Neurocomputing 314 (2018) 284–298

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Local and distributed concept representation via colimits: An example

Michael J. Healy*, Thomas P. Caudell

University of New Mexico, United States

ARTICLE INFO

Article history:
Received 2 December 2017
Revised 6 April 2018
Accepted 17 June 2018
Available online 30 June 2018

Communicated by Dr Wei He

Keywords:
Category
Colimit
Concept
Diagram
Distributed
Local

ABSTRACT

An issue in the representation by a neural network of objects detected by its sensors is whether such representations are based at a single neural cell or require a distributed subnetwork. We perform an analysis near the sensor level of a simple geometric shape and show that given the appropriate mathematical structure, representations can be both local and distributed. Moreover, as simulations demonstrate, using the proposed structures in neural network design leads to improved performance in object discrimination.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

In this paper we apply a mathematical model to address the issue of local versus distributed concept representation in neural networks. Can single neural cells represent objects or events (hereafter we shall use the term "objects" for both) or do objects require a population of cells? At one extreme, "local" means that *only* a single cell is required to represent an object, and at the other extreme "distributed" means that there is *no* single cell capable of signifying an object's presence. Is there a middle ground between these two extremes? Another issue is of no less importance, since the notion of representation inevitably leads to the question of whether neural networks are capable of symbolic processing. "Representation" implies some form of description of the objects represented by the structure and activity of a neural system, suggesting the more specific term "concept representation". This in turn is suggestive of symbolic structures such as syntactic formulae. But claiming that a neural network is capable of learning and manipulating symbolic expressions begs the question of whether any neural network has the properties required for manipulating symbolic representations. We mention this because the discussion of representation involves symbols, such as in language or graphics. Due to space limitations, however, we do not treat this second issue here. An analysis of these problems—exclusively defining connectionist systems as distributed, and the notion of representations as symbolic concepts—is presented in [1], with a later review of localist versus distributed representations presented in [2].

1.1. Is representation local or distributed?

The notion that representations of perceived objects or situations are encoded uniquely by single neurons in the brain is referred to as the "grandmother cell" or "localist" view. The complementary notion that representation requires the activity of a multi-cell population is known as the "distributed" or "population coding" view. Establishing which of these views is correct has been an issue in the related fields of neuroscience and neural network computer modeling. Some studies in neural science have provided evidence for the representation of physical objects, visual scenes and other sensorial stimuli by individual neurons in the cerebral cortex and associated areas such as the hippocampus [1–4]. Individual hippocampal neurons have been observed to fire in correspondence with specific combinations of an animal's behavior and location in an environment [5], which would seem to implicate single-cell encoding within a network of alternative representations of behaviors-in-place. On the other hand, Huth et al. [6] have found evidence that representations of stimuli such as objects and scenes occur in a continuum across the brain, such that entities sharing many properties are spaced closely together in the neural tissue while dissimilar entities are represented at distant areas. In this view, whole populations of cells in small brain regions are required for the representation of specific categories and, moreover, a population may represent more than one category, and the

* Corresponding author.
E-mail addresses: mjhealy@unm.edu (M.J. Healy), tcaudell@unm.edu (T.P. Caudell).

https://doi.org/10.1016/j.neucom.2018.06.052
0925-2312/© 2018 Elsevier B.V. All rights reserved.

representations change in a continuous fashion as one transitions across a sheet of cortical tissue. A variant of this view has it that the statistical properties of variations in activity across a cell population constitute the representations of different entities [7] by the population as a whole. Stansbury et al. have shown that natural scenes can be associated with individual voxels (hence, relatively small cell populations) in visual cortex based upon object co-occurrence statistics and voxel fMRI signals recorded when the scenes are presented to subjects [8]. Their results suggest that multiple scenes can be associated with a given voxel and also that the brain exploits scene statistical properties at multiple levels of abstraction.

In a recent survey of neural models, Bowers states [2] that connectionists who favor the distributed view often regard the localist view as the extreme position that a single neuron exclusively encodes an observed instance of an object in some configuration ("my grandmother, knitting by the fire"), with no support from additional neural cells—the "Grandmother Cell view". However, although there is much recent evidence for highly selective firing in some neurons in humans, there is no evidence for this extreme view. At the opposite extreme is the distributed connectionist network in which no cell represents symbolic information, that is, a meaningful item of information that may be expressed in human-understandable symbols. Only the activation of an entire subnetwork of cells represents a meaningful item in this view; the individual cells are sometimes said to be encoding "sub-symbolic" information.

Finally, there is a view that lies between the localist and distributed views, in which a single object or category is encoded by a cell population, but the population acts as if it were a single cell. This type of representation is distributed, but acts as a local representation because the entire population is active when, and only when, the aforementioned object or category is being represented; thus, it acts in a manner disjoint from the rest of the network and can be characterized symbolically. Apparently, the evidence supports both the localist and distributed views of object and category representation in the human brain. Indeed, some have suggested that population coding is really the interaction of local representations which provide inputs to a single cell that has a local representation that is based upon those of the other cells. Damasio in a classic paper proposed a model of the convergence across multiple brain regions of representations at local sites to form more complex multimodal or multi-featural representations further downstream in the processing of sensory data [9]. In [10], Barsalou proposed a detailed model combining representations recursively and proposed it as a neural network symbol system (the Perceptual Symbol System), emphasizing its cognitive representational capacities. Neither of these proposals commits to one or the other extreme, local or distributed, but both have the flavor of an intermediate position. Recent applications of text processing for document ranking, language translation and other tasks harness local-distributed artificial neural networks [11,12].

1.2. The contents herein

Our position lies within the intermediate view. In past writings, we have introduced the Categorical Neural Semantic Theory (CNST) (see for example [13–15]), a mathematical theory of the semantics of neural structure and activity based upon category theory. On the one hand, the CNST models a representation as local, represented by a single cell. However, this cell serves as a central site of organization for the represented object's separately-represented components, and also as a unique *indicator* for the object. To effect this, the indicator cell interacts with an interconnected array of cells which collectively provide the semantic details of the indicator cell's representation. In this sense, the representation is distributed as well as local. Tracing back through a hierarchy of such structures grounds these local-distributed representations at the level of sensor, actuator or autonomous pattern-generating cells depending upon the network's functional capabilities. The remaining sections of this paper are dedicated to making this case with an example at the sensor level. The resulting neural structure will be seen as a system of representations that is both local and distributed.

The organization of the paper is as follows. In Section 2 we introduce our main example and begin the discussion with local and distributed neural structures. Section 3 deepens the analysis with the mathematical considerations involved. Sections 4 and 5 explore the main theme of this paper, showing how one of the category-theoretic constructs used in the CNST, the *colimit*, addresses the problem by building complex representations as blended combinations of simpler ones. This results in a type of structure that can be used for understanding an object as an organized structure, a particular assembly of its components. We demonstrate the advantage of this analysis through computer simulations. Section 6 provides a discussion and the conclusion.

2. A neural vision system and its semantics

2.1. A single cell and concept representation

Consider a neural network equipped with a vision sensor. The design goal for the neural network is simple: to discriminate a particular kind of geometric object, a triangle, from other objects purely based upon brightness stimuli detected by the sensor elements, where each element supplies a pixel of the image. As a baseline for analysis, the simplest possible neural network would be a single neuron T. The sensor communicates with T through input connections emanating from the sensor elements. The receptive field (RF) of T is the entire scope of the vision sensor. If T receives an aggregate stimulus through its input connections whose sum exceeds a threshold value T becomes active, supposedly signaling the presence of a triangle image in its RF.

We use the term *concept representation* for our semantic analysis of the neural network's representational capability, our ability to interrogate the activity of neural cells to formulate a description of the stimuli that cause the activity. Whether T can recognize a stimulus as a triangle as opposed to other visual stimuli might be considered a criterion for T representing the triangle concept. Recognition requires the ability to distinguish a stimulus as a specific familiar object or situation, and this suggests some use by the neural architecture of information implicit in the organization of its stimuli. Beyond recognition, a full representation requires the ability of the neural network to infer properties of a stimulus such as inferring that a particular arrangement of objects constitutes a more complex object. For example, a triangle has three angles and three sides joined in a specific arrangement. It is possible to recognize such a structure implicitly through connection-weight adaptation by having been presented with closely similar as well as contrasting stimuli; to make greater use of a stimulus through, for example, a more explicit decoding or "parsing" of its constituents into objects and relations is another matter. Recall (as opposed to recognition) might be said to utilize the latter capability, with association between items as a search tool.

In fact, the semantic content of T's representation is minimal and the concept we would use to express it reflects that. We might express it as "$\theta > \theta_T$", where θ is a number expressing the brightness of the stimulus and θ_T is an activation threshold for T. The sensor for T might as well consist of a single photocell aggregating the light from the visual field. Thus, we know two things: first, if T represents a concept it is certainly a local representation since T is the only neuron and second, this single neuron

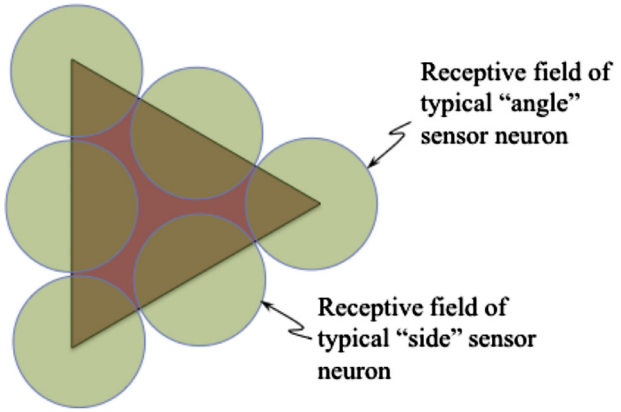


Fig. 1. The receptive fields of sensory neurons that detect the components of a triangle in the visual field.

represents only the simple concept of the brightness stimulus, not a triangle.

2.2. An initial multi-cell architecture

To satisfy the design goal of the discrimination of a stimulus object as a triangle, we can try adding more neurons to the neural network to separately and uniquely represent the components of the desired triangle and signal their presence to T. If T has a high enough activation threshold, the presence of all constituents would be required to cause T to become active. In this way the neural network would be a step closer to representing the concept of a triangle because it is essentially parsing its inputs into the components that make up a triangle. This of course assumes that it can discriminate the components, or visual features, of the triangle from other objects.

An improved neural network design provides six "feature-detector" neurons, three specialized for edges and three specialized for vertices or angles,[1] each with its own sensor element. Fig. 1 shows a triangle whose sides each consist of a single edge. The feature-detector RFs lie largely within the RF of the neuron T, which now receives its input via connections emanating from them. If these are binary with outputs of 0 or 1, the connection weights are excitatory with unit strengths, and the neuron T has a stimulus threshold value greater than 5 but less than 6, then using a simple, rate-coded neuron model in which T sums its inputs and tests the sum against its threshold T will be activated if and only if all six of its inputs are active.

Of course, there are several problems with this. First, the detection problem has been moved only a step back from T, for without more structure in their input stimuli the feature-detectors are responding only as single brightness stimuli. Second, merely connecting feature detectors as inputs to T says nothing about how the features are arranged within the image, even if they did correctly detect their intended features. A triangle-detector neuron T must not only correctly detect and identify the six elements of a triangle, it must also combine them in an arrangement that says "triangle". But as Fig. 2 shows, the representation is incorrect regardless of the fact that the appropriate features are located in the RF of T. This is to say nothing of the problems posed by environmental contingencies such as mist and other sources of shape distortion, a slight overlap with a nearby object, or a noisy image.

[1] Here, the terms "vertices" and "angles" are used interchangeably albeit an angle is the area between two lines meeting at a vertex.



Fig. 2. Three shapes using correctly-located angles and sides that are not triangles.

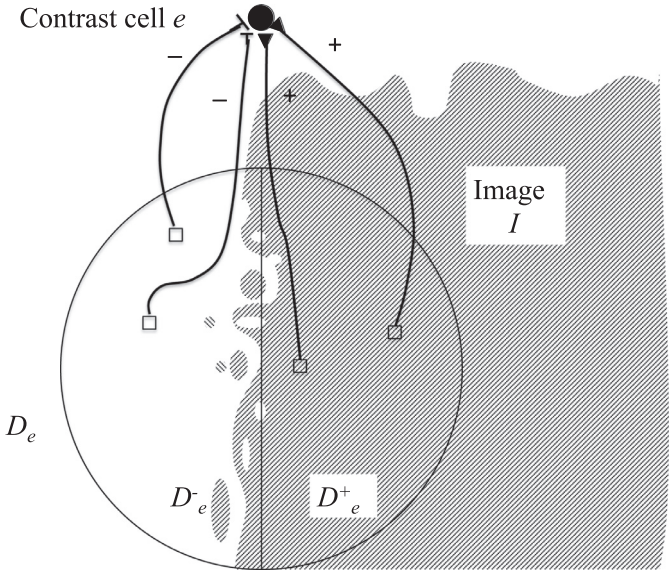


Fig. 3. An edge-detector cell, or contrast cell, with its sensory inputs.

3. Semantics—going deeper

3.1. Edge detection

Fig. 3 shows a hypothetical neural system for edge-detection, the simplest problem to solve in discriminating between visual shapes. This architecture consists of a cell e with an RF D_e in the sensor pixel array. This set has been subdivided into two regions or subsets, D_e^+ and D_e^-. Two of the potentially many pixels (shown as very small squares) from each region along with their weighted connections to e are shown. The connections from the pixels in D_e^+ to the edge-detector cell e are excitatory, indicated by plus signs where they synapse upon e. The connections from the pixels in D_e^- to e are inhibitory, indicated by minus signs. The resultant input sum processed by e is

$$\theta_e = \Sigma_{j\in D_e^+} w_j s_j - \Sigma_{j\in D_e^-} w_j s_j \quad (0 \le s_j \le 1)\ , \tag{1}$$

where the w_j are unsigned connection weight magnitudes and the s_j are pixel brightness values, with $0 \le s_j \le 1$. The upper bound of unity is to avoid saturation.

In the spirit of simplicity, assume for now that the signal function ϕ_e for e is a rectified linear function whose constant of proportionality is unity,

$$\phi_e(\theta_e) = [\theta_e - \theta_T]_+, \tag{2}$$

where the cell's activation threshold is θ_T with $\theta_T \ge 0$ and the function $[x]_+$ is defined as

$$[x]_+ = \begin{cases} x, & x > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Further simplifying, let $\theta_T = 0$ so that a net positive value in the input summation activates e.

The edge detector functionality is to be such that if an image almost covers D_e^+ but shows up hardly at all in D_e^-, e will be activated and emit a strong output signal, signaling an edge occurring along the midline separating D_e^+ from D_e^-. Fig. 3 illustrates this for

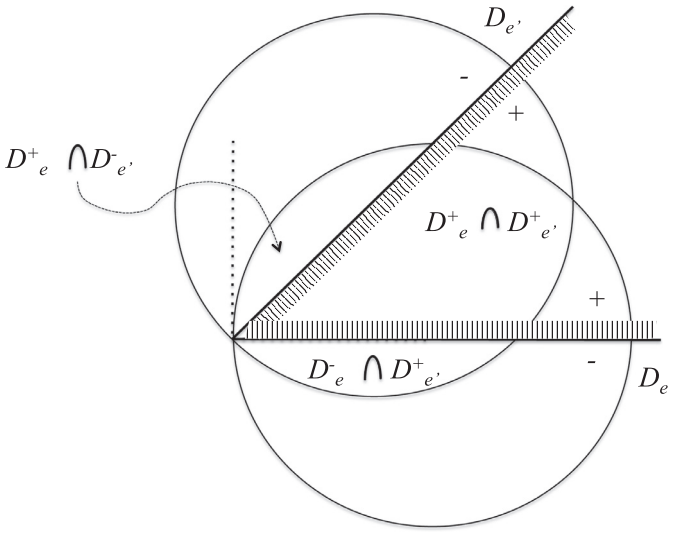


Fig. 4. Angle detection based upon the overlapping RFs of two edge detectors.

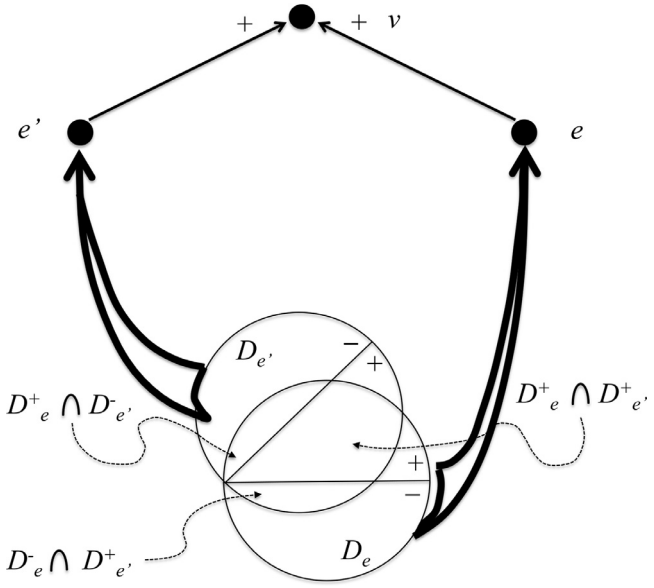


Fig. 5. A putative architecture for angle detection using the outputs of two edge detectors.

an image I (for ease of illustration, in this figure the shaded region contains the brightness stimuli while the clear region indicates an absence of stimuli). Of course, the balance of excitatory versus inhibitory inputs required to activate e can vary widely depending upon the weights w_j. We can simplify the analysis somewhat by assuming that there is little or no intrinsic bias toward any particular set of pixels in the transmission of stimuli. Therefore, we assume that the magnitudes of the weights over D_e^+ are uniform and normalized, and the same for D_e^-. The simplest type of normalization is an ℓ_1-like summation with the value unity, where in this case $\Sigma_{j\in D_e^+} w_j = 1 = \Sigma_{j\in D_e^-} w_j$. We also assume that both the excitatory and inhibitory regions have the same number of pixels (in the simulations of Section 5 this held only to an approximation). Given these assumptions, the edge-detector output will be restricted within the range

$$0 \le [\theta_e]_+ \le 1 \ .$$

In fact, if each pixel's output value is a binary 0 or 1, merely subtracting the area covered by an image I within D_e^- from that within D_e^+ in an appropriately-scaled map of the visual field represents the value of the input sum θ_e for I.

Let us assume heretofore that the signal function for all cells is

$$\phi(\theta) = [\theta]_+,$$

as with e. This provides a baseline, making the simplest reasonable assumptions for the subsequent analysis. If we are successful at deriving working architectures, nonzero thresholds and/or nonlinearities can be left as an option available for either "fine-tuning" an artificial neural network or enhancing the realism of a biological model.

3.2. Vertex detection (angle detection)

A simple design for vertex- or angle-detection cells can be attempted by combining edge-detection cells. Imagine two e-cells e and e' with overlapping RFs, with the midlines of the RFs meeting at an angle at an intersection point on their circumferences as shown in Fig. 4. Set-theoretically, an angle in the Euclidean plane is defined as the set of points between two rays emanating from a vertex. Using pixels in place of mathematical points, the angle is the intersection of the sets of excitatory pixels, $D_e^+ \cap D_{e'}^+$, for two edge detectors e and e' whose orientations represent the directions of the two rays as shown in Fig. 4. The midlines of D_e and $D_{e'}$ enclose the intersection $D_e^+ \cap D_{e'}^+$ of excitatory pixels. Formally, the unbounded region extending this enclosed region to the right is the angle, indicated by shading in Fig. 4.

To detect the angle, the v-cell takes its input directly from the two e-cells (see Fig. 5). The two inputs from the e-cells to v are excitatory, effectively relaying the competition between the excitatory and inhibitory pixels that defines each e-cell's activity. With this design we intend that if e and e' are both sufficiently active, the angle depicted in Fig. 4 has been detected at the shaded region, otherwise not. The input sum for v in terms of e and e' is

$$\theta_v = w_e [\theta_e]_+ + w_{e'} [\theta_{e'}]_+ \ . \tag{3}$$

Here, w_e and $w_{e'}$ are unsigned weight magnitudes. Assuming the same signal function as for e-cells, the v-cell produces signals $[\theta_v]_+$ based upon (2).

The cell v simply adds the inputs from its two e-cells and tests the sum against zero, allowing a signal from either e or e' to activate v. But this is clearly inadequate for angle detection, for a balance of pixel activity in favor of either D_e^+ alone or $D_{e'}^+$ alone could be mistaken for an angle. The v-cell can respond to many other shapes as exemplified by the one in Fig. 6. Therefore, to avoid false positives, either the threshold $\theta_{T,v}$ must be greater than zero or ϕ must be different for v, having some highly nonlinear form. We need the v-cell output to be nonzero only if *both* of the e-cell inputs to v are nonzero.

For simplicity, suppose that the v-cell's inputs have equal weights, with $w_e = 1/2 = w_{e'}$ in (3). Then,

$$\theta_v = \frac{1}{2}[\theta_e]_+ + \frac{1}{2}[\theta_{e'}]_+ \ .$$

If ϕ_v had a threshold $\theta_T \ge 0$, an angle detection would require that $\theta_v > \theta_{T,v}$, requiring in turn that

$$[\theta_e]_+ + [\theta_{e'}]_+ > 2\theta_{T,v}. \tag{4}$$

Recalling that the e-cell outputs are bounded above by unity, we must have it that if for example $[\theta_e]_+ = 1$ while $[\theta_{e'}]_+ = 0$, then $\theta_v \le \theta_{T,v}$ so that $[\theta_v]_+ = 0$. From (4), this would imply $1 + 0 \le$

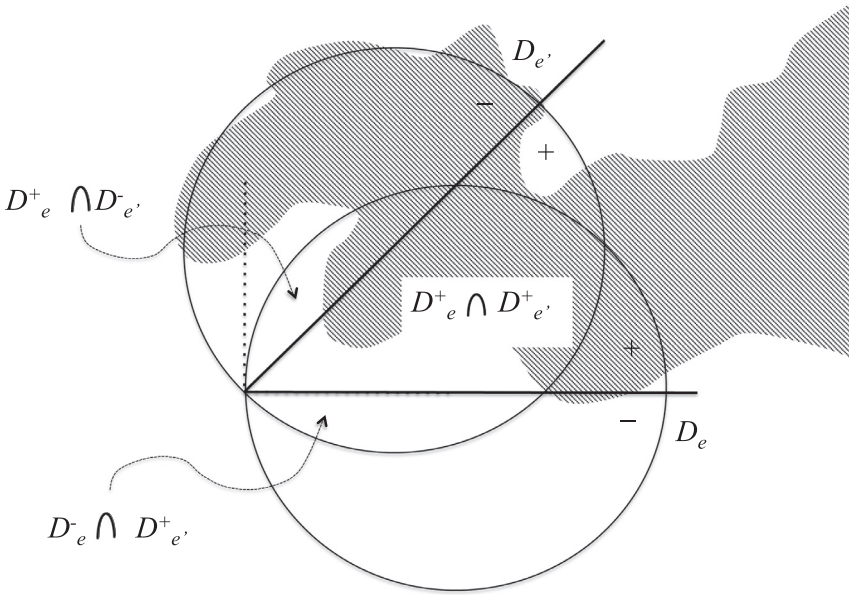


Fig. 6. An image shape that is a poor match for an angle-detector.

$2\theta_{T,v}$, or

$$\theta_{T,v} \geq \frac{1}{2}.$$

With its threshold value in this range, activating v would require that

$$[\theta_e]_+ + [\theta_{e'}]_+ > 1. \tag{5}$$

Enforcing the shape control sufficient to ensure a reasonable approximation to the angle that v is supposed to represent requires that $\theta_{T,v}$ have a high value, and the question becomes one of just how high this must be to force the image shape to be a reasonable approximation to the specified angle. Alternatively, ϕ_{edge} could have a high threshold, but this is not a solution because it would make edge detection difficult without really solving the problem. The root of the problem is that the simple neural network design so far proposed has no explicit mechanism for shape control. Too much is being expected of ϕ_{edge}, ϕ_{angle} and the parameters $\theta_{T,e}, \theta_{T,v}, w_e$, and $w_{e'}$.

4. Semantic analysis applied to neural network design

The underlying problem with the architecture in the previous section is that connecting e and e' to v in effect simply combines the separate influences of the two sets D_e and $D_{e'}$, with each set acting independently upon v. This ignores the fact that the sets may have a nonempty intersection $D_e \cap D_{e'}$ in which each pixel plays a dual role in its influence upon v since it acts separately through e and e'. A part of this intersection is $D^+_e \cap D^+_{e'}$, in which each pixel's excitatory effect upon v is doubled when both e and e' are active because as an element of D^+_e it has an *excitatory* effect upon e, and, hence, v through the excitatory connection with weight w_e, and as an element of $D^+_{e'}$ it also has an additional excitatory effect upon v. On the other hand, an active pixel in the region $D^+_e \cap D^-_{e'}$ has no influence when both e and e' are active because as an element of D^+_e it has an excitatory effect upon v but as an element of $D^-_{e'}$ it has an *inhibitory* effect of equal magnitude upon v. This means that if the weights w_e and $w_{e'}$ are equal the influence of the pixel upon v is effectively cancelled to zero. A similar statement applies to a pixel that is a member of $D^-_e \cap D^+_{e'}$. We can summarize this in the set-theoretic model by stating that this architecture combines the two sets D_e and $D_{e'}$ in a *disjoint union* $D_e + D_{e'}$ because it allows each pixel in the intersection $D_e \cap D_{e'}$ to play two independent roles. The linearity of ϕ_{edge} is not the problem; similar statements about the *effects* of the intersections would be true if ϕ_{edge} were nonlinear. And as previously suggested, forcing the form of ϕ_{edge} to help in shape control for angle detection is an added burden upon its role in edge detection. In summary, the expression $D_e + D_{e'}$ does not express the *intended* semantics of angle detection with the design illustrated in Fig. 5, but it does express something about what the design *actually does*.

4.1. Reconsidering the connectionist structure

The previous example shows in mathematical detail that an object, even one as simple as an edge over a small region, is more than a mere collection of its parts: A proper description of it includes information about its structure, how its parts are joined to comprise the whole. To this end, let us reformulate the inputs to the v-cell to include the missing structural information. The resulting input sum is meant to discriminate between angles and non-angles without requiring the demanding adjustments to the neural network signal functions of the preceding formulation. If successful at detecting the desired angle, the use of thresholds and/or nonlinearities will be an option available for "fine-tuning" the neural network, as previously proposed.

Continuing with the rectified-linear signal functions, let all thresholds be zero, $\theta_T = 0$. There are two cases to consider:

- Case 1. Each e-cell has a net positive input and is therefore active according to (2).
- Case 2. Only one of the e-cells, e say, is active.

We first express the quantities θ_e and $\theta_{e'}$ in (5) in greater detail by expanding the summation quantities in terms of the subregions $D^+_e \cap D^+_{e'}$, $D^+_e \cap D^-_{e'}$, $D^+_e \cap D^0_{e'}$, $D^-_e \cap D^+_{e'}$, $D^-_e \cap D^0_{e'}$, $D^-_e \cap D^-_{e'}$, $D^0_e \cap D^+_{e'}$ and $D^0_e \cap D^-_{e'}$. The notation for the subregions of the intersection of RFs is as shown in Figs. 4 and 6, but with the additional inclusion of D^0_e and $D^0_{e'}$, which are the regions complementary to D_e and $D_{e'}$, respectively, within $D_e \cup D_{e'}$. Assuming as before that there is no bias in favor of either e or e' so

that $w_e = w_{e'}$, we replace (5) with the following:

$$\begin{aligned}\theta_v &= w_e\,([\theta_e]_+ + [\theta_{e'}]_+)\\ &= w_e\Big([\Sigma_{j\in D_e^+\cap D_{e'}^+} w_j s_j + \Sigma_{j\in D_e^+\cap D_{e'}^-} w_j s_j + \Sigma_{j\in D_e^+\cap D_{e'}^0} w_j s_j\\ &\quad - \Sigma_{j\in D_e^-\cap D_{e'}^+} w_j s_j - \Sigma_{j\in D_e^-\cap D_{e'}^0} w_j s_j - \Sigma_{j\in D_e^-\cap D_{e'}^-} w_j s_j]_+\\ &\quad +[\Sigma_{j\in D_e^+\cap D_{e'}^+} w_j s_j + \Sigma_{j\in D_e^-\cap D_{e'}^+} w_j s_j + \Sigma_{j\in D_e^0\cap D_{e'}^+} w_j s_j\\ &\quad - \Sigma_{j\in D_e^+\cap D_{e'}^-} w_j s_j - \Sigma_{j\in D_e^0\cap D_{e'}^-} w_j s_j - \Sigma_{j\in D_e^-\cap D_{e'}^-} w_j s_j]_+\Big), \quad (6)\end{aligned}$$

where the s_j are as usual the current activities of the pixels and the w_j are the weights, normalized for each region D_e^+ or D_e^- according to the equation $\Sigma_{j\in D_e^+} w_j = 1 = \Sigma_{j\in D_e^-} w_j$ from Section 3.1. Assuming further a uniform weight distribution and equal numbers of cells $\mathrm{card}(D_e^+)$ and $\mathrm{card}(D_e^-)$ in D_e^+ and D_e^- just for simplicity in this discussion, the weight values are $w_j = \frac{1}{\mathrm{card}(D_e^+)} = \frac{1}{\mathrm{card}(D_e^-)}$ (again, this is an approximation and the simulations use corrected weights).

At this point, a change in notation will yield greater clarity. In a weighted summation with weights w_j of the activities s_j of the pixels represented by the elements of a set X, let us substitute the expression ΣX for $\Sigma_{j\in X} w_j s_j$. With this simplification, we can rewrite (6) as

$$\begin{aligned}\theta_v = w_e\,\big([&\Sigma D_e^+ \cap D_{e'}^+ + \Sigma D_e^+ \cap D_{e'}^- + \Sigma D_e^+ \cap D_{e'}^0\\ &- \Sigma D_e^- \cap D_{e'}^+ - \Sigma D_e^- \cap D_{e'}^0 - \Sigma D_e^- \cap D_{e'}^-]_+\\ &+[\Sigma D_e^+ \cap D_{e'}^+ + \Sigma D_e^- \cap D_{e'}^+ + \Sigma D_e^0 \cap D_{e'}^+\\ &- \Sigma D_e^+ \cap D_{e'}^- - \Sigma D_e^0 \cap D_{e'}^- - \Sigma D_e^- \cap D_{e'}^-]_+\big). \quad (6')\end{aligned}$$

Notice that $D_e^- \cap D_{e'}^- = \emptyset$ for angles of size not greater than 180°, and, hence, its contribution to the summations is zero. Now for Case 1: since both cells are active, we can drop the rectification symbols for the two signal functions, group like summation terms, and cancel like terms with opposite signs, yielding

$$\begin{aligned}\theta_v = w_e\,(2\cdot\Sigma D_e^+ \cap D_{e'}^+ + \Sigma D_e^+ \cap D_{e'}^0 + \Sigma D_e^0 \cap D_{e'}^+\\ - \Sigma D_e^- \cap D_{e'}^0 - \Sigma D_e^0 \cap D_{e'}^-). \quad (7)\end{aligned}$$

(7) expresses in detail that in the disjoint union the effects of the pixels in $D_e^+ \cap D_{e'}^+$ are doubled and the effects of those in $D_e^+ \cap D_{e'}^-$ and $D_e^- \cap D_{e'}^+$ are cancelled out. As for Case 2: only one of the cells is active, e say. Then,

$$\begin{aligned}\theta_v = w_e\,(&\Sigma D_e^+ \cap D_{e'}^+ + \Sigma D_e^+ \cap D_{e'}^- + \Sigma D_e^+ \cap D_{e'}^0\\ &- \Sigma D_e^- \cap D_{e'}^+ - \Sigma D_e^- \cap D_{e'}^0). \quad (7')\end{aligned}$$

(7′) expresses in detail the effects of e acting alone upon v.

4.2. A change to the architecture

Now consider the following change in the neural network design. A new type of cell called a u-cell is shown in Fig. 7 at a level intermediate between the pixels and the e-cells. The effect of the new architecture upon cell v in angle detection is illustrated in Fig. 8. The inputs to the u-cells u_1, u_2 and u_3 emanate directly from the pixels in their RFs—the two crescent-shaped regions and the arrowhead region within D_e and $D_{e'}$—and are uniformly excitatory. Input connections to the e-cells emanate from these new cells and are excitatory or inhibitory as required. Thus, the RF D_{u_1} of the cell u_1 is the pixel set $D_e^+ \cap D_{e'}^+$ and the u_1-cell supplies excitatory input to its e-cell targets. On the other hand, the RF D_{u_2} of u_2 is $D_e^+ \cap D_{e'}^-$, where u_2 supplies excitatory input to the e-cell and inhibitory input to the e'-cell. Similarly, the RF D_{u_3} of u_3 is $D_e^- \cap D_{e'}^+$, supplying input to the e-cell and excitatory input to the e'-cell. For now one may think of the pixels supplying the remainder of the inputs to the e-cells, but see the last paragraph

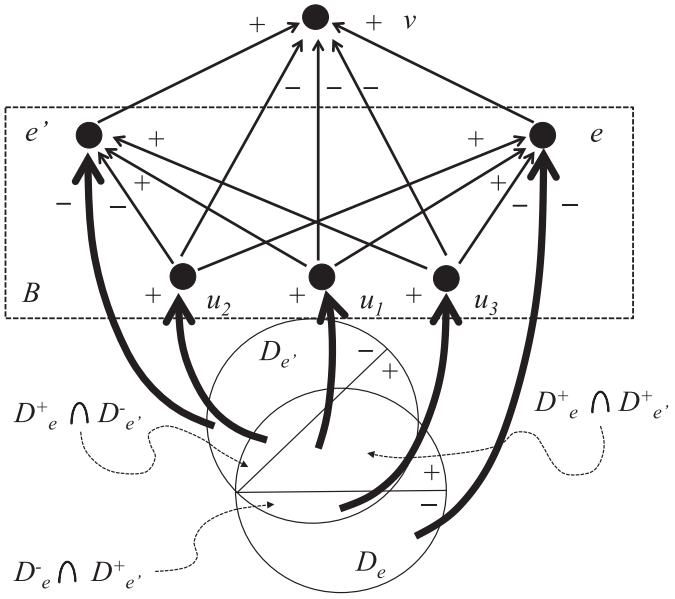


Fig. 7. A modified architecture for angle detection.

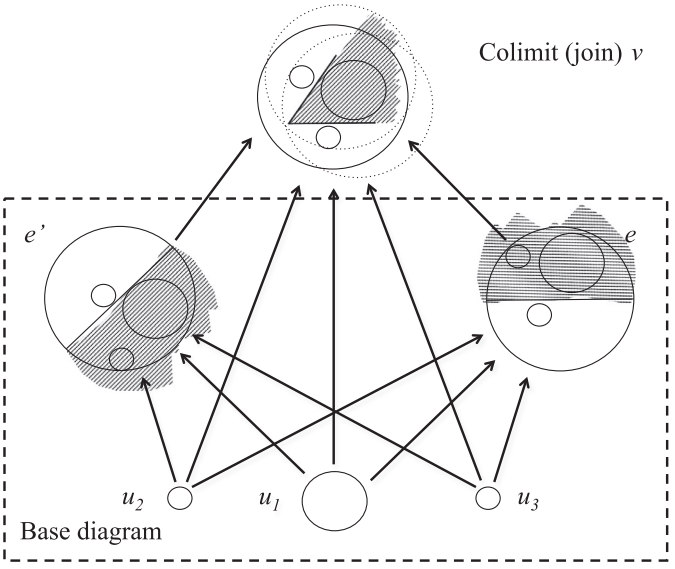


Fig. 8. A colimit for a base diagram of sets and functions with an angle as colimit object. Here, the small circles are simply an aid for tracking the inputs from the u-cells.

of the following section. To simplify the presentation, the RF sets such as those in the diagram in Fig. 8 will hence forth be labeled with the cells whose RFs are the sets. For example, the label u_1 will used in place of D_{u_1} in the remaining figures, although not in the text.

As was noted in the last section, the three sets $D_e^+ \cap D_{e'}^+$, $D_e^+ \cap D_{e'}^-$ and $D_e^- \cap D_{e'}^+$ are represented twice in the disjoint union architecture by the duplication of excitatory inputs from the cells e and e'. However, now the u-cells are available and enable us to add a modification. Through an additional set of exclusively inhibitory connections directly to the v-cell, the u-cells correct for the disjoint representation by supplying the missing information about the intersection $D_e \cap D_{e'}$. This increases the sensitivity of the angle detector to shape distortions in Case 1 above by (1) allowing any shape distortions occurring in $D_e^+ \cap D_{e'}^-$ and $D_e^- \cap D_{e'}^+$ to lower the activity of v and (2) further lowering the overall activity of v by discounting for one copy of its positive input from $D_e^+ \cap D_{e'}^+$. It increases sensitivity in Case 2 by simply lowering the

activity of v. In both cases, the extra inhibitory input from the u-cells in effect supplies a variable threshold for the v-cell. The angle detector remains sensitive to shape distortions from inactive pixels within $D_e^+ \cap D_{e'}^+$ because there remains a net of one excitatory input per pixel from that region in the desired Case 1. Now, however, each inactive pixel in this region has a greater relative effect in lowering the activation level of v. For the other two intersection regions, the effect of the inhibitory input to each e-cell is doubled, again lowering the activation of v. In Case 2, the effects are similar, and a significant amount of either an inappropriate image, image distortion or noise can cause v to become inactive. As a result, the new architecture makes nonlinear signal functions and nonzero thresholds an option, not a necessity.

4.3. A colimit architecture

The formalization via sets illustrated in Fig. 8 enables an examination of the semantics of the new architecture illustrated in Fig. 7. This contrasts the functionality of the new architecture with the disjoint union architecture illustrated in Fig. 5. The shaded regions indicate the excitatory inputs to each e-cell when it acts alone as an edge detector, and the net excitatory input to the angle-detector v-cell from the e and e' cells. Also acting upon the v-cell are inhibitory inputs via direct pathways from the u-cells. The formalization via sets does not convey the excitatory and inhibitory natures of the connection pathways. It is for this reason that Fig. 8 must be enhanced with shading to represent the regions of net excitatory input to each cell, with white in the crescent-shaped regions representing the regions of net inhibitory input. Formally, the interplay of excitatory and inhibitory inputs is expressed in the input summations. The illustration with the shading added provides a conceptual, semantic view of the functionality of the architecture illustrated in Fig. 7.

In the mathematical discipline of category theory, Fig. 8 illustrates a *diagram* that defines a *colimit*. The sets are *objects* and the functions which map pixels in one set to pixels in another are the arrows, or *morphisms*, in a *category*. An introduction to category theory is contained in several of our papers and references therein; see for example [13–15]. The sets and functions of the diagram are illustrated by the circular disks and the arrows joining them; in the category of sets and functions, these morphisms are *inclusion functions* representing the subset relation. The inclusions define $D_{u_1}, D_{u_2}, D_{u_3}$ as subsets of D_e and $D_{e'}$ and $D_{u_1}, D_{u_2}, D_{u_3}, D_e$ and $D_{e'}$ as subsets of D_v. The smaller disks inside the disks representing D_e and $D_{e'}$ indicate the locations of D_{u_1}, D_{u_2} and D_{u_3}. The edge-detector disks are shown inside D_v as lightly-dashed circles containing the smaller disks which again locate $D_{u_1}, D_{u_2}, D_{u_3}$, which sets cover the intersection $D_e \cap D_{e'}$. A separate disk for the angle-detector v is superimposed over the union of the dashed disks to emphasize that the result is an RF for a single angle-detector cell, with its shaded "arrowhead" region illustrating the intersection of the excitatory RF region in $D_e \cap D_{e'}$.

The entire figure shows a cone-like structure called a *cocone* attached to the RFs inside the dashed-rectangle box, which encloses the *base diagram* of the colimit. The set D_v together with $D_{u_1}, D_{u_2}, D_{u_3}$ and D_e and $D_{e'}$, and the arrows joining them, comprise the *defining diagram* of the colimit. The base diagram can be thought of as a *specification* for constructing the "Colimit (join) v". In the colimit construction, the union of D_e and $D_{e'}$ is properly expressed by *blending* them so that they are merged where they share the images of the u-cell RFs that represent the three parts of the intersection $D_e \cap D_{e'}$.

The claim that the colimit expresses the intersection of the two edge-detector RFs correctly within the union follows from the following property of the cocone when combined with the base diagram, a property known as *commutativity*: choosing any one of D_{u_1}, D_{u_2} or D_{u_3} and following all three pathways from that set through the mapping arrows to D_v results in one and the same composite function regardless of the pathway followed; that is, the three mappings given by tracing the three different pathways are equivalent. One of the three pathways consists of a single arrow mapping the u-cell's pixels directly into D_v. The other two pathways map each of the u-cell's pixels through D_e and $D_{e'}$ and on into D_v. In doing so, the base diagram arrows based at the u-cell are composed with arrows from the cone-like structure. The three pathways, two of them forming the function compositions, yield one and the same function, the same function represented by the single arrow directly from the u-cell to v. This states mathematically that each u-cell intersection region is represented only once in the union of D_e and $D_{e'}$, for the commutativity guarantees that all pathways from any D_{u_i} to D_v ($i \in \{1,2,3\}$) correspond to a single function.

It is important at this point to defray a reasonable objection to our introduction of the u-cells as having such specialized small RFs as $D_e^+ \cap D_{e'}^+$, $D_e^+ \cap D_{e'}^-$ and $D_e^- \cap D_{e'}^+$, the arrowhead- and crescent-shaped regions. This would require specialized u-cells for blending for practically every neural network structure that represents a colimit. In fact, our specialized u-cells are merely a gross simplification for clarity in this discussion. The three u-cells shown in Fig. 7 in actuality are substituting for many u-cells covering those regions. Collectively these provide the appropriate inputs to the e, e' and v cells. In fact, overlapping u-cells provide a uniform covering of the entire visual space and supply excitatory or inhibitory output as required, while receiving uniformly excitatory input from their RF pixels. Collectively, then, the u-cells obviate the need for direct input from the pixels to cells such as e, e' and v. Each u-cell has an RF in the form of a disc-shaped region, and clusters of u-cells' RFs form the specialized regions such as those used in the example. Thus, specialized blending cells are not required for colimits at the sensor level.

4.4. The colimit as a local-distributed concept representation

The colimit is one of the key categorical structures used in the CNST. Colimits exist for all diagrams, finite and infinite, in many categories based upon a theorem in category theory. Functors and colimits are discussed in books on category theory (a comprehensive formal treatment is given in [16]; a more accessible source for the non-mathematician is [17]) and in some of our previous papers (see for example [13]).

We formalize a category of concepts as a category **Concept** of theories expressed in a symbolic or formal logic and morphisms between these theories. A morphism or arrow $m : T \longrightarrow S$ with *domain* T and *codomain* S embeds the theory T into the theory S as follows: first, it maps the symbols of T to either symbols or formulas of S. Second, the mapping is validity-preserving, meaning that the truth of a formula in T, transformed by the substitution of new symbols or formulas for the symbols of T, is preserved in its transformed form in S. In a sense, the morphism m expresses the fact that T is expressed as a sub-theory of S through its inclusion within the symbolic expressions in S. For example, a concept (a theory) expressing the properties of an edge (which we illustrated in terms of pixel sets) is a part of the concept of an angle (or a side); if we let T be the concept that expresses the edge e and if S expresses the angle v, a morphism $m : T \longrightarrow S$ expresses just *how* T is a part of S, or e is a part of v. But there is another edge e' associated with v and, hence, another concept T' which is embedded within S, and this is expressed through a different theory morphism $m' : T' \longrightarrow S$.

Now notice that in the general case, the intersection of the pixel sets expressed in T and T' might well be embedded separately in S because each of the two morphisms m and m' expresses only

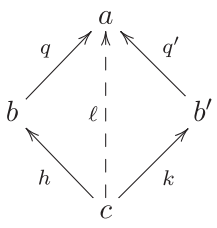
the relationship of its domain, T or T', to S. This is evident in our neural network example as the separate connections from the edge-cells e and e' to the angle cell v, through which v receives separate input stimuli from e and e'. But e and e' have nothing to do with each other except that their pixel sets contain the same subset, their intersection. Let the cell u represent this intersection by virtue of the fact that it receives input from its pixels. Then u is active when the appropriate combination of pixels in e is active and, separately, u is active when e' is active, but there is no intrinsic requirement that e and e' must be active simultaneously. If we let U be a concept that represents u, we can express the fact that u is a part of e by a theory morphism $n : U \longrightarrow T$, and, separately, that u is a part of e' by a theory morphism $n' : U \longrightarrow T'$. Separately, then, there exist two composition morphisms mapping the theory U into the angle theory S, $m \circ n : U \longrightarrow S$ and $m' \circ n' : U \longrightarrow S$.

Notice that in general, it can happen that $m \circ n \neq m' \circ n'$. In the edge and angle example, the two connection pathways from u to v through e and e' are separate and, in general, will not be active simultaneously. Moreover, when they are active simultaneously the pixel set of u is represented twice through excitatory connections described previously in our example. Thus it cannot be said that u is correctly represented as the intersection of the sets of pixels of e and e' as seen in the activity of v. Note that this is an example of the fact that the separate concept embeddings of U within S through T and T', $m \circ n : U \longrightarrow S$ and $m' \circ n' : U \longrightarrow S$, are not equal in general. As a consequence, U has two separate manifestations within S; they have the same basic structure but different symbols and possibly some different formula substitutions, expressing different parts of the theory S.

To see this in the general case, consider two morphisms from one object c to another object a in a category $\mathbf{C}$, $f : c \longrightarrow a$ and $g : c \longrightarrow a$. The two morphisms share the same domain c and codomain a, but otherwise are not required to have anything in common—they represent two relationships between c and a and, hence, are not equal in the general case. An example in other categories are functions with domain c and codomain a in **Set**, the category whose objects are sets and whose functions are mappings from one set to another such that each element of the domain has a unique image in the codomain. Some other examples are categories of groups or rings in which the mappings are homomorphisms, categories of topological spaces in which the mappings are continuous functions, and a category with the natural numbers 1, 2, 3, ...as objects in which the morphisms are not mappings, but divisibility indicators |, where $c|a$ if and only if c divides a with no remainder, yielding another natural number.

Now consider a diagram consisting of four morphisms $h : c \longrightarrow b$, $k : c \longrightarrow b'$, $q : b \longrightarrow a$ and $q' : b' \longrightarrow a$. The four morphisms can be pictured as arrows forming a diamond shape with the morphisms as sides and the objects c, b, b' and a at the vertices. As with the morphisms in our theory category example, the arrows along the two pathways from c to a have compositions, which are arrows we shall call f and g, where $f = q \circ h : c \longrightarrow a$ and $g = q' \circ k : c \longrightarrow a$. In general, f and g are not equal.

However, suppose the diamond-shaped diagram commutes: $f = g$. Then, the compositions would be one and the same morphism with domain c and codomain a, $q \circ h = q' \circ k$. Let this single morphism be ℓ, $q \circ h = \ell = q' \circ k$:



That is, the diagram, were it to commute, would express the single relationship $\ell : c \longrightarrow a$ of c to a resulting from the two compositions $q \circ h$ and $q' \circ k$ involving the objects b and b' as intermediates. A commutative diagram is one of the properties of a colimit of b and b' along a shared object c, with colimit object a. The other property of a colimit is that the cocone formed by the morphisms $q : b \longrightarrow a$ and $q' : b' \longrightarrow a$ has the property of *initiality*. The practical consequence of initiality for our purposes is that the commutative diagram expresses all there is to know about the relationship of c to a involving b and b'. Notice that this depends upon the interaction of the morphisms $h : c \longrightarrow b$, $k : c \longrightarrow b'$, $q : b \longrightarrow a$ and $q' : b' \longrightarrow a$; this implies that colimits are a very special case. It is a fact, however, that colimits arise quite naturally in many categories, and in categories such as the theory category **Concept** colimit theories can be constructed by an algorithm. The notion of colimit for diamond-shaped diagrams generalizes to arbitrary diagrams which superimpose commutative diamonds to form more complex diagrams.

In the more general theory category example, suppose the diagram were to commute, $m \circ n = r = m' \circ n'$, with $r : U \longrightarrow S$ the single morphism derived from the diagram which maps U into S. If in addition to commuting the arrows $m : T \longrightarrow S$ and $m' : T' \longrightarrow S$ form an initial cocone for the diagram consisting of $n : U \longrightarrow T$ and $n' : U \longrightarrow T'$, then the diamond-shaped diagram defines the theory S as a colimit of T and T' along the shared sub-theory U. In other words, the union of T and T' to form S contains a single copy of U.

In our earlier papers such as the one cited at the beginning of this section we have defined neural categories for neural neural networks at a given stage of connection-weight adaptation. We also define functors that map the theory category **Concept** into the neural categories. Functors preserve compositional structure and, hence, they map commutative diagrams to commutative diagrams. Certain functors also preserve initiality and, hence, preserve colimits. Based upon this discussion and the discussions of previous sections (where concepts about structures formed over visual sensors are illustrated with mappings of pixel sets), we assert that simply combining objects does not properly represent them. The appropriate combination of objects requires blending them along shared subobjects. This combined object is a colimit of its parts.

It is possible to express the foregoing with more mathematical rigor than used here. However, the purpose of this paper is to reach an audience concerned not with the deep mathematical detail, but with the issue of local versus distributed representation in neural networks.

5. Triangle detection

A colimit can also use u-cells to blend three angles to make a triangle. As mentioned in Section 3.2 and illustrated in Fig. 4, an angle is defined set-theoretically as the set of points between two rays emanating from a common point. Using this definition, a triangle is defined as the intersection of three angles whose rays intersect pairwise to form the sides. As with the angle colimits, the mathematical set of points defining a triangle is represented by a colimit that defines a composite of three properly-oriented angle-detectors. The angle orientations are determined by the requirement that their sides be aligned. The sides in turn are formed from edge detectors, as many as required for the size of the triangle in the visual space. A construction for this using three edge-detectors in addition to those forming the three angles is shown in Fig. 9.

The three angles are labelled v_a, v_b, and v_c and the added edges are labelled s_{ab}, s_{bc}, and s_{ca} in Fig. 9. The angles v_a and v_c are adjacent to side s_{ca}. Thus, as concepts in the concept category, v_a and s_{ca} are blended by a concept which we express as the in-

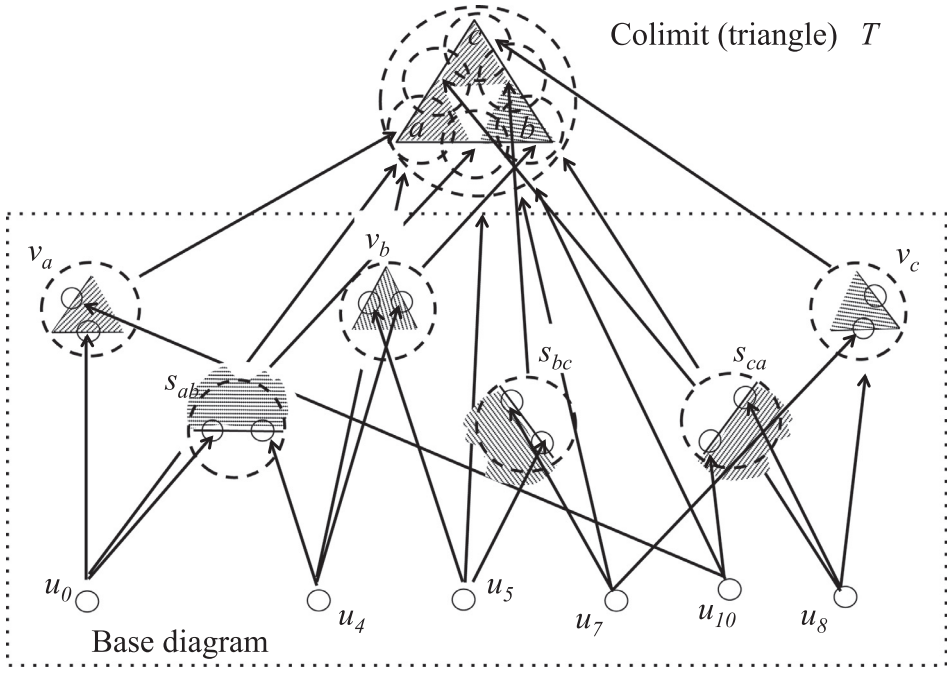


Fig. 9. A triangle as a colimit formed by blending angles along common edges. Here, the small circles are simply an aid for tracking the inputs from the u-cells.

tersection of the RF of the cell representing v_a, which we call cell v_a, and the RF of the immediately adjacent edge-cell of s_{ca} (a side such as s_{ca} is expressed as a colimit of one, two or more colinear edges). Similarly, the side s_{ca} is blended with the angle v_c at the intersection of its edge adjacent to the angle v_c with the angle v_c itself. As with the blending of edges in an angle colimit, the cells in the neural network represent the sides and angles and the blending regions, the latter via u-cells. Recall that the pixel inputs to all u-cells are uniformly excitatory. The feedforward connections between u-cells and their target v- and s-cells in the defining diagram match the arrows in Fig. 9; however, these connections are a mixture of excitatory and inhibitory connections. The rule for determining which sign a connection has is as with the angle colimit and its accompanying figure: excitatory u-cell connections project to shaded regions in the RFs of their target cells and the inhibitory connections project to the regions left blank in the figures. The feedforward connections from the v- and s-cells to the t-cell, which also match the arrows in the figure, are excitatory. The long-range connections projecting directly from the u-cells to the t-cell are inhibitory, as are the direct projections from the u-cells to the v- and s-cells (not shown, but see Section 5.1) in their respective colimit defining diagrams. In fact, the same u-cells and their projections to the edge-cells are re-used. The commutativities of the different diagrams which use these connections is consistent with this, for when the triangle colimit diagram is active, the diagrams for the three angles (and the side colimits as well) will also be active.

In the blending of s_{ca} and v_c, for example, excitatory and inhibitory connections project to those cells from the u-cell u_8. The cell u_8 also has an inhibitory connection projecting directly to the triangle colimit cell t to enforce commutativity. This scheme is repeated for all angles and sides in the defining diagram for the triangle. Maintaining our use of linear signal functions as a baseline architectural feature, the net effect of the u-cells' inhibitory feedforward projections to the t-cell is to subtract one copy of each blending u-cell in the total input summation of the t-cell. This is the same set-theoretic view posed in the discussion of the angle neural structure to describe the manner in which the commutativity of the defining diagram ensures a correct representation of the colimit object. Because the set intersection of the excitatory subsets of the six RFs of $v_a, v_b, v_c, s_{ab}, s_{bc}$, and s_{ca} defines the triangle, the colimit object t in the neural category is a formal representation of the triangle.

The triangle detector simulations described in Section 5 show how the colimit t-cell architecture performs in comparison with a disjoint union t-cell architecture. Both use the same v-cells together with additional edge detectors for the sides (s-cells), but the disjoint architecture for the t-cell does not include the colimit t-cell's blending cells and their accompanying connections. Based upon a preliminary performance comparison, the angle and side detector cells (v and s cells) for both the disjoint and colimit triangle detectors are based solely upon the colimit architecture. Note that the triangle representation for the simulations contains multiple edge detectors along each side, separately blended as colimits to form the sides. Shape distortions were created to test the triangle detectors in these simulations.

Before proceeding, it is worth mentioning that the neural network computations in these simulations are feedforward only using predetermined connection-weight values and, hence, are of low complexity. From an input triangle image through the neural network layers which compute u, e, v and s cell activations and, finally, the t-cell activation value, each cell computes the rectified linear function of summed, connection-weighted inputs as its output through simple multiply-and-add operations as described in the text (see Eqs. (1) and (2) and the subsequent discussion).

5.1. Triangle detector simulations

5.1.1. Situated triangle disjoint union and colimit networks

We have carried out numerical simulations on artificial neural networks for both the disjoint union and colimit architectures to quantify and compare their detection properties. The neural simulation models were of low complexity in order to provide a clear comparison of their design principles. Hence, the calculations involve rate-coded neural cell outputs transmitted to neurons farther along in a feedforward neural network (feedback was not considered here because only a straightforward test of triangle image detection was the object). The neuron response is a rectified linear

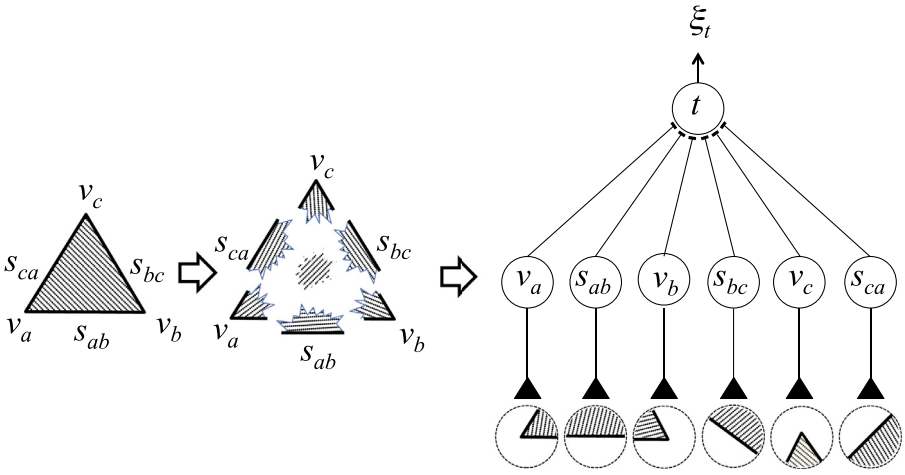


Fig. 10. A design layout for the disjoint union triangle cell network.

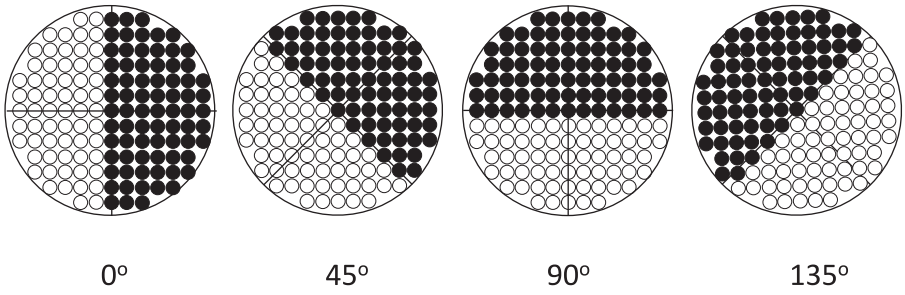


Fig. 11. Examples of receptive fields of diameter 13 pixels for edge cells tuned for several contrast directions. Dark circles represent excitatory weights while open ones represent inhibitory weights. In this example, each RF contains approximately 75 excitatory and 62 inhibitory connections from a total of 137 pixels. In all cases, excitatory and inhibitory RF subfields are independently normalized to +/−1, respectively.

function of its connection-weighted input sum as described in the text. The t-cell's detection capability was tested by applying a simple Heaviside function with a specified threshold level to its output (see Fig. 16 in the Results section).

The stimulus-based activities of the cells transmitted through the neural network's weighted connections represent the angle, side and triangle colimit defining diagrams responding to a variety of visual images. The images are an array of numerical values simulating the brightness values of the pixels in the focal plane of a sensor, which we envision as a rectangular grid of photoreceptor neurons. The neural cell outputs vary linearly as a function of their input stimuli from zero up to a maximum intensity at which they saturate, that is, reach an upper bound value. In this analysis we arbitrarily assume the maximum to be unity. Fig. 3 illustrates a region of the focal plane that is centered on a vertical straight-line edge of some object which is not quite perfect.

The triangle cell t is tuned to detect a situated triangle, one which has specific angles and a specific size at a specific location in the visual space. This is accomplished by combining the v and s cell outputs of the situated side and angle detector colimit subnetworks for the desired triangle. This is illustrated for a disjoint union triangle concept in Fig. 10 where v_a, v_b and v_c represent the angle-detector cells and s_{ab}, s_{bc} and s_{ca} represent the side-detector cells. A reading of the illustration from left to right suggests the process by which the neural network deconstructs the triangle image: First it identifies its situated parts and then combines these to form the triangle representation.

The side and angle cells are driven by subnetworks that combine the outputs of edge detector cells. In the simulations, the edge cells implement a symmetric difference quotient (to compute image intensity gradients). Fig. 11 shows examples of circular RF weight patterns for edge cells that are tuned for maximum output for four image contrast orientations. The diameter of the RF sets the spatial scale at which the neuron reaches full sensitivity. Image features smaller than the RF diameter in the excitatory half of the RF will produce decreased activation of the edge cell even if the inhibitory half has no stimuli. The RFs of the side and angle cells are also illustrated as disks in Fig. 10, where the shaded region in each disk shows the represented angle (for v_a, v_b, v_c) or side (for s_{ab}, s_{bc}, s_{ca}). As mentioned, the v and s cells are supported by subnetworks in the colimit configuration (see Fig. 7).

Colimits for the sides of a triangle have not yet been described. Like v-cells, s-cell colimit defining diagrams map to neural category colimits, which correspond in a neural network to subnetworks in which the overlaps between adjacent e-cell RFs form the u-cell RFs. For example, the total activation for cell s_{ab} in Fig. 12 many be written as

$$\theta_{s_{ab}} = \Sigma_{i=1}^{4} w_{s_{ab}e_i}[\theta_{e_i}]_+ - \Sigma_{j=1}^{3} w_{s_{ab}u_j}[\theta_{u_j}]_+ \quad (8)$$

where $w_{s_{ab}e_i} = 1/4$ in this example for side ab. In general, stimulus weights such as $w_{s_{ab}e_i}$ are chosen to normalize the total primary synaptic input to unity for maximum stimulus, assuming each input stimulus is limited by the value unity. Were disjoint union s-cells used, the value of $w_{s_{ab}u_j}$ would be set to zero. For the colimit form, $w_{s_{ab}u_j} = w_{s_{ab}e_i} (1 \le i \le 4; 1 \le j \le 3)$ to achieve the blending function by subtracting the duplicate stimulus contributions to the activation of s_{ab} from the overlap in the e-cell RFs, which are also the u-cell RFs. This is the same strategy used earlier in the neural network architecture for an angle colimit.

As the object of the experiments was to evaluate the relative performance of disjoint union and colimit t-cells, both architectures were tried. Fig. 10 illustrates the layout of the disjoint union

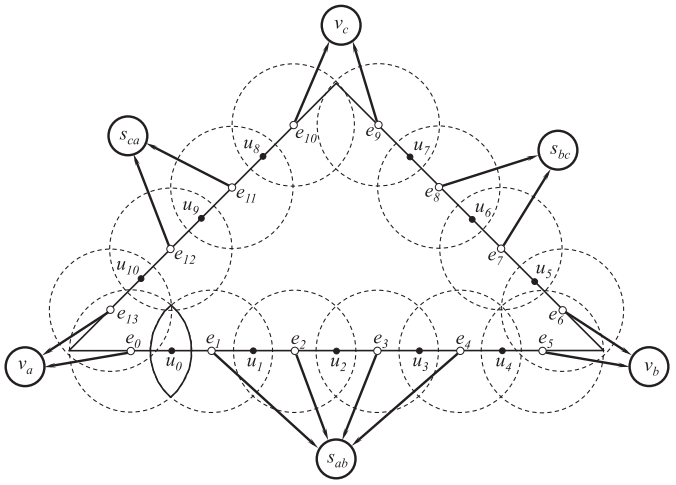


Fig. 12. Details of the s- and v-cell RFs overlaid on a triangle image. The dashed circles are the borders of the edge-cell RFs that are blended to form them as colimits. The small open circles within them indicate the centers of each edge cell's RF. The e-cells $e_0 \ldots e_5$ are oriented at 90°, $e_6 \ldots e_9$ are oriented at 135°, and $e_{10} \ldots e_{13}$ are oriented at 45°. The small filled circles indicate the centers of the u-cell RFs; one of these RFs is the solid outline surrounding u_0.

simulations, and the colimit layout with blending u-cells is shown in Fig. 12. In Fig. 9 the blending regions are shown as the intersection of edge-cell RFs, with a shaded half-region to indicate excitatory input to the target e-cells and a clear half-region for the inhibitory inputs, and a separate u-cell is shown for each half-region. For computational simplicity in the simulations, the pairs of excitatory/inhibitory-input u-cells are consolidated into single u-cells as shown in Fig. 12. For example, u_0 here replaces the u_1 and u_2 from Fig. 9. For either the disjoint union or colimit version, the total activation for the triangle cell many be written as,

$$\begin{aligned}\theta_t = {} & w_{ts}([\theta_{s_{ab}}]_+ + [\theta_{s_{bc}}]_+ + [\theta_{s_{ca}}]_+) + w_{tv}([\theta_{v_a}]_+ + [\theta_{v_b}]_+ + [\theta_{v_c}]_+) \\ & - w_{tu_{0,4}}([\theta_{u_0}]_+ + [\theta_{u_4}]_+) - w_{tu_{5,7}}([\theta_{u_5}]_+ \\ & + [\theta_{u_7}]_+) - w_{tu_{8,10}}([\theta_{u_8}]_+ + [\theta_{u_{10}}]_+), \end{aligned} \tag{9}$$

where $w_{ts} = w_{tv} = \frac{1}{6}$ in this example. The disjoint union version requires the u-cell weights in (9) to be zero, $w_{tu_{0,4}} = w_{tu_{5,7}} = w_{tu_{8,10}} = 0$. The selection of the weights for the colimit version is more complicated because the u-cell activation potentials have different cumulative weightings along three connection pathways into the triangle cell, two excitatory pathways through the e-cells whose overlaps they represent and the other a direct inhibitory pathway synapsing upon the t-cell. For example, the cell u_0 has the three paths:

1. $u_0 \to v_a \to t$,
2. $u_0 \to s_{ab} \to t$, and
3. $u_0 \to t$.

with net weightings:

1. $w_{tv_au_0} = w_{tv}w_{v_ae_0} \approx 0.167 \times 0.5 \approx 0.0833$,
2. $w_{ts_{ab}u_0} = w_{ts}w_{s_{ab}e_1} \approx 0.167 \times 0.25 \approx 0.0417$, and
3. w_{tu_0}.

The weight value w_{tu_0} for the direct, inhibitory $u_0 \to t$ connection requires a modelling decision brought about by the differences between the cumulative weights along the two excitatory pathways. Presumably, a biological system would determine the weight values for the direct, inhibitory connection through long-term potentiation of the synapse during adaptation. For the numerical experiments reported here for our artificial neural network, we choose to use the maximum of the two excitatory cumulative weight magnitudes. Using this scheme, the $u_0 \to t$ blending weight is calculated as

$$w_{tu_0} = \max(w_{tv_au_0}, w_{ts_{ab}u_0}). \tag{10}$$

In addition, because of this path dependent ambiguity in the choice of weight values affecting all of the direct connections to the t-cell emanating from the u-cells, the numerical experiments take the liberty of exploring the effect of their variation on triangle object detection.

5.2. Results

Two numerical experiments were performed. In these experiments, we measured the performance of the network's object detection as the magnitude of the t-cell response for both high-quality triangle images of the correct size and shape centered in the RF of the t-cell, versus the response magnitude when the image is distorted. In the first experiment the triangle images were centered in the t-cell RF but were successively distorted by noise, regarded as random on-off pixel toggling in a black-and-white (1 – 0) image. In the second experiment the triangle images were successively distorted through translations to the left, right, up, and down in the t-cell RF. These translations could arise from abrupt, yet small, motions of the triangle object in the visual field, perhaps due to sensor oscillations. A high-performing neural network is one for which any distortion in an image reduces the t-cell's response magnitude in comparison with the response for no distortion. Thus, the degree to which the network's activity is reduced as a function of object distortion gives a quantitative measure of its object detection capability. Functional relationships of this kind will be referred to as tuning curves to be consistent with the literature. The following two sections give the details of the experiments and their results.

5.2.1. Noise tuning curve numerical experiments

To compute the noise tuning curve for the triangle cell t, we stimulated our artificial neural network triangle-detector with images centered in the RF of t. Each image consists of the same 734-by-734-pixel triangle corrupted with random noise. There are many ways to model noise processes, but in the following simulations we treat it as pixel noise. To simplify the examples given in the discussion and also our simulations, we consider images having binary-valued pixels. A binary value of 0 indicates no stimulus at that pixel and a value of 1 indicates a stimulus. In Fig. 13 a 0 is shown as a diminutive pixel, a point centered on a white background, whereas a 1 is shown as a black-filled square. A simple noise model for binary images randomly toggles a pixel's value between zero and one. The probability p of toggling any individual pixel in an image will be referred to as the "noise level"; a "noise-free image" would be one in which no pixels are toggled, $p = 0.0$, resulting in a filled triangle of 1 s surrounded by a field of 0 s. Fig. 13 shows examples of a single noise trial on a small vertical-edge patch with two pixel toggle probabilities, $p = 0.1$ and $p = 0.5$.

To calculate a noise level tuning curve for the triangle-detection cell t, we preformed a simplified Monte Carlo simulation over a sample of pixel-toggled 0–1 images. We varied the noise level (pixel toggle probability) p over a discrete range of values, and at each p value the cell's output was averaged over N trials. A trial is defined as the stimulation of the cell by an image drawn from our noise model. We varied the noise level p from 0.0 to 0.5 in increments of 0.005, a total 101 noise level values. At each noise level, we computed the average t-cell output value and standard deviation over $N = 500$ trials. To explore the effect of blending in the triangle cell defining diagram on triangle detection, we introduce a *triangle blending weight gain factor* f and vary it while

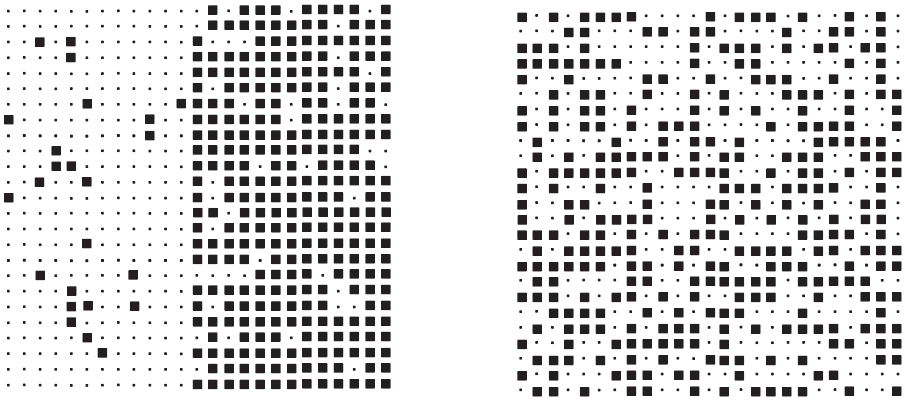


Fig. 13. A vertical-edge image with two pixel toggle probabilities, (a) $p = 0.1$ and (b) $p = 0.5$.

