Pattern Recognition Vol 15, No 6, pp 455–469, 1982
Printed in Great Britain

0031-3203/82/060455-15 $03.00/0
Pergamon Press Ltd

NEOCOGNITRON: A NEW ALGORITHM FOR PATTERN RECOGNITION TOLERANT OF DEFORMATIONS AND SHIFTS IN POSITION

KUNIHIKO FUKUSHIMA and SEI MIYAKE
NHK Broadcasting Science Research Laboratories, 1-10-11, Kinuta, Setagaya, Tokyo 157, Japan

(*Received* 15 *May* 1981, *in revised form* 27 *October* 1981, *received for publication* 23 *December* 1981)

Abstract—Suggested by the structure of the visual nervous system, a new algorithm is proposed for pattern recognition. This algorithm can be realized with a multilayered network consisting of neuron-like cells. The network, "neocognitron", is self-organized by unsupervised learning, and acquires the ability to recognize stimulus patterns according to the differences in their shapes. Any patterns which we human beings judge to be alike are also judged to be of the same category by the neocognitron. The neocognitron recognizes stimulus patterns correctly without being affected by shifts in position or even by considerable distortions in shape of the stimulus patterns.

Visual pattern recognition
Unsupervised learning
Neural network model
Deformation-resistant
Self-organization
Visual nervous system
Position-invariant
Multilayered network
Simulation

1. INTRODUCTION

Most of the methods for pattern recognition, especially template matching techniques, are oversensitive to shifts in position and distortions in shape of the stimulus patterns, and it is necessary to normalize the position and the shape of the stimulus pattern beforehand. A good method for normalization, however, has not been developed as yet. Therefore, the finding of an algorithm for pattern recognition which can cope with shifts in position and distortions in shape of the stimulus patterns has long been desired.

In this paper, we propose a new algorithm* which gives an important solution to this problem. The new algorithm proposed here can be realized with a multilayered network consisting of neuron-like cells.

The network is called a "neocognitron". It is self-organized by unsupervised learning and acquires the ability for correct pattern recognition.

Historically, the three-layered perceptron proposed by Rosenblatt[4] is a famous example of networks of neuron-like cells capable of pattern recognition. For some time after the perceptron was first proposed, great hope was felt for its capability for pattern recognition and a great deal of research was done on it. With the progress of this research, however, it was gradually revealed that the capability of the perceptron was not so great as had been expected.[5]

The perceptron consists of only three layers of cells, but it is well known that the capability of a multilayered network can be greatly enlarged if the number of layers is increased. The algorithm of self-organization employed in the perceptron, however, cannot be successfully applied for the self-organization of a multilayered network. The three-layered perceptron is trained by supervised learning, a "learning-with-a-teacher" process, and only the deepest-layer cells have their input interconnections reinforced on instructions from a "teacher". The "teacher" knows the category to which each of the training patterns should be classified, and teaches each of the deepest-layer cells what its desired output is. If we want to train a multilayered network, in which not only the deepest-layer cells but also the intermediate-layer cells should have their input interconnections reinforced, the conventional training method used for the perceptron is not suited. This is because we do not know how the "teacher" should give instructions to the intermediate-layer cells, whose desired responses cannot simply be determined only from the information about the category to which the stimulus pattern should be classified.

One of the authors, Fukushima, formerly proposed a new algorithm for self-organization which can be effectively applied even to a multilayered network. This algorithm is based on the principle that only maximum-output cells have their input intercon-

* The basic idea of the new algorithm proposed here was first reported by one of the authors as a neural network model for a mechanism of visual pattern recognition in the brain.[1,2] In this paper, we discuss the problem from an engineering point of view, and concentrate our discussion on the application of the algorithm in pattern recognition and learning. Part of this paper was reported at the 5th ICPR[3] by the authors, and the results of a computer simulation were presented as a movie.

nections reinforced, and no instructions from a "teacher" are necessary The "cognitron"[6 7] is a multi-layered network, in which this algorithm is employed The self-organization of the cognitron is performed by unsupervised learning, a "learning-without-a-teacher" process A computer simulation demonstrated that the cognitron had a much higher capability for pattern recognition than the three-layered perceptron

However, the cognitron, as with the three-layered perceptron, does not have the capability to correctly recognize position-shifted or shape-distorted patterns the same pattern presented at a different position is usually recognized as a different pattern by the conventional cognitron

The "neocognitron", which will be discussed in this paper, is an improved version of the conventional cognitron and has the capability to recognize stimulus patterns correctly, even if the patterns are shifted in position or distorted in shape

The self-organization of the neocognitron is also performed by unsupervised learning only repetitive presentation of a set of stimulus patterns is necessary for the self-organization of the neocognitron and no information about the categories to which these patterns should be classified is needed The neocognitron acquires the ability to classify and correctly recognize these patterns by itself, according to the differences in their shapes any pattern which we human beings judge to be alike are also judged to be of the same category by the neocognitron The neocognitron recognizes stimulus patterns correctly without being affected by shifts in position or even by considerable distortions in shape of the stimulus patterns

