A Neural Network for Attentional Spotlight*

Wee Kheng Leow and Risto Miikkulainen

Technical Report AI91-165
Department of Computer Sciences,
University of Texas at Austin, Austin, Texas 78712

leow@cs.utexas.edu, risto@cs.utexas.edu

Abstract

According to *space-based theory*, visual attention is limited to a local region in space called the *attentional field*. Visual information within the attentional field is enhanced for further processing while information outside is suppressed. There is evidence that enhancement and suppression are achieved with dynamic weighting of network activity. This paper discusses a neural network that generates the appropriate weights, called the *attentional spotlight*, given the size and the position of the intended attentional field. The network has three layers. A *shunting feedback network* serves as the output layer and performs a critical task which cannot be accomplished by feedforward networks.

1 Introduction

Selective visual attention is an important mechanism in the human visual system. With visual attention, the system can allocate its limited resources to the processing of the most important visual information and avoid getting overwhelmed by unimportant information. Psychological experiments indicate that attention can be directed to different parts of a stimulus and attention shift is independent of the movement of the eyes (Treisman et al., 1983; Duncan, 1984; Eriksen and Yeh, 1985). In other words, visual attention is a different process from visual saccade.

Selective visual attention is important in computer vision system as well. To process large amount of visual information in a short period of time, the processes in a computer vision system need to be parallelized (Uhr, 1980; Feldman, 1985). However, complete parallelism is practically impossible because it would require an enormous amount of processors and connections (Tsotsos, 1988). One solution is to process only a small portion of the data in parallel (through selective attention) and sequentially shift the parallel processing to other portions.

There are two complementary psychological models of human visual attention. According to *object-based model* (Kahneman and Henik, 1977; Merikle, 1980; Treisman et al.,

*A more concise version is published in *Proceedings of the International Conference on Neural Networks (Singapore)*, 1991, 436–441.

1983; Duncan, 1984), visual attention is always focused on a single object or a coherent group of visual information. Consequently, multiple judgements on a single object can be made simultaneously, but multiple judgements on different objects cannot. In *space-based model* (LaBerge, 1983; Eriksen and Yeh, 1985; Eriksen and St. James, 1986; LaBerge and Brown, 1986; Anderson, 1990), visual attention is focused on a region localized in space known as the *attentional field*. Information within the attentional field is enhanced for further processing, whereas information outside is largely suppressed and ignored. There is neurophysiological evidence that selective enhancement and suppression are achieved by a gating mechanism that dynamically weights the inputs to neurons (Crick, 1984; Moran and Desimone, 1985; LaBerge, 1990). There is also evidence that the position and the size of the attentional field may vary according to the nature of the visual stimuli and the visual tasks, as well as with the subject's expectations (LaBerge, 1983; Eriksen and Yeh, 1985; Eriksen and St. James, 1986).

Several computational models of visual attention have been proposed (Mozer, 1988; Ahmad and Omohundro, 1990b; Ahmad and Omohundro, 1990a; Mozer and Behrmann, 1990; Sandon, 1990). The mechanism of Mozer and Behrmann uses an optimization process to form a continuous region (not necessarily local) of neural activities consistent with bottom-up inputs and top-down information. Activities outside the region are suppressed. Since the top-down information contains the images of objects, their system is compatible with the object-based psychological model. In Sandon's system, inputs compete locally for attention. The system does not implement the object-based model per se, but appears more compatible with it than with the space-based model. The attention mechanism of Ahmad and Omohundro is consistent with the space-based psychological model. In their system, neural units are laid out in a retinotopic map. Each unit receives the values of the size and the coordinates of the intended attentional field, and computes an appropriate weighting factor for its retinotopic position. Such explicit representation of coordinates in the activity values has two shortcomings. First, biological visual systems do not appear to encode coordinates explicitly. Instead, they represent positions by the locations of active cells in retinotopic maps, i.e. by *value-unit encoding* (Barlow, 1972; Ballard, 1987). Second, explicit encoding may be problematic for hardware implementation. The operational range and precision of devices[1] determine the maximum number of locations that can be encoded, which in turn limits the size of the largest map. Judging by the number of neurons in biological maps and the amount of noise in neural activity, one may speculate that value-unit encoding could well be nature's solution to the hardware problem. Our spotlight-generation network encodes the position of the attentional field using value units and the size of the field using activation value. Explicit encoding of size is acceptable because the largest possible attentional field depends on the largest amount of information that can be processed in parallel, instead of the size of the input layer.

Before describing the network architecture and the simulation results, we will analyze the space-based model of visual attention from the computational point of view.

[1] All analog devices have a finite range over which they can operate properly. In order to encode discrete locations, the analog signal must be discretized, and the accuracy of this operation depends on the amount of noise in the signal. For digital devices, the operational range and precision are determined by the number of bits used for encoding.

2 Computational Model of Space-Based Visual Attention

Two fundamental processes are involved in the space-based model: the selection of the size and the position of the attentional field, and the generation of the appropriate weighting factors. The selection process depends on both bottom-up information and top-down knowledge. Psychological research suggests that many factors, such as the nature of the visual stimuli and the visual tasks, as well as the subject's expectations, affect the size and the position of the attentional field (LaBerge, 1983; Eriksen and Yeh, 1985; Eriksen and St. James, 1986). After deciding on where to focus attention, the selection mechanism passes the information to the generation mechanism. The generation mechanism produces the appropriate weighting factors which filter the information flowing between the modules. Information within the attentional field is strongly weighted while other information is suppressed.

This paper concentrates on the design of the generation mechanism. The computational task can be formulated as follows:

> Given the size and the position of the desired attentional field, compute the weighting factors for space-based visual attention.

We shall call the weighting factors the *attentional spotlight*.[2] A typical implementation of the visual system consists of modules organized in maps. The output units of a module connect to the input units at the corresponding positions in another module. The output units of the generation mechanism connect to the pathways between modules, weighting the signal flow. Consequently, the generation mechanism is most naturally organized in maps as well. This architecture is consistent with neurophysiological mechanisms of visual attention (Crick, 1984; Moran and Desimone, 1985; LaBerge, 1990) and is also used in most of the artificial mechanisms discussed in Section 1.

Our generation mechanism is based on the following design considerations:

1. The position of the spotlight is indicated by a positive input value at the desired location and zero values elsewhere.

2. The size of the spotlight increases monotonically with the input value.

3. The strength of the spotlight should be roughly uniform in the central region (information in this region is roughly equally important), and should decrease gradually with distance from the center of the spotlight (Eriksen and St. James, 1986; Anderson, 1990). In other words, a cross-section of the spotlight should have approximately a bell-shaped or inverted-U-shaped profile (Fig. 1). The position and the size of the spotlight are defined as the coordinates of its center and the radius of its base.

Our network is used as a component in a vision system for image understanding. The system extracts features such as strength and orientation of intensity contrasts and encodes them in maps. Based on bottom-up features and top-down information, the system focuses its attention on different portions of the maps, performs feature integration, and recognizes objects. Only the generation mechanism is discussed in this paper. The selection mechanism and the overall design of the system will be described in later reports.

[2]The word *spotlight* usually refers to the *spotlight model* (Duncan, 1984; Eriksen and St. James, 1986) of visual attention. We extend its meaning to refer to the weighting factors.

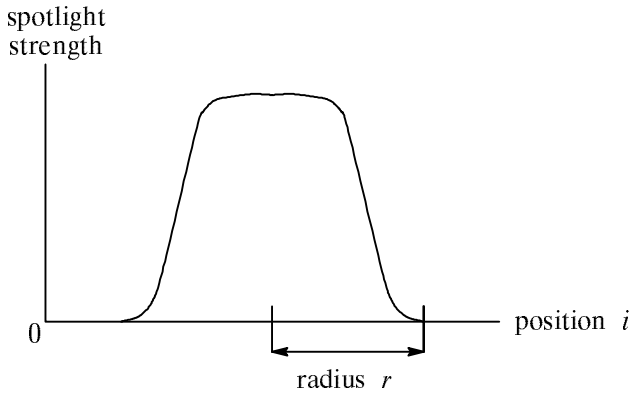


Figure 1: Bell-shaped profile of the attentional spotlight.

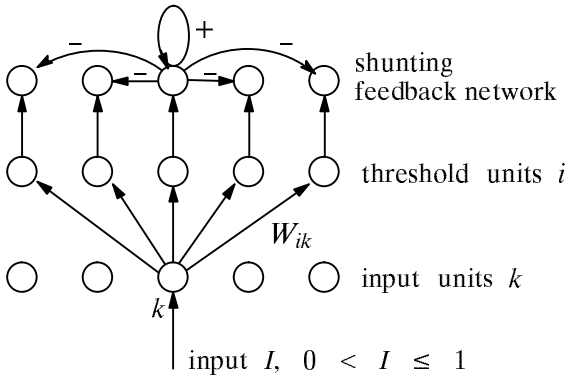


Figure 2: The spotlight generation network. Connections from input unit k distribute the input value I to threshold units i according to the weight distribution W_{ik}. Threshold units make one-to-one connections with units in the shunting feedback layer. Each shunting feedback unit receives feedback from every unit in the layer (not all connections are shown).

3 The Spotlight-Generation Network

In the following discussion, the generation network is assumed to be one-dimensional for simplicity and clarity. Extension of the discussion to two-dimensional network is straightforward.

3.1 Network Architecture

What is required in a network to produce the bell-shaped profile? The task can be divided into three subtasks. Recall that only one of the input units has a positive value while all others have zero values (Design Consideration 1). The first task is to distribute the non-zero input activity over a group of units in the next layer (Fig. 2). The distribution weights W_{ik} should decrease with distance from the center of the group. One possible form is the triangular distribution (Fig. 3):

$$W_{ik} = \frac{1}{R}(R - |i-k|), \quad 0 \leq |i-k| \leq R, \tag{1}$$

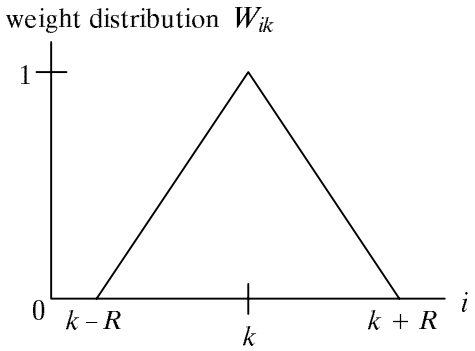


Figure 3: Weight distribution function W_{ik} centered at position k. R is a positive constant.

where R is a positive constant parameter. In this case, $0 \leq W_{ik} \leq 1$.

The second task is to enforce the monotonic relationship between the radius r of the spotlight and the strength of the input I (Design Consideration 2). This task is performed by thresholding the distributed activity:

$$b_i = \left[\sum_k W_{ik}\, a_k - \theta_r \right]^+ , \tag{2}$$

where b_i and a_k are activities of threshold unit i and input unit k, respectively; θ_r is a global positive constant threshold, and $[x]^+ = \max(x, 0)$. Suppose that input unit c is non-zero, i.e. $a_c = I$. Then, the activities of the threshold units are

$$b_i = [W_{ic} I - \theta_r]^+ . \tag{3}$$

In effect, the threshold θ_r cuts the triangular input pattern to produce an output pattern of smaller radius r (Fig. 4b) which is also the radius of the final spotlight. From the figure, we can see that the base of the triangle increases monotonically with I. The equation of r in terms of I is derived in Appendix B.

The third task is to flatten the central region of the activity pattern (Design Consideration 3). To do this properly, a feedback network is required. If a feedforward network is used, it must consist of units whose activities saturate at large input values. For example, the threshold units can be connected to sigmoid units of the form

$$e_i = \frac{1}{1 + e^{-\gamma b_i}}, \tag{4}$$

where e_i is the activity of the sigmoid unit and γ is a positive constant which determines the slope of the sigmoid. For sufficiently large b_i, the sigmoid units saturate and flatten the central region (Fig. 4c). However, saturation will not occur for small b_i which result from small input strength I. Consequently, the spotlight will have a peak at the center (Fig. 4c). Increasing the slope parameter γ to force more rapid saturation will not eliminate the peaking effect although it will reduce the range of I where the peaks occur. Increasing γ also steepens the boundaries of the spotlight, conflicting the requirement that they should decrease gradually (Design Consideration 3). The peaking problem is solved by a *shunting feedback network*.

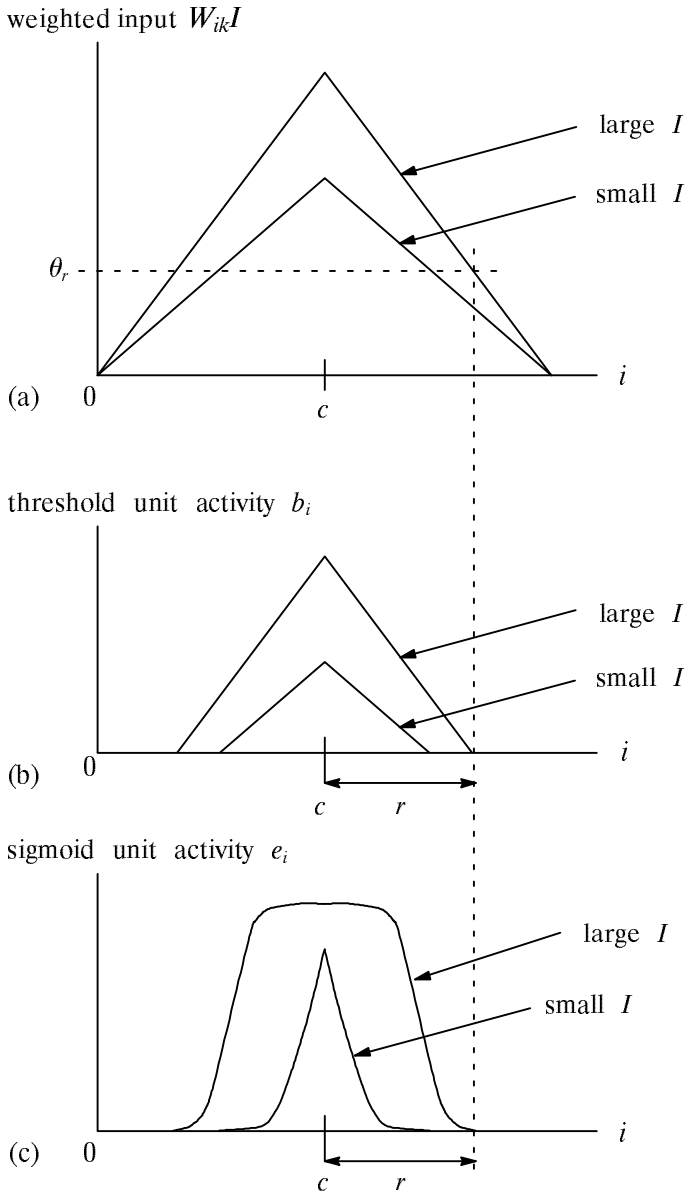


Figure 4: (a), (b) Response of the first two layers of the spotlight generation network. (a) Weighted input pattern $W_{ik}I$ received by threshold units. (b) Activity b_i obtained by thresholding the weighted input pattern against θ_r. Radius r of non-zero activity increases monotonically with respect to input strength I. (c) Response of a feedforward network with sigmoidal output units. The sigmoid units saturate when b_i is sufficiently large, and a spotlight with a flat central region is produced. With sufficiently small I, however, none of the b_i's reaches saturation level, and a peak is produced.

3.2 The Shunting Feedback Network

A shunting feedback network (Grossberg, 1973; Grossberg, 1988) consists of units whose outputs feed back to themselves (Fig. 2). The feedback is weighted by the activities of the receiving units (hence the term *shunting*). An example of the dynamics of the network is given by:

$$\dot{e}_i \equiv \frac{de_i}{dt} = -Ae_i + (B - e_i)f(e_i) - (C + e_i)\sum_{k \neq i} f(e_k) + J_i \tag{5}$$

where e_i is the activity of unit i; A, B and C are positive constant parameters, $J_i \geq 0$ is the input to unit i, and $f(e_i)$ is an arbitrary output function. The term $-Ae_i$ represents decay with rate A. $(B - e_i)f(e_i)$ is the excitatory self-feedback term, and $-(C + e_i)\sum_{k \neq i} f(e_k)$ is the total inhibitory feedback term. Both the self-feedback signal $f(e_i)$ and the inhibitory feedback signal $f(e_k)$ are shunted, i.e. weighted by e_i. The inputs J_i's are only switched on very briefly to set the initial values of e_i. After the J_i's have been switched off (i.e. $J_i = 0$), the network gradually converges to an equilibrium state.

The network given by Eq. 5 exhibits many interesting properties (Grossberg, 1973). For example, the activity e_i is always bounded between $-C$ and B regardless of the size of the input J_i. The derivative $\dot{e}_i$ changes sign when e_i reaches the bounds which prevents e_i from getting beyond the bounds. It has also been shown that under certain mild conditions, the activities can reverberate (i.e. be sustained) indefinitely even after the inputs are switched off. When the network reaches an equilibrium state, its total activity is automatically normalized to a value that depends on the network parameters and the relative strength of the inputs (i.e. $J_i / \sum_i J_i$).

The type of the output function $f(e_i)$ is critical to the performance of the network (Grossberg, 1973). If $f(e_i)$ is linear, the equilibrium relative activities (i.e. $e_i / \sum_i e_i$) will be identical to the relative inputs. If $f(e_i)$ is slower than linear, the differences between the inputs will dissipate and the equilibrium activities will be uniform. If $f(e_i)$ is faster than linear, the differences between the inputs will be enhanced and only the largest activity will survive. If $f(e_i)$ is a combination of linear, slower-than-linear and faster-than-linear functions, then the equilibrium activities will be determined by the complex interactions between the different types of functions. For example, suppose that $f(e_i)$ is sigmoidal, i.e. faster than linear for small e_i, approximately linear for e_i in the middle range, and slower than linear for large e_i. Then, a *quenching threshold* (determined by network parameters) exists such that relative activities below the threshold are quenched (become ≤ 0), and relative activities above the threshold are contrast-enhanced and sustained (Grossberg, 1973).

In our network, $C = 0$ and $J_i = b_i$, and equation (5) becomes:

$$\dot{e}_i = -Ae_i + (B - e_i)f(e_i) - e_i \sum_{k \neq i} f(e_k) + b_i. \tag{6}$$

In this case, e_i is bounded between 0 and B. The output function $f(e_i)$ has the following form (Fig. 5):

$$f(e_i) = e_i\, g(e_i) \tag{7}$$

$$g(e_i) = \begin{cases} D & \text{if } 0 \leq e_i < \theta_e \\ D_0 + \dfrac{D - D_0}{\theta_e - B}(e_i - B) & \text{if } \theta_e \leq e_i \leq B \end{cases} \tag{8}$$

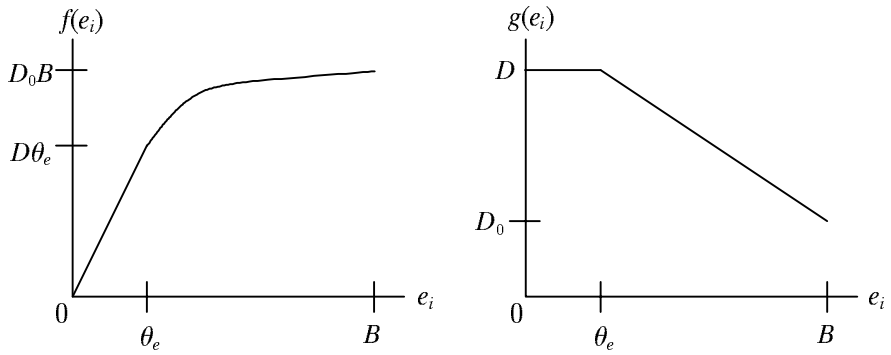


Figure 5: Output function $f(e_i) = e_i\, g(e_i)$. $f(e_i)$ is linear between 0 and θ_e, and slower-than-linear between θ_e and B.

where D and D_0 are positive constants with $D > D_0$, and θ_e is a positive constant parameter which influences the shape of the spotlight (see Appendix C). $f(e_i)$ is linear for $0 \leq e_i < \theta_e$ and slower than linear for $\theta_e \leq e_i \leq B$. With this type of $f(e_i)$, two critical values of relative activity determine the equilibrium state of the shunting feedback layer (Grossberg, 1973). One of them is the *fair distribution limit*: if all relative activities are below this limit, then they will remain unchanged at equilibrium (effect of the linear portion of $f(e_i)$). The other critical value is the *uniformization threshold*: if all relative activities of the units are above this threshold, then the equilibrium activities will be uniform (effect of the slower-than-linear portion of $f(e_i)$). In general, not all relative activities are below the fair distribution limit, and not all of them are above the uniformization threshold. In this case, our simulation results suggest that a *partial uniformization threshold* exists somewhere between the critical values. Relative activities above this threshold (the central region of the spotlight) become uniform, whereas the sub-threshold relative activities (the boundary region) retain their original patterns. Since the fair distribution limit can be computed easily (Grossberg, 1973), we use that as the lower bound for the partial uniformization threshold. In the simulations, this threshold is then adjusted upwards to obtain the desired shape for the spotlight.

The response of the shunting feedback layer is determined by the relative values (i.e. the shape) of its initial activity rather than the absolute values. Since its input comes from the threshold layer, whose output shape is independent of the input strength I, the shape of its response is the same for all values of I. Hence, the peaking effect observed in feedforward networks is avoided. A more detailed mathematical treatment of the network dynamics is given in the appendices.

4 Simulations

The spotlight generation network was simulated numerically using forward Euler method (Wilson and Bowser, 1989) with $\Delta t = 0.02$. The threshold parameter θ_r (of the threshold units, Eq. 2) was set at 0.5. Thus, the smallest spotlight is produced with I just above 0.5 (see Eqs. 1 and 3). When $I \leq 0.5$, no spotlight is generated since there is no activity in the threshold and shunting feedback layers. $I = 1.0$ produces the largest spotlight. In general,

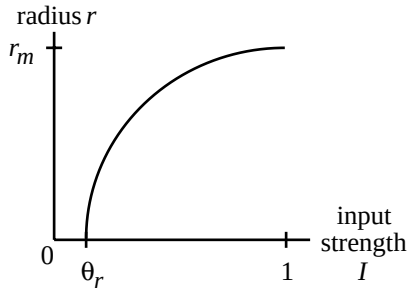


Figure 6: Radius r of the spotlight as a function of input strength I. r increases monotonically with I.

the radius r of the spotlight is given by

$$r = R(1 - \frac{\theta_r}{I}), \tag{9}$$

i.e. r increases monotonically with I, as illustrated in Fig. 6. In the simulations, $R = 40$ and r ranged between 0 and 20 units. The derivation of Eq. 4 and the values of the other network parameters are given in the appendices.

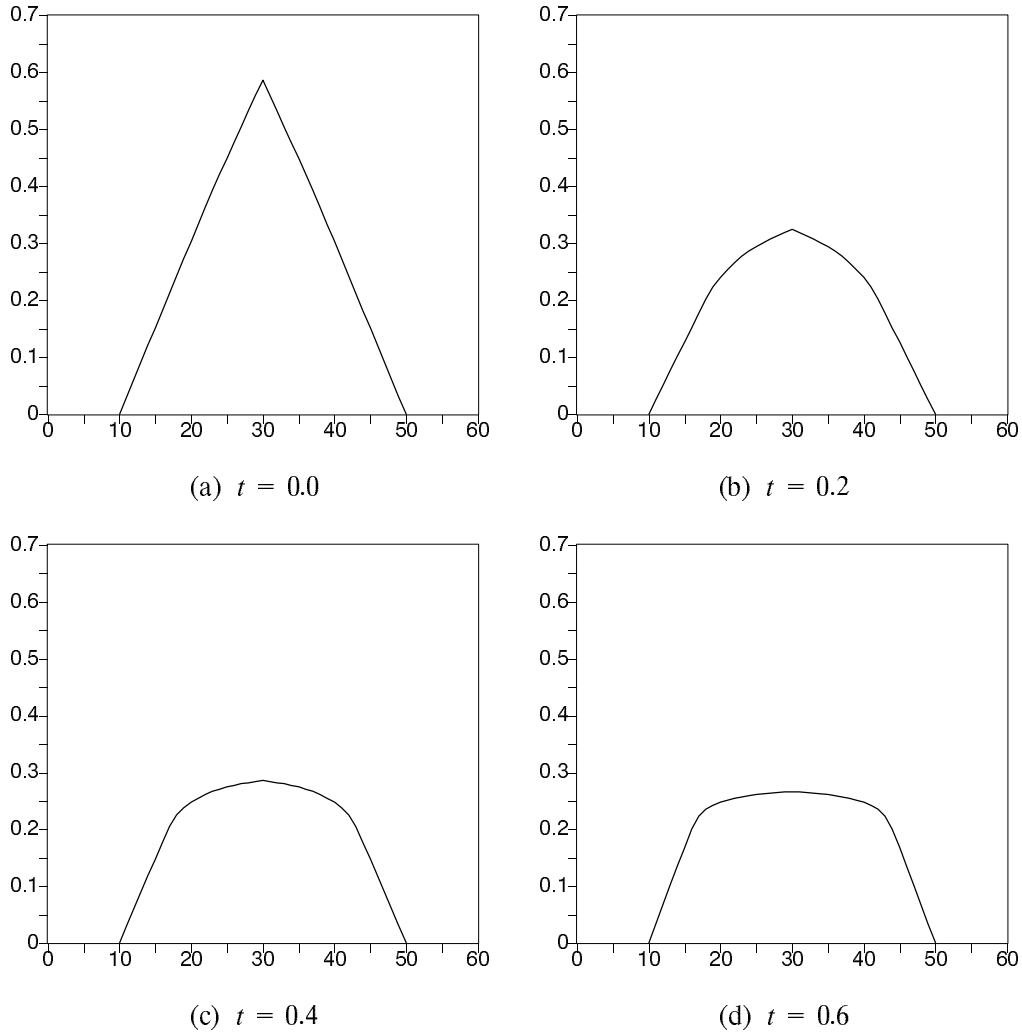
Simulation results for different values of I are shown in Figs. 7, 8 and 9. Figure 7 illustrates the transformation of the activity of the shunting feedback layer through time. The input strength I in this case was 1.0, and consequently, the radius r of the spotlight was 20. Over time, the central region of the spotlight became roughly uniform while the boundary regions retained their gradual decrease in activity.

Figures 7, 8 and 9 show how the radius of the spotlight can be controlled through the input strength. For stronger input, the network generates a larger spotlight. Note that even when the input is as small as 0.55, the central region of the spotlight is still uniform.

Figures 7, 8 and 9 indicate that narrower spotlights are taller. This phenomenon is due to the normalization of total activity (Section 3.2), i.e., the area under the curve is the same for different radii. Consequently, the relative strength of the spotlight, instead of its absolute strength, should be used in weighting the information flowing between visual modules. On the other hand, if the input layer of the receiving module is another shunting network (feedforward or feedback), then it automatically uses the relative input, and renormalization is not required.

There are two qualitatively different equilibrium states. If the parameter θ_e is large enough (about 0.5 in our simulation), the spotlight at equilibrium has a flat central region with gradually sloping boundaries (trapezoidal shape). This observation indicates that the partial uniformization threshold indeed exists. The weight distribution function W_{ik} determines the shape of the boundary regions since those shapes are retained in the spotlight. On the other hand, if θ_e is small, then the (total) uniformization threshold tends to be lower than the relative activities and the equilibrium spotlight becomes uniform with sharp boundaries (rectangular shape).

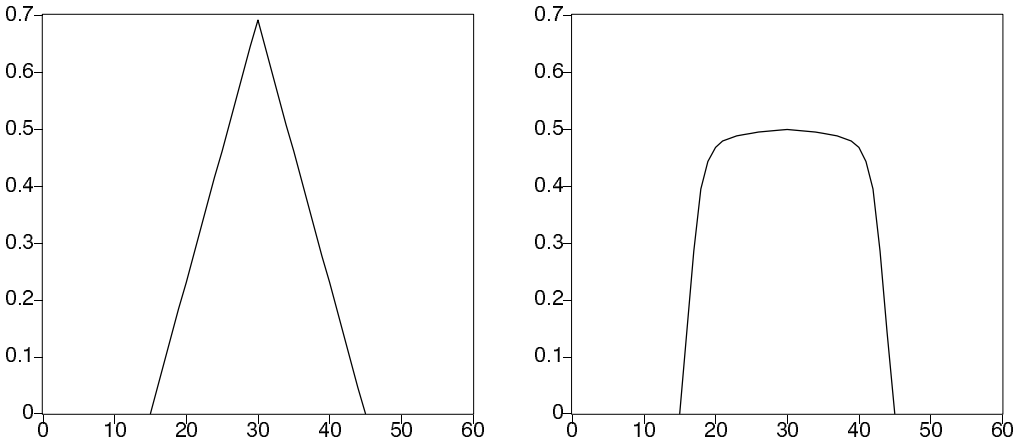
In both cases, as the shunting feedback network relaxes, the intermediate shape of the spotlight has a curved central region (Fig. 7). An additional gating layer could be added to the network to control the *gating time* at which the output of the shunting feedback layer



(a) $t = 0.0$ (b) $t = 0.2$

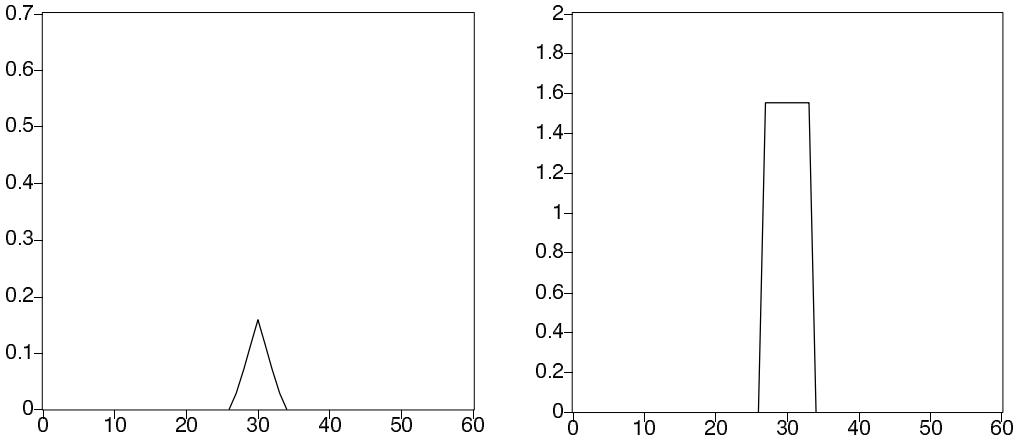
(c) $t = 0.4$ (d) $t = 0.6$

Figure 7: Response of the shunting feedback network with $\Delta t = 0.02$ and $I = 1.0$. (a)–(d) show the transformation of activity pattern over time.



(a) $t = 0.0$ (b) $t = 0.6$

Figure 8: Response of the shunting feedback network with $\Delta t = 0.02$ and $I = 0.8$.



(a) $t = 0.0$ (b) $t = 0.6$

Figure 9: Response of the shunting feedback network with $\Delta t = 0.02$ and $I = 0.55$.

is released. In other words, the network can output spotlights of different shapes depending on whether the output is taken from the equilibrium state or an intermediate state. The appropriate gating time can be easily determined experimentally.

5 Conclusions

The architecture and dynamics of a neural network that generates an attentional spotlight have been discussed. In the input to the network, the position of the attentional field is encoded using value units and the size of the field using activation value. We have demonstrated that feedforward networks cannot generate inverted-U-shaped spotlights with small radii: when the input is small, a sharp peak is produced at the center. This peaking effect is eliminated with shunting feedback. The simulation results show that a partial uniformization threshold exists. This threshold can be adjusted to produce trapezoidal or rectangular spotlights at equilibrium. In addition, spotlights with curved central regions and gradually sloping boundaries can be produced as intermediate states of the network.

References

Ahmad, S., and Omohundro, S. (1990a). Equilateral triangles: A challenge for connectionist vision. In *Proceedings of 12th Annual Conference of Cognitive Science Society.*

Ahmad, S., and Omohundro, S. (1990b). A network for extracting the locations of point clusters using selective attention. Technical Report TR-90-011, International Computer Science Institute, Berkeley, CA.

Anderson, G. J. (1990). Focused attention in three-dimensional space. *Perception & Psychophysics*, 47:112–120.

Ballard, D. H. (1987). Cortical connections and parallel processing: Structure and function. In Arbib, M. A., and Hanson, A. R., editors, *Vision, Brain, and Cooperative Computation*, 563–621. Cambridge, Massachusetts: MIT Press.

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, 1:371–394.

Crick, F. (1984). The function of the thalamic reticular complex: The searchlight hypothesis. *Proceedings of National Academy of Sciences USA*, 81:4586–4590.

Duncan, J. (1984). Selective attention and the organization of visual information. *Experimental Psychology: General*, 113:501–517.

Eriksen, C. W., and St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, 40(4):225–240.

Eriksen, C. W., and Yeh, Y. Y. (1985). Allocation of attention in the visual field. *Experimental Psychology: Human Perception and Performance*, 11(5):583–597.

Feldman, J. A. (1985). Four frames suffice: A provisional model of vision and space. *Behavioral and Brain Sciences*, 8:265–313.

Grossberg, S. (1973). Contour enhancement, short term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, 52:213–257.

Grossberg, S. (1988). *Neural Networks and Natural Intelligence*. Cambridge, Massachusetts: MIT Press.

Kahneman, D., and Henik, A. (1977). Effects of visual grouping on immediate recall and selective attention. In Dornie, S., editor, *Attention and Performance VI*, 17–61. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

LaBerge, D. (1983). Spatial extent of attention to letters in words. *Experimental Psychology: Human Perception and Performance*, 9:371–179.

LaBerge, D. (1990). Thalamic and cortical mechanisms of attention suggested by recent positron emission tomographic experiments. *Cognitive Neuroscience*, 2(4):358–372.

LaBerge, D., and Brown, V. (1986). Variations in size of the visual field in which targets are presented: An attentional range effect. *Perception & Psychophysics*, 40(3):188–200.

Merikle, P. M. (1980). Selection from visual persistence by perceptual groups and category membership. *Experimental Psychology: General*, 109:279–295.

Moran, J., and Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, 229:782–784.

Mozer, M. C. (1988). A connectionist model of selective attention in visual perception. In *Proceedings of 10th Annual Conference of Cognitive Science Society.*

Mozer, M. C., and Behrmann, M. (1990). On the interaction of selective attention and lexical knowledge: A connectionist account of neglect dyslexia. *Cognitive Neuroscience*, 2(2):96–123.

Sandon, P. A. (1990). Simulating visual attention. *Cognitive Neuroscience*, 2(3):213–231.

Treisman, A. M., Kahneman, D., and Burkell, J. (1983). Perceptual objects and the cost of filtering. *Perception & Psychophysics*, 33:527–532.

Tsotsos, J. K. (1988). A 'complexity level' analysis of immediate vision. *International Journal of Computer Vision*, 1(4):303–320.

Uhr, L. M. (1980). Psychological motivation and underlying concepts. In Tanimoto, S. L., and Klinger, A., editors, *Structured Computer Vision*. New York: Academic Press.

Wilson, M. A., and Bowser, J. M. (1989). The simulation of large-scale neural networks. In Koch, C., and Segev, I., editors, *Methods in Neuronal Modeling*. Cambridge, Massachusetts: MIT Press.

Appendix A: Network Definitions

The network for generating the attentional spotlight has three layers (Fig. 2). The dynamics of these layers are described by the following equations:

Input layer:

$$a_i = I_i, \tag{10}$$

where a_i is the activity of unit i, and I_i is the intensity of the input to unit i. Only one of the inputs is non-zero (between 0 and 1); all others are zero.

Threshold layer:

$$b_i = \left[\sum_k W_{ik} a_k - \theta_r \right]^+, \tag{11}$$

where b_i and a_k are the activities of threshold unit i and input unit k, respectively; W_{ik} is the weight on the connection from input unit k to threshold unit i, θ_r is a global positive constant threshold that influences the radius r of the spotlight, and $[x]^+ = x$ if $x > 0$, and 0 otherwise. In the simulation, a triangular distribution function is used for the weights (Fig. 3):

$$W_{ik} = \frac{1}{R}(R - |i-k|),\ 0 \le |i-k| \le R, \tag{12}$$

where R is a positive constant and $0 \le W_{ik} \le 1$.

