ESD-TR-88-322

Project Report
NN-2

Prospects for Classifying Complex Imagery Using a Self-Organizing Neural Network

M.M. Menon
K.G. Heinemann

11 January 1989

Lincoln Laboratory

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LEXINGTON, MASSACHUSETTS

Prepared for the Tactical Technology Office
of the U.S. Defense Advanced Research Projects Agency (DARPA/TTO)
with the support of the Balanced Technology Initiative Program
under Electronic Systems Division Contract F19628-85-C-0002.

The work reported in this document was performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology. This work was sponsored by the Tactical Technology Office of the U.S. Defense Advanced Research Projects Agency (DARPA/TTO) with the support of the Balanced Technology Initiative Program under Air Force Contract F19628-85-C-0002.

The views and conclusions contained in this document are those of the contractor and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the United States Government.

The ESD Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Hugh L. Southall

Hugh L. Southall, Lt. Col., USAF
Chief, ESD Lincoln Laboratory Project Office

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

PROSPECTS FOR CLASSIFYING COMPLEX IMAGERY USING A SELF-ORGANIZING NEURAL NETWORK

M.M. MENON
K.G. HEINEMANN
Group 53

PROJECT REPORT NN-2

11 JANUARY 1989

LEXINGTON MASSACHUSETTS

ABSTRACT

The objective of this study is to evaluate the performance of Fukushima's Neocognitron model [2] when it is applied to complex imagery. In his original report, Fukushima demonstrated that this system could discriminate between simple alphabetical characters represented in fields of 16 by 16 pixels, and that shift invariance can be achieved through a proper choice of design parameters. The present work describes results for expanded Neocognitron architectures operating on complex images of 128 by 128 pixels. These neural network systems were simulated on a VAX-8600 minicomputer. Wire frame models of three different vehicles were used to test the properties which Fukushima had demonstrated. The expanded Neocognitron systems were able to classify these objects and to identify their critical features. After training, each object was placed at different positions in the plane, and the Neocognitron's shift invariance property was tested. With complex (128x128) imagery, it was difficult to achieve proper classification and maintain shift invariance using only a few levels. In another experiment, the Neocognitron trained on polar transforms of objects in the training set. Objects in the training set were rotated, and polar transforms of the rotated images were submitted as input. In this manner, the Neocognitron's shift invariance was exploited to recognize rotated imagery. These investigations gave insight into the role of various model parameters and their proper values, as well as demonstrating the model's applicability to complex images.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

1. INTRODUCTION

The Neocognitron of Fukushima [2] is a massively parallel multilevel neural network system which performs visual pattern recognition. Its architecture models the anatomy of the human retina in a qualitative way. This system also resembles the Adaptive Resonance model of Carpenter and Grossberg [1] in that it is self organizing and operates without a "teacher". The Neocognitron has a demonstrated capability to discriminate alphabetical characters stored in a matrix of 16x16 pixels. Performance on hand written characters in a 19x19 matrix was demonstrated by Fukushima [3]. A more recent study by Stoner and Schilke [4] has confirmed the model's ability to classify dot-matrix characters. While many accurate character recognition algorithms already exist, the Neocognitron is noteworthy because it handles positional shifts and moderate deformations in the shapes of input characters. These properties suggest that Fukushima's model might be very useful in solving more demanding machine vision problems. Work at the Massachusetts Institute of Technology Lincoln Laboratory has produced a simulation of the Neocognitron on a serial machine. This program has operated successfully on wire frame images embedded in a matrix of 128x128 pixels. The model was able to classify images by extracting features from the input images and retaining only those whose response was above the average. Results from different Neocognitron systems showed that its shift tolerance depends on the number of levels used. A four level system was unable to classify patterns uniquely and tolerate shifts with an input plane of 128x128 pixels. However, a single level version was found that did classify properly and provide shift invariance at the same time. The shift tolerance property can be exploited to cope with other kinds of variation by submitting appropriate transforms of the imagery as input. This approach has been investigated by applying a polar transform to achieve automatic recognition of rotated images.

2. MODEL DESCRIPTION

The Neocognitron is a structured network of analog processing units which receive and transmit zero or positive valued analog signals. This network contains four distinct kinds of computational elements called S-cells, C-cells, V_s-cells, and V_C-cells. Each class of processor is defined by the types of cells which provide its input and a specific mathematical operation which determines the strength of its output.

The output from an individual processor generates input signals for certain other nodes after passing through a set of weighted connections. Each of these communication channels multiplies the transmitting unit's output by a specific connection strength (weight) and presents that product as an input for the receiving unit. The weight for a given connection can take on any positive value, so the effect of a specific unit's output may vary considerably from one node to the next.

Cells in the Neocognitron generally receive inputs from a number of different nodes and respond to the total received signal, but signals from the different types of processors are summed separately, because they affect the response in different ways. For a given unit, different patterns of output at the source nodes will produce varying levels of total input. This behavior arises, because the specific pattern of connection weights will amplify some of the individual source signals more than others. The total input will be particularly high when the source nodes send strong signals along paths with large weights, and it will decrease as strong signals are shifted to paths with smaller weights or the paths with large weights carry smaller signals. Thus, communication through the weighted connections enables the processors to detect differences in the pattern of transmitted signals. An analog transfer function then produces corresponding variations in the response level.

The Neocognitron's processing elements are organized into a hierarchical series of levels, where units of each type appear at every level. All these levels share a common structure wherein the different types of cells are segregated into distinct layers, and signals traverse these layers in the same order. A schematic representation of this architecture is shown in Figure 2-1, where an image comes in at the left and data flows to the right. A layer of V_s-cells and a layer of V_C-cells also exists at each level, but these have been omitted in order to simplify the diagram. Output from any given level serves as input for the next one, until a layer representing the final classification categories is reached.

The system is strictly a feed-forward network where signals originate at an initial input layer and propagate towards the final output layer. A hierarchical structure is produced by connecting the cells in a "fan out" pattern, so that the number of units gradually decreases as signals propagate into deeper levels of the system. Under this connection scheme, each unit receives input signals from specific small regions on the layers which immediately precede it. However, the number of indirect connections between a processor and more distant predecessors grows significantly as the number of intervening layers increases. For any particular cell, the complete set of input sources on an earlier layer will be referred to as the cell's "receptive field" on that layer. Since processors at deeper levels gain access to progressively larger portions of the input patterns, they can respond to progressively more complicated features, and simpler features will be detected over a progressively larger receptive field. The final output layer consists of cells whose receptive field covers the entire input layer. This hierarchical structure contributes to the Neocognitron's capacity for shift invariant pattern recognition.

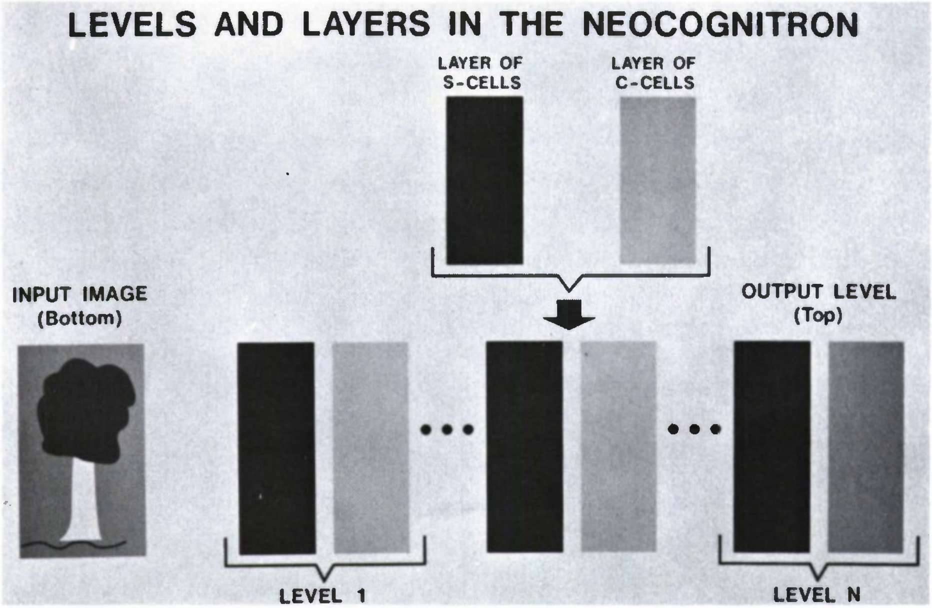


Figure 2-1. Multilevel feed-forward architecture of the Neocognitron.

In order to completely explain the property of shift invariance, one must consider the structure of an individual level. The S-cells and the C-cells on any given level are organized into a number of subgroups which will be called "S-planes" or "C-planes" according to the type of processor which is involved. V_s-cells and V_C-cells also are grouped into planes, but there is only one V_s-plane and one V_C-plane on any level. These cell planes are treated as two dimensional matrices where the location of an individual element is specified by a pair of column and row coordinates.

The relationships between planes, layers, and individual cells is illustrated in Figure 2-2. All the elements in a given plane share a single pattern of connection weights on their input channels. Consequently, a specific pattern of transmitted signals will elicit the same response from any element which observes that pattern exactly. While the input field of an individual processor covers only a small portion of the source layer, the fields of adjacent cells can be positioned in a way which insures that the entire source layer is covered. If the number of cells in the plane, the size of their input fields, and the offset between these fields are correctly matched, one can guarantee that some cell will show the optimum response when a specific pattern appears anywhere in the source layer. Hence, the behavior of these cell planes provides a massively parallel technique for shift invariant feature detection. This architectural feature is the fundamental mechanism responsible for the Neocognitron's tolerance of positional shifts.

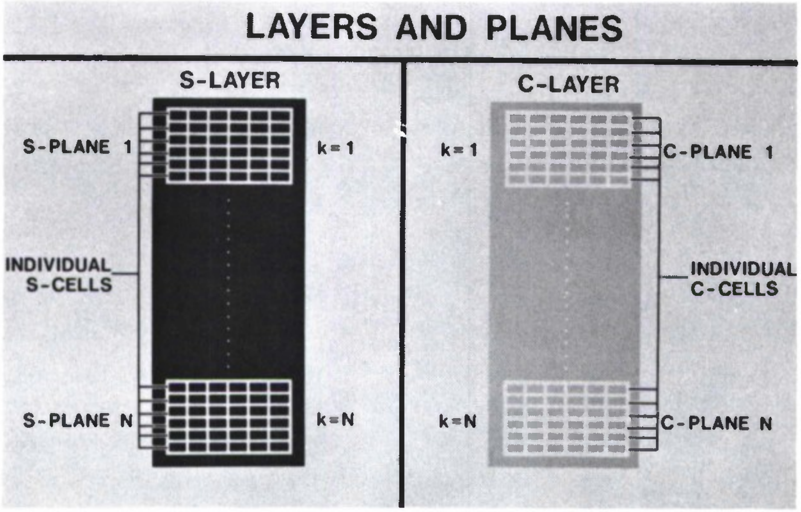


Figure 2-2. Detailed organization of the Neocognitron.

The Neocognitron acquires its ability to classify patterns because each level contains a number of separate S-planes and C-planes. These two structures must always be paired with one another, so a given level has the same number of each type. However, the number of paired cell planes can vary from one level to the next. Each of the S-planes has a distinct pattern of input connection weights, but the C-planes on any particular level share one pattern in common.

The weights which feed into the S-planes have a special role, because they change as the system learns. All the other connection weights are built into the design of a specific Neocognitron architecture, and they cannot be modified. As the system learns to discriminate between diverse input images, the S-planes become sensitive to different spatial arrangements of the source signals. However, units in any given plane will receive small input signals from almost any pattern that happens to occur. Difficulties could arise if all these signals were allowed to propagate deeper into the system. Some of the very weak signals could be greatly amplified when they pass through connections with large weights, and the results might convey some very misleading information.

In order to avoid this problem, the Neocognitron incorporates mechanisms which suppress the transmission of insignificant input signals. Interactions between the different types of processors work in concert with their particular response functions to provide a form of adaptive filtering. This design prevents the S-cells and C-cells from responding unless the pattern dependent input signal exceeds an independent estimate of the "typical" incoming signal strength.

A brief discussion of the different processors' actual operating characteristics and their interconnections will help to illustrate and clarify these general principles.

2.1 CELLS IN THE S-LAYER AND THE V$_c$-LAYER

The S-cells in a given level obtain information about the previous one through two separate input mechanisms. Units in the first S-layer respond to the initial input signals, while those on subsequent levels receive input from C-planes on the preceding level. Direct connections from C-cells to S-cells carry excitatory signals which act to increase the S-cell's output. The S-cells also receive an inhibitory input which reduces the output signal through a shunting effect. This inhibitory signal ultimately comes from the same C-cells which produce the excitatory ones, but a layer of V$_c$-cells intervenes to perform some additional processing.

Any given level contains a number of S-planes and a single V$_c$-plane which all share the same geometric structure. The units at a given position in any of these planes share the same input fields, which extend over a specific set of adjacent coordinates in the preceding C-planes. Figure 2-3 illustrates the configuration of direct connections going from a C-layer to a particular S-plane. As a result of this connection scheme, S-cells and V$_c$-cells receive input from small regions on all of the C-planes. This arrangement enables the S-cells to recognize groupings of features that might have been detected in earlier stages of processing. If k refers to the k-th S-plane in level l and n refers to a specific position in that S-plane, the response of the corresponding S-cell is given by:

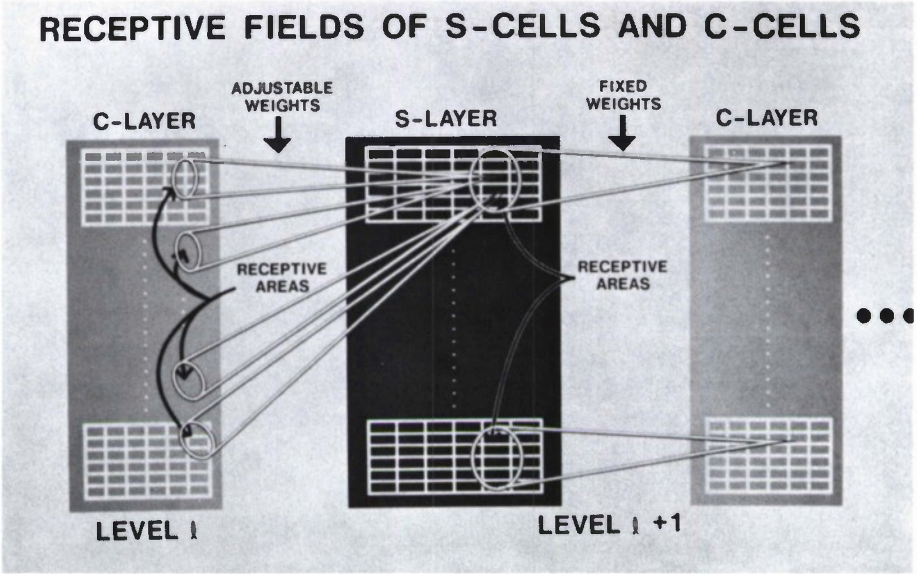
$$Us_l(k_l, n) = r_l \times f(A) \tag{2.1}$$

where

$$A = \left[\frac{1 + \sum_{k_{l-1}=1}^{K_{l-1}} \sum_{v \epsilon S_l} a_l(k_{l-1}, v, k_l) \times Uc_{l-1}(k_{l-1}, n+v)}{1 + \frac{r_l}{l+r_l} \times b_l(k_l) \times Vc_l(n)} - 1 \right]. \tag{2.2}$$

Expression $Uc_{l-1}(k_{l-1}, n+v)$ represents an excitatory signal coming from the unit at position $n+v$ on C-plane k_{l-1}, and $Vc_l(n)$ represents the inhibitory input. In Equation 2.2, $a_l(k_{l-1}, v, k_l)$ are the connection weights for excitatory input, and v designates a relative position inside the region of input. Positional variations of the weight values give rise to the "weight patterns" which were discussed earlier. These weight distributions differ from one source plane (k_{l-1}) to the next, so that the S-cells can recognize combinations of different source patterns. Note that $a_l(k_{l-1}, v, k_l)$ does not depend on the S-cell position, n, because all members of a given plane k_l have the same distribution of input weights. In the Equation 2.2, the inner sum computes the total excitatory input from a specific C-plane, and the outer sum adds together the contributions from different planes.

