mathematics," *J. Theoret. Biol.*, vol. 36, pp. 9–22, 1972.
[28] S. W. Kuffler and J. G. Nicholls, *From Neuron to Brain*. Sunderland, MA: Sinauer Assoc., 1976.
[29] J. P. LaSalle, "An invariance principle in the theory of stability," in *Differential Equations and Dynamical Systems*, J. K. Hale and J. P. LaSalle, Eds. New York: Academic, 1967.
[30] ____, "Stability theory for ordinary differential equations," *J. Differential Equations*, vol. 4, pp. 57–65, 1968.
[31] D. Levine and S. Grossberg, "On visual illusions in neural networks: Line neutralization, tilt aftereffect, and angle expansion," *J. Theoret. Biol.*, vol. 61, pp. 477–504, 1976.
[32] R. H. MacArthur, "Species packing and competitive equilibrium for many species," *Theoret. Population Biol.*, vol. 1, pp. 1–11, 1970.
[33] R. M. May and W. J. Leonard, "Nonlinear aspects of competition between three species," *SIAM J. Appl. Math.*, vol. 29, pp. 243–253, 1975.
[34] F. Ratliff, *Mach Bands: Quantitative Studies of Neural Networks in the Retina*. San Francisco, CA: Holden-Day, 1965.
[35] W. Rudin, *Function Theory on the Unit Ball of* $\mathbb{C}^n$. New York: Springer-Verlag, 1980.
[36] P. Schuster, K. Sigmund, and R. Wolff, "On ω-limits for competition between three species," *SIAM J. Appl. Math.*, vol. 37, pp. 49–54, 1979.
[37] Y. Takeuchi, N. Adachi, and H. Tokumaru, "The stability of generalized Volterra equations," *J. Math. Anal. Appl.*, vol. 62, pp. 453–473, 1978.

Neocognitron: A Neural Network Model for a Mechanism of Visual Pattern Recognition

KUNIHIKO FUKUSHIMA, SEI MIYAKE, AND TAKAYUKI ITO

Abstract—**A neural network model, called a "neocognitron," for a mechanism of visual pattern recognition was proposed earlier, and the result of computer simulation for a small-scale network was shown. A neocognitron with a larger-scale network is now simulated on a digital computer and is shown to have a great capability for visual pattern recognition: The neocognitron's ability to recognize handwritten Arabic numerals, even with considerable deformations in shape, is demonstrated. The neocognitron is a multilayered network consisting of a cascaded connection of many layers of cells. The information of the stimulus pattern given to the input layer is processed step by step in each stage of the multilayered network. A cell in a deeper layer generally has a tendency to respond selectively to a more complicated feature of the stimulus patterns and, at the same time, has a larger receptive field and is less sensitive to shifts in position of the stimulus patterns. Thus each cell of the deepest layer of the network responds selectively to a specific stimulus pattern and is not affected by the distortion in shape or the shift in position of the pattern. The synapses between the cells in the network are modifiable, and the neocognitron has a function of learning. A learning-with-a-teacher process is used to reinforce these modifiable synapses in the new model, instead of the learning-without-a-teacher process which was applied to the previous small-scale model.**

I. Introduction

THE NEURAL mechanism of visual pattern recognition in the brain is little known, and revealing it by conventional physiological experiments alone seems to be almost impossible. So, we take a slightly different approach to this problem. If we could make a neural network model which has the same capability for pattern recognition as a human being, it would give us a powerful clue to the understanding of the neural mechanism in the brain. In this paper, we discuss how to synthesize a neural network model in order to endow it with pattern recognition capability like that of a human being.

Several models were proposed with this intention [1]–[6]. In synthesizing such models, one of the most difficult problems is to design the networks so as to show position- and deformation-invariant responses. Some of these conventional models fail to recognize patterns which are shifted in position or deformed in shape. Although the four-layer perceptron [2] shows a kind of position-invariant responses, it works correctly only when the distance of shift is equal to one of the several specific values which are determined during the training of the network.

A few years ago, the authors [7], [8] proposed a multilayered neural network model, called a "neocognitron," which is capable of recognizing stimulus patterns correctly without being affected by any shift in position or even by considerable distortion in shape of the patterns. The result of computer simulation of a neocognitron with a small-scale network was reported there.

In this present paper, a neocognitron with a larger-scale network is simulated on a minicomputer PDP-11/34 and is shown to have a great capability for visual pattern recognition. The new model consists of nine layers of cells, while

Manuscript received August, 1, 1982; revised April 4, 1983.
The authors are with the NHK Broadcasting Science Research Laboratories, 1-10-11, Kinuta, Setagaya, Tokyo 157, Japan.

0018-9472/83/0900-0826$01.00

mathematics," *J. Theoret. Biol.*, vol. 36, pp. 9–22, 1972.
[28] S. W. Kuffler and J. G. Nicholls, *From Neuron to Brain*. Sunderland, MA: Sinauer Assoc., 1976.
[29] J. P. LaSalle, "An invariance principle in the theory of stability," in *Differential Equations and Dynamical Systems*, J. K. Hale and J. P. LaSalle, Eds. New York: Academic, 1967.
[30] ____, "Stability theory for ordinary differential equations," *J. Differential Equations*, vol. 4, pp. 57–65, 1968.
[31] D. Levine and S. Grossberg, "On visual illusions in neural networks: Line neutralization, tilt aftereffect, and angle expansion," *J. Theoret. Biol.*, vol. 61, pp. 477–504, 1976.
[32] R. H. MacArthur, "Species packing and competitive equilibrium for many species," *Theoret. Population Biol.*, vol. 1, pp. 1–11, 1970.
[33] R. M. May and W. J. Leonard, "Nonlinear aspects of competition between three species," *SIAM J. Appl. Math.*, vol. 29, pp. 243–253, 1975.
[34] F. Ratliff, *Mach Bands: Quantitative Studies of Neural Networks in the Retina*. San Francisco, CA: Holden-Day, 1965.
[35] W. Rudin, *Function Theory on the Unit Ball of* $\mathbb{C}^n$. New York: Springer-Verlag, 1980.
[36] P. Schuster, K. Sigmund, and R. Wolff, "On ω-limits for competition between three species," *SIAM J. Appl. Math.*, vol. 37, pp. 49–54, 1979.
[37] Y. Takeuchi, N. Adachi, and H. Tokumaru, "The stability of generalized Volterra equations," *J. Math. Anal. Appl.*, vol. 62, pp. 453–473, 1978.

Neocognitron: A Neural Network Model for a Mechanism of Visual Pattern Recognition

KUNIHIKO FUKUSHIMA, SEI MIYAKE, AND TAKAYUKI ITO

Abstract—**A neural network model, called a "neocognitron," for a mechanism of visual pattern recognition was proposed earlier, and the result of computer simulation for a small-scale network was shown. A neocognitron with a larger-scale network is now simulated on a digital computer and is shown to have a great capability for visual pattern recognition: The neocognitron's ability to recognize handwritten Arabic numerals, even with considerable deformations in shape, is demonstrated. The neocognitron is a multilayered network consisting of a cascaded connection of many layers of cells. The information of the stimulus pattern given to the input layer is processed step by step in each stage of the multilayered network. A cell in a deeper layer generally has a tendency to respond selectively to a more complicated feature of the stimulus patterns and, at the same time, has a larger receptive field and is less sensitive to shifts in position of the stimulus patterns. Thus each cell of the deepest layer of the network responds selectively to a specific stimulus pattern and is not affected by the distortion in shape or the shift in position of the pattern. The synapses between the cells in the network are modifiable, and the neocognitron has a function of learning. A learning-with-a-teacher process is used to reinforce these modifiable synapses in the new model, instead of the learning-without-a-teacher process which was applied to the previous small-scale model.**

I. Introduction

THE NEURAL mechanism of visual pattern recognition in the brain is little known, and revealing it by conventional physiological experiments alone seems to be almost impossible. So, we take a slightly different approach to this problem. If we could make a neural network model which has the same capability for pattern recognition as a human being, it would give us a powerful clue to the understanding of the neural mechanism in the brain. In this paper, we discuss how to synthesize a neural network model in order to endow it with pattern recognition capability like that of a human being.

Several models were proposed with this intention [1]–[6]. In synthesizing such models, one of the most difficult problems is to design the networks so as to show position- and deformation-invariant responses. Some of these conventional models fail to recognize patterns which are shifted in position or deformed in shape. Although the four-layer perceptron [2] shows a kind of position-invariant responses, it works correctly only when the distance of shift is equal to one of the several specific values which are determined during the training of the network.

A few years ago, the authors [7], [8] proposed a multilayered neural network model, called a "neocognitron," which is capable of recognizing stimulus patterns correctly without being affected by any shift in position or even by considerable distortion in shape of the patterns. The result of computer simulation of a neocognitron with a small-scale network was reported there.

In this present paper, a neocognitron with a larger-scale network is simulated on a minicomputer PDP-11/34 and is shown to have a great capability for visual pattern recognition. The new model consists of nine layers of cells, while

Manuscript received August, 1, 1982; revised April 4, 1983.
The authors are with the NHK Broadcasting Science Research Laboratories, 1-10-11, Kinuta, Setagaya, Tokyo 157, Japan.

0018-9472/83/0900-0826$01.00

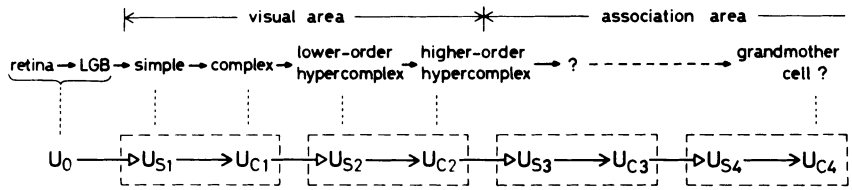

Fig. 1. Comparison between hierarchical model by Hubel and Wiesel and structure of neural network of neocognitron.

the previous model consisted of seven layers. We demonstrate that the new model can be trained to recognize handwritten Arabic numerals even with considerable deformations in shape.

We use a learning-with-a-teacher process for the reinforcement of the modifiable synapses in the new large-scale model, instead of the learning-without-a-teacher process applied to the previous model. In this paper, we focus on the mechanism for pattern recognition rather than that for self-organization.

II. Structure of the Network

The neocognition is a multilayered network with a hierarchical structure similar to the hierarchical model for the visual system proposed by Hubel and Wiesel [9], [10]. As shown in Fig. 1, the neocognitron is composed of a cascaded connection of a number of modular structures preceded by an input layer U_0 consisting of photoreceptor array. Each of the modular structures is composed of two layers of cells, namely, a layer U_S consisting of S cells, and a layer U_C consisting of C cells. The layers U_S and U_C in the lth module are denoted by U_{Sl} and U_{Cl}, respectively. An S cell has a response characteristic similar to a simple cell or a lower order hypercomplex cell according to the classification by Hubel and Wiesel, while a C cell resembles a complex cell or a higher order hypercomplex cells. In this network, a cell in a higher stage generally has a tendency to respond selectively to a more complicated feature of the stimulus pattern and, at the same time, has a larger receptive field and is more insensitive to the shift in position of the stimulus pattern.

Each S cell has modifiable input synapses which are reinforced with learning and acquires an ability to extract a specific stimulus feature. That is, an S cell comes to respond only to a specific stimulus feature and not to respond to other features.

Each C cell has afferent synapses leading from a group of S cells which have receptive fields of similar characteristics at approximately the same position on the input layer. This means that all of the presynaptic S cells are to extract almost the same stimulus feature but from slightly different positions on the input layer. The efficiencies of the synapses are determined in such a way that the C cell will be activated whenever at least one of its presynaptic S cells is active. Hence, even if a stimulus pattern which has elicited a large response from the C cell is shifted a little in position, the C cell will keep responding as before, because another presynaptic S cell will become active instead of the first one. In other words, a C cell responds to the same stimulus feature as its presynaptic S cells do but is less sensitive to the shift in position of the stimulus feature.

S cells or C cells in any single layer are sorted into subgroups according to the optimum stimulus features of their receptive fields. Since the cells in each subgroup are set in a two-dimensional array, we call the subgroup as a "cell plane." We will also use the terminology S plane and C plane to represent the cell planes consisting of S cells and C cells, respectively. All the cells in a single cell plane have input synapses of the same spatial distribution, and only the positions of the presynaptic cells are shifted in parallel depending on the position of the postsynaptic cells. Even in the process of learning, in which the efficiencies of the synapses are modified, the modification is performed always under this restriction.

Fig. 2 is a schematic diagram illustrating the synaptic connections between layers. Each tetragon drawn with heavy lines represents an S plane or a C plane, and each vertical tetragon drawn with thin lines, in which S planes or C planes are enclosed, represents an S layer or a C layer.

In Fig. 2, for the sake of simplicity, only one cell is shown in each cell plane. Each of these cells receives input synapses from the cells within the area enclosed by the ellipse in its preceding layer. All the other cells in the same cell plane have input synapses of the same spatial distribution, and only the positions of the presynaptic cells are shifted in parallel from cell to cell. Hence all the cells in a single cell plane have receptive fields of the same function but at different positions.

Since the cells in the network are interconnected in a cascade as shown in Fig. 2, the deeper the layer is, the larger becomes the receptive field of each cell of that layer. The density of the cells in each cell plane is so determined as to decrease in accordance with the increase of the size of the receptive fields. The number of cells in each layer is shown at the bottom of Fig. 2. In the deepest module, the receptive field of each C cell becomes so large as to cover the whole input layer, and each C plane is so determined as to have only one C cell. Fig. 3 illustrates concretely how the cells of each cell plane are interconnected to the cells of other cell planes.

S cells and C cells are excitatory cells. Although it is not shown in Figs. 2 and 3, we have inhibitory V_C cells in C layers.

Here, we will describe the outputs of these cells with numerical expressions. All the cells employed in the neocognitron are of analog type; that is, the input and output signals of the cells take nonnegative analog values proportional to the instantaneous firing frequencies of

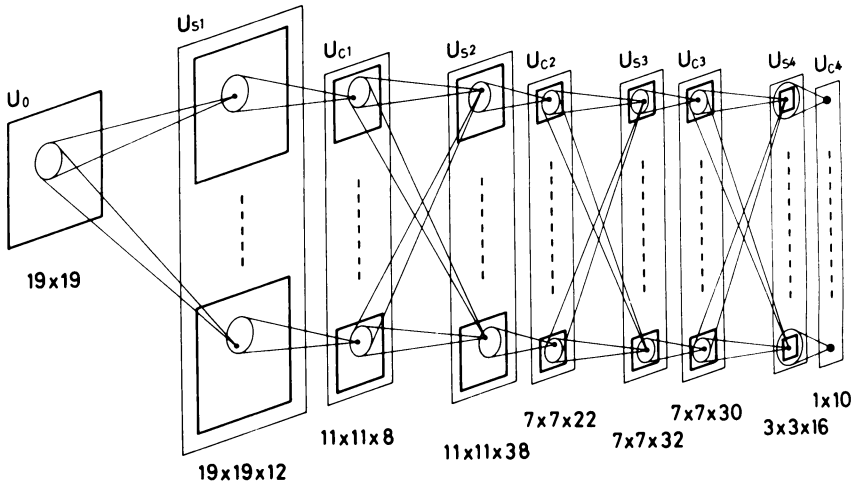

Fig. 2. Schematic diagram illustrating synaptic connections between layers in neocognitron.

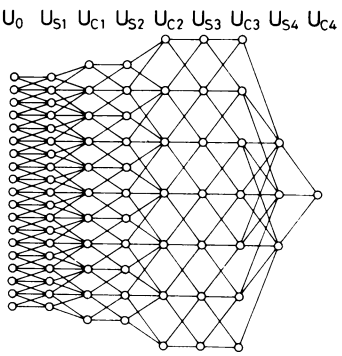

Fig. 3. One-dimensional view of interconnections between cells of different cell planes. Only one cell plane is drawn in each layer.

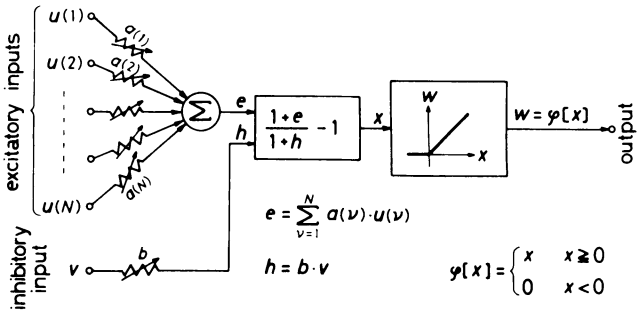

Fig. 4. Input-to-output characteristics of S cell: typical example of cells employed in neocognitron.

actual biological neurons. The output of a photoreceptor is denoted by $u_0(\boldsymbol{n})$ where $\boldsymbol{n}$ represents the two-dimensional coordinates indicating the location of the cell. We will use notations $U_{Sl}(k, \boldsymbol{n})$ to represent the output of an S cell in the kth S plane in the lth module, and $u_{Cl}(k, \boldsymbol{n})$ to represent the output of a C cell in the kth C plane in that module, where $\boldsymbol{n}$ is the two-dimensional coordinates representing the position of these cells' receptive fields on the input layer.

As shown in Fig. 4, S cells have inhibitory inputs with shunting mechanism. Incidentally, S cells have the same characteristics as the excitatory cells employed in the conventional cognitron [5], [6]. The output of an S cell of the kth S plane in the lth module is given by

$$U_{Sl}(k, \boldsymbol{n}) = r_l \cdot \phi\left(\frac{1 + \sum_{\kappa=1}^{K_{Cl-1}} \sum_{\nu \in A_l} a_l(\kappa, \nu, k) \cdot u_{Cl-1}(\kappa, \boldsymbol{n} + \nu)}{1 + \frac{r_l}{1 + r_l} \cdot b_l(k) \cdot v_{Cl-1}(\boldsymbol{n})} - 1\right), \qquad k = 1, 2, \cdots, K_{Sl} \tag{1}$$

where $\phi[x] = \max(x, 0)$. In the case of $l = 1$ in (1), $u_{Cl-1}(\kappa, \boldsymbol{n})$ stands for $u_0(\boldsymbol{n})$, and we have $K_{Cl-1} = 1$.

Here, $a_l(\kappa, \nu, k)$ and $b_l(k)$ represent the efficiencies of the excitatory and inhibitory modifiable synapses, respectively. As described before, all the S cells in the same S plane are assumed to have an identical set of afferent synapses. Hence $a_l(\kappa, \nu, k)$ and $b_l(k)$ do not contain any argument representing the position $\boldsymbol{n}$ of the receptive field of cell $u_{Sl}(k, \boldsymbol{n})$.

Parameter r_l in (1) controls the intensity of the inhibition. The larger the value of r_l is, the more selective becomes the cell's response to its specific feature. Their values are $r_1 = 1.7$, $r_2 = 4.0$, $r_3 = 1.5$, and $r_4 = 1.0$. (A detailed discussion on the response of S cells will be given in Section III-B.)

The inhibitory cell $v_{Cl-1}(\boldsymbol{n})$, which is sending an inhibitory signal to cell $U_{Sl}(k, \boldsymbol{n})$, receives afferent synapses from the same group of cells as $u_{Sl}(k, \boldsymbol{n})$ does and yields an output proportional to the weighted root mean square of its inputs:

$$v_{Cl-1}(\boldsymbol{n}) = \sqrt{\sum_{\kappa=1}^{K_{Cl-1}} \sum_{\nu \in A_l} c_{l-1}(\nu) \cdot u_{Cl-1}^2(\kappa, \boldsymbol{n} + \nu)}\,. \tag{2}$$

The efficiencies of the unmodifiable synapses $c_{l-1}(\nu)$ are determined so as to decrease monotonically with respect to $|\nu|$ and to satisfy

$$K_{Cl-1} \cdot \sum_{\nu \in A_l} c_{l-1}(\nu) = 1. \tag{3}$$

The size of the connection area A_l of these cells is set to be

small in the first module and to increase with the depth l as illustrated in Fig. 3.

The output of a C cell of the kth C plane in the lth module is given by

$$u_{Cl}(k, \boldsymbol{n}) = \psi\left(\sum_{\kappa=1}^{K_{Sl}} j_l(\kappa, k) \sum_{\boldsymbol{\nu} \in D_l} d_l(\boldsymbol{\nu}, k) \cdot u_{Sl}(\kappa, \boldsymbol{n} + \boldsymbol{\nu}) \right), \quad \cdot k = 1, 2, \cdots, K_{Cl}, \quad (4)$$

where

$$\psi[x] = \begin{cases} x/(\alpha_l + x), & (x \geqq 0); \\ 0, & (x < 0). \end{cases}$$

The parameter α_l is a positive constant which determines the degree of saturation of the output. Their values are $\alpha_1 = \alpha_2 = \alpha_3 = 0.25$, and $\alpha_4 = 1.0$.

In (4), $d_l(\boldsymbol{\nu}, k)$ represents the efficiencies of the excitatory synapses leading from S cells, and $j_l(\kappa, k)$ takes value one or zero depending on whether synaptic connections really exist from the κth S plane to the kth C plane. The value of $d_l(\boldsymbol{\nu}, k)$ is determined so as to decrease monotonically with respect to $|\boldsymbol{\nu}|$ and is independent of k except for $l = 1$. The size of the connection area D_l is set to be small in the first module and to increase with depth l as illustrated in Fig. 3.

The process of pattern recognition in this multilayered network can be briefly summarized as follows. The stimulus pattern is first observed within a narrow range by each of the S cells in the first module, and several features of the stimulus pattern are extracted. In the next module, these features are combined by observation over a little larger range, and higher order features are extracted. Operations of this kind are repeatedly applied through a cascaded connection of a number of modules. In each stage of these operations, a small amount of positional error is tolerated. The operation by which positional errors are tolerated little by little, not at a single stage, playes an important role in endowing the network with an ability to recognize even distorted patterns.

III. Synaptic Connections Between Cells

The synaptic connections in the new model of the neocognitron are reinforced by means of a supervised learning, that is, a learning-with-a-teacher process. During the training process, the network is presented with a set of training patterns to the input layer, together with the instructions which cells in the network should come to respond to each of the training patterns. This algorithm is different from that used in the previous model [7], [8]. In the new model, the algorithm for the reinforcement of synapses is determined from a standpoint of an engineering application to a design of a pattern recognizer rather than from that of pure biological modeling. That is, the algorithm is determined with the criterion of obtaining a better performance in handwritten character recognition.

A. Reinforcement of the Input Synapses of S Cells

The reinforcement of the synaptic connections are performed in sequence from the distal to the deeper layers. That is, the reinforcement of the input synapses of the lth layer is performed after completion of the reinforcement of up to the $(l-1)$th layer.

A number of cell planes are in an S layer. These cell planes are reinforced one at a time. In order to reinforce a cell plane, the "teacher" presents a training pattern to the input layer, and at the same time chooses one S cell which should work as the "representative" from that cell plane. The representative cell works like a seed in the crystal growth. The input synapses to the representative cell are reinforced depending on the stimuli given to these synapses. That is, only the synapses through which nonzero signals are coming are reinforced. As the result, the representative cell acquires a selective responsiveness to the training pattern which is now presented to the input layer. All the other cells in that cell plane have their input synapses reinforced in the identical manner as their representative.

This algorithm can be expressed as follows. Let cell $u_{Sl}(\hat{k}, \hat{\boldsymbol{n}})$ be the representative. The modifiable synapses $a_l(\kappa, \boldsymbol{\nu}, \hat{k})$ and $b_l(\hat{k})$, which are afferent to the S cells of this S plane, are reinforced by the amount shown below:

$$\Delta a_l(\kappa, \boldsymbol{\nu}, \hat{k}) = q_l \cdot c_{l-1}(\boldsymbol{\nu}) \cdot u_{Cl-1}(\kappa, \hat{\boldsymbol{n}} + \boldsymbol{\nu}), \quad (5)$$

$$\Delta b_l(\hat{k}) = q_l \cdot v_{Cl-1}(\hat{\boldsymbol{n}}), \quad (6)$$

where q_l is a positive constant which determines the amount of reinforcement. The initial values of these modifiable synapses are all zero.

We can choose any cell of a cell plane as the representative, and the choice of the representative does not have so much effect on the result of training, provided that the training pattern is presented at a proper position in the respective field of the representative. Hence, in the computer simulation discussed later, we always choose the cell situated at the center of each cell plane as the representative.

In the computer simulation, the number of training patterns given to each cell plane is from one to four, depending on the required allowance to the deformation of the stimulus features. (See the following section for more discussions.)

B. Analysis of the Response of an S Cell

In this section, we discuss how each S cell is trained to respond selectively to differences in stimulus patterns. Since the structure between two adjoining modules is similar in all parts of the network, we observe the response of an arbitrary S cell $U_{S1}(k, \boldsymbol{n})$ of layer U_{S1} as a typical example. Fig. 5 shows the synaptic connections converging to such a cell. For the sake of simplicity, we will omit the suffixes S, $l = 1$ and the arguments k, $\boldsymbol{n}$ and represent the response of this cell simply by u. Similarly, we will use the notation v for the output of the inhibitory cell $v_{C0}(\boldsymbol{n})$, which sends an

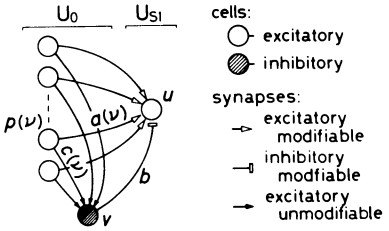

Fig. 5. Synaptic connections converging to S cell.

inhibitory signal to cell u. For the other variables, the arguments k and $\boldsymbol{n}$ and suffixes S, C, l, and $l-1$ will also be omitted.

Let $p(\nu)$ be the response of the cells of layer U_0 situated in the connection area of cell u, so that

$$p(\nu) = u_0(\boldsymbol{n} + \nu). \tag{7}$$

In other words, $p(\nu)$ is the stimulus pattern (or feature) presented to the receptive field of cell u.

With this notation, (1) and (2) can be written

$$u = r \cdot \phi\left(\frac{1 + \Sigma_\nu a(\nu) \cdot p(\nu)}{1 + \frac{r}{1+r} \cdot b \cdot v} - 1 \right) \tag{8}$$

$$v = \sqrt{\Sigma_\nu c(\nu) \cdot p^2(\nu)}\,. \tag{9}$$

When cell u is chosen as the representative, the amounts of reinforcement of the modifiable synapses are derived from (5) and (6), that is,

$$\Delta a(\nu) = q \cdot c(\nu) \cdot p(\nu), \tag{10}$$

$$\Delta b = q \cdot v. \tag{11}$$

Let s be defined by

$$s = \frac{\Sigma_\nu a(\nu) \cdot p(\nu)}{b \cdot v}. \tag{12}$$

Then (8) reduces approximately to

$$u \simeq r \cdot \phi\left(\frac{r+1}{r} \cdot s - 1 \right), \tag{13}$$

provided that $a(\nu)$ and b are sufficiently large.

Let a stimulus pattern $p(\nu) = P(\nu)$ be presented, and let cell u be chosen as the representative. Then, from (5) and (6), we obtain

$$a(\nu) = q \cdot c(\nu) \cdot P(\nu), \tag{14}$$

$$b = q\sqrt{\Sigma_\nu c(\nu) \cdot P^2(\nu)}\,. \tag{15}$$

Substituting (9), (14), and (15) into (12), we obtain

$$s = \frac{\Sigma_\nu c(\nu) \cdot P(\nu) \cdot p(\nu)}{\sqrt{\Sigma_\nu c(\nu) \cdot P^2(\nu)} \cdot \sqrt{\Sigma_\nu c(\nu) \cdot p^2(\nu)}}. \tag{16}$$

If we regard $p(\nu)$ and $P(\nu)$ as vectors, (16) can be interpreted as the (weighted) inner product of the two vectors normalized by the norms of both vectors. In other words, s gives the cosine of the angle between the two vectors $p(\nu)$ and $P(\nu)$ in the multidimensional vector space. Therefore, we have $s = 1$ only when $p(\nu) = P(\nu)$, and we have $s < 1$ for all patterns such as $p(\nu) \neq P(\nu)$. This means that s becomes maximum for the training pattern and becomes smaller for any other patterns.

If parameter q is large enough, (13) holds. When an arbitrary pattern $p(\nu)$ is presented, and if it satisfies $s > r/(r+1)$, we have $u > 0$ by (13). Conversely, for a pattern which makes $s \leqq r/(r+1)$, cell u does not respond. We can interpret by saying that cell u judges the similarity between patterns $p(\nu)$ and $P(\nu)$ using the criterion defined by (16) and that it responds only to patterns judged to be similar to $P(\nu)$. Incidentally, if $p(\nu) = P(\nu)$, we have $s = 1$ and consequently $u \simeq 1$.

Since the value $r/(r+1)$ tends to one with increase of r, a larger value of r makes the cell's response more selective to one specific pattern or feature. In other words, a large value of r endows the cell with a high ability to discriminate patterns of different classes. However, a higher selectivity of the cell's response is not always desirable, because it decreases the ability to tolerate the deformation of patterns. Hence the value of r should be determined at a point of compromise between these two contradictory conditions.

In the above analysis, we supposed that cell u is trained only for one particular pattern $P(\nu)$. When cell u has been trained to two patterns, say, to $P_1(\nu)$ and $P_2(\nu)$, $P(\nu)$ in the above discussions should be replaced with $\{P_1(\nu) + P_2(\nu)\}$. Hence cell u acquires a tendency to respond equally to both $P_1(\nu)$ and $P_2(\nu)$. This, however, depends on the value of r, and also on the similarity between $P_1(\nu)$ and $P_2(\nu)$. If the difference between $P_1(\nu)$ and $P_2(\nu)$ is too large, or if the value of r is too large, cell u comes to respond neither to $P_1(\nu)$ nor $P_2(\nu)$.

The above discussion is not restricted to S cells of layer U_{S1}. Each S cell in succeeding modules shows a similar type of response, if we regard the response of the C cells in its connection area in the preceding layer as its input pattern.

C. Layers U_{SI} and U_{CI}

Layer U_{S1} has 12 cell planes, and each cell plane contains the same number of cells as layer U_0, that is, 19×19 (see Figs. 2 and 3). These S cells have their input synapses reinforced so as to acquire the ability to detect line components of various orientations.

The training patterns which are used for training the 12 cell planes are displayed in column a_1 in Fig. 6. This figure shows, for example, that the cells of the first cell plane are trained to detect a horizontal line component. We can also interpret this by saying that the patterns in column a_1 show the structure of the receptive fields of the cells of layer U_{S1}.

Since the spread of the excitatory input synapses $a_1(\kappa, \nu, k)$ of each S cell (i.e., the connection area A_1 in (1) and (2)) is as small as 3×3, cases exist where two different cell planes should be prepared for detecting a line of a particular orientation. For example, in Fig. 6, the second and third cell planes of layer U_{S1} have the receptive fields of the same preferred orientation but of different structures. Hence the outputs from such pairs of cell planes are

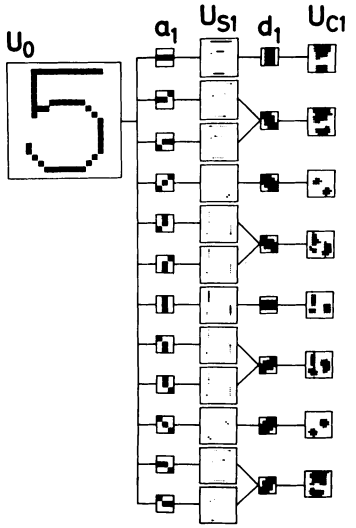

Fig. 6. Example of response of cells of layers U_0, U_{S1}, and U_{C1}, and synaptic connections between them.

joined together at the input stage of layer U_{C1} as shown in Fig. 6. The parameter $j_1(\kappa, k)$ in (4) takes value one or zero, depending on this joining condition. For instance, $j_1(\kappa, 2) = 1$ for $\kappa = 2$ and 3, and $j_1(\kappa, 2) = 0$ for other κ. Because of this joining process, layer U_{C1} contains only eight cell planes.

Parameter r_1 in (1), which determines the selectivity of an S cell, is set at value of 1.7. Since the stimulus feature which is used for training an S cell contains three active elements, the S cell with $r_1 = 1.7$ yields nonzero output for a stimulus feature contaminated with up to two additive elements of noise, or one additive and one subtractive elements. However, it does not respond to a stimulus feature with two subtractive elements of noise or more. (These results can be obtained from the analysis in Section III-B as well as from the computer simulation.)

The spatial distribution of the input synapses $d_1(\nu, k)$ of a C cell (i.e., the connection area D_1 in (4)) is 5×5 in size, but all of these 5×5 synapses are not effective. As shown in Fig. 6, the effective part of the distribution is elongated to the direction perpendicular to its preferred orientation and is compressed in the direction of its preferred orientation.

Since each C cell receives excitatory signals from a number of S cells, it usually responds similarly as its neighboring C cells. Hence it is possible to reduce the number of C cells in each cell plane compared to that of S cells. The density of cells in each cell plane of layer U_{C1} are thinned out by two to one compared to that of layer U_{S1} both in horizontal and vertical directions. Thus as shown in Fig. 3, the number of cells in each cell plane is reduced to 11×11 in layer U_{C1}.

D. Layers U_{S2} and U_{C2}

Layer U_{S2} has 38 cell planes, and each cell plane contains 11×11 S cells. Layer U_{C2} has only 22 cell planes, because the outputs from some of the cell planes of layer U_{S2} are joined together at the input stage of layer U_{C2}.

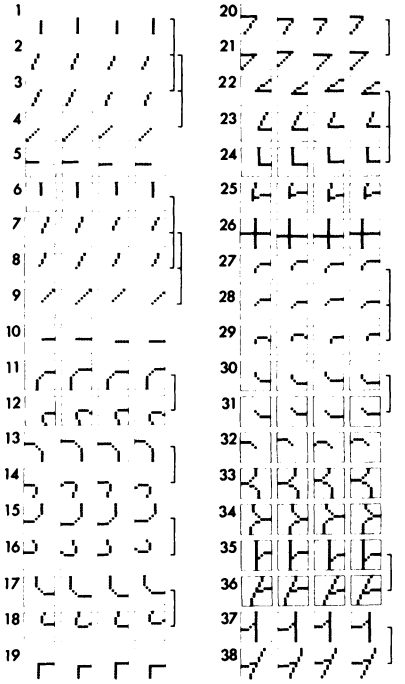

Fig. 7. Training patterns used to train 38 cell planes of layer U_{S2}. Way of joining at input stage of layer U_{C2} is also shown to right of each group of training patterns.

Each cell plane of layer U_{C2} contains 7×7 C cells because of the two to one thinning-out of the cell density as was shown in Fig. 3.

Each S cell of layer U_{S2} has modifiable excitatory input synapses of 3×3 spatial distribution. Since the preceding layer U_{C1} has eight cell planes, the total number of the excitatory input synapses to each S cell is $3 \times 3 \times 8$. All of these synapses are not reinforced by learning, but most of them usually stay at the initial value of zero. The input synapses to each C cell of layer U_{C2} have spatial distribution of 5×5.

Figure 7 shows the training patterns used for training the 38 cell planes of layer U_{S2}. Four training patterns, in which the same stimulus feature is shifted in parallel to each other by one element in both horizontal and vertical directions, are used to train each cell plane. The reason why the use of four patterns are necessary is discussed below.

Because the cells of this layer are thinned out by two to one compared to those of layer U_0 in both horizontal and vertical directions, each S cell should take charge of extracting a specific stimulus feature from four different positions on layer U_0. In this network, the two to one thinning-out of the cells is already made at the stage of layer U_{C1}, from which the relevant S cells receive synaptic connections. The effect of this thinning-out is not so small, and a somewhat different spatial response might appear in the preceding layer U_{C1} when the stimulus pattern is shifted in position by one element. Hence each cell-plane of layer U_{S2} should be trained with four different patterns beforehand so as to come to respond equally to them.

In this experiment, we intend to train the neocognitron so as to recognize handwritten Arabic numerals. When

patterns are written by hand, the stimulus features in the patterns usually suffer from considerable deformations depending on the writers. However, the way of deformation is not at random but usually has some tendency. Some of such deformed features are detected separately in a number of cell planes of layer U_{S2} and are combined together at the input stage of layer U_{C2}. In Fig. 7, the lines drawn to the right of the 36 groups of training patterns indicate how this joining is made.

E. Layers U_{S3} and U_{C3}

Layer U_{S3} has 32 cell planes, and U_{C3} has 30 cell planes. The number of cells in each cell plane is 7×7 for both layers U_{S3} and U_{C3}. Thinning-out is not performed between these layers. Each S cell has $3 \times 3 \times 22$ modifiable excitatory input synapses, and the input synapses of each C cell have spatial distribution of 3×3.

Fig. 8 shows the training patterns used for training the 32 cell planes of layer U_{S3} and also shows how the outputs from these cell planes are joined together at the input stage of layer U_{C3}. Most of these training patterns consist of some parts of the standard numeral patterns which are to be taught to this network.

As is seen in Fig. 8, only two or three different patterns are used to train each cell plane of layer U_{S3}. They are deformed in shape or varied in size to each other. In the case of this layer, it is not necessary to present all of the deformed patterns which should be detected by the cell plane. Presentation of only a few number of typical patterns is enough for the training of each cell plane, because a considerable amount of deformation has already been absorbed before this stage.

F. Layers U_{S4} and U_{C4}

Layer U_{S4} has 16 cell planes, and each cell plane has 3×3 S cells. Each of these S cells has $5 \times 5 \times 30$ modifiable input synapses. Although the number of cells in each cell plane is reduced in layer U_{S4} from that in the preceding layer U_{C3}, no thinning out is made between these layers. Only the cells near the periphery of the cell planes of layer U_{S4} are omitted, because they are of little use for the recognition of the whole input pattern (see Fig. 3).

The 16 cell planes of layer U_{S4} are trained with the 16 sets of patterns as shown in Fig. 9 and are joined together into ten cell planes at layer U_{C4}. Each cell plane of layer U_{C4} has only one cell which has input synapses of 3×3 spatial distribution.

The ten cells of layer U_{C4} have one-to-one correspondence with ten Arabic numerals. In Figs. 10–12, which will be discussed later, they are arranged vertically from zero to nine in order at the rightmost column. Among these cells, a mechanism of lateral inhibition exists, although it is omitted in (4).

For some of the numerals, more than one quite different styles of writing are accustomed to be used. For each of these numerals, two S planes are prepared and are trained independently with two typical patterns of different styles

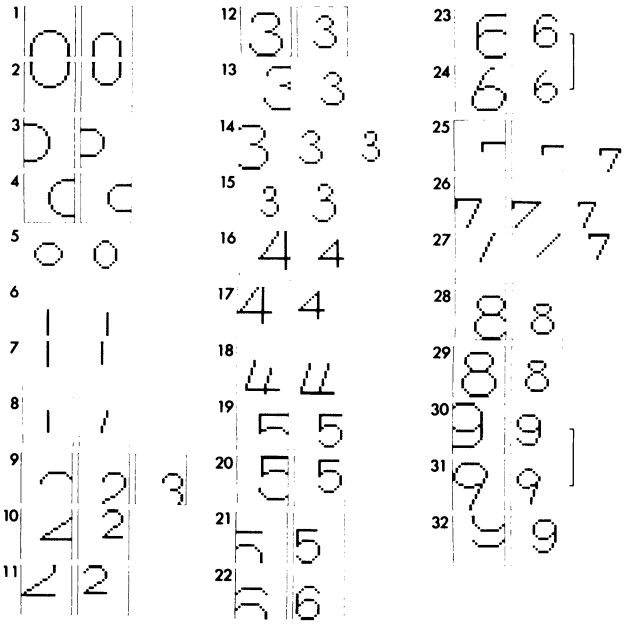

Fig. 8. Training patterns used to train 32 cell planes of layer U_{S3} and way of joining at input stage of layer U_{C3}.

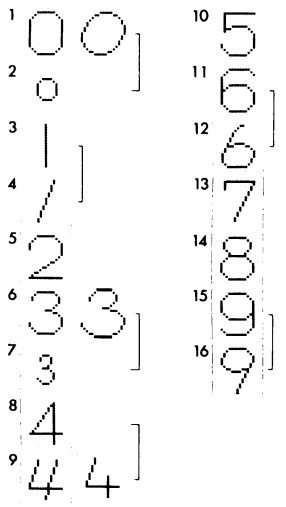

Fig. 9. Training patterns used to train 16 cell planes of layer U_{S4} and way of joining at input stage of layer U_{C4}.

as shown in Fig. 9, and their outputs are joined together at the input stage of layer U_{C4}.

IV. Response of the Network

The neocognitron, which has been trained with the procedure discussed in the previous chapter, is tested with various input patterns. Fig. 10 shows the response of the cells in the network to one of the patterns used for training the network. It is seen that only cell 2 of layer U_{C4} yields an output. This means that the neocognitron recognizes the input pattern correctly. Even if the input pattern is deformed from the training pattern as much as shown in Fig. 11, the neocognitron recognizes it correctly.

In the case of Fig. 12, two of the cells of layer U_{C4} respond; that is, a large output is obtained from cell 5 and

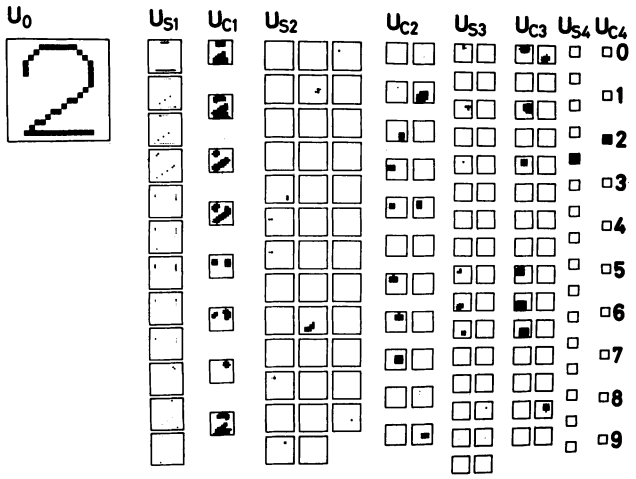

Fig. 10. Response of cells in network to one of training patterns "2."

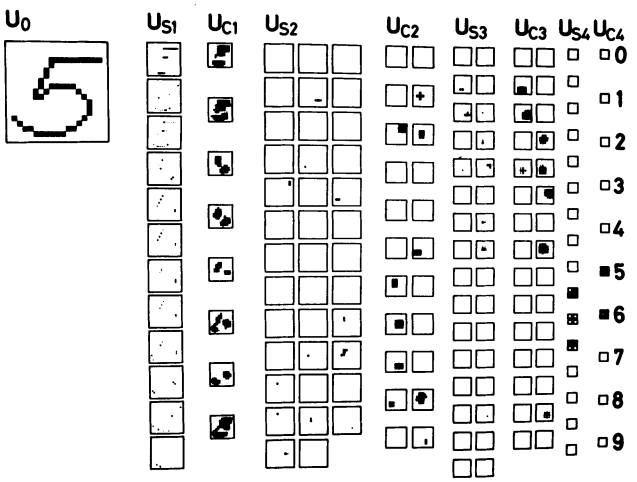

Fig. 12. Response of cells in network to deformed pattern, which elicits response from two cells of layer U_{C4}.

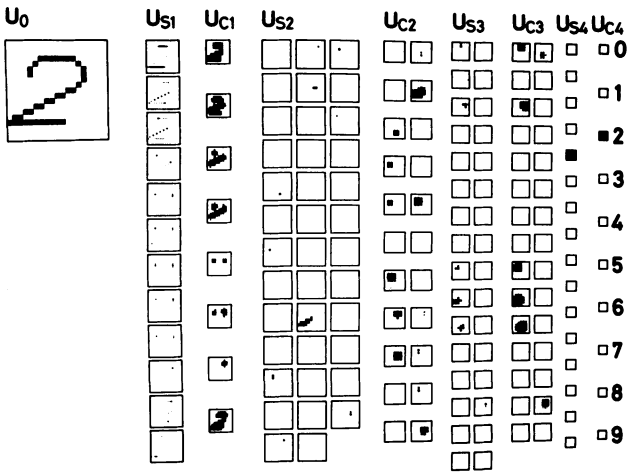

Fig. 11. Response of cells in network to deformed pattern.

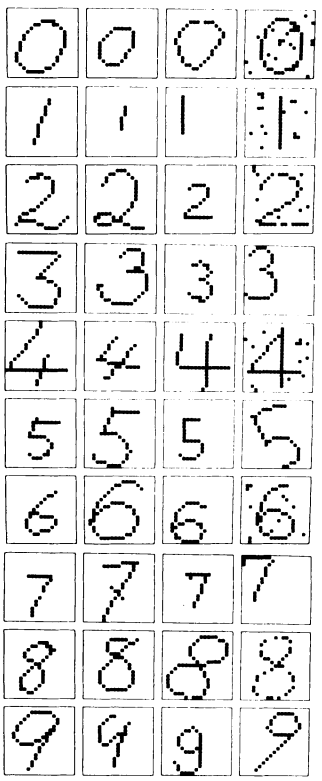

Fig. 13. Some examples of deformed numerals which neocognitron recognizes correctly.

a small output from cell 6. This means that the neocognitron correctly judges that the input pattern is 5, but also admits that the input pattern slightly resembles 6.

Fig. 13 shows some examples of the stimulus patterns which the neocognitron correctly recognizes. On the other hand, Fig. 14 shows some examples of the patterns which cannot be correctly recognized. Some of these patterns elicit no response from any of the cells of layer U_{C4}, and the others elicit responses from wrong cells of layer U_{C4}.

V. Discussion

We have demonstrated that the neocognitron recognizes handwritten numerals of various styles of penmanship correctly, even if they are considerably distorted in shape. Although the result is shown for the recognition of Arabic numerals, the neocognitron can be trained to recognize other set of patterns such as alphabet, geometrical shapes, or others.

The number of cell planes of each layer should be changed adaptively, depending on the set of patterns which the neocognitron should learn to recognize. The program for the computer simulation is made in such a way that the number of cell planes can be chosen freely and can readily be increased when necessary.

Although each S cell has a large number of modifiable input synapses, all of them are not generally reinforced by learning. On the contrary, most of them remain at the initial state in which their efficiencies are zero. Furthermore, the modifiable synapses tend to be reinforced in clusters. In the computer program, we made full use of these characteristics of the synapses and reduced the re-

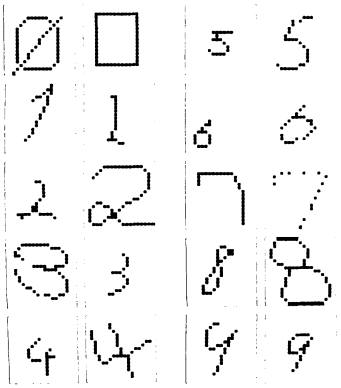

Fig. 14. Some examples of distorted patterns which are not correctly recognized.

quired memory capacity and increased the computation speed by eliminating unnecessary calculations.

In the simulated model, we made two to one thinning-out in several parts of the network in order to increase the computation speed. The thinning-out between layers U_{S1} and U_{C1}, however, was too coarse compared to the 5×5 spread of the input synapses of the cells of layer U_{C1}. As a result, we felt a little difficulty in training the network, and we had to use four different training patterns for each cell plane of layer U_{S2}. If we do not make the thinning-out at this stage, we can possibly improve the capability of the network further.

Acknowledgment

The authors are very grateful to Mr. Toshinori Hirano for his assistance in making the computer program.

References

[1] F. Rosenblatt, *Principles of Neurodynamics*. Washington, DC: Spartan, 1962.
[2] H. D. Block, B. W. Knight, and F. Rosenblatt, "Analysis of a four-layer series-coupled perceptron. II," *Rev. Mod. Phys.*, vol. 34, pp. 135–152, Jan. 1962.
[3] H. Marko and H. Giebel, "Recognition of handwritten characters with a system of homogeneous layers," *Nachrichtentechnische Zeitschrift*, vol. 23, pp. 455–459, Sept. 1970.
[4] H. Marko, "A biological approach to pattern recognition," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-4, pp. 34–39, Jan. 1974.
[5] K. Fukushima, "Cognitron: A self-organizing mulitlayered neural network," *Biol. Cybern.*, vol. 20, pp. 121–136, Nov. 1975.
[6] ——, "Cognitron: A self-organizing multilayered neural network model," NHK Tech. Monograph, no. 30, Jan. 1981.
[7] ——, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.*, vol. 36, pp. 193–202, Apr. 1980.
[8] K. Fukushima and S. Miyake, "Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position," *Pattern Recognition*, vol. 15, pp. 455–469, 1982.
[9] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in cat's visual cortex," *J. Physiol.* (London), vol. 160, pp. 106–154, Jan. 1962.
[10] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture in two nonstriate visual area (18 and 19) of the cat," *J. Neurophysiol.*, vol. 28, pp. 229–289, 1965.

Neuronlike Adaptive Elements That Can Solve Difficult Learning Control Problems

ANDREW G. BARTO, MEMBER, IEEE, RICHARD S. SUTTON, AND CHARLES W. ANDERSON

***Abstract*—It is shown how a system consisting of two neuronlike adaptive elements can solve a difficult leaning control problem. The task is to balance a pole that is hinged to a movable cart by applying forces to the cart's base. It is assumed that the equations of motion of the cart–pole system are not known and that the only feedback evaluating performance is a failure signal that occurs when the pole falls past a certain angle from the vertical, or the cart reaches an end of a track. This evaluative feedback is of much lower quality than is required by standard adaptive control techniques. It is argued that the learning problems faced by adaptive elements that are components of adaptive networks are at least as difficult as this version of the pole-balancing problem. The learning system consists of a single *associative search element* (ASE) and a single *adaptive critic element* (ACE). In the course of learning to balance the pole, the ASE constructs associations between input and output by searching under the influence of reinforcement feedback, and the ACE constructs a more informative evaluation function than reinforcement feedback alone can provide. The differences between this approach and other attempts to solve problems using neuronlike elements are discussed, as is the relation of this work to classical and instrumental conditioning in animal learning studies and its possible implications for research in the neurosciences.**

Manuscript received August 1, 1982; revised April 20, 1983. This work was supported by AFOSR and the Air Force Wright Aeronautical Laboratory under Contract F33615-80-C-1088.

The authors are with the Department of Computer and Information Science, University of Massachusetts, Amherst, MA 01003.

0018-9472/83/0900-0834$01.00 ©1983 IEEE

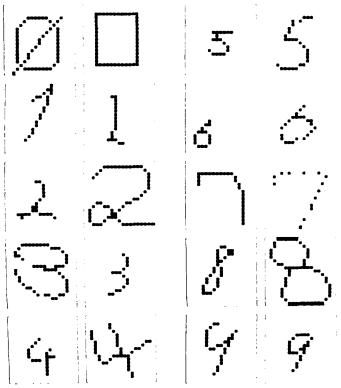

Fig. 14. Some examples of distorted patterns which are not correctly recognized.

quired memory capacity and increased the computation speed by eliminating unnecessary calculations.

In the simulated model, we made two to one thinning-out in several parts of the network in order to increase the computation speed. The thinning-out between layers U_{S1} and U_{C1}, however, was too coarse compared to the 5×5 spread of the input synapses of the cells of layer U_{C1}. As a result, we felt a little difficulty in training the network, and we had to use four different training patterns for each cell plane of layer U_{S2}. If we do not make the thinning-out at this stage, we can possibly improve the capability of the network further.

Acknowledgment

The authors are very grateful to Mr. Toshinori Hirano for his assistance in making the computer program.

References

[1] F. Rosenblatt, *Principles of Neurodynamics*. Washington, DC: Spartan, 1962.
[2] H. D. Block, B. W. Knight, and F. Rosenblatt, "Analysis of a four-layer series-coupled perceptron. II," *Rev. Mod. Phys.*, vol. 34, pp. 135–152, Jan. 1962.
[3] H. Marko and H. Giebel, "Recognition of handwritten characters with a system of homogeneous layers," *Nachrichtentechnische Zeitschrift*, vol. 23, pp. 455–459, Sept. 1970.
[4] H. Marko, "A biological approach to pattern recognition," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-4, pp. 34–39, Jan. 1974.
[5] K. Fukushima, "Cognitron: A self-organizing mulitlayered neural network," *Biol. Cybern.*, vol. 20, pp. 121–136, Nov. 1975.
[6] ——, "Cognitron: A self-organizing multilayered neural network model," NHK Tech. Monograph, no. 30, Jan. 1981.
[7] ——, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.*, vol. 36, pp. 193–202, Apr. 1980.
[8] K. Fukushima and S. Miyake, "Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position," *Pattern Recognition*, vol. 15, pp. 455–469, 1982.
[9] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in cat's visual cortex," *J. Physiol.* (London), vol. 160, pp. 106–154, Jan. 1962.
[10] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture in two nonstriate visual area (18 and 19) of the cat," *J. Neurophysiol.*, vol. 28, pp. 229–289, 1965.

Neuronlike Adaptive Elements That Can Solve Difficult Learning Control Problems

ANDREW G. BARTO, MEMBER, IEEE, RICHARD S. SUTTON, AND CHARLES W. ANDERSON

***Abstract*—It is shown how a system consisting of two neuronlike adaptive elements can solve a difficult leaning control problem. The task is to balance a pole that is hinged to a movable cart by applying forces to the cart's base. It is assumed that the equations of motion of the cart–pole system are not known and that the only feedback evaluating performance is a failure signal that occurs when the pole falls past a certain angle from the vertical, or the cart reaches an end of a track. This evaluative feedback is of much lower quality than is required by standard adaptive control techniques. It is argued that the learning problems faced by adaptive elements that are components of adaptive networks are at least as difficult as this version of the pole-balancing problem. The learning system consists of a single *associative search element* (ASE) and a single *adaptive critic element* (ACE). In the course of learning to balance the pole, the ASE constructs associations between input and output by searching under the influence of reinforcement feedback, and the ACE constructs a more informative evaluation function than reinforcement feedback alone can provide. The differences between this approach and other attempts to solve problems using neuronlike elements are discussed, as is the relation of this work to classical and instrumental conditioning in animal learning studies and its possible implications for research in the neurosciences.**

Manuscript received August 1, 1982; revised April 20, 1983. This work was supported by AFOSR and the Air Force Wright Aeronautical Laboratory under Contract F33615-80-C-1088.

The authors are with the Department of Computer and Information Science, University of Massachusetts, Amherst, MA 01003.

0018-9472/83/0900-0834$01.00 ©1983 IEEE