For naturalistic objects such as faces, however, Chen et al. report that amodal completion takes longer than 200 ms (Chen et al. 2009), well beyond the previous estimates based on simpler shapes and tasks.

2.2 From Amodal Completion to Recognition of Occluded Objects

Many studies of amodal completion have used simple shapes and contours, as outlined in the previous section. When recognizing whole objects, psychophysical studies using simple shapes and neurophysiological studies describing the linear filters in primary visual areas led to the basic building blocks to develop deep models of visual recognition capable of detecting complex shapes. Inspired by the
Fig. 2 Robust object categorization despite strong occlusion. a) Experiment timeline. Partial images similar to the ones shown in Fig. If were presented during exposure times ranging from 33 to 150 ms. Subjects performed a five alternative forced-choice categorization task determining whether the image contained animals, chairs, faces, fruits, or vehicles. b) Performance as a function of percentage occlusion across n = 14 subjects for various exposure times. Error bars denote SEM. Dashed line indicates chance performance.

success of this approach, we assume that contour completion is one of the initial steps involved in interpreting complex objects that are partially occluded.

Recognition remains robust to partial occlusion for complex objects. We used naturalistic objects that were occluded by presenting information through “bubbles” (Gosselin and Schyns 2001) (Fig. If). After a variable exposure time from 33 to 150 ms, subjects performed a five alternative forced-choice categorization task. Recognition was robust even when 80–90% of the object was occluded across the various exposure times (Fig. 2). Similar results were obtained by (Wyatte et al. 2012a). As illustrated by Bregman (1981), the presence of an occluder during object completion aids recognition performance. For example, one study presented natural objects that were either occluded (Fig. 1e, left) or where the same object parts were deleted (Fig. 1e, right). Recognition was significantly impaired in the deleted part case compared to the occluded part case, but only when using high percentage of occlusion (> 75% missing pixels) (Johnson and Olshausen 2005).

Intuitively, we would expect successful recognition to depend on the exact features shown. This intuition was quantitatively measured by (Gosselin and Schyns 2001): the facial features critical to recognition varied depending on the task (e.g. the eyes and eyebrows were more relevant for gender discrimination and the mouth provided more information when evaluating expressiveness). Similar conclusions were reached when using the same paradigm to evaluate recognition in monkeys (Nielsen et al. 2006).
3 Neural Representation of Occluded Objects

A series of scalp electroencephalography (EEG) studies have measured the latency at which responses differ between occluded objects and suitable control images. Using simple geometric stimuli, differences between occluded shapes and notched shapes emerged at 140–240 ms (Murray 2004). Using more naturalistic stimuli (e.g. Figure 1e), other investigators report differential activity in the 130–220 ms (Chen et al. 2010) and 150–200 ms (Johnson and Olshausen 2005) ranges. In a more difficult task with fragmented line drawings that are progressively completed, Doniger et al. reported that differences are only observed in the 200–250 ms response window. Even though these studies use different stimuli and make different comparisons, they consistently conclude that amodal completion effects manifest within 130–250 ms.

Several neuroimaging (Hegde et al. 2008; Komatsu 2006; Lerner et al. 2004, 2002; Olson et al. 2004; Rauschenberger et al. 2004) and scalp EEG (Chen et al. 2010; Doniger et al. 2000; Johnson and Olshausen 2005) studies with more complex objects have contrasted activity changes between an occluded object and an appropriately scrambled counterpart (e.g. Fig. 1c, d). In these stimuli, the low-level features are maintained but disruption in their geometric arrangement destroys the percept. For example, investigators have reported differential activity in the lateral-occipital complex between occluded line drawings and their scrambled counterparts (Lerner et al. 2002). The authors reason that, since the occluded images elicit a larger response in the lateral-occipital complex (LOC) than scrambled images, the LOC could be involved in object completion. It should be noted that LOC also demonstrates increased activity to whole objects compared to scrambled versions of those objects without any occlusion (Grill-Spector et al. 2001). Thus, the increased responses to whole objects may not be necessarily related to object completion mechanisms per se, but rather neural activity related to perceptual recognition.

Similarly, EEG and intracranial studies compared line drawings against their fragmented counterparts (Doniger et al. 2000; Sehatpour et al. 2008) to measure the timing and brain regions involved in object completion. Sehatpour et al. worked with epilepsy patients who have intracranial electrodes implanted for clinical purposes. This approach allows a rare opportunity to record directly from human brain. The authors take advantage of simultaneous recordings from multiple brain regions to show that line fragments elicited greater coherence in the LOC-Prefrontal Cortex-Hippocampus network compared to scrambled line fragments. They suggest that this network synchrony is responsible for the perceptual line closure of objects.

In order to understand the neural mechanisms orchestrating object completion, it is also critical to examine the neural architectures that could implement the computational solutions suggested in the previous section. Essential aspects of object completion can be traced back to the earliest stages in visual processing. An early study demonstrated that neurons in area V2 show selective responses to illusory contours (Peterhans and von der Heydt 1991; von der Heydt et al. 1984). Other
work has demonstrated that even V1 neurons can respond to occluded shapes. One study recorded single cells in macaque V1 when presented with occluded moving bars (Sugita, 1999). Approximately 12% of orientation-selective cells responded to the moving oriented bar even when it was occluded, thus potentially underlying the phenomenology of amodal completion. These cells responded strongly only when the occluder was presented in front of the moving bar (positive disparity), and not at zero or negative disparity. Notably, responses to the occluded bar were not different from those obtained when presenting the bar alone. These results have led to the suggestion that amodal completion is achieved by contextual modulation from outside the classical receptive field. While other studies have suggested that contextual modulation occurs with a delay of 50–70 ms with respect to the onset of the visually evoked responses (Bakin et al., 2000; Zipser et al., 1996), Sugita did not observe any latency delays for the amodally-completed response. The author suggests that these contextual modulations may come from lateral connections or feedback from proximal areas. In another study, responses to illusory contours in V1 were delayed by about 55 ms compared to the response to real contours (Lee and Nguyen, 2001). Importantly, illusory contour responses appeared first in V2 before emerging in V1, suggesting that modal completion in V1 might require feedback modulation from V2. Complementing these studies, psychophysical studies on the effect of inferred depth and apparent motion on the perception of occluded surfaces also conclude that amodal completion effects manifests in early visual processing (Shimojo and Nakayama, 1990a, b).

These neurophysiology studies have focused on the occlusion or inducing of linear contours, where the inducers are close in proximity to the classical receptive field. However in natural vision we complete curvilinear contours over distances much longer than the width of classical V1 receptive fields. Often in these cases, correct completion of an object depends on the global context in which the object is embedded. Future studies are needed to examine whether and when V1 neurons respond to completed contours of varying curvature, length, and context.

As outlined above, V1 neurons feed onto a cascade of semi-hierarchical processing steps through V2 and V4, culminating in the inferior temporal cortex (ITC) (Felleman and Van Essen, 1991). How do these higher visual areas respond to occluded shapes? Few studies have examined the responses in intermediate visual areas to occluded shapes. A recent elegant study has begun to fill in this gap by characterizing how macaque V4 neurons respond to different curvatures when they are partially occluded by dots (Kosai et al., 2014). The authors report that neurons can maintain selectivity within a range of occlusion. While the response onset times of these neurons were not delayed by occlusion, the latency at which selectivity arose was delayed by hundreds of milliseconds.

Kovacs et al. found that visually selective responses to complex shapes in ITC were similar between whole shapes and occluded shapes defined by adding noise, occluders or deleting shape parts (Kovács et al., 1995). Although selectivity to complex shapes was retained despite up to ~50% occlusion, the absolute magnitude of the responses was modulated linearly with the amount of occlusion. Contrary to what Kosai et al. found in V4, these authors observed delays of up to 50 ms
in the response latency of occluded shapes. While it is tempting to attribute this discrepancy to differences in processing between V4 and IT, we note that the stimuli and occluding patterns used are different between the two studies.

Nielsen et al. extended this work by examining the responses of IT neurons to occluded objects embedded in naturalistic stimuli (Fig. 1f) (Nielsen et al. 2006a). Using the bubbles paradigm (Gosselin and Schyns 2001), the authors defined parts of an image that provided more diagnostic value (i.e. provided information that aided recognition) versus other non-diagnostic parts. The authors first demonstrated that monkeys and humans show striking behavioral similarities in terms of what object parts are considered diagnostic (Nielsen et al. 2006b). For occluded scenes containing diagnostic parts, both firing rates and local field potentials in ITC remained largely invariant to significant amounts of occlusion, in contrast to the findings of the Kovacs study with simpler stimuli (Kovacs et al. 1995). However, for scenes that contained only non-diagnostic parts, the results from the Kovacs study were reproduced—the firing rate varied linearly with the amount of occlusion. This comparison also serves as a cautionary tale against extrapolating results based on geometric shapes to the processing of more naturalistic stimuli because the details of which features are revealed can play a very important role in dictating the effects of occlusion. Issa et al. reached similar conclusions when demonstrating that ITC responses selective to faces were particularly sensitive to occlusion of certain parts (one eye) and that those parts could drive the responses almost as well as the whole face (Issa and Dicarlo 2012). These results suggest that the robustness of the neural representation to missing parts depends on the diagnosticity of the visible features.

Tang et al. used intracranial recordings to evaluate how and when visually selective responses to occluded objects emerge. Naturalistic objects were presented through the bubbles paradigm (Gosselin and Schyns 2001) in a task similar to the one illustrated in Fig. 2. The use of objects seen through bubbles evaluates the core ability to spatially integrate multiple parts to subserve recognition. Figure 3 (left) shows the responses from an example electrode in the fusiform gyrus that displayed a strong response to face stimuli. Remarkably, the electrode showed similar responses to images that displayed only $\sim$11% of the object content (Fig. 3, columns 2–6). Even in cases where the images shared essentially no common pixel (e.g. columns 3 and 5), the responses remained similar. Overall, the magnitude of the responses did not vary with the amount of visible pixels. Yet, the responses to occluded shapes were not identical to those obtained upon presenting whole images. A notable difference was that the responses to occluded objects were significantly delayed (compare the position of the arrows in Fig. 3). Selective neural response emerged with a delay of $\sim$100 ms. These delays were also apparent when using a machine learning approach to decode the category or identity of the objects from the physiological responses to the whole or occluded images (Fig. 4).

Image processing does not end with visual cortex. Information from visual cortex is conveyed to frontal cortex and to medial temporal lobe structures including the amygdala, hippocampus and entorhinal cortex. One group recorded single unit activity in the amygdala of human epilepsy patients, and found that
neurons were surprisingly sensitive to even small degrees of occlusion (Rutishauser et al. 2011). Their firing rates varied non-linearly with the amount of occlusion and the responses to image parts did not necessarily bear a resemblance to the responses to the whole images. These non-visual medial temporal lobe neurons, in effect, ‘lost’ the robustness developed in the ventral visual stream, in that individual parts were not sufficient to drive the response to the level observed for whole faces.

In addition to the type of spatial integration demonstrated in the studies above, the visual system is able to integrate information over time. Temporal integration is particularly prominent in examples of action recognition. Yet, in some cases, different parts of the same object may appear in a dynamic fashion over time. At the
behavioral level, presenting object parts asynchronously significantly disrupts object recognition performance, even when the temporal lag is as short as 16 ms (Singer and Kreiman 2014). This disruption by asynchronous presentation is also evident at the physiological level (Singer et al. 2015). Thus, the ability to spatially integrate parts into a whole for recognition is quite sensitive to deviations from the synchronous presentation of those parts.

In sum, early visual areas show evidence of contour completion in the presence of both occluded and illusory contours when the corresponding edges are in close spatial proximity. In some, but not all cases, these contour completion responses show a delay with respect to both responses to real contours and responses in higher visual areas. In higher visual areas responsible for object recognition, physiological signals show strong robustness to large degrees of occlusion, consistent with behavioral recognition performance, and these physiological signals also show a significant delay. The robustness in the physiological responses and the dynamic delays are consistent with the behavioral observations. These delays are interpreted as originating from the involvement of recurrent and/or top-down connections during the process of object completion. Given these initial steps in understanding plausible neural circuits underlying recognition of occluded objects, we turn our attention to describing the possible biological algorithms instantiated by these neural signals.

4 Computational Models of Occluded Object Recognition

There has been significant progress over the last decade in developing computational models of object recognition (Deco and Rolls 2004; DiCarlo et al. 2012; Kreiman 2013; Riesenhuber and Poggio 1999; Serre et al. 2007a). To a first approximation, these models propose a hierarchical sequence of linear filtering and non-linear max operations inspired by the basic principles giving rise to simple and complex cells in primary visual cortex (Hubel and Wiesel 1962). Concatenating multiple such operations gave rise to some of the initial models for object recognition (Fukushima 1980). Recently, these ideas have also seen wide adoption in the computer science literature in the form of deep convolutional neural networks (e.g. (Krizhevsky et al. 2012) among many others). Both biologically inspired models and deep convolutional neural networks (CNNs) geared towards performance on benchmark datasets share similar core architectures.
4.1 Performance of Feed-Forward Models in Recognizing Occluded Objects

The canonical steps in feed-forward computational models is inspired by the observation of simple and complex cells in primary visual cortex of anesthetized cats. In their classic study, Hubel and Wiesel discovered ‘simple’ cells tuned to bars oriented at a particular orientation (Hubel and Wiesel 1959). They also described ‘complex’ cells, which were also tuned to a preferred orientation, but exhibited a degree of tolerance to the spatial translation of the stimulus. They hypothesized that to generate this spatial invariance, the complex cells pool over simple cells whose receptive files tile the visual space with a max-like operation. This complex cell would then respond to an oriented bar regardless of its spatial location. Both hierarchical models of biological vision such as HMAX (Riesenhuber and Poggio 1999; Serre et al. 2007b) and CNNs are composed of alternating layers of tuning and pooling with increasingly more complicated tuning functions as one ascends this hierarchy. Whereas biologically-inspired models such as HMAX have about 4 layers, state-of-the-art computer vision models have moved to complex topologies with up to 20 layers and different mixtures of tuning and pooling layers (e.g. Russakovsky et al. 2015). Performance of feed-forward models such as HMAX on object recognition datasets match the pattern of human performance (Serre et al. 2007b). Additionally, the activity of individual layers in deep learning networks can capture the variance of neurons in the corresponding layers in macaque cortex (Cadieu et al. 2014; Yamins et al. 2014).

While these feed-forward architectures are designed to build tolerance to image transformations such as position and scale changes, they are not necessarily robust to the removal or occlusion of object features. Indeed, Fig. 5 shows the performance of an HMAX-like architecture in recognizing the same occluded objects form Figs. 2, 3 and 4. Small amounts of occlusion do not impair performance. However, performance drops rapidly with increasing occlusion, much more rapidly than human performance (see “behavior” line in Fig. 5 and compare Fig. 2 versus Fig. 5. Experimental simulations with other similar models confirm that that both HMAX and CNN models are challenged by recognition of occluded objects (Pepik et al. 2015; Wyatte et al. 2012b). Unlike position or scale transformations, the underlying representation in these models is not robust to occlusion. Feed-forward networks do not have explicit mechanisms to compensate for the missing features of occluded objects. In addition, because these models do not distinguish between the occluder and the object, the occluder can introduce spurious features that, through spatial pooling, are mixed with the object features. With small amounts of occlusion, the remaining features may be sufficient to lead to successful classification. However, with increasing levels of occlusion, the lack of sufficient information and completion mechanisms lead to a significant impairment in performance.
4.2 Beyond Feed-Forward Models

Models that incorporate additional computations beyond feed-forward architectures can be subdivided into several categories. A group of computational models describes the process by which contours are amodally completed (Ullman 1976; Yuille and Kersten 2006). These models typically rely on an axiomatic set of desirable qualities that completed curves must satisfy (such as minimum total bending energy).

Another set of ideas argues that creating an understanding of the surfaces in a scene is critical to object recognition for occluded objects. Nakayama (1995) proposes a theory where surface representation is constructed via feedback in early visual cortex by learning associations between a viewed image and the underlying surface representations. The authors argue that this intermediate surface representation is vital for subsequent recognition for occluded objects and mediates many important functions in texture segregation and visual search.

A different model proposes that neural representations of surfaces are created in three stages based on the low-level features (Sajda and Finkel 1995). First, points belonging to the same contour are bound together, followed by a process that determines the surfaces, and finally the surfaces are ordered by depth. Both feed-forward and feedback connections subserve to communicate between these three stages.
Several theories based on the role of feedback connections emphasize inference, but these ideas have largely not been operationalized into object recognition models. Predictive coding models are generative models of object recognition (Rao and Ballard 1999). In these models, higher visual areas send their predictions to lower levels, which then return only the mismatch between the predicted activity and the actual activity. This creates an efficient system where each layer only sends forwards signals that deviate from the receiving layer’s predictions. The higher layers then attempt to generate the correct hypothesis of the image by reducing the incoming prediction errors. A related model proposes that visual cortex is essentially performing Bayesian inference where feed-forward inputs combine with top-down priors for recognition (Lee and Mumford 2003; Yuille and Kersten 2006).

As pointed out by (Wyatte et al. 2014), predictive coding models would expect occluded images to lead to increased activity in visual cortex, since the first generated hypotheses would have a very large prediction errors due to occlusion. Over time, we would expect this activity to subside as the system converged on an accurate hypothesis. Neurophysiological studies, however, find decreased or unchanged activity throughout visual cortex (Kovacs et al. 1995; Nielsen et al. 2006a; Tang et al. 2014).

Several models that more directly examine object recognition deal with recognition from partial information, and do not incorporate any of the amodal completion mechanisms previously described. These models take advantage of the extensive feedback and recurrent connections in visual cortex. While the role of these connections in attentional modulation has been extensively studied, their contribution to object recognition remains unclear. A particularly prominent and attractive class of models that can perform pattern completion is the all-to-all connectivity architectures such as Hopfield networks (Hopfield 1982). The Hopfield network generates attractors for previously learned patterns in such a way that if the network is initialized with partial information, it can dynamically evolve towards the right attractor. Interestingly, this type of dynamical convergence towards the attractor state could account for the type of delays observed in the behavioral and physiological experiments. This general principle is operationalized by a neural network model that combines bottom-up input with top-down signals carrying previously learned patterns to complete occluded objects (Fukushima 2005). For occluded patterns that are novel, this network attempts to interpolate from visible edges. The author applies this model to complete occluded letters of the alphabet. This concept has been extended to naturalistic objects with a feed-forward model that is augmented with recurrent feedback (Wyatte et al. 2012a). This recurrent feedback served to strengthen the feed-forward signals that were diminished from the occluded image.

Given that strong behavioral and neurophysiological evidence exists for amodal completion in human brain, and that surface representations are important for organizing the visual scene, theories of object recognition would be remiss to exclude these features in favor of purely feature-based recognition. An important step towards models that fully capture natural biological vision would be to integrate traditional feed-forward models with feedback mechanisms, including amodal
completion, surface generation, and top-down modulation based on priors and context. The challenge of recognizing occluded objects stands as the first test of these future integrative theories.

References


Recognition of Occluded Objects


Wyatte D, Jilk DJ, O’Reilly RC (2014) Early recurrent feedback facilitates visual object recognition under challenging conditions. Front Psychol 5:674
Author Query Form

<table>
<thead>
<tr>
<th>Query Refs.</th>
<th>Details Required</th>
<th>Author’s Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQ1</td>
<td>Please confirm if the section headings identified are correct.</td>
<td></td>
</tr>
<tr>
<td>AQ2</td>
<td>References ‘Nielsen et al. (2006), Serre et al. (2007)’ are cited in the text but not provided in the reference list. Please provide the respective references in the list or delete these citations.</td>
<td></td>
</tr>
<tr>
<td>AQ3</td>
<td>As Refs. [35] of Kovács (1995) and [36] of Kovács (1995) are same, we have deleted the duplicate references and renumbered accordingly. Please check and confirm.</td>
<td></td>
</tr>
</tbody>
</table>
Please correct and return this set

Please use the proof correction marks shown below for all alterations and corrections. If you wish to return your proof by fax you should ensure that all amendments are written clearly in dark ink and are made well within the page margins.

<table>
<thead>
<tr>
<th>Instruction to printer</th>
<th>Textual mark</th>
<th>Marginal mark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leave unchanged</td>
<td>⋮ ⋮ under matter to remain</td>
<td>ː</td>
</tr>
<tr>
<td>Insert in text the matter</td>
<td>/ through single character, rule or underline or / through all characters to be deleted</td>
<td>ː or Ⓞ</td>
</tr>
<tr>
<td>indicated in the margin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delete</td>
<td></td>
<td>ː or Ⓞ</td>
</tr>
<tr>
<td>Substitute character or substitute part of one or more word(s)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change to italics</td>
<td>— under matter to be changed</td>
<td></td>
</tr>
<tr>
<td>Change to capitals</td>
<td>≡ under matter to be changed</td>
<td></td>
</tr>
<tr>
<td>Change to small capitals</td>
<td>≡ under matter to be changed</td>
<td></td>
</tr>
<tr>
<td>Change to bold type</td>
<td>≃ under matter to be changed</td>
<td></td>
</tr>
<tr>
<td>Change to bold italic</td>
<td>≃ under matter to be changed</td>
<td></td>
</tr>
<tr>
<td>Change to lower case</td>
<td>Encircle matter to be changed</td>
<td></td>
</tr>
<tr>
<td>Change italic to upright type</td>
<td>(As above)</td>
<td></td>
</tr>
<tr>
<td>Change bold to non-bold type</td>
<td>(As above)</td>
<td></td>
</tr>
<tr>
<td>Insert ‘superior’ character</td>
<td>/ through character or / through letter or / through characters</td>
<td></td>
</tr>
<tr>
<td></td>
<td>where required</td>
<td></td>
</tr>
<tr>
<td>Insert ‘inferior’ character</td>
<td>(As above)</td>
<td></td>
</tr>
<tr>
<td>Insert full stop</td>
<td>(As above)</td>
<td></td>
</tr>
<tr>
<td>Insert comma</td>
<td>(As above)</td>
<td></td>
</tr>
<tr>
<td>Insert single quotation marks</td>
<td>(As above)</td>
<td></td>
</tr>
<tr>
<td>Insert double quotation marks</td>
<td>(As above)</td>
<td></td>
</tr>
<tr>
<td>Insert hyphen</td>
<td>(As above)</td>
<td></td>
</tr>
<tr>
<td>Start new paragraph</td>
<td>(As above)</td>
<td></td>
</tr>
<tr>
<td>No new paragraph</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transpose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Close up</td>
<td>linking characters</td>
<td></td>
</tr>
<tr>
<td>Insert or substitute space</td>
<td>/ through character or / through letter or / through characters</td>
<td></td>
</tr>
<tr>
<td>between characters or words</td>
<td>where required</td>
<td></td>
</tr>
<tr>
<td>Reduce space between characters or words</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>