The neocognitron has a hierarchical structure The information of the stimulus pattern given to the input layer of the neocognitron is processed step by step in each stage of the multilayered network a cell in a deeper stage 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 in the deepest stage responds only to a specific stimulus pattern without being affected by the position or the size of the stimulus patterns

2 CELLS EMPLOYED IN THE NEOCOGNITRON

Before discussing the structure of the neocognitron, let us discuss the characteristics of the cells employed in it

All the cells employed in the neocognitron are of analog type i e, the input and output signals of the cells take non-negative analog values Each cell has characteristics analogous to a biological neuron, if we consider that the output signal of the cell corresponds to the instantaneous firing frequency of the actual biological neuron

In the neocognitron, we use four different kinds of cells, i e, S-cells, C-cells, V_S-cells and V_C-cells As a typical example of these cells, we will first discuss the characteristics of an S-cell

As shown in Fig 1, an S-cell has a lot of input terminals, either excitatory of inhibitory If the cell receives signals from excitatory input terminals, the output of the cell will increase On the other hand, a signal from an inhibitory input terminal will suppress the output Each input terminal has its own interconnecting coefficient whose value is positive Although the cell has only one output terminal, it can send signals to a number of input terminals of other cells

An S-cell has an inhibitory input which causes a shunting effect[6 7] Let $u(1)$, $u(2)$, , $u(N)$ be the excitatory inputs and v be the inhibitory input The output w of this S-cell is defined by

$$w = \varphi\left[\frac{1 + \sum_{\nu=1}^{N} a(\nu)\ u(\nu)}{1 + b\ v} - 1\right], \tag{1}$$

where $a(\nu)$ and b represent the excitatory and inhibitory interconnecting coefficients, respectively The function $\varphi[\ ]$ is defined by the following equation

$$\varphi[x] = \begin{cases} x & (x \geqq 0) \\ 0 & (x < 0) \end{cases}, \tag{2}$$

Let us further discuss the characteristics of this cell Let e be the sum of all the excitatory inputs weighted with the interconnecting coefficients and h the inhibitory input multiplied by the interconnecting coefficient, i e,

$$e = \sum_{\nu=1}^{N} a(\nu)\ u(\nu), \tag{3}$$

$$h = b\ v \tag{4}$$

Equation (1) can now be written as

$$w = \varphi\left[\frac{1+e}{1+h} - 1\right] = \varphi\left[\frac{e-h}{1+h}\right] \tag{5}$$

When the inhibitory input is small ($h \ll 1$), we have $w = \varphi[e - h]$, which coincides with the characteristics

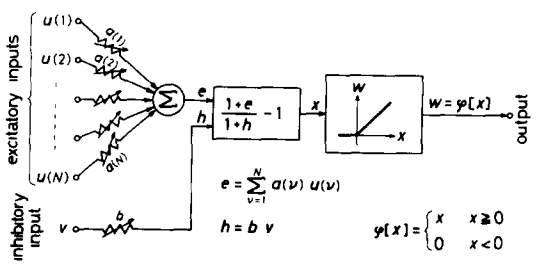

Fig 1 Input-to-output characteristics of an S-cell A typical, example of the cells employed in the neocognitron

* The "receptive field" of a cell is defined as an area on the input layer such that a stimulus presented there gives some effect on the response of the cell In other words, it is the area from which some information is transmitted, either directly or indirectly, to the cell

of the conventional analog-threshold-element [8] For a system like the neocognitron, where the interconnecting coefficients $a(v)$ and b increase unboundedly as learning progresses, the employment of elements such as analog-threshold-elements is not suitable because their outputs may increase without bound In the cell employed here, however, if the interconnecting coefficients increase and we have $e \gg 1$ and $h \gg 1$, equation (5) reduces approximately to $w = \varphi[e/h - 1]$, where the output w is determined by the ratio e/h and not by the difference $e - h$ Therefore, even if the interconnecting coefficients increase with learning, the output of the cell approaches a certain value without divergence, so long as both the excitatory interconnecting coefficients $a(v)$ and the inhibitory interconnecting coefficient b increase at the same rate

Let us observe the input-to-output relation of the cell in the case where the excitatory and inhibitory inputs vary in proportion If we write

$$e = \varepsilon x, \quad h = \eta x,$$

and if $\varepsilon > \eta$ holds, equation (5) can be transformed as

$$w = (\varepsilon - \eta)\frac{x}{1 + \eta x} = \frac{\varepsilon - \eta}{2\eta}\left\{1 + \tanh\left(\frac{1}{2}\log \eta x\right)\right\} \quad (6)$$

This input-to-output relation coincides with the logarithmic relation expressed by Weber-Fechner's low on which an S-shaped saturation, expressed by tanh is superposed The same expression is often used as an empirical formula in neurophysiology and psychology to approximate the nonlinear input-to-output relations of the sensory systems of animals

The cells other than S-cells also have characteristics similar to those of S-cells The input-to-output characteristics of a C-cell are obtained from equation (5) if we replace $\varphi[\ ]$ with $\psi[\ ]$, where $\psi[\ ]$ is a saturation function defined by

$$\psi[x] = \begin{cases} \dfrac{x}{\alpha + x} & (x \geqq 0) \\ 0 & (x < 0) \end{cases} \quad (7)$$

The parameter α is a positