Laterally connected neural field provides precise centroid estimates

Sebastian Waz (swaz@uci.edu)
Department of Cognitive Sciences, University of California - Irvine
Irvine, CA 92697

Charles Chubb (c.chubb@uci.edu)
Department of Cognitive Sciences, University of California - Irvine
Irvine, CA 92697

Abstract

The convolution operator is an essential tool in digital signal processing. Among many other things, it grants tremendous perceptual ability to a popular class of deep learning models known as convolutional neural networks; however, the convolution operator's neural implementation typically requires dense connectivity, so it also diminishes the biological plausibility of such models. We found that the convolution of an input vector with a parabolic function can be performed without dense connectivity. Specifically, this operation can be performed by a laterally connected (i.e., biologically plausible) neural network that evolves in continuous time (a "neural field"). This particular convolution is shown to have useful properties for centroid estimation. Rapid and precise centroid estimation is known to take place early in the human visual system (Drew, Chubb, & Sperling, 2010) (Sun, Chubb, Wright, & Sperling, 2016), but to date, this computation lacks an adequate biologically plausible neural model. Using the aforementioned results, we present such a model.

Keywords: convolution, neural field, centroid

Methods

A concise description of the model will be facilitated by some preliminary definitions.

Definitions

Shift function Let $\mathbb{Z}$ be the set of all integers. For any function $f : \mathbb{Z} \to \mathbb{R}$ and $k \in \mathbb{Z}$, define the function $S_k(f)$ by setting

$$S_k(f)[x] = f[x-k] \tag{1}$$

for all $x \in \mathbb{Z}$. S_k is called the shift transformation with displacement k.

Convolution matrices It will be convenient to use $0, 1, \cdots, N-1$ to index the rows and columns of an $N \times N$ matrix.

We use the term "convolution matrix" to refer to any $N \times N$ matrix ϕ whose r^{th} row is identical to the 0^{th} row of ϕ circularly shifted by r columns. Specifically, for $r = 1, 2, \cdots, N-1$

$$\phi_{r,c} = \phi_{0,(c-r)^\dagger} \tag{2}$$

where

$$n^\dagger = \begin{cases} n & \text{if } n \geq 0 \\ N+n & \text{otherwise.} \end{cases} \tag{3}$$

Block convolution matrices Further, it will be convenient to use $0, 1, \cdots, M-1$ to index the partitions of an $NM \times NM$ block matrix partitioned into $N \times N$ blocks.

We use the term "block convolution matrix" to refer to any such $NM \times NM$ block matrix Φ whose submatrices are convolution matrices as defined in (2) and whose r^{th} row partition is identical to the 0^{th} row partition of Φ circularly shifted by rN columns. Specifically, for $r = 1, 2, \cdots, M-1$

$$\Phi_{r,c} = \Phi_{0,(c-r)^\dagger} \tag{4}$$

where $\Phi_{r,c}$ is an $N \times N$ convolution matrix and

$$m^\dagger = \begin{cases} m & \text{if } m \geq 0 \\ M+m & \text{otherwise.} \end{cases} \tag{5}$$

A matrix of this form may be viewed as a linear transformation that convolves a 2-D $N \times M$ input (recast as an $NM \times 1$ input vector) with a 2-D periodic function.

We use $\mathcal{X}$ for a particular block convolution matrix whose 0^{th} row elements are equal to the column indices of their $N \times N$ submatrix. Thus, for $N = 4$ and $M = 4$,

$$\mathcal{X}_{0,*} = \begin{bmatrix} 0 & 1 & 2 & 3 & 0 & 1 & 2 & 3 & 0 & 1 & 2 & 3 & 0 & 1 & 2 & 3 \\ 3 & 0 & 1 & 2 & 3 & 0 & 1 & 2 & 3 & 0 & 1 & 2 & 3 & 0 & 1 & 2 \\ 2 & 3 & 0 & 1 & 2 & 3 & 0 & 1 & 2 & 3 & 0 & 1 & 2 & 3 & 0 & 1 \\ 1 & 2 & 3 & 0 & 1 & 2 & 3 & 0 & 1 & 2 & 3 & 0 & 1 & 2 & 3 & 0 \end{bmatrix} \tag{6}$$

We use $\mathcal{Y}$ for a particular block convolution matrix whose 0^{th} row elements are equal to the indices of their column partition. Thus, for $N = 4$ and $M = 4$,

$$\mathcal{Y}_{0,*} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 3 & 3 & 3 & 3 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 3 & 3 & 3 & 3 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 3 & 3 & 3 & 3 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 3 & 3 & 3 & 3 \end{bmatrix} \tag{7}$$

The model

Architecture Our model consists of (1) a primary set of toroidally connected neurons and (2) an auxiliary set of toroidally connected neurons with one-to-one correspondence

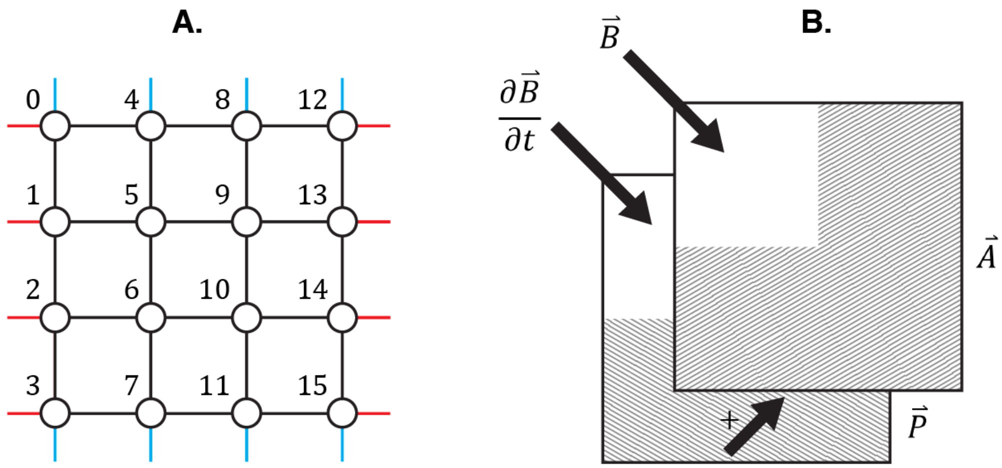

Figure 1: A. Diagram of lateral connectivity for an instance of the model where $N = 4$ and $M = 4$. All connections shown are bidirectional with excitatory weight $\frac{1}{4}$. This connectivity holds for neurons in both $\vec{A}$ and $\vec{P}$. Numbers indicate the index of a neuron within a vector. Red weights complete each row cycle, and blue weights complete each column cycle such that the neuron at index 0 is connected to neurons at indices 3 and 12 (as well as 1 and 4). B. Schematic diagram of overall connectivity in the model. Each square represents a set of neurons $\vec{A}$ or $\vec{P}$ with internal connectivity shown in Figure 1.A. The bold arrows labeled $\vec{B}$ and $\frac{\partial \vec{B}}{\partial t}$ represent input to $\vec{A}$ and $\vec{P}$ respectively. Textured areas of each square represent quadrants where input is forbidden. The bold arrow labeled "$+$" represents one-to-one feedforward connectivity from neurons in $\vec{P}$ to neurons in $\vec{A}$.

to the primary set. These two sets of neurons constitute a dynamical system whose evolution over time can be described by a pair of differential equations:

$$\begin{aligned} \frac{\partial \vec{A}}{\partial t} &= W\vec{A} + \vec{P} - \vec{B} \\ \frac{\partial \vec{P}}{\partial t} &= W\vec{P} - \frac{\partial \vec{B}}{\partial t} \end{aligned} \tag{8}$$

In the above notation, $\vec{A}$ and $\vec{P}$ track the activation of neurons in the primary set and auxiliary set, respectively. The vector $\vec{B}$ tracks the input values to the system, and we think of $\vec{B}$ as a spatiotopic representation of light intensities or Weber contrasts. The matrix W gives the connectivity between neurons. We specify W by first defining a submatrix $\mathcal{W}$.

Let $\mathcal{W}$ be the convolution matrix whose 0^{th} row is defined as follows for all columns c:

$$\mathcal{W}_{0,c} = \begin{cases} -1 & \text{if } c = 0, \\ \frac{1}{4} & \text{if } c = 1 \text{ or } c = N-1, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

Then we specify W to be the block convolution matrix whose 0^{th} row partition is defined as follows for all column partitions c:

$$W_{0,c} = \begin{cases} \mathcal{W} & \text{if } c = 0, \\ \frac{1}{4}I & \text{if } c = 1 \text{ or } c = M-1, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

When we say that the neurons in $\vec{A}$ and $\vec{P}$ are toroidally connected, we mean that each row of W has exactly five non-zero elements that connect a neuron to itself, its two row neighbors, and its two column neighbors. The M diagonal submatrices of W each contain the entire within-cycle connectivity of a single cycle of N neurons connected poloidally. There are M such cycles, and the cycles themselves are connected toroidally.

The indices of neurons in the model correspond to equidistant points in a periodic domain. If we apply the same circular shift to $\vec{A}$, $\vec{P}$, and $\vec{B}$, we do not change the relationship between points. This property leads to degenerate input cases such that the input is cumulatively non-zero, but a centroid for the input cannot be computed (e.g., the case where input values are evenly distributed across the space). To prevent degenerate cases from appearing, we restrict the input to one contiguous quadrant of the periodic domain and require the model to have an even period in both dimensions. Thus, for a period length $N \in 2\mathbb{Z}$ and period width $M \in 2\mathbb{Z}$, valid input must have the form

$$\begin{aligned} &B[n,m] \in \mathbb{R} \text{ for } n = 0, 1, \cdots \frac{N}{2} - 1;\ m = 0, 1, \cdots \frac{M}{2} - 1 \\ &B[n,m] = 0 \text{ for } n = \frac{N}{2}, \cdots, N-1;\ m = \frac{M}{2}, \cdots, M-1. \end{aligned} \tag{11}$$

Dynamics Consider the 0^{th} element of the dot product $W\vec{A}$: this element is the sum of a decay term $-A_0$ (the decay of the 0^{th} neuron's activation in proportion to its present activation) and four quarter-weighted inputs A_1, A_{N-1}, A_M, and $A_{N(M-1)}$ (the excitation of the 0^{th} neuron by its neighbors). The decay term produces a tendency in the 0^{th} neuron toward the resting activation, 0. Since the sum of four quarter-weighted inputs is equal to their mean, W altogether produces a tendency in the 0^{th} neuron toward the mean activation of its neighbors. The same is true of W with respect to all neurons in $\vec{A}$ and $\vec{P}$.

It is also worthwhile to note that the sum of any row or column in W is equal to 0. This means that the within-vector *diffusion* of activation never changes the cumulative activation of the cycle (hence the emphasis on "diffusion"). That is,

$$\begin{aligned} &\sum_{n=0}^{NM-1} \frac{\partial A_n}{\partial t} = 0 + \sum_{n=0}^{NM-1} P_n - \sum_{n=0}^{NM-1} B_n \\ \text{and } &\sum_{n=0}^{NM-1} \frac{\partial P_n}{\partial t} = 0 + \sum_{n=0}^{NM-1} \frac{\partial B_n}{\partial t}. \end{aligned} \tag{12}$$

Thus the terms $W\vec{A}$ and $W\vec{P}$ in (8) preserve the cumulative activation of their respective vectors but redistribute the activation as to smooth any local activation differentials, pushing $\vec{A}$ and $\vec{P}$ toward uniform activation.

Naturally, we are interested in the steady state of the system given some constant input. To understand what this looks like, it is helpful to first consider the steady state of $\vec{P}$. Given $\vec{P} = \vec{0}$ at $t = 0$, $\vec{P}$ will remain a zero vector until there is some change in the input. Let's consider the case that the input begins changing, and at time t_1, the new input pattern stabilizes. Then the cumulative activation in $\vec{P}$ at time t_1 can be derived in the following way:

$$\begin{aligned} \sum_{n=0}^{NM-1} \int_0^{t_1} \frac{\partial P_n}{\partial t} dt &= 0 + \sum_{n=0}^{NM-1} \int_0^{t_1} \frac{\partial B_n}{\partial t} dt \\ \sum_{n=0}^{NM-1} P_n \Big|_0^{t_1} &= \sum_{n=0}^{NM-1} B_n \Big|_0^{t_1} \\ \sum_{n=0}^{NM-1} P_{n,t_1} &= \sum_{n=0}^{NM-1} B_{n,t_1} \end{aligned} \tag{13}$$

Thus at all times $t_1 \geq 0$, the cumulative activation of $\vec{P}$ is equivalent to the cumulative activation of the instantaneous input. As $\vec{P}$ changes, there will be activation differentials in $\vec{P}$ around the input elements that are changing; this differential will be smoothed by the action of the diffusion term $W\vec{P}$ in (8). Thus, on constant input, $\vec{P}$ will converge on a uniform vector where each element is equal to the average activation of the input $\vec{B}$.

As expressed in (8), the change in activation in $\vec{A}$ over time depends on the instantaneous state of $\vec{P}$. Once again, it useful to consider the cumulative within-vector change. Combining (8) and (13), we see that the cumulative voltage in $\vec{A}$ is always 0. Thus any change in the activation from baseline in $\vec{A}$ is accompanied by an equal and opposite change elsewhere in $\vec{A}$. Change is induced by $\vec{B}$ which has a direct inhibitory effect on $\vec{A}$ and an indirect excitatory effect via $\vec{P}$.

Physical intuition For an intuitive description of the outward behavior of our model, an inner tube is a useful physical metaphor. Our metaphorical inner tube has two basic attributes, membrane shape and internal pressure. These correspond to values in $\vec{A}$ and $\vec{P}$, respectively. We can provide "input" to the inner tube by pinching it at several points with varying intensity; similarly, an input image makes an impression on $\vec{A}$ by clamping the activation of the neurons underneath it. By pinching the inner tube, we reduce its volume and increase its internal pressure, inducing distension elsewhere; similarly, an input image increases the cumulative activation in $\vec{P}$, inducing a uniform positive offset across $\vec{A}$. The activation differentials in $\vec{A}$ produced by these forces is smoothed by the action of the diffusion term $W\vec{A}$; in the case of the inner tube, smoothing is the result of tension on the membrane and internal diffusion of air.

At the steady state, the combination of these forces produce an activation pattern that is negative around the input image and positive elsewhere. It can be shown that, in the "readout" quadrant (i.e., the quadrant opposite the input quadrant) the resultant activation pattern is equal to a convolution of the image with a negative parabolic function. This will be stated formally in (19). Moreover, the extremum of this convolution identifies the centroid of the image, as will be stated in (15). Thus, at the steady state, the index of maximum activation along $\vec{A}$ will always be antipodal to the centroid of the input image.

Convolution with a shifted parabolic function

Consider $f**g$ for any function $f : \mathbb{Z}^2 \to \mathbb{R}^2$ with domain $([0,N-1],[0,M-1])$ and periodic function $g[x,y]$ that takes the value $-\left(x-\frac{N}{2}\right)^2 - \left(y-\frac{M}{2}\right)^2$ across $x = 0,\cdots,N-1$ and $y = 0,\cdots,M-1$:

$$\begin{aligned} &f**g[x,y] = \\ &\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} -f[n,m]\left[\left(x-n-\frac{N}{2}\right)^2 + \left(y-m-\frac{M}{2}\right)^2\right] \end{aligned} \tag{14}$$

Theorem

$$\begin{aligned} &\text{If } \sum_{n=0}^{N-1}\sum_{m=0}^{M-1} f[n,m] > 0, \\ &\qquad \underset{x}{\operatorname{argmax}}\{f**g\} = \frac{\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} nf[n,m]}{\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} f[n,m]} + \frac{N}{2}, \text{ and} \\ &\qquad \underset{y}{\operatorname{argmax}}\{f**g\} = \frac{\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} mf[n,m]}{\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} f[n,m]} + \frac{M}{2}. \\ &\text{If } \sum_{n=0}^{N-1}\sum_{m=0}^{M-1} f[n,m] < 0, \\ &\qquad \underset{x}{\operatorname{argmin}}\{f**g\} = \frac{\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} nf[n,m]}{\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} f[n,m]} + \frac{N}{2}, \text{ and} \\ &\qquad \underset{y}{\operatorname{argmin}}\{f**g\} = \frac{\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} mf[n,m]}{\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} f[n,m]} + \frac{M}{2}. \end{aligned} \tag{15}$$

These equations can be derived from the 1^{st} and 2^{nd} partial derivatives of (14) with respect to x and with respect to y.

In the model, we take $f = \vec{B}$. Moreover, let us define a convolution matrix C that performs the convolution of an input vector with g such that $C\vec{B} = f**g$. We specify C by first defining a submatrix $\mathcal{C}$. Let $\mathcal{C}$ be the convolution matrix whose 0^{th} row is defined as follows for all columns c:

$$\mathcal{C}_{0,c} = -\left(c - \frac{N}{2}\right)^2. \tag{16}$$

Then let C be the block convolution matrix whose 0^{th} row partition is defined as follows for all column partitions c:

$$C_{0,c} = \mathcal{C} - \left(c - \frac{M}{2}\right)^2, \tag{17}$$

or (equivalently)

$$C = S_{\frac{N}{2}}(f_1)[\mathcal{X}] + S_{\frac{M}{2}}(f_2)[\mathcal{Y}] \tag{18}$$

for periodic function $f_1[x]$ that takes the value $-x^2$ across $x = -\frac{N}{2}, \cdots, \frac{N}{2} - 1$ and periodic function $f_2[x]$ that takes the value $-x^2$ across $x = -\frac{M}{2}, \cdots, \frac{M}{2} - 1$.

It can be shown that at the steady state,

$$\vec{A} = \frac{1}{NM} C\left(\vec{B} + (k-b)\vec{1}\right) \tag{19}$$

for scalars k and b.

That is, at the steady state, the activation in $\vec{A}$ is equal to an affine transformation of $\vec{B}$ convolved with $-\left(x - \frac{N}{2}\right)^2 - \left(y - \frac{M}{2}\right)^2$. Since $N > 0$ and $M > 0$,

$$\text{argmax}\left\{\frac{1}{NM} C\left(\vec{B} + (k-b)\vec{1}\right)\right\} = \text{argmax}\{C\vec{B}\} = \text{argmax}\{\vec{A}\}. \tag{20}$$

Thus, from (15) we know that at the steady state, the x-index of the extremum in $\vec{A}$ is located a half-period in the x-dimension from the centroid of the input (and similarly for the y-index of the extremum in the y-dimension). An analogous statement is true of the index of the minimum.

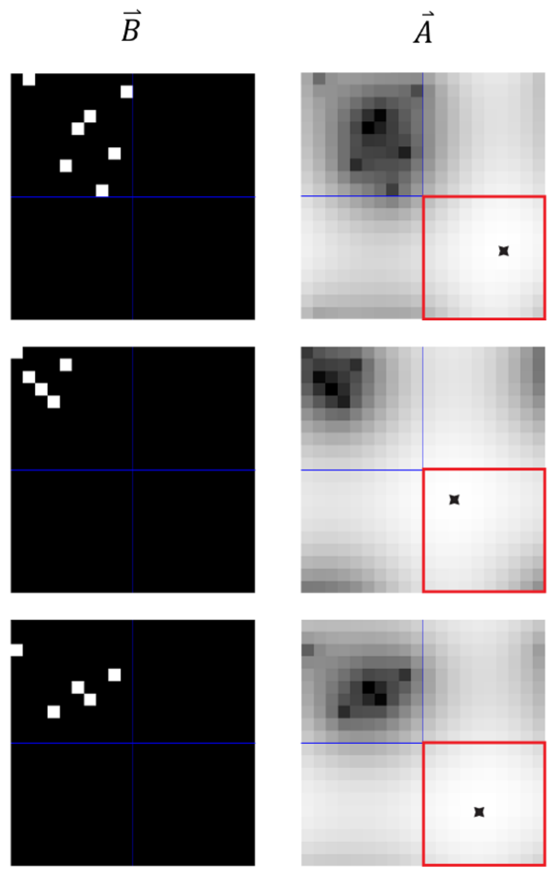

Figure 2: Input and steady state activation for three different dot stimuli given an instance of the model where $N = 20$ and $M = 20$. Values for input and activation range from 0 to 1, corresponding to black and white respectively. The input is restricted to the top left quadrant. The readout quadrant is outlined in red. The index of max activation in $\vec{A}$ is marked by an X. Top: true centroid: $(6.71, 5.29)$; estimate: $(17.00, 15.00)$. Middle: true centroid: $(3.00, 3.00)$; estimate: $(13.00, 13.00)$. Bottom: true centroid: $(5.40, 5.80)$; estimate: $(15.00, 16.00)$. Indices are enumerated left-to-right and top-to-bottom. Note that a difference of $(10, 10)$ between the true centroid and the estimate should be observed for a matrix of these dimensions.

Results

Simulation As stated above, it can be shown that the centroid estimates provided by the model are guaranteed to be precise. Thus, Figure 2 suffices to illustrate the model's ability to estimate the centroid of a sparse dot cloud stimulus, as is used in the study of attentional filters (Drew et al., 2010) (Sun et al., 2016).

References

Drew, S. A., Chubb, C. F., & Sperling, G. (2010). Precise attention filters for weber contrast derived from centroid estimations. *Journal of Vision*, *10*(10), 20–20.

Sun, P., Chubb, C., Wright, C. E., & Sperling, G. (2016). Human attention filters for single colors. *Proceedings of the National Academy of Sciences*, *113*(43), E6712-E6720.