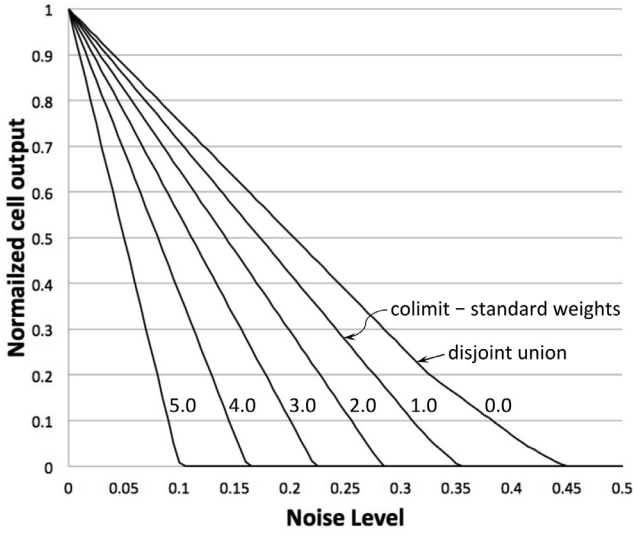


Fig. 14. Triangle noise level tuning curves. The numerical label on each curve displays the triangle cell weight gain factor for that curve. The gain factors 0 and 1 represent the triangle disjoint union (no blending) and colimit (blended) defining diagram configurations, respectively. Gain factors greater than 1 represent enhanced colimit configurations, where the u-cell to t-cell direct connection-weight values are magnified over those for the chosen, standard colimit values. Each point on the curves is the average of 500 trials. The averaged output standard deviations at the sample points on the curves are actually smaller than the vertical thickness of the curves as plotted here.

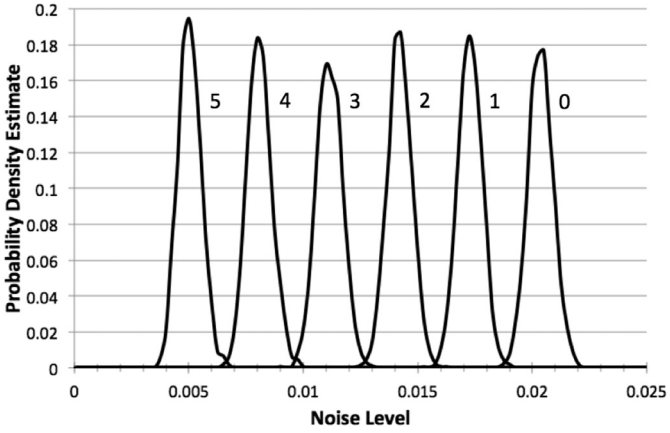


Fig. 15. Numerically computed density functions that give estimates of the probability that the t-cell output will be in the neighborhood of the activation threshold, as a function of noise level and for each of the 6 blending weight gain factors.

repeating selected simulations (to be specified). For example, the $u_0 \rightarrow t$ blending weight will now be calculated as

$$w_{tu_0} = f \times \max(w_{tv_a u_0}, w_{ts_{ab} u_0}) \tag{11}$$

where $f \geq 0$. Setting the weight gain factor $f = 0$ produces the disjoint union configuration (no blending), $f = 1$ produces a colimit configuration which serves as a standard, and $f > 1$ produces an enhanced colimit configuration. It was thought not necessary to try values between 0 and 1; in fact, only whole-number values were tried as shown in the figures.

The tuning curve plots in Fig. 14 show the results of the Monte Carlo simulation. Each curve shows the triangle t-cell output averaged over 500 trials as a function of noise level at a given weight gain factor. The gain factor f varies from 0 to 5 in increments of unity as shown by the curve labels. At each value of the weight gain factor the t cell output is normalized so that it is unity at the zero noise level. The reasoning here is that we wish to measure the effect of the gain factor upon the variation in cell output independently of the output absolute magnitude. Note that the cell output varies almost linearly at each gain factor value, is higher at lower noise levels, and decreases more rapidly with noise level at the higher weight gain factor values. The steeper the curve, the lower the highest noise level at which the t-cell's weighted input sum would exceed the t-cell's activation threshold and therefore the more sensitive the cell's noise discrimination.

We conducted statistical tests on the significance of the difference in performance between the disjoint union and colimit t-cell architectures in the noise tuning curve simulation experiments. First, note that as stated in the caption for Fig. 14, the standard deviation for the average normalized t-cell output as a function of noise level and blending weight gain factors f ranging from 0 to 5 is smaller than the thickness of the plotted tuning curves. For the subsequent analysis, we focused upon an arbitrarily chosen (but fairly demanding) t-cell detection threshold of 0.95 on a scale from 0 to 1. Using the standard t-test, the difference in threshold crossing of the tuning curves at the threshold value 0.95 for $f = 0$ (disjoint union architecture) and $f = 1$ (our standard colimit architecture) is statistically significant with $p < 0.0001$. To make the difference in noise discrimination clearer, we have computed numerical density functions that are estimates of the probability that the t-cell output will be in the neighborhood of the activation threshold 0.95, as a function of noise level and for each of the 6 blending weight gain factors. See Fig. 15.

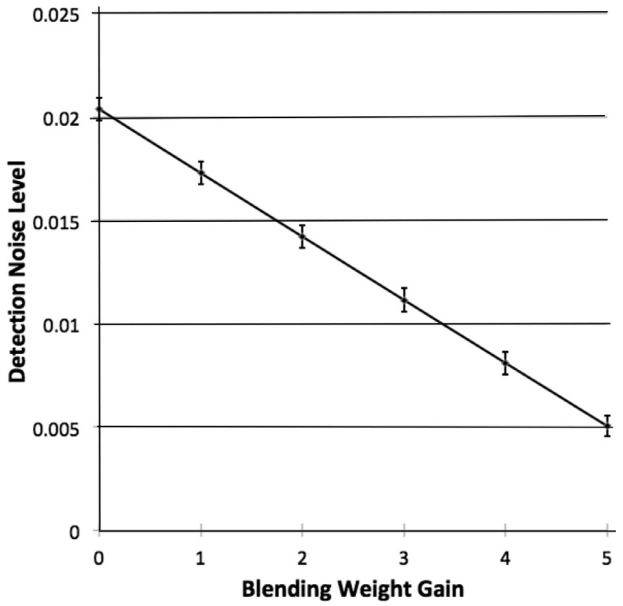


Fig. 16. The mean noise level values p which, when exceeded, will cause the triangle cell output 500-trial average to drop below an output threshold of 0.95, plotted as a function of the triangle blending weight gain factor f. The vertical bars at the points represent the standard deviations of the level mean values.

Table 1
Summary of Experiment 1 results. These are the mean and standard deviation over the noise levels for the detection probability estimates plotted in Fig. 15. Notice that the means shift to smaller values at the larger blending weight gain factors, corresponding to a greater sensitivity to noise (lower probability of t-cell activation at higher noise levels).

Blending weight gain factor	Detection mean	Detection standard deviation
0	0.020362	0.000553
1	0.017279	0.000534
2	0.014204	0.000542
3	0.011138	0.000563
4	0.008085	0.000547
5	0.005039	0.000508

Analyzing the distributions for weight gain factor values $f = 0$ and $f = 1$ in Fig. 15, we observe a very small amount of overlap in their tails. The qualitative interpretation of this is as follows: the probability that the $f = 0$ (disjoint union architecture) and the $f = 1$ (standard colimit architecture) triangle networks can jointly exceed the discrimination threshold of 0.95 is extremely small, and the colimit architecture will do this only within a lower range of noise level values. The t-test quantifies this, yielding the value $p < 0.0001$. Table 1 further supports the conclusion. Therefore we can safely state that the colimit architecture yields a more sensitive discrimination in the presence of noise than does the disjoint union architecture.

To quantify triangle detection under noisy conditions, we applied the detection threshold to the tuning curves, where a Heaviside step function transforms the t-cell 500-trial average output to the value zero when below the threshold and to unity when above the threshold at each noise level. The plot in Fig. 16 summarizes the results from all of the tuning curves, showing the noise level below which the t-cell's output exceeds the 0.95 threshold value, as a function of the weight gain factor f. On this curve, a lower noise level value implies object discrimination with greater sensitivity. This is because the t-cell output is a decreasing function of noise level since image noise interferes with object detection. Therefore, a t-cell which exceeds its detection threshold only for noise magnitudes below a lower noise level is more sensitive than a cell which exceeds its threshold for noise magnitudes below a higher noise level. Fig. 16 illustrates this effect over the range of weight gain factors investigated.

All computations for the creation of the noisy images, the simulations, and subsequent statistical analysis (results below) were performed on an Apple MacBook Pro running OS 10.13.3 at 2.9 GHz. Approx 60 minutes were required to run all $6 \times 101 \times 500 = 303{,}000$ trials (the 6 f values, 101 noise levels, 500 neural network simulations for each combination of f value and noise level). Each trial included the creation of a noisy image and the forward propagation of the image data through the neural network. The performance statistics (the average and standard deviation of the t-cell output over 500 trials) were calculated at each combination of f value and noise level. The cumulative 60 minutes to run the trials includes the $6 \times 101 = 606$ calculations of the performance statistics.

5.2.2. Translation tuning curve numerical experiments

To compute the translation tuning curves for the triangle cell t, we stimulate the neural network with triangle images translated in the i (vertical) and j (horizontal) directions, with (j, i) pixel coordinates measured with respect to a translation of $(0, 0)$ (no translation) centered at the triangle cell's RF. In order to measure solely the effect of translation, these images are uncorrupted by noise or other distortions and therefore multiple trials per translated image are not required. The tuning curve plots in Fig. 17 show the results upon the t-cell output of triangle image shifts in the two directions as functions of translations of i and j pixels in the respective plots. As with the noise experiment, the gain u-cell-to t-cell weight gain factor f varies from 0 to 5 in increments of unity. As with the noise curves, the t-cell output is normalized such that the zero translation output is unity.

As with the noise analysis above, we apply a detection threshold to the tuning curves. In this case, the tuning curves show the effect of translations of the triangle image from the standard position centered in the t-cell's RF. Again we summarize this effect by applying a Heaviside step function to the t-cell output. The Heaviside function transforms the t-cell's output to the value zero when below the threshold and to unity when above the threshold at each translation along either the i (vertical) or j (horizontal) direction. The plot in Fig. 18 summarizes the results from all of the translation tuning curves, showing the translation at which the cells output exceeds an arbitrary threshold of 0.95, as a function of the weight gain factor f. Again, a lower value implies better object discrimination, in this case because the t-cell output is a decreasing function of the i or j translation away from the center of the RF. Therefore, a t-cell which exceeds its detection threshold for translations less than a lower value is more sensitive than a cell which exceeds its threshold for all translations less than a higher value.

6. Discussion and conclusion

We have proposed that the representation of sensed or imagined objects by a neural network lies in neural structures that are both local and distributed. A colimit has this type of structure: it consists of a network that forms a commutative diagram of interconnected cells. That is, all of the cells in the structure are simultaneously active when and only when the presence of its represented object is evident. The diagram, called the defining diagram, has a cell indicating the representation at the apex of a connectionist cone-like structure, where the components, or constituents or parts, of the represented object are in turn represented by cells at the origins of the feedforward connections that define the cone.

An apical cell representing a disjoint union of a collection of objects also has a conelike structure with connections from its

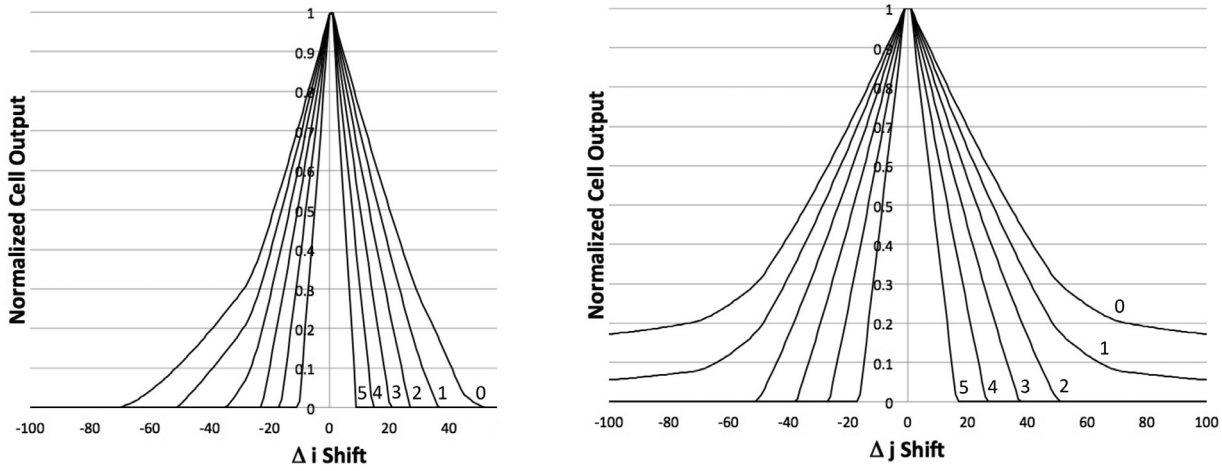


Fig. 17. Triangle cell translation tuning curves where the curve labels display the weight gain factors. Gain factors 0 and 1 represent the triangle disjoint union and colimit configurations, respectively. Gain factors greater that 1 represent enhanced colimit configurations.

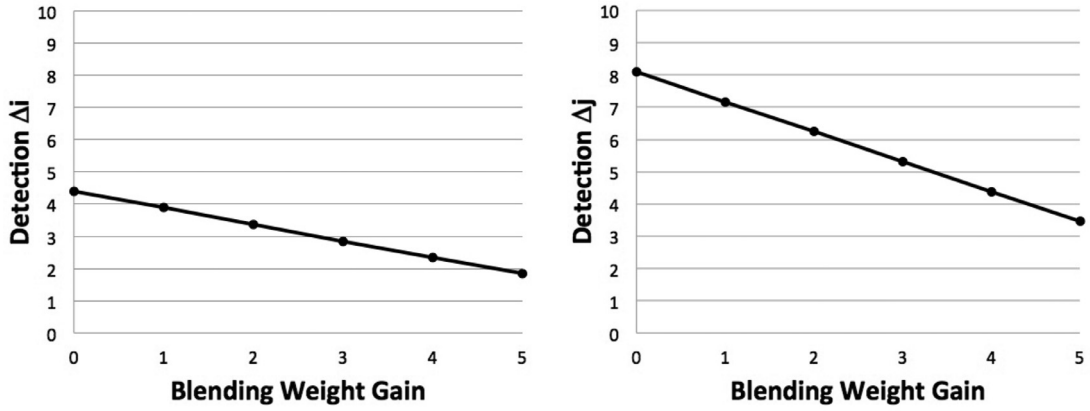


Fig. 18. The translation shifts which, when exceeded, will cause the triangle cell output to drop below an output threshold of 0.95, plotted as a function of the triangle blending weight gain factor *f*.

component-representing cells. However, a disjoint union does not express the information about how the objects are combined; the apical object representation is simply a collection of unrelated objects, with what would be their shared parts regarded differently in each. It is difficult to see this as a local-distributed representation.

As shown in Figs. 8 and 9 on the other hand, a colimit defining diagram has significant additional structure which gives it the capability required to be a valid object representation. The significance of this structure is in its activity states, which are instances of a commutative diagram that expresses the blending of component objects along shared subobjects to form a more complex object. Thus, the defining diagram of a colimit contains not only the representations of the colimit objects's components, but also the manner in which they are joined, or *blended*, in forming the colimit object. Our claim, then, is that an object representation by a neural network is both local and distributed and this is a consequence of the blending of component representations via colimits. While the sensed-object representation is distributed in a commutative diagram of components and blending object representations, it allows holistic access by virtue of its local structure, the colimit cell at the apex.

We suggest that the representations in biological neural networks, at least in mammals and perhaps in birds, are at once local and distributed as described. Expressing object representation via colimits has at least three major advantages, both for the neuroscientist and for the neural network itself. First, it allows neuroscientists to parse the object representation to understand how it is put together (in other words, to determine its full semantics). Second, it confers upon the neural network the ability to both manifest concept (or object) stability and to exploit parsed representations which are active in its working memory. Put another way, this mathematical model of local-distributed neural networks explicates the functional architecture of sensed- (or imagined-) object detection. It states exactly that specific neural structures representing an object in terms of its component objects *and their compositional relationships* must be involved when the objects in question are detected or are called up in mental imagery. Finally, the simulations reveal a performance advantage for colimit neural circuits. The simulated example of a triangle detector network with a colimit configuration exhibits superior discrimination over the disjoint union network configuration for both the noise and translation perturbations. We believe that these results begin to expose a deeper property of local-distributed object representations that relates to biological neural networks. Colimit networks have a greater complexity than the disjoint configuration because of the inclusion of blending object representations enforced by the commutative diagram property. On the other hand, perhaps the extra biological cost of these more complex neural circuits is offset by the fitness advantages of greater object discrimination.

In the full theoretical analysis on which we base our assertions, other local-distributed object representations exist as limits, math-

ematically the dual notion to colimits. Along with colimit formation during synaptic adaptation, defining successively more complex concepts as neural processing progresses further downstream from the sensor level, limits are formed by "pulling back" on concept diagrams to form diagrams which represent abstractions. An abstraction expresses the information common to different concepts. In previous papers such as those mentioned, we have discussed both limits and colimits and have further discussed the necessity of feedback connections operating in parallel with the feedforward connections in their diagrams.

We argue that colimits and limits serve as a local-distributed knowledge system, formed adaptively, that expresses the sensory and autonomously-synthesized information represented by the structure and activity of a neural network. In the mammalian brain, particularly in humans, a region called the medial temporal lobe (MTL) exists evidently at the apex of a hierarchy of representations [18,19]. We hypothesize that colimits, limits and other categorical constructs can be a significant aid in analyzing the hierarchy of representation in the MTL as well as other brain regions.

We also argue that categorical constructs can be useful in bringing mathematical rigor to the analysis and design of neural networks based upon the ability to express the structures needed for concept representation. Ultimately, this can be useful in the design of artificial neural networks capable of serving as artificial brains for autonomous robots. Although it was not our intent to address the practically of our methods for neural network design and analysis in this paper, we mention a selection of our previous work in this regard. We have discussed neural design principles in [20,21]. In [14] we present an application in hyperspectral imaging from satellite data.

References

[1] J.S. Bowers, Challenging the widespread assumption that connectionism and distributed representations go hand-in-hand, Cognit. Psychol. 45 (2002) 413–445.
[2] J.S. Bowers, Grandmother cells and localist representations: a review of current thinking, Lang. Cognit. Neurosci. 32 (3) (2017) 257–273.
[3] R.Q. Quiroga, A. Kraskov, C. Koch, I. Fried, Explicit encoding of multimodal percepts by single neurons in the human brain, Curr. Biol. 19 (2009) 1308–1313.
[4] A. Roy, A theory of the brain: localist representation is used widely in the brain, Front. Psychol. 3 (2012) 1–4.
[5] E.R. Wood, P.A. Dudchenko, R.J. Robitsek, H. Eichenbaum, Hippocampal neurons encode information about different types of memory episodes occurring in the same location, Neuron 27 (2000) 623–633.
[6] A.G. Huth, S. Nishimoto, A.T. Vu, J.L. Gallant, A continuous semantic space describes the representation of thousands of object and action categories across the human brain, Neuron 76 (6) (2012) 1210–1224.
[7] E.T. Rolls, A. Treves, M.J. Tovee, The representational capacity of the distributed encoding of information provided by populations of neurons in primate temporal visual cortex, Exp. Brain Res. 114 (1997) 149–162.
[8] D.E. Stansbury, T. Naselaris, J.L. Gallant, Natural scene statistics account for the representation of scene categories in human visual cortex, Neuron 79 (2013) 1025–1034.
[9] A. Damasio, Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition, Cognition 33 (1989) 25–62.
[10] L.W. Barsalou, Perceptual symbol systems, Behav. Brain Sci. 22 (1999) 577–660.
[11] B. Mitra, F. Diaz, N. Craswell, Learning to Match using Local and Distributed Representations of Text for Web Search, arXiv:1610.08136v1, 2016.
[12] L. Pang, Y. Lan, J. Guo, J. Xu, S. Wan, X. Cheng, Text matching as image recognition, in: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI Press, Palo Alto, CA, 2016.
[13] M.J. Healy, T.P. Caudell, Ontologies and worlds in category theory: implications for neural systems, Axiomathes 16 (1–2) (2006) 165–214.
[14] M.J. Healy, R.D. Olinger, R.J. Young, S.E. Taylor, T.P. Caudell, K.W. Larson, Applying category theory to improve the performance of a neural architecture, Neurocomputing 72 (2009) 3158–3173.
[15] M.J. Healy, T.P. Caudell, Temporal semantics: an extended definition for neural morphisms, in: Proceedings of the 2009 IEEE International Joint Conference on Neural Networks, INNS, 2009. P. paper no. 278.
[16] S.M. Lane, Categories for the Working Mathematician, Springer-Verlag, 1971.
[17] M. Barr, C. Wells, Category Theory for Computing Science, third ed., Les Publications CRM, 1999.
[18] P. Lavenex, D.G. Amaral, Hippocampal-neocortical interaction: a hierarchy of associativity, Hippocampus 10 (2000) 420–430.
[19] F. Mormann, S. Kornblith, R.Q. Quiroga, A. Kraskov, M. Cerf, I. Fried, C. Koch, Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe, J. Neurosci. 28 (2008) 8865–8872.
[20] M.J. Healy, T.P. Caudell, Aphasic compressed representations: a functorial semantic design principle for coupled ART networks, in: Proceedings of the 2002 IEEE International Joint Conference on Neural Networks (IJCNN '02), INNS, IEEE Press, Honolulu, 2002, pp. 2656–2661. (CD-ROM proceedings).
[21] M.J. Healy, T.P. Caudell, From categorical semantics to neural network design, in: Proceedings of the 2003 IEEE International Joint Conference on Neural Networks, IJCNN, IEEE Press, Portland, OR, USA, 2003, pp. 1981–1986. (CD-ROM proceedings), INNS.



Michael J. Healy received a B.A degree in mathematics from Eastern Washington University at Cheney, Washington in 1965 and a M.S. in mathematics from the University of Idaho in 1967. From 1967 through 2001 he worked for The Boeing Company, Seattle. From 1967 through 1989 he performed software R&D for nonlinear programming applications and was the optimization focal point for the Boeing mathematical software libraries. Beginning in 1989, he performed research in neural networks, machine learning, formal verification, and applications of category theory in software synthesis, knowledge based systems (KBS) engineering (KBSE), and knowledge representation and ontologies for KBS interoperability. He retired from Boing in 2002. He joined the Electrical and Computer Engineering Department (ECE) at the University of New Mexico as a Research Scholar in 2000, and has since pursued his research interests in the semantics of neural networks and cognitive neural systems, knowledge representation, and ontologies for computational systems. Category theory continues to be a central mathematical discipline for his research.



Prof. Thomas Preston Caudell received his Ph.D. in Physics with a minor in Astronomy from the University of Arizona, Tucson in 1980. After his postdoctoral work, in 1984 he became a Senior Staff Physicist at the Hughes Artificial Intelligence Center in Malibu CA where he did artificial intelligence, virtual reality, and neural network research. In 1989 he joined The Boeing Company in Seattle, WA as a Senior Principal Scientist, where he did work in neural networks and augmented reality, coining the term in 1990. He joined the University of New Mexico in 1994 where he was Professor of Electrical and Computer Engineering with secondary appointments in Computer Science & Psychology until 2014 when he retired. He is currently a Emeritus Professor in the Department of Electrical and Computer Engineering as well as Research Professor in the Department of Mechanical Engineering. The focus of his research remains in neural networks. Specifically he is addressing neuro-biologically motivated, mathematical formulated theories and simulation models of neural systems. His research is highly interdisciplinary, involving collaborations with neuroscientists, psychologists, mathematicians, and computer scientists.