Shunting feedback layer:

$$\dot{e}_i = -Ae_i + (B - e_i) f(e_i) - e_i \sum_{k \neq i} f(e_k) + b_i, \tag{13}$$

where e_i is the activity of unit i, A and B are positive constant parameters, and $f(e_i)$ is the output function. The activity e_i is bounded between 0 and B. The function $f(e_i)$ has the following form (Fig. 5):

$$f(e_i) = e_i\, g(e_i) \tag{14}$$

$$g(e_i) = \begin{cases} D & \text{if } 0 \le e_i < \theta_e \\ D_0 + \frac{D - D_0}{\theta_e - B}(e_i - B) & \text{if } \theta_e \le e_i \le B, \end{cases} \tag{15}$$

where D and D_0 are positive constants such that $D > D_0$, and θ_e is a positive constant parameter which influences the shape of the spotlight. The value of D_0 (Eq. 15) is not critical as it affects mainly the rate of convergence. It can be as small as 0.1 or as large as 5.0.

Appendix B: Operation of the Network

Let a_c be set to I, $0 < I \le 1$, very briefly just before time $t = 0$. At this brief instance, input I is passed to the shunting feedback layer:

$$e_i(0) \equiv e_i(t=0) = b_i = [W_{ic} I - \theta_r]^+, \tag{16}$$

At time $t = 0$, the input is switched off. The shunting feedback layer gradually approaches the equilibrium state. For notational simplicity, let $j = |i - c|$. Then,

$$W_j \equiv W_{j0} = \frac{1}{R}(R - j), \ 0 \le j \le R, \text{ and} \tag{17}$$

$$e_j(0) = \left[\frac{I}{R}(R - j) - \theta_r\right]^+. \tag{18}$$

The radius of the spotlight, r, is determined by setting $\theta_r = W_r I$ (Fig. 4b), giving

$$r = R(1 - \frac{\theta_r}{I}). \tag{19}$$

This equation shows that r increases monotonically with respect to I (Fig. 6). The maximum radius r_m is obtained when $I = 1$, i.e.,

$$r_m = R(1 - \theta_r). \tag{20}$$

Appendix C: Setting the Partial Uniformization Threshold

The fair distribution limit L_f is given by (Grossberg, 1973)

$$L_f = \frac{\theta_e}{\max(B - A/D, E(0))} \tag{21}$$

where $E(0)$ is the total activity of the network at time $t = 0$. If we ensure that $B - A/D$ is always greater than $E(0)$ regardless of I, then L_f is given by

$$L_f = \frac{\theta_e}{B - A/D}. \tag{22}$$

To do so, we have to first obtain the maximum of $E(0)$. $E(0)$ is the sum of activities $e_j(0)$, i.e., from Eq. 18 and 19,

$$E(0) = 2\sum_{j=0}^{r} e_j(0) - e_0(0) = \frac{I}{R}r^2 \tag{23}$$

Since both $E(0)$ and r are maximum when $I = 1$, the maximum of $E(0)$ is r_m^2/R. Therefore, the network parameters must be chosen such that $B - A/D > r_m^2/R$. The relative activity of each unit at time $t = 0$ with $I = 1$ is

$$E_j(0) = \frac{e_j(0)}{E(0)} = \begin{cases} \frac{r_m - j}{r_m^2} & \text{if } 0 \le j \le r_m \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

In the simulation, we set $\theta_r = 0.5$ and $R = 40$, giving $r_m = 20$ and $r_m^2/R = 10$. Taking $B = 12$ and $A = D = 10$ ensures that $B - A/D = 11 > r_m^2/R$ and satisfies the condition for applying Eq. 22. Let ρ denote the maximum radius of the central region.

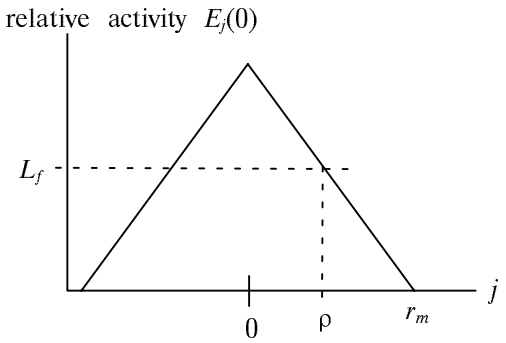


Figure 10: Relationship between the fair distribution limit L_f and the maximum radius ρ of the central region.

To produce an approximate bell-shaped profile, set L_f (the lower bound of the partial uniformization threshold) such that the relative activities $E_j(0)$ beyond the range of ρ are below L_f (Fig. 10). Let $\rho = 0.4R$. Then, with Eqs. 22 and 24, we obtain the value for θ_e as follows:

$$L_f = \frac{\theta_e}{11} = E_\rho(0) = \frac{4}{400}, \text{ and thus,} \tag{25}$$

$$\theta_e = 0.11 \tag{26}$$

Simulation results indicate that when $\theta_e >= 0.5$, the equilibrium state of the shunting feedback layer has a flat central region with boundary regions decreasing gradually in strength (trapezoidal shape). With smaller θ_e, the equilibrium distribution tends to be uniform (rectangular shape). In all cases, the intermediate states have curved central regions.