The inhibitory effect works through this expression's denominator, where the connection weight $b_l(k_l)$ multiplies the output from a single node on the V$_c$ plane. This inhibitory signal comes from the V$_c$ cell at location n , which corresponds to the S-cell's position in its own plane. The V$_c$-cell at those coordinates receives input from the same C-cells which are exciting the S-cell, and it responds by computing a weighted root-mean-square:



94743-3

Figure 2-3. Interconnection architecture of the Neocognitron.

$$Vc_l(n) = \sqrt{\sum_{k_{l-1}=1}^{K_{l-1}} \sum_{v\epsilon S_l} c_l(v) \times Uc^2_{l-1}(k_{l-1}, n+v)} \tag{2.3}$$

where $c_l(v)$ represents the input connection strength for a particular position v in that cell's input field. These weights can follow any distribution which decreases monotonically as the magnitude of v increases, and they must be normalized so that their sum is exactly equal to unity, i.e.,

$$\sum_{k_{l-1}=1}^{K_{l-1}} \sum_{v\epsilon S_l} c_l(v) = 1. \tag{2.4}$$

In the present study, the $c_l(v)$ were defined by a decaying exponential distribution:

$$c_l(v) = \frac{1}{C(l)} \alpha_l^{r'(v)} \tag{2.5}$$

where $r'(v)$ is the normalized distance between location v and the center of the input region ($0 \leq r' \leq 1$). The parameter α_l is a small constant ($\alpha_l < 1$) which determines how quickly these weights fall off as $r'(v)$ increases. Consequently, weights at the edge of $S_l(r' = 1)$ are equal to a fraction α_l of the value at the center ($r' = 0$). The expression $C(l)$ is a normalizing constant:

$$C(l) = \sum_{k_{l-1}=1}^{K_{l-1}} \sum_{v\epsilon S_l} \alpha_l^{r'(v)} \tag{2.6}$$

which insures that Equation 2.4 will be satisfied. All the V_c-cells in a given plane (and level) use the same pattern of input connections, but α_l is free to assume a different value for each level.

The weighted root-mean-square signal from a V_c-cell propagates to all S-cells at the same coordinates n. However, the weights on these connections, $b_l(k_l)$, are all independent, so the actual inhibitory effect will differ from one S-plane to the next. In addition, the denominator of the S-cell response function contains a factor $r_l/1 + r_l$, which further modulates the inhibition. This factor can provide any degree of attenuation as the parameter r_l goes from 0 to ∞, and it has great sensitivity at the low end of its dynamic range. Note that r_l also appears as a multiplicative factor in the S-cell response function. It is given this additional role to curb growth in the final output when high attenuation (low r_l) makes the inhibition ineffective. The values for r_l are set by the system designer, and the subscript indicates that these values can be different at each level. Hence, the action of these parameters enables the system designer to control the overall influence of the weighted root-mean-square input at each level of the system. In order to prevent division by zero when inhibitory input is totally absent, the attenuated signal is incremented by one. This solution conveniently neutralizes the denominator when there is no inhibitory input.

As discussed previously, an S-cell's excitatory input measures the degree of similarity between a particular arrangement of source signals and a feature represented by the distribution of input weights a_l. The quotient in Equation 2.2 compares the actual excitatory input with some fraction of the weighted root-mean-square source signal. The resulting ratio is decreased by one to determine which of the two input signals is greater. A positive difference indicates that the excitatory signal is greater, because the previous ratio exceeded unity, and a negative difference indicates that the inhibitory signal was greater. The function "f" which operates on this result is the linear threshold function:

$$f(x) = \begin{cases} x & (x \geq 0) \\ 0 & (x < 0). \end{cases} \tag{2.7}$$

Consequently, an S-cell responds only if the excitatory input exceeds the inhibitory input, and the transmitted signal is proportional to the relative difference. The double sum in the numerator is incremented by one to produce proper behavior (zero response) when excitatory input is absent.

The Neocognitron learns to discriminate between different patterns of input by updating the adjustable weights $a_l(k_{l-1}, v, k_l)$ and $b_l(k_l)$ in Equation 2.2. Weights for the excitatory connections (a_l) start off with small values that allow different S-planes to produce distinct responses to an arbitrary input pattern. The inhibitory weights (b_l) are set to zero initially. Increments for both types of weight are determined by finding those S-cells which show the greatest response with respect to a certain set of the others. These units are selected by imagining that all S-planes on a given level are stacked vertically.

Many overlapping columns are defined in this stack, where a column goes through the same set of spatial positions in each S-plane. The learning procedure examines each column and records the position and plane where the S-cell response is strongest. This analysis is carried out for all possible columns, so that the entire S-layer is considered. If two or more maxima occur in one

S-plane, the strongest one of those is retained and the others are discarded. Hence, this selection procedure locates the strongest response in each S-plane, but the maximum for a given plane can be rejected if it is overshadowed by the output from a nearby cell in some other plane. If this procedure selects a representative for S-plane $\hat{k}_l$ at position $\hat{n}$, then the input weights for that plane are reinforced according to the rules:

$$\Delta a_l(k_{l-1}, v, \hat{k}_l) = q_l \times c_{l-1}(v) \times Uc_{l-1}(k_{l-1}, \hat{n} + v) \tag{2.8}$$

$$\Delta b_l(\hat{k}_l) = q_l \times Vc_{l-1}(\hat{n}). \tag{2.9}$$

None of the weights are reinforced if all the columns produce the same response. The parameter q_l is a gain factor that controls the rate of learning at each level, and it usually becomes larger as one progresses to higher levels. Since the increment for a given excitatory weight is proportional to input from the C-cell, only those connections carrying strong input signals are substantially reinforced. Consequently, the most significant modifications occur for connections where the input and the output are both relatively strong. This behavior is similar to Hebbian learning without a decay term. Note that the Neocognitron could be operated in a supervised learning ("with a teacher") mode by specifying the plane $\hat{k}_l$ and location $\hat{n}$ to be used at each level for a given input image. Further refinements are possible by including decay terms in the learning rule, but only Equations (2.8 , 2.9) were implemented in the present work.

2.2 CELLS IN THE C-LAYER AND THE V$_s$-LAYER

The interactions and operational characteristics of the C-cells and the V$_s$-cells function in a manner that is very similar to the subsystem of S-cells and V$_c$-cells. These processors also take a given collection of source features and perform a comparison of two metrics. The result again determines whether information about that feature set will be passed on to higher level classifiers. Units in the C-layers and the V$_s$-layer have input fields on the preceding S-planes. The exact locations covered by a specific field are related to the position of the receiving unit, just as before. However, elements in a given C-plane receive excitatory inputs from only one particular S-plane, and each S-plane communicates with only one C-plane. This design principle is depicted in Figure 2-3, and it is responsible for the pairing of cell planes that was mentioned above. V$_s$-cells receive input from all the preceding S-planes and generate an inhibitory signal. The C-cells compare these excitatory and inhibitory inputs by applying the same shunting mechanism which an S-cell uses:

$$Uc_l(k_l, n) = g\left[\frac{1 + \sum_{v \epsilon D_l} d_l(v) \times Us_l(k_l, n+v)}{1 + Vs_l(n)} - 1\right] \tag{2.10}$$

where D_l is the region of input on the S-layer, Us is the S-cell output from position $n + v$ on S-plane k_l, and d_l is the input connection weight at relative position v in the region of input. This expression is quite similar to the S-cell response function given in Equations (2.1 , 2.2), but the excitatory component in the numerator includes contributions from only one S-plane, and the connection weights for inhibitory input are set to unity. The excitatory connection weights d_l have fixed values that are determined according to the same general principles used for the weights $c_l(v)$ in Equation 2.3. In practice, setting the d_l to be uniform across the receptive field has proven to be adequate.

Cells in the V_s plane produce the inhibitory signal, Vs_l, which is a weighted arithmetic mean of the S-cell outputs:

$$Vs_l(n) = \frac{1}{K_l} \sum_{k_{l=1}}^{K_l} \sum_{v \epsilon D_l} d_l(v) \times Us_l(k_l, n + v). \tag{2.11}$$

This V_s-cell response function computes the average S-cell output over the regions of input (D_l) for position n on all of the S-planes (K_l).

Since the numerator and the denominator in Equation 2.10 both contain the same set of connection weights $d_l(v)$, the essential effect of Equations 2.10 and 2.11 is to compare particular features on the S-layer with an average for all the S-planes. The C-cell response function in Equation 2.10 computes an adjusted signal ratio and passes it through the nonlinear saturation function:

$$g(x) = \begin{cases} \frac{x}{\beta + x} & (x \geq 0) \\ 0 & (x < 0) \end{cases} \tag{2.12}$$

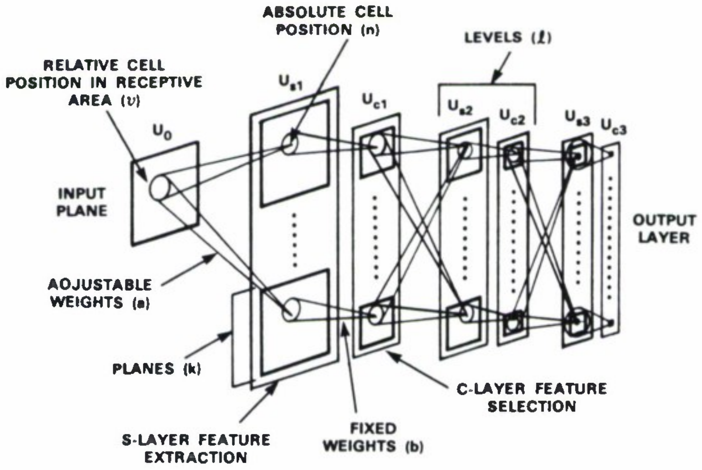
where β defines the degree of saturation. This parameter is typically chosen to be 0.5 for all levels. As a result of this processing, the C-cells select only those features which have a stronger response than the overall average. If every S-plane contained identical features, then none of the C-cells would respond. However, similar behavior ensues from the processing which occurs in S-cells, i.e., neither type of unit responds unless the input from a particular pattern is stronger than a measure of the "typical" source signal. In addition, both types of cell monotonically increase the strength of their response as the relative difference between the two inputs becomes greater.

2.3 OVERALL STRUCTURE AND FUNCTION

Most implementations of the Neocognitron are multilevel systems with an S-layer and a C-layer at each level. An example of the complete architecture is given in Figure 2-4, where l refers to the level, k refers a specific plane on a given layer, n refers to the absolute cell position within a plane, and v refers to the relative position within a receptive field. When the system operates, different patterns of connectivity decompose the input image into distinct spatial features. A learning rule then reinforces those patterns which produce the greatest response. As this type of system learns, different S-cells become sensitive to distinct combinations of features in the input plane.

The C-layer examines all feature groupings and rejects those that yield weak or mediocre output. Successive levels act to recognize increasingly complex feature groupings. On the top C-layer, each cell comprises an entire plane. Each C-cell in this final layer produces its maximum response only when a specific input image is presented.

NEOCOGNITRON
(Fukushima)



90150-1

Figure 2-4. Architecture of the Neocognitron.

3. RESULTS AND DISCUSSION

The Neocognitron's functional description contains many parameters which can be adjusted independently at each level. Cell planes in successive layers contain progressively fewer units as signals travel from the initial input layer to the top C-layer. The best sizes for input regions depend on the type of features in the input. Input images with few internal features need an initial region of input which covers the entire object so that the overall shape can be discerned. Smaller regions of input can be used for imagery containing many distinct features. Typically, the regions of input range in size from about 11x11 to 3x3 for wire frame objects in a 128x128 image plane, and the first S-layer's region of input on the initial image plane is largest. The acuity of discrimination is controlled by the selectivity parameter, r_l, in Equations (2.1, 2.2). Large values of r_l in the lower levels causes the Neocognitron to extract subtle features that can differentiate highly similar images. Smaller values of r_l result in the assignment of those same images to one class. When the value of r_l is too large, all patterns will be classified together, because the number of extracted features is too small. Typically, a successful scheme chooses the first level's r_l so that each input image produces a unique set of features, and subsequent levels then have decreasing values of r_l. The input weights for the V_c-plane (Equations 2.5, 2.6) also play a role in the feature extraction process. More features are extracted as the distribution of these weights becomes steeper, because fewer C-cells make substantial contributions to the inhibition. Finally, the rate of learning, q_l, increases from lower to higher levels in a manner which allows feature sensitive cells to develop slowly at all levels.

The input weights were initialized with small values that vary slightly between different S and C-planes. A slight dependence on the plane number insures that each plane will have a unique response. This variation is necessary, because C-cells will not respond if all the S-planes produce identical features. As an alternative, one could initialize the weights with biases towards certain expected features, e.g., set the weights so that the system responds to horizontal and vertical edges.

3.1 CLASSIFICATION

A four level Neocognitron structure was trained to classify three different wire frame objects. The training set, shown in Figure 3-1, consists of binary images which are stored in matrices of 128x128 pixels. Structural parameters for the different levels are listed in Table 3-1. A four level configuration with six planes per layer was used, providing a theoretical storage capacity of six different patterns. Only three different images actually were presented to this system, because the ability to uniquely classify inputs degrades (very sensitive to the choice of parameters) when the number of input patterns exceeds 50% of the number of planes. Training was conducted by presenting each of the images in turn (1 ⇒ 2 ⇒ 3), and twenty iterations were required to stabilize the weights. The category of classification corresponds to the top (level 4) C-Cell with the maximum response. Results for the images in Figure 3-1 are recorded in Table 3-2. The results show that the first input image (jeep) produced a maximum response in the sixth top C-layer cell, while images 2 (truck) and 3 (tank) produced a maximum response in cells 2 and 3 respectively. As discussed

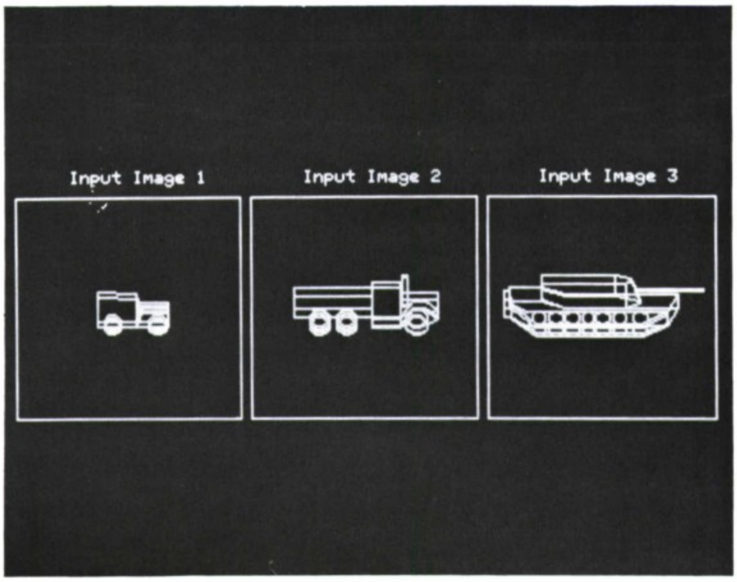


Figure 3-1. Input images used to train the four level Neocognitron.

		S-plane	C-plane	S-rec.	C-rec.			S-col.
Level	No. Planes	Size	Size	Area	Area	r_l	q_l	Size
1	6	121x121	62x62	7x7	5x5	4	1	4x4
2	6	57x57	30x30	5x5	5x5	4	4	4x4
3	6	25x25	13x13	5x5	5x5	3	6	4x4
4	6	8x8	1x1	5x5	8x8	2	20	4x4

TABLE 3-1.

Parameters for Training on Three Input Images

above, the Neocognitron classifies input patterns by choosing unique features to characterize each one. Figures 3-2 through 3-7 show the features extracted by all six planes in the first S-layer for each of the input images in Figure 3-1. These features are simply the output from the cells in S-planes of the first level, represented on an 8 bit grey scale to aid in visualization. The output strengths are coded with intensity, i.e., the maximum response is white and zero response is black. In Figures 3-2 and 3-3 each S-plane is extracting different sets of features from the first input image. Plane 4 in Figure 3-3 seems to emphasize horizontal features, while plane 6 responds mostly to vertical features. The response in planes 4 and 6 for the other input images (Figures 3-5 and 3-7), shows that these planes concentrate mostly on horizontal and vertical features. Note that in the other planes the system has selected many common features such as lines, corners, circles, and there is overlap between the different planes. The key point is that the S-cell outputs in these planes represent a set of features that uniquely classifies all the inputs. The features are not necessarily the ones that a human observer would choose as a basis for classification. The response patterns in the

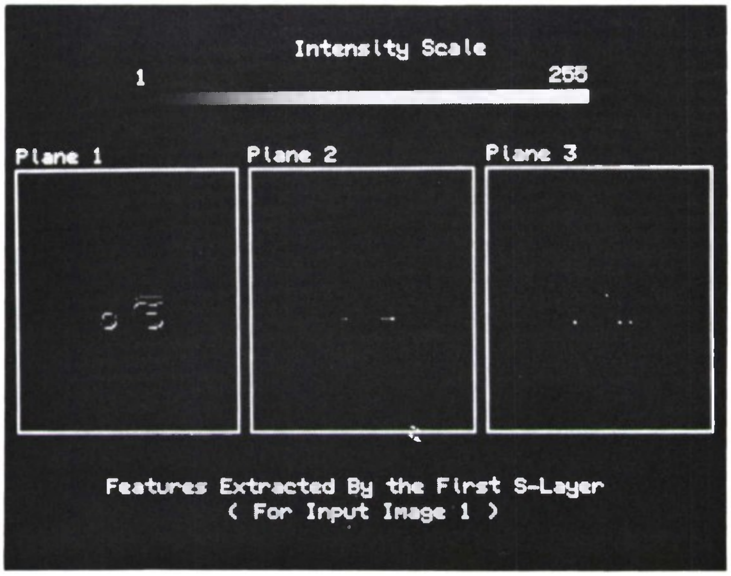


Figure 3-2. Extracted Features for S-planes (1, 2, 3) on the first level for image 1 in Figure 3-1 (using parameters in Table 3-1).

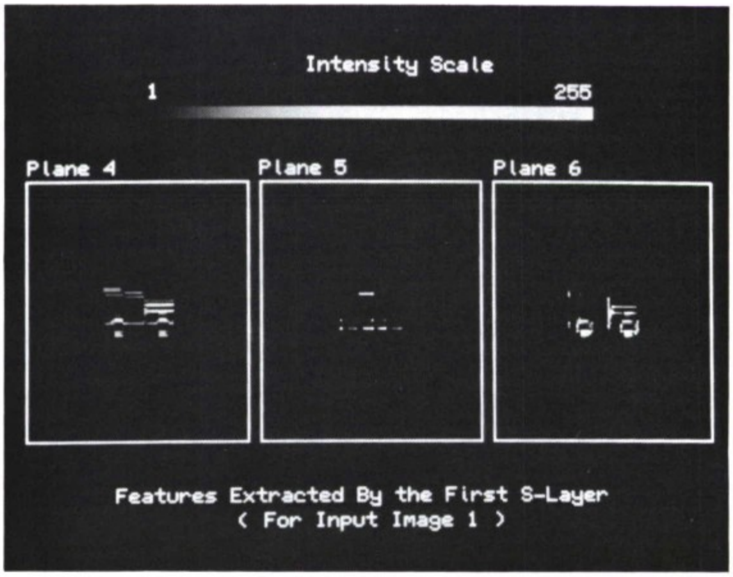


Figure 3-3. Extracted Features for S-planes (4, 5, 6) on the first level for image 1 in Figure 3-1 (using parameters in Table 3-1).

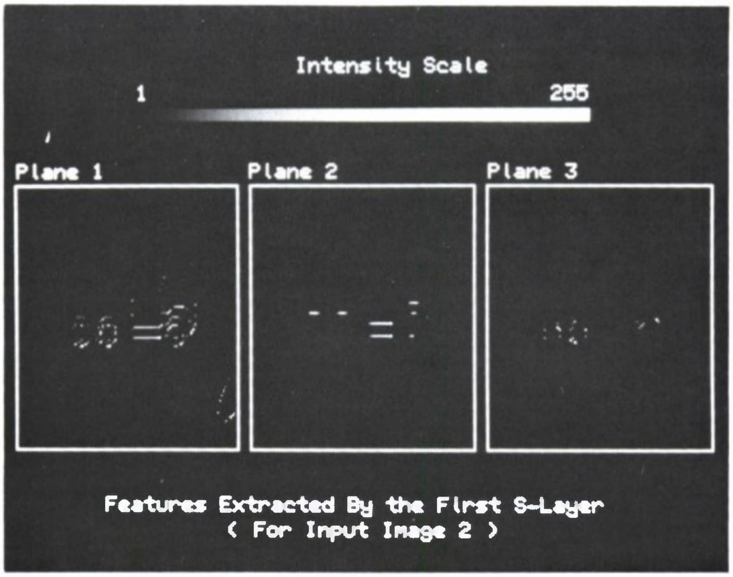


Figure 3-4. Extracted features for S-planes (1, 2, 3) on the first level for image 2 in Figure 3-1 (using parameters in Table 3-1).

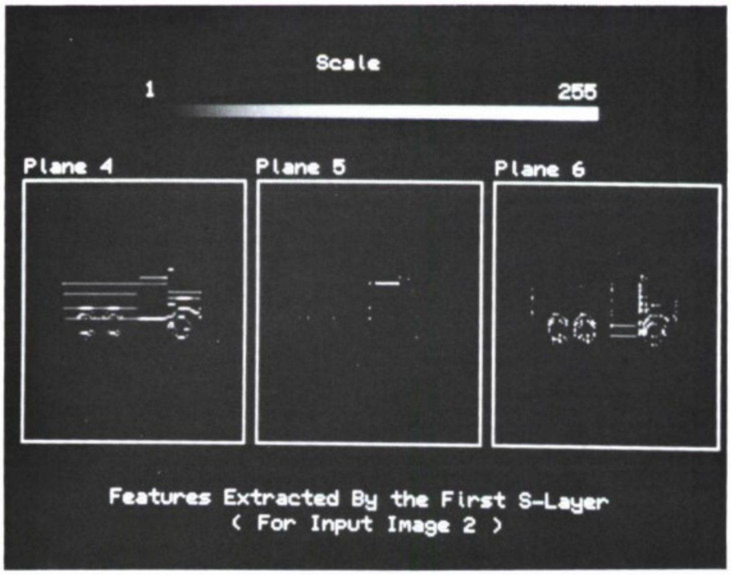


Figure 3-5. Extracted features for S-planes (4, 5, 6) on the first level for image 2 in Figure 3-1 (using parameters in Table 3-1).

Input Image	C-plane with Max Response
1	6
2	2
3	3

TABLE 3-2.

Trained Results (Using Parameters in Table 3-1)

feature planes bear a close resemblance to the input images because the two phenomena occur on adjacent levels. Features on higher levels generally are less similar to the input patterns, since they represent combinations of simpler features from the preceding planes. Though a human observer would have difficulty in interpreting features on higher levels, they act as abstract signatures that characterize particular input patterns. The selection of features is controlled by parameter r_l in Equations (2.1, 2.2). Increasing this parameter from a value of 4 to a value of 12 on the first level causes the system to become more critical and choose fewer features (Figures 3-8 and 3-9). The classification procedure also is influenced by the distribution of inhibitory connection weights c_l over an S-cell's region of input. A large value of α_l in Equation 2.5 results in a small inhibitory radius where very few of the C-cells make substantial contributions to the inhibition. Small values of α_l increase the inhibitory radius so that a densely populated region of input will produce significant inhibition to counter the excitation.

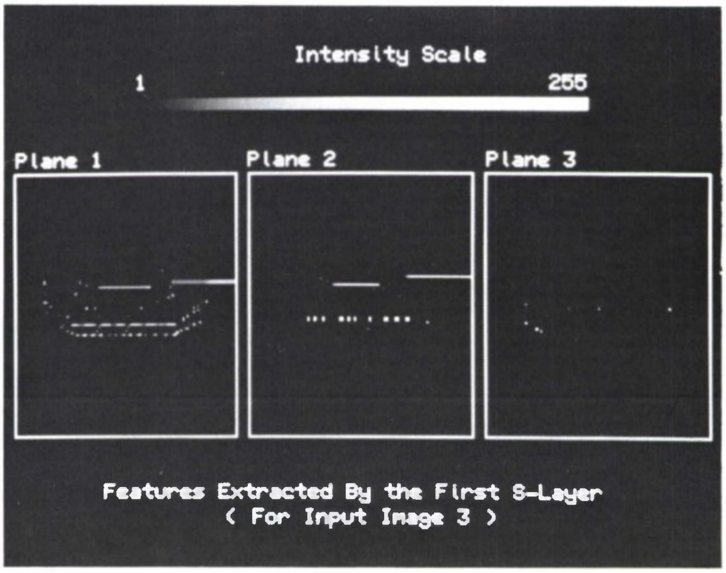


Figure 3-6. Extracted Features for S-planes (1, 2, 3) on the first level for image 3 in Figure 3-1 (using parameters in Table 3-1).

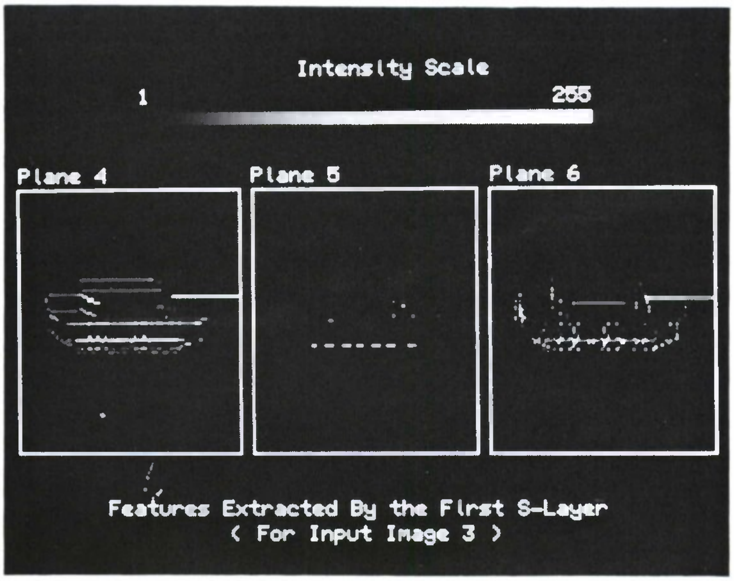


Figure 3-7. Extracted Features for S - planes (4, 5, 6) on the first level for image 3 in Figure 3-1 (using the parameters in Table 3-1).

The order of pattern presentation influences the actual mapping between individual input images and specific output C-cells, but the system reaches a stable configuration in all cases. In a self organizing system, the details of the final mapping are not important as long as it is stable. All the simulations performed here exhibited this behavior. A wide range of system parameters have been used, and the model has always settled into a state where the top level C-cells produce consistent outputs on every iteration.

3.2 SHIFT INVARIANCE

The Neocognitron's shift invariance property was studied by presenting a set of images showing each vehicle at four different positions in the 128 by 128 matrix. Figure 3-10 shows the shifted versions of image 1; note that large displacements are used (100% of the object width). Initial results indicated that the system described in Table 3-1 was not shift invariant for all three patterns. Several of the shifted images were misclassified. Further testing of the model showed that the rate of decay for the C-cells' input weights has a major effect. There seems to be a tradeoff between the ability to classify and shift invariance where the actual point of compromise depends on certain properties of the C-cells' weighting function. Uniform weights (no weighting) make the system shift invariant, but patterns cannot be classified consistently. In general, reliable classification requires a very steep weighting function.

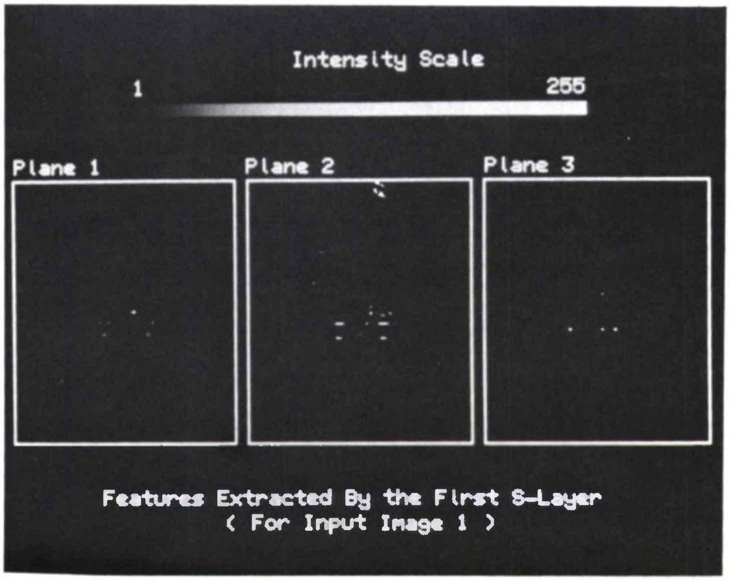


Figure 3-8. Extracted features for S-planes (1, 2, 3) on the first level for image 1 in Figure 3-1 (with higher selectivity, $r_l = 12$).

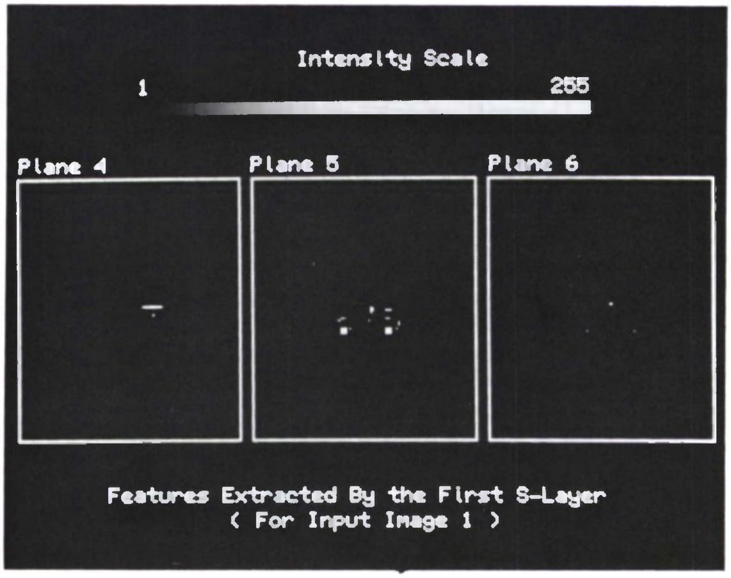


Figure 3-9. Extracted features for S-planes (4, 5, 6) on the first level for image 1 in Figure 3-1 (with higher selectivity, $r_l = 12$).

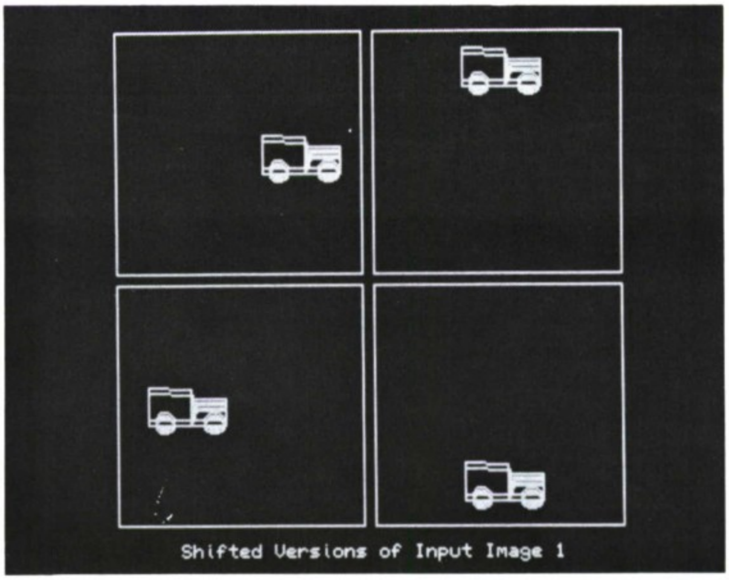


Figure 3-10. Shifted versions of image 1 in Figure 3-1.

The degree of shift tolerance also varies with the amount of overlap that exists between adjacent regions of input. It is essential that the regions of input for S-cells have a great deal of overlap, because these cells respond to very precise features. Actual occurrences of a selected feature can be missed if they fail to coincide with the available regions of input. A high degree of overlap is less important for the input regions of C-cells. However, copious overlap at any layer increases the number of levels which are needed before the cell population falls to one per plane in the final C-layer. For the case of a 128x128 input, the Neocognitron must have at least 15 levels to allow adequate overlap. In the examples that Fukushima has published, alphabetic characters from a 16x16 input layer were used, so only 3 to 4 levels were required. The processing time for a 15 level simulation on serial hardware is prohibitive (about 10 hours on a VAX-8600 computer to classify three patterns). Hence, implementation in parallel hardware seems to be the only viable solution for processing large input images.

One simple architecture has been found which classifies patterns reliably and maintains shift invariance: a single level system with one S-layer and one C-layer. Planes on the S-layer are made identical in size to the input plane (128x128), and the C-planes contain a single cell each. Input regions for the C-cells cover 128x128 pixels, and uniform weighting is used within this field. The selectivity parameter r_l in Equations 2.1, 2.2 is made sufficiently large to uniquely classify all the patterns. Using this scheme, the Neocognitron could classify the three images in Figure 3-1 in a completely shift invariant manner. This design has a drawback, because the the C-layer's selection of features may be based on the sheer number of pixels in an object, rather than their detailed arrangement. In a multi-level system, the features are gradually extracted and selected, so that more subtle variations in the imagery can be resolved.

3.3 PERFORMANCE WITH NOISE

The effect of noise was investigated by submitting corrupted images to a system that had been trained on a noise free data set. Uniform noise was introduced into the input images by replacing a certain portion of the pixels with random values. For the system described in Table 3-1, the ability to recognize patterns becomes unreliable when more than 30% of the pixels are corrupted. In general, this threshold will depend on the value used for the sensitivity parameter, r_l, during training; tolerance to noise is increased with lower sensitivity. Essentially, the Neocognitron calculates the inner product of a specific subpattern and the excitatory weights for an S-cell. The result is compared to a threshold determined by the selectivity parameter (r_l) and the inhibitory weights. Hence, the response with noise should be similar to that of a matched filter. In many systems, the distribution of raw pixel values is modified by some form of preprocessing, such as a median filter. Under these circumstances, the noise will no longer be distributed uniformly across the image. Further study is needed to evaluate the effect of such preprocessing on the model's performance with noisy imagery.

3.4 OBJECT ROTATION

Another form of preprocessing might enable the Neocognitron to recognize a particular type of scene variation by transforming the differences between images into positional shifts. This concept has been successfully applied to the problem of recognizing rotated objects. A simple one level Neocognitron was trained on polar transforms of the images in Figure 3-1. The three vehicles were then rotated through various angles, and polar transforms of the resulting images were submitted as input. Since the polar transform operation converts angular orientations into a linear coordinate, the different orientations for a given object were mapped into a positional shift with wrap-around. Use of a 128x128 input plane provides an angular resolution of about three degrees. The system successfully discriminated three images with no rotation, 7 degrees rotation, and 15 degrees rotation. Polar transforms of the input patterns are shown in Figure 3-11, while Figures 3-12 and 3-13 show features which the system extracted for the first image. Note that plane 4 in Figure 3-13 has no response at all, indicating that no features were extracted. Self organizing systems offer the advantage that they can identify the significant features in any representation of the image "without a teacher".

The preprocessing stage should locate the object's center of mass and perform the transform with respect to that position. This step minimizes distortions in the transformed image, e.g., truncation of the image due to edge effects. A difference in scale between the stored and presented patterns would present another problem. This difficulty can be overcome by using a combination log-polar transform which translates scale change to a shift on one axis and rotation to a shift on the other.

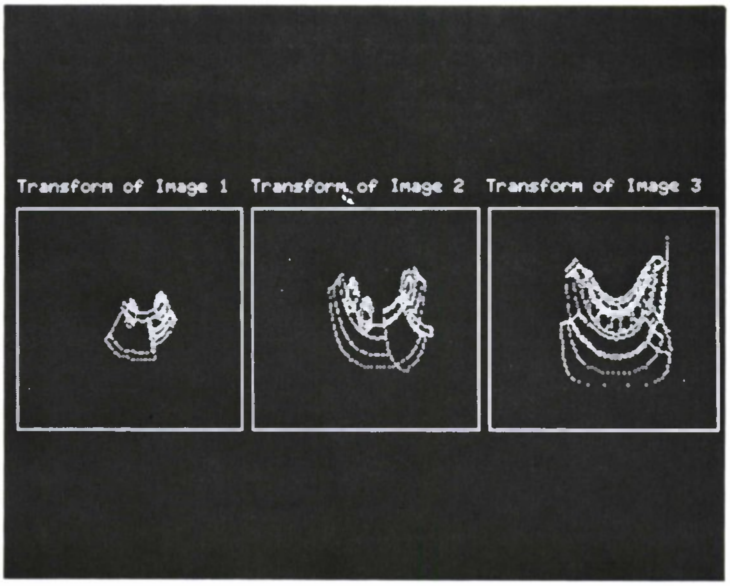


Figure 3-11. Polar transforms of the images in Figure 3-1.

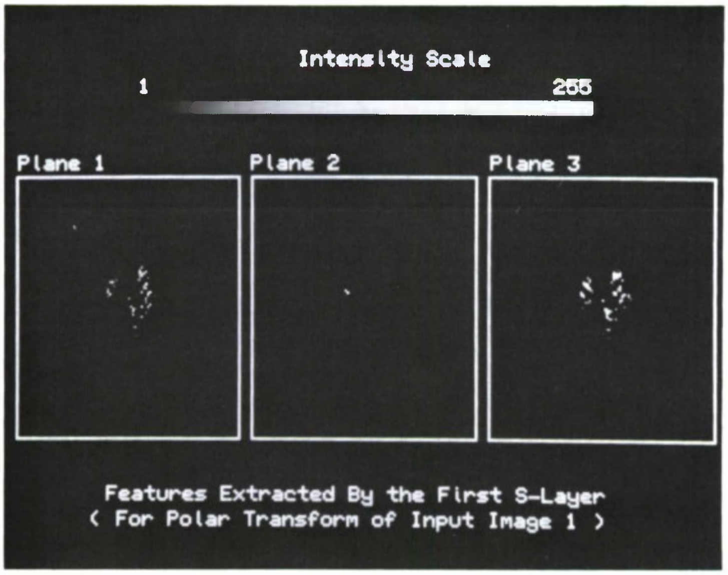
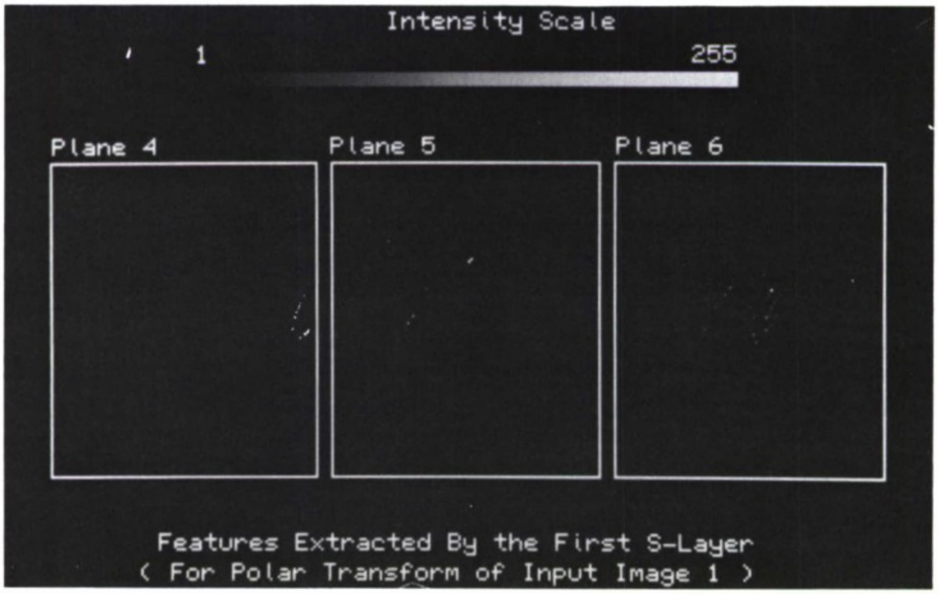

Intensity Scale
1
255
Plane 1
Plane 2
Plane 3
Features Extracted By the First S-Layer
(For Polar Transform of Input Image 1)

Figure 3-12. Extracted features for S-planes (1, 2, 3) on the first level for transforms of image 1 in Figure 3-1 (one level system).



Intensity Scale
1
255
Plane 4
Plane 5
Plane 6
Features Extracted By the First S-Layer
(For Polar Transform of Input Image 1)

Figure 3-13. Extracted features for S-planes (4, 5, 6) on the first level for transforms of image 1 in Figure 3-1 (one level system).

4. CONCLUSIONS

The Neocognitron model has been adapted to operate on complex wire frame imagery and the performance of this implementation was evaluated. This study found that a four level system with six S-planes and six C-planes per level can classify three different wire frame images The model works in two stages which are associated with S and C-layers. Cells in different S-planes extract distinct features, and weights for each plane are reinforced according to the pattern of input which produces the strongest response. The C-layer compares the response of individual S-planes with an overall average and rejects those features whose response falls below this average. This procedure is repeated across many levels until a single cell at the top C-layer responds to a specific input pattern. The first S-layer decomposes the input image into many small features and increasingly complex feature groupings are recognized in subsequent levels until the entire image has been characterized. Since the regions of input typically overlap, a number of cells respond to each group of features, and slightly deformed or shifted versions still produce significant responses.

Our results indicate that in order to achieve full shift tolerance, the Neocognitron must be designed with a high degree of overlap between adjacent regions of input. As a consequence, the cell layers should be thinned quite gradually. A four level system is not sufficient to provide shift invariance for 128x128 imagery. We estimate that an input plane of this size will require a system of at least 15 levels, and a system of this magnitude is best implemented in parallel hardware. A one level system with a 128x128 region of input was found to classify patterns in a shift tolerant manner for the three image training sequence used in our simulations. However, this is not an acceptable solution when a large number of patterns must be classified. In general, pattern classification using the Neocognitron will be practical with a multi-level architecture on a parallel machine.

The robustness of the model was tested by degrading the original training set with uniform noise. Correct classification could be maintained with up to 30% of the input pixels corrupted. A one level system was used to recognize rotated objects by using the polar transform as a preprocessor in training and in subsequent trials. This type of system correctly identified all rotated versions of the imagery.

Further work is needed to develop a systematic method for parameter selection. One promising approach is to incorporate a global feedback mechanism which would adjust the gain (q_l), and selectivity (r_l) at each level. The Neocognitrons's sensitivity to parameter changes also needs to be studied more extensively. This analysis would identify the set of parameters values which provides correct classification for the widest possible variety of input imagery. It is important to recognize that investigation of global feedback mechanisms and model sensitivity for large images (128x128 pixels and larger) can only be realized in a parallel hardware implementation.

REFERENCES

1. G. A. Carpenter and S. Grossberg, "Brain structure, learning and memory," in *AAAS Symposium Series*, (J. Davis, R. Newburgh, and E. Wegman, eds.), pp. 1–49, 1985.

2. K. Fukushima, "Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological Cybernetics*, vol. 36, pp. 193–202, 1980.

3. K. Fukushima, S. Miyake, and T. Ito, "Neocognitron: a neural network model for a mechanism of visual pattern recognition," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-13, pp. 826–834, 1983.

4. W. Stoner and T. M. Schilke, "Pattern recognition with a neural net," *Real Time Signal Processing IX*, vol. 698, pp. 171–181, 1986.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION
Unclassified

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE

3. DISTRIBUTION/AVAILABILITY OF REPORT
Approved for public release; distribution is unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S)
Project Report NN-2

5. MONITORING ORGANIZATION REPORT NUMBER(S)
ESD-TR-88-322

6a. NAME OF PERFORMING ORGANIZATION
Lincoln Laboratory, MIT

6b. OFFICE SYMBOL *(If applicable)*

7a. NAME OF MONITORING ORGANIZATION
Electronic Systems Division

6c. ADDRESS *(City, State, and Zip Code)*
P.O. Box 73
Lexington, MA 02173-0073

7b. ADDRESS *(City, State, and Zip Code)*
Hanscom AFB, MA 01731

8a. NAME OF FUNDING/SPONSORING ORGANIZATION
Defense Advanced Research Projects Agency

8b. OFFICE SYMBOL *(If applicable)*
DARPA/TTO

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

8c. ADDRESS *(City, State, and Zip Code)*
1400 Wilson Boulevard
Arlington, VA 22209

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
	320		

11. TITLE *(Include Security Classification)*
Prospects for Classifying Complex Imagery Using a Self-Organizing Neural Network

12. PERSONAL AUTHOR(S)
Murali M. Menon and Karl G. Heinemann

13a. TYPE OF REPORT
Project Report

13b. TIME COVERED
FROM ________ TO ________

14. DATE OF REPORT *(Year, Month, Day)*
11 January 1989

15. PAGE COUNT
36

16. SUPPLEMENTARY NOTATION
None

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*

self-organizing
neural network
shift invariance
rotation invariance
image classification
receptive fields
unsupervised learning
multilayer architecture
parallel computing

19. ABSTRACT *Continue on reverse if necessary and identify by block number)*

The objective of this study is to evaluate the performance of Fukushima's Neocognitron model when it is applied to complex imagery. In his original report, Fukushima demonstrated that this system could discriminate between simple alphabetical characters represented in fields of 16 by 16 pixels, and that shift invariance can be achieved through a proper choice of design parameters. The present work describes results for expanded Neocognitron architectures operating on complex images of 128 by 128 pixels. These neural network systems were simulated on a VAX-8600 minicomputer. Wire frame models of three different vehicles were used to test the properties which Fukushima had demonstrated. The expanded Neocognitron systems were able to classify these objects and to identify their critical features. After training, each object was placed at different positions in the plane, and the Neocognitron's shift invariance property was tested. With complex (128 × 128) imagery, it was difficult to achieve proper classification and maintain shift invariance using only a few levels. In another experiment, the Neocognitron trained on polar transforms of objects in the training set. Objects in the training set were rotated, and polar transforms of the rotated images were submitted as input. In this manner, the Neocognitron's shift invariance was exploited to recognize rotated imagery. These investigations gave insight into the role of various model parameters and their proper values, as well as demonstrating the model's applicability to complex images.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT
□ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. □ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION
Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL
Lt. Col. Hugh L. Southall, USAF

22b. TELEPHONE *(Include Area Code)*
(617) 981-2330

22c. OFFICE SYMBOL
ESD/TML

83 APR edition may be used until exhausted.
All other editions are obsolete.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE