Experimental Results Indicate that Generation, Local Receptive Fields and Global Convergence Improve Perceptual Learning in Connectionist Networks

Vasant Honavar and Leonard Uhr

Computer Sciences Technical Report #805

November 1988

Experimental Results Indicate that Generation, Local Receptive Fields and Global Convergence Improve Perceptual Learning in Connectionist Networks

Vasant Honavar and Leonard Uhr
Computer Sciences Department
University of Wisconsin-Madison

Abstract

This paper presents and compares results for three types of connectionist networks:

[A] Multi-layered converging networks of neuron-like units, with each unit connected to a small randomly chosen subset of units in the adjacent layers, that learn by re-weighting of their links;

[B] Networks of neuron-like units structured into successively larger modules under brain-like topological constraints (such as layered, converging-diverging heterarchies and local receptive fields) that learn by re-weighting of their links;

[C] Networks with brain-like structures that learn by generation-discovery, which involves the growth of links and recruiting of units in addition to re-weighting of links.

Preliminary empirical results from simulation of these networks for perceptual recognition tasks show large improvements in learning from using brain-like structures (e.g., local receptive fields, global convergence) over networks that lack such structure; further substantial improvements in learning result from the use of generation in addition to reweighting of links. We examine some of the implications of these results for perceptual learning in connectionist networks.

This work was partially supported by the National Science Foundation and the University of Wisconsin-Madison.

Introduction

Connectionist networks are graphs of linked nodes of the following sort: Each node is a simple neuron-like unit. Each link has a weight associated with it. The net input to a node is a weighted sum of the outputs of the nodes that are actively firing into it. Each node applies some form of non-linear function (such as the threshold or the sigmoid) to its net input and sends the result to other nodes to which it is connected via its output links.

It is easy to show that networks of threshold units are universal computing engines (McCulloch, 1943) in the sense that there exist such (sufficiently large) networks that can compute any function computable by a Turing machine or a system of Post Productions; but the problem of finding the necessary, sufficiently powerful, efficient and robust networks for perceptual recognition tasks remains, just as it does no matter how we try to embody intelligent processes.

Learning in connectionist networks can involve modification of any of the following:

[1] Processing functions of the nodes (e.g., changes in the threshold or the output function),

[2] The weights associated with the links,

[3] The topology of the network (addition and deletion of links and nodes), and

[4] The learning rules themselves.

Most of the work on learning in connectionist networks to date has concentrated on [2]. Several algorithms for changing weights associated with the links are available (Hinton, 1987). Some of them utilize feedback that allows the network to compute the error between its output and the desired output and use the back-propagated error to change the weights, e.g., the generalized delta rule (Rumelhart, 1986). Some use a kind of reinforcement learning that enables the network to utilize feedback in the form of a *reward* for good actions or a *penalty* for bad ones. If a unit can learn to increase the frequency of reward from a noisy *critic*, it can act cooperatively with other units in the network to improve the performance of the entire network (Barto, 1985). Some use a form of association learning i.e., a link between two units is strengthened if both of them fire at the same time. Such a scheme tends to sharpen the unit's predisposition *without any external feedback*, getting its firing to become better and better correlated with a cluster of stimulus patterns (Hebb, 1949).

A learning scheme for [3] that employs a mechanism for growth of links and recruiting of nodes guided by regulatory mechanisms designed to discover minimally complex networks has been described in (Honavar, 1988).

Complexity Issues

Given the Turing equivalence of (sufficiently large) connectionist networks, the problem of building such networks for perceptual recognition tasks is reduced to one of discovering the design principles that yield economically feasible structures (for machine perception) and/or biologically plausible structures (for brain modeling). We briefly examine the complexity of perceptual recognition, and list some observations as to the physics of the environment and the structure of the brain that could potentially help us in deriving such design principles.

The complexity of recognition is $O(V^N)$ for an N-pixel image where each pixel can range through V values. This means that to handle the general recognition problem, including the worst case, a network needs at least V^N nodes, each linked (either directly or via intermediate nodes in layers or some other structure) to all nodes in the input retina. This of course is combinatorially explosive, and our real problem is the expected case, that is, recognition of real-world images. The human brain and its visual system is clearly capable of perception of real-world objects in real-time, yet it does rather poorly at the worst case, e.g., discriminating images that differ by a few randomly placed pixels. For the expected case E, the number of nodes needed is clearly within feasible bounds; otherwise nature could not have evolved brains capable of successful recognition.

If the structure of the human brain and the visual system is any indication, the necessary number of nodes, N_E, is still almost certainly extremely large, and the necessary topology G_E, of the network is far from random. A great deal is known about the human brain (Peters, 1986; Uhr, 1986; Crick, 1986; Zeki, 1988; DeYoe, 1988; Livingstone, 1988). Neurons predominantly interact with near-neighbors and are organized into fairly ordered structures (columns, hypercolumns, areas); Yet a great deal is unknown about how the neurons get allocated for computing specific functions, and how the detailed topology of the network of neurons emerges as a result of learning through constant exposure to the environment.

If the desired perceptual recognition abilities are to be attained by a connectionist network through re-weighting of its links alone, it must be initialized to contain *a sufficient number of appropriately linked nodes.* The only way to guarantee that this kind of net-

work has enough nodes, each with the necessary links, is either to build them in, using a priori knowledge, or to make some guess as to N_E - and use a substantially larger number of nodes and links than that to be on the safe side. To handle the full vision problem the only completely safe thing to do would appear to be to use V^N nodes, each with N links - but this is impossibly large to actually implement.

Generation involving the addition of nodes and links enables a network to modify its topology, and appears to offer a way out of this dilemma. Given mechanisms to generate new nodes as needed, the network can gradually grow, until the number of nodes approaches N_E and the network topology approaches G_E - whatever N_E and G_E may be. Thus there is no need to estimate E; this is done constructively by the network itself.

Recent neuroanatomical and neurophysiological evidence indicates that the arrangement of synaptic connections in the mature nervous system can undergo striking changes during normal functioning - structural changes that alter the number and/or pattern of synaptic connections (Greenough, 1987; Greenough, 1988).

Rather than hope that some particular random or pre-programmed connectivity will work, or pay the excessive costs of complete connectivity, a system that generates can, under the implicit guidance of the environment's inputs and feedback, move toward sufficient connectivity. Generation works best hand-in-hand with, supplementing, the fine-tuning of functions provided by re-weighting of links. In addition, mechanisms that break links, effectively eliminating generations that turn out to be poor, are desirable to keep nets within reasonable size bounds. Some of these issues, as well as a specific learning scheme combining generation and re-weighting, have been examined in (Honavar, 1988).

Connectionist Network Structures Compared Experimentally

Several alternative multi-layered connectionist network structures and learning mechanisms are possible. Our goal is to compare some of these alternatives in terms of the sizes of networks required as well as the number of training presentations needed for achieving correct recognition of simple patterns. The multi-layered converging network structures studied include those that learn by re-weighting of their links (with no generation), using several types of connectivity - random, as well as restricted to near-neighbors, and those that learn by a combination of generation and re-weighting (where generation takes place within the constraints of near-neighbor connectivity). A summary of these network structures is given in figure 1. The objective of these comparison experiments is to examine and draw some insights into the usefulness of both topological constraints (e.g., local receptive fields) and also generation for connectionist networks that learn to recognize patterns.

Network structure	Receptive field	Generation	Built-in edges
[CP.R--]	Random	No	-
[CP.L--]	Local	No	No
[CP.L-E]	Local	No	Yes
[CP.LG-]	Local	Yes	No
[CP.LGE]	Local	Yes	Yes

Figure 1: *Summary of multi-layered, feed-forward, converging network structures; CP stands for the connectionist pyramids; R for random and L for local receptive fields; G for generation; E for built-in edge-detectors; a - in a given position indicates the absence of the corresponding network property; all use reweighting of links as a learning mechanism; only the last two use generation in addition to reweighting.*

Connectionist Networks That Learn by Re-Weighting

Several multi-layer, converging connectionist networks (using the same number of nodes and links in all the cases) were built, with the following structure:

Layer L contains 1/4th the number of node-clusters found in the adjacent layer $L-1$. Each node at layer L contains 4 times as many nodes per cluster as in the layer $L-1$. Each node in a node cluster at layer L receives input from 2-tuples of nodes drawn from 4 node clusters in layer $L-1$. In the current implementation, layer 2 is an exception in that each node in layer 2 receives input from 9 nodes in the input layer. This forms an overall Pyramid-like converging-diverging structure (see figure 2). In all the simulations described in this paper, the input layer (the retina) is a 32x32 array of pixels.

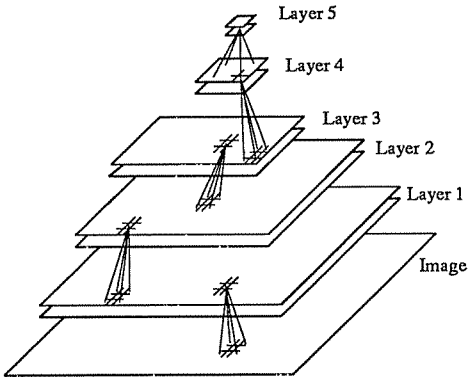

Figure 2: *A Converging pyramid-like structure: Each point in a layer has a cluster of nodes; Each node in a cluster computes a simple function over the outputs of nodes in the node-clusters in a small neighborhood in the layer below.*

Three variants of the basic multi-layered, converging network described above were implemented:

[CP.L-E]
: With local receptive fields preserving topographic mapping between layers: each node in layer L is linked to nodes in the 4 node clusters spatially located directly below it in layer $L-1$; layer 1 contains 8 pre-wired edge detectors (these are extremely simplified versions of the local spot and edge detectors found in the retina and primary visual area (V1) of living primate brains),

[CP.L--]
: Same as [CP.L-E] above, but without the built-in edge detectors in layer 1, and

[CP.R--]
: With random receptive fields: Each node in layer L is linked to nodes in 4 randomly chosen node clusters in layer $L-1$.

In all the simulations, 8 detectors (either pre-wired or learnable) were provided at layer 1. All the weights other than those corresponding to the built-in edge detectors were assigned randomly.

Clearly, there are other variants of the same basic structure that could be studied; but we chose the ones described here because they are useful in evaluating the usefulness of locality of receptive fields in perceptual learning.

In all cases, learning involved re-weighting links as a function of the back-propagated error signal. Suppose a pattern class C_W is implied by the network with a weight W_W, and the pattern class indicated by the feedback, C_R is implied with a weight W_R, the amount of reweighting at the output layer is given by $(K\times(W_W-W_R))$ where K is a parameter related to the rate of learning. Our current implementation has K set equal to 0.25. This weight change is distributed equally among all the links firing into the node implying C_W. At internal nodes, the weight changes are computed in a similar fashion.

Connectionist Networks That Learn By Generation and Discovery As Well As Reweighting

Connectionist network structures that learn by generation and re-weighting of links and recruiting of new nodes from a pool of unused nodes were studied. The topological constraints on the network structure are the same as those present in [CP.L--] described earlier. However, the networks that learn by generation as well as reweighting differ in that they start with a pool of *uncommitted* nodes and no pre-wired links. Generation grows new links and adds new nodes to the network from the pool of nodes as the network learns aided by feedback. The weights associated with the links are changed using the same reweighting mechanism as the one used in [CP.L--]. A particular implementation of generation and reweighting of this sort is described in more detail in (Honavar, 1988).

Because generation takes place within the topological constraints of the layered, logarithmically converging organization as well as the local receptive fields, the networks that are discovered through generation and reweighting (e.g., [CP.LGE] and [CP.LG-]) are topologically similar to [CP.L--] and [CP.L-E]. But unlike [CP.L--], the number of nodes per node-cluster at a given layer in [CP.LG-] and [CP.LGE], or the connectivity between node clusters in adjacent layers, is not preprogrammed; it is determined dynamically through learning.

Runs were made with pre-wired edge-detectors in the first layer - [CP.LGE], and without any pre-wired nodes (i.e., having all the nodes added to the network as part of the learning process) - [CP.LG-]. In both these cases, the reweighting of nodes as a function of feedback proceeds according to the same reweighting rule as the one used in [CP.L--] and [CP.L-E]. In addition, the network occasionally generates a new node, when it determines this to be appropriate - on the basis of information provided by sub-networks that monitor the network's performance on each pattern class on which it is being trained. The design of these sub-structures is motivated by the need to discover the *simplest* networks capable of the desired accuracy of recognition. A particular implementation of such structures is explained in detail elsewhere (Honavar, 1988).

The rationale behind the design is as follows: Continue to reweight existing links so long as the network's performance is improving. When it is observed that the network's performance has leveled off (before reaching the desired accuracy of recognition), generate a new node. This is accomplished easily by a simple network of neuron-like units, using local computations that are performed incrementally following each training presentation.

Generation proceeds as follows: In the 1st layer, a 3-by-3 sub-array is extracted from the raw input image (this is done only when feedback indicates an error was made, and the history of the recent past indicates that performance is levelling off rather than improving. The 9 links from this sub-array fire into a new node placed directly *above* it in the next layer.

The extraction is got from a *busy* part of the input image, one where the network judges there may be useful information. The present simple system insists that a gradient be present; but potentially more powerful mechanisms for evaluating the candidate extraction are being investigated.

In layers other than the 1st, extraction randomly links into a new node from 2 nodes that actively responded to the present (incorrectly identified) input image in the 2-by-2 of node-clusters directly below it in the previous layer.

Whenever a node is generated it is put into a node-cluster at that location, and also at every other location in that layer of the network - as though it is immediately broadcast (either laterally through that layer or up to the

apex of the pyramid and then showered back down). All the links added to the network through generation will now get tuned through reweighting as a function of feedback.

Experimental Results

Several test runs were made to compare multi-layered connectionist network structures ([CP.R--], [CP.L--], [CP.L-E], [CP.LG-] and [CP.LGE]). Simple 2-dimensional patterns such as letters of the alphabet (T, D, E) and simple objects (apple, cup, banana) were used for training the networks. The training and test sets were obtained by randomly dividing the set of drawings of each pattern (provided by 3 different volunteers) into two subsets. The drawings were made using the *Xgremlin* graphics utility on a Digital VAXstation-3200, in a 24x24 subarray of a 32x32 grid. A sample set of patterns is shown in figure 3, and an indication of the variation among instances of a pattern class is given by the instances shown in figure 4. Figure 5 gives a summary of the pattern classes used in the runs, i.e., (T, D, E), (apple, banana, cup) and the combined set (T, D, E, apple, banana, cup).

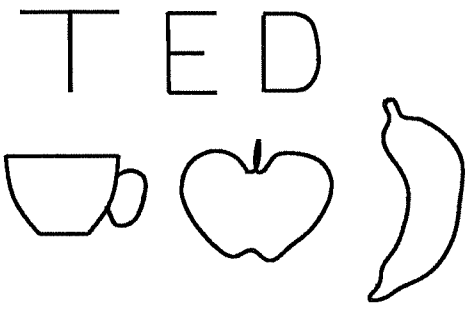

Figure 3: *Sample images used in the simulation of learning*

A *run* consists of several *epochs* of training interspersed with epochs of testing, repeated until the desired acuuracy of recognition (currently set to 100 percent) is attained or the performance clearly levels off, as indicated

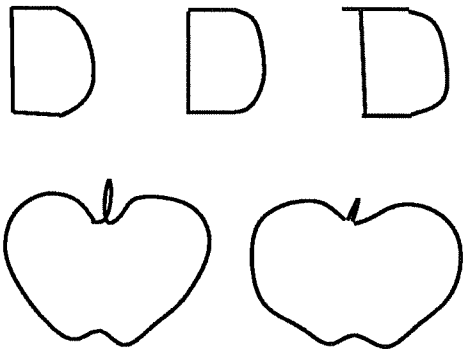

Figure 4: *Sample instances of two of the object classes used in the training and test sets.*

Pattern Set	# of classes	# of train-ing instances per classs	# of test instances per class
T, D, E	3	4	3
apple, cup, banana	3	4	4
T, D, E, apple, cup, banana	6	4	3

Figure 5: *Summary of pattern sets used in the experiments: the pattern sets were obtained from instances provided by 3 volunteers. Training and test instances for each class were obtained by randomly partitioning the set of instances for a given class into two subsets - one for training and the other for testing.*

by the learning curve. An *epoch* of training (or testing) involves cycling through the entire training set (or test set) once, in some arbitrary order. The runs for the structures [CP.R--], [CP.L-E] and [CP.L--] were made with several

different percentages of possible connections, having fixed the number of nodes at each location in the first layer to 8, each with 9 connections (these percentages are indicated next to the corresponding learning curves in the figures 6A through 6C).

In all cases, [CP.LGE] (pyramid convergence, locality, generation, built-in edge detectors) gave the best results, followed by [CP.LG-] (pyramid convergence, locality, no built-in edge detectors). These were both substantially better than the networks [CP.L-E] (pyramid convergence, locality, and built-in edge detectors), which in turn were substantially better than [CP.L--] (pyramid convergence, locality, no built-in edge detectors).

The figures 6A through 6E show the results of these runs on the pattern set (T, D, E). The results with pattern sets (apple, cup, banana) were qualitatively similar in all the cases (the runs were slightly longer (took about 10% more epochs); about 10% more links were generated in [CP.LGE] and [CP.LG-]).

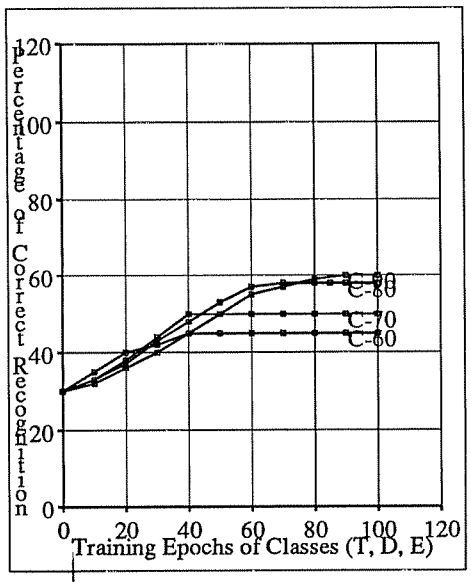

Figure 6A: [CP.R--]: *Pyramid convergence, no local receptive fields*

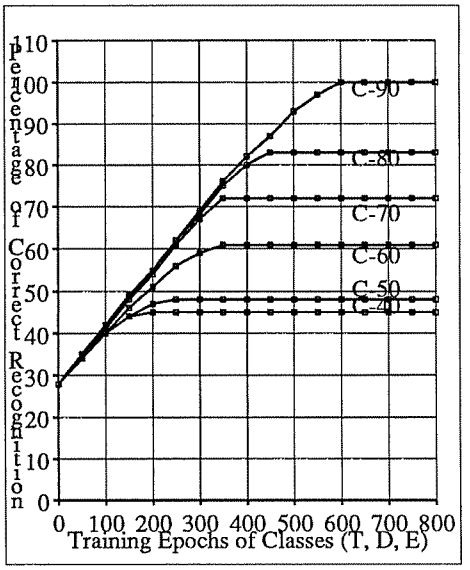

Figure 6B: [CP.L--]: *Pyramid convergence, local receptive fields, no built-in edges*

The networks [CP.R--] (random connectivity between layers, logarithmic convergence) failed to improve beyond 60% correct recognition (see figure 6A) given the same maximum number of connections that were used in [CP.L--] structures The networks [CP.L--] attained 100% accuracy of recognition with approximately 16k links (see figure 6B), which were distributed equally between layers (1,2), (2,3), (3,4) and (4,5) in about 600 epochs of training, whereas the networks [CP.L-E] attained the same performance with the same network size, in about 90 epochs of training (see figure 6C).

The network [CP.LG-] attained 100% accuracy of recognition in 26 epochs of training and at about 8000 links (14 new transforms were generated and they were replicated at each location in the corresponding layers). The network [CP.LGE] reached 100% correct recognition in about 8 epochs of training (see figure 6E) and at about 6000 links (6 new transforms were generated and they were replicated at each location in the

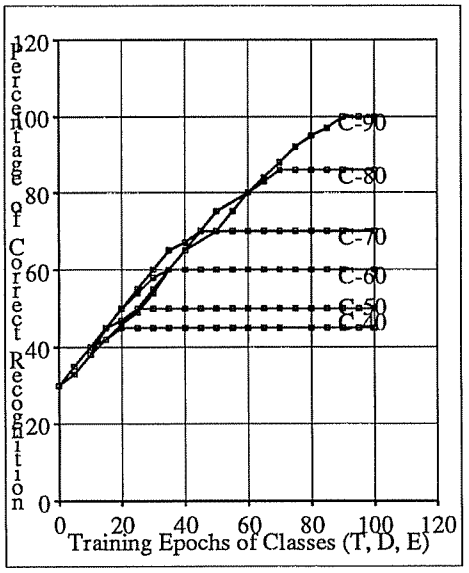

Figure 6C: [CP.L-E]: *Pyramid convergence, local receptive fields, built-in edges*

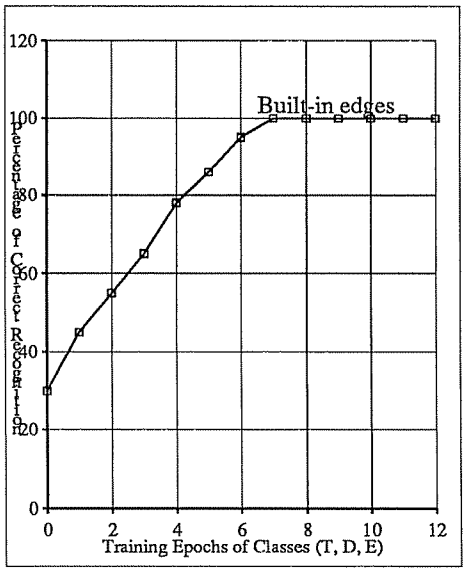

Figure 6E: [CP.LGE]: *Pyramid convergence, generation, local receptive fields, with built-in edges*

corresponding layers).

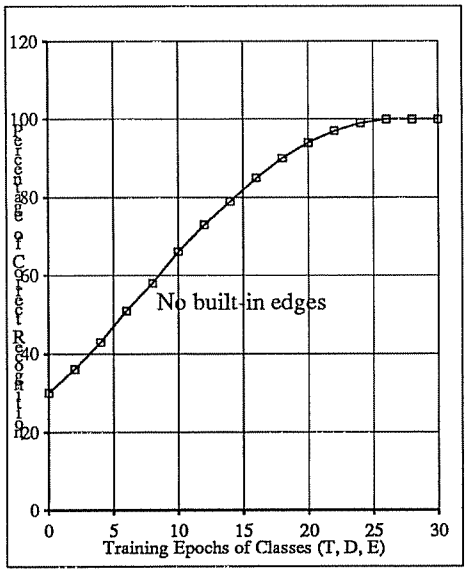

Figure 6D: [CP.LG-]: *Pyramid convergence, generation, local receptive fields, without built-in edges*

The runs were repeated for [CP.LG-] and [CP.LGE] with 6 pattern classes (T, D, E, apple, banana, cup) and the results were qualitatively similar, but there were more generations (about twice as many) at the higher layers resulting in approximately 10k and 8k links respectively, and about twice as many epochs of training were needed to attain 100% accuracy of recognition. The exact numbers reported here should not be given too much importance; however the results do suggest that other factors being constant, generation and local structure significantly improve learning, both in terms of the number of training epochs needed as well as the size of the networks necessary to attain the desired accuracy of recognition.

Discussion and Summary

The results presented in this paper suggest that the incorporation of brain-like constraints on network structure can substantially reduce the complexity of connectionist networks for perceptual learning and improve the speed of learning. Retinotopic mapping and near-neighbor connectivity exploit spatio-temporal contiguity in the environment. The choice of constraints on network connectivity

is important: random connectivity is unlikely to work in most practical problems. Similar conclusions were reached in another study that used recurrent back-propagation (Pineda, 1987) to train a connectionist network to solve random-dot stereograms (Quian, 1988). Constraints on the network topology determine the space of the transforms that may be learned, and bias the network so that the learning of certain relations is favored. Topographic mapping and local receptive fields favor the discovery of relations between the subpatterns that are imaged onto the neighboring regions of the retina. Multi-layered pyramid-like hierarchical architectures embodying such topological constraints have been studied rather extensively for image processing and computer vision (Uhr, 1972; Uhr, 1987; Burt, 1984; Rosenfeld, 1983; Dyer, 1987; Li, 1987). For an examination of such architectures as connectionist networks, see (Honavar, 1987). The results presented here indicate that such structures are also useful in networks that learn (as opposed to being carefully preprogrammed) to recognize objects.

Our results suggest that the addition of mechanisms such as generation, that enable the network to grow new links as needed, under guidance from feedback, aided by network structures that enable it to monitor its own performance over time, yield further improvements in learning. Although generation has been used in the experiments described in this paper in conjunction with a form of error back-propagation for reweighting of links, local receptive fields and global convergence, it could potentially be used with other reweighting rules and other network structures.

Generation ensures that successively more complex non-linear relations between features in the input patterns are discovered at higher layers, to be assessed by the new transforms that are added. Thus, generation in a layered, converging network structure biases the system so that: learning of simpler features preceeds the learning of more complex relations; and successively more global relations are learned at successively higher layers. An examination of the transforms generated in the network simulations supports this intuition.

The extraction-generation programs described here do not discard bad nodes or place any limit on the number of nodes generated. Neither capability was needed for the test runs reported here, since these programs learned to recognize the pattern-sets they were tested on in relatively small numbers of training epochs. But to handle larger sets of more complex patterns, the ability to discard is almost certainly necessary; otherwise the network will get bogged down with many poor or worthless links. Thus, additional subnetworks that constantly evaluate and maintain records of the usefulness of functions computed by individual nodes or small groups of nodes would be useful in determining which nodes (and hence, transforms) to discard, so that they can be reused to compute new, and potentially better functions.

There are a number of promising improvements to be made, including the addition of networks that make better assessments of potential generations, that learn to improve upon these assessments, that evaluate the generations for their usefulness for recognition, that discard poor generations to make room for new ones, that narrow and broaden the tolerance-threshold for matching, and that generate sets of alternate possible transforms that are placed in competition with one another. There are a number of other issues to be investigated, including the development of good sub-networks that realize functions for deciding whether to further re-weight or to generate, the optimal number of nodes in a node-cluster, and the usefulness of putting nodes within a cluster into direct competition.

There are several aspects of networks that learn by generation and reweighting that may be worth examining in detail, individually

as well as collectively. In what follows, we outline some of these issues briefly.

The extent of generalization, i.e., building of meaningful internal representations by discarding uninteresting details, is an important property of connectionist systems that learn. More compact representations result from better generalization. There is reason to believe that the extent of generalization in connectionist networks is sensitive to the number of hidden units as well as the connectivity (Hinton, 1987a). If the hidden units (or connections) are too many, the network may generalize rather poorly; if they are too few, the network may never learn. Thus, finding the optimal number of hidden units and/or weights is of interest. One approach involves the introduction of local and distributed bottlenecks on the number of units and/or links that tend to drive certain weights to zero (Kruschke, 1988). Generation and deletion of links can be seen in this context as providing mechanisms that dynamically determine the optimal numbers of hidden units and connections needed in the network. Thus, such networks may exhibit good generalization properties as well. Generation makes possible the linking up of an adequate number of units to solve a given problem; minimal generation favors the discovery of the smallest necessary number, hence better generalization. We intend to examine this conjecture experimentally.

Sub-structures that, as appropriate, maintain, update, and transmit information, recording the network's performance over time (e.g., using portions of the learning curve, to trigger generation) offer several promising mechanisms that may be worth examining. Such structures can be used to alter learning strategies, rates of learning, and thresholds of firing, each of which has an impact on the *plasticity* of the network. Future work will address some of these issues.

Networks that generate, reweight, and discard transforms need to have mechanisms that keep a reasonable balance between them. Reweighting of links by small amounts changes the pattern descrimination properties of the network gradually. If the network topology is held constant, reweighting tends to minimise the error between the actual and the desired outputs of the network for the various pattern classes. However, there is a risk of getting caught in the local minima (Rumelhart, 1986) of the error function. Generation and discarding of transforms can be thought of as providing the network some means of getting out of such local minima. When the accuracy of recognition levels off and fails to improve with reweighting alone, it suggests that the network may have hit a local minimum of the error function. Future work will examine several mechanisms that might be useful in striking just the right balance between reweighting on one hand, and generation and discarding of transforms on the other.

Intuition suggests that good system performance requires a proper match between the *entropy* of the source of external stimuli and the connectivity, both between the source and the system (Abu-Mostafa, 1988) as well as within the system itself. Since generation relies on the environmental stimuli to develop the connectivity of the system, the resulting network is likely to have a better match with the entropy of the environment, than a network that starts out with a random subset of the possible connections and is constrained not to change the initial topology, and whose learning is restricted to reweighting of links alone.

The network structures that we have studied so far have been restricted to be of the feed-forward type. Future work may explore the usefulness of relaxing this constraint and using a form of recurrent back-propagation (Pineda, 1987) for reweighting of links.

Most of the research on learning in connectionist networks has concentrated on schemes that modify weights in a static topology. Recent anatomical and physiological stu-

dies suggest that learning may involve alteration of the number as well as the pattern of synaptic interconnections in the brain, in addition to the well-accepted mechanisms for changes in synaptic strengths (Greenough, 1987; Greenough, 1988). The results presented in this paper suggest that there may be promising improvements to be realized using additional learning mechanisms that dynamically alter the network topology (e.g., generation), suitable constraints on the network structure for particular domains (such as local receptive fields and global convergence for vision) and regulatory mechanisms that alter the plasticity of the network, choose between different learning strategies, and so on. Extensive and systematic evaluations of networks incorporating one of more of these features for perceptual learning of pattern sets of varying complexity are needed in order determine how they perform individually as well as collectively. The experiments and results discussed in this paper constitute at best, a preliminary exploration of only a few aspects of the problem. Future work will examine these issues in greater detail.

References

Abu-Mostafa, Y., "Connectivity versus entropy," in *Neural Information Processing systems: Natural and Synthetic*, ed. D. Anderson, American Institute of Physics, New York, 1988.

Barto, A. G. and Anandan, P., "Pattern recognizing stochastic learning automata," *IEEE Trans. Systems, Man and Cybernetics*, vol. 15, pp. 360-375, 1985.

Burt, P. J., "The pyramid as a structure for efficient computation," in *Multiresolution Image Processing and Analysis*, ed. A. Rosenfeld, Springer-Verlag, Berlin, 1984.

Crick, F. H. C. and Asanuma, C., "Certain aspects of the anatomy and physiology of the cerebral cortex," in *Parallel Distributed Processing, vol. 2: Psychological and Biological Models* , The MIT Press, Cambridge, Massachusetts, 1986.

DeYoe, E. A. and Van Essen, D. C., "Concurrent processing streams in monkey visual cortex," *Trends in Neuroscience*, vol. 11-5, pp. 219-226, May 1988.

Dyer, C. R., "Multiscale image understanding," in *Parallel Computer Vision*, ed. L. Uhr, Academic Press, New York, 1987.

Greenough, W. T., Black, J. E., and Wallace, C. S., "Experience and brain development," *Child Development*, vol. 58, pp. 539-559, 1987.

Greenough, W. T. and Bailey, C. H., "The anatomy of a memory: convergence of results across a diversity of tests," *Trends in Neuroscience*, vol. 11, pp. 142-147, April 1988.

Hebb, D. O., *The Organization of Behavior,* Wiley, New York, 1949.

Hinton, G. E., "Connectionist learning procedures," *Technical Report CMU-CS-87-115*, Computer Science Department, Carnegie Mellon University, Pittsburgh, Pennsylvania, 1987.

Hinton, G. E., "Learning translation invariant recognition in a massively parallel network," in *PARLE: Parallel Architectures and Languages, Europe. Lecture Notes in Computer Science*, ed. G. Goos, J. Hartmanis, Springer-Verlag, Berlin, 1987.

Honavar, V. and Uhr, L., "Recognition Cones: A neuronal architecture for perception and learning," *Computer Sciences Technical Report #717*, Computer Sciences Department, University of Wisconsin-Madison, Madison, Wisconsin, September 1987.

Honavar, V. and Uhr, L., "A network of neuron-like units that learns to perceive by generation as well as reweighting of its links," in *Proceedings of the 1988 Connectionist Models Summer School*, ed. G. E. Hinton, T. J. Sejnowski and D. S. Touretzky, Morgan Kaufmann, San Mateo, CA, 1988.

Kruschke, J. K., "Creating local and distributed bottlenecks in hidden layers of back-propagation networks," in *Proceedings of the 1988 Connectionist Models Summer School*, ed. G. E. Hinton, T. J. Sejnowski and D. S. Touretzky, Morgan Kaufmann, San Mateo, CA.

Li, Z. N. and Uhr, L., "Pyramid vision using key features to integrate image-driven bottom-up and model-driven top-down processes," *Systems, Man and Cybernetics*, vol. 17, pp. 250-263 , March 1987.

Livingstone, M. and Hubel, D., "Segregation of form, color, movement, and depth: anatomy, physiology, and perception," *Science*, vol. 240, pp. 740-749, May 1988.

McCulloch, W. S. and Pitts, W. H., "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, vol. 5, pp. 115-133, University of Chicago Press, Chicago, 1943.

Peters, A. and Jones, E. G. (eds.), *Cerebral Cortex: Vol. 3. Visual Cortex,* Plenum, New York, 1986.

Pineda, F. J., "Generalization of back-propagation to recurrent neural networks," *Physical Review Letters*, vol. 59, pp. 2229-2232, 1987.

Qian, N. and Sejnowski, T. J., "Learning to solve random-dot stereograms of dense and transparent surfaces with recurrent backpropagation," in *Proceedings of the 1988 Connectionist Models Summer School*, ed. G. E. Hinton, T. J. Sejnowski and D. S. Touretzky, Morgan Kaufmann, San Mateo, CA, 1988.

Rosenfeld, A., "Pyramids: multiresolution image analysis," *Proceedings of the Third Scandinavian Conference on Image Analysis*, pp. 23-28, July 1983.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J., "Learning internal representations by error propagation," in *Parallel Distributed Processing vol. 1: Foundations*, The MIT Press, Cambridge, Massachusetts, 1986.

Uhr, L., "Layered recognition cone networks that preprocess, classify, and describe," *IEEE Transactions on Computers*, vol. 21, pp. 758-768, 1972.

Uhr, L., "Toward a computational information processing model of object perception," *Computer Sciences Technical Report #651*, Computer Sciences Department, University of Wisconsin-Madison, Madison, Wisconsin, July 1986.

Uhr, L., "Highly parallel, hierarchical, recognition cone perceptual structures," in *Parallel Computer Vision*, ed. L. Uhr, Academic Press, New York, 1987.

Zeki, S. and Shipp, S., "The functional logic of cortical connections," *Nature*, vol. 335, pp. 311-317, September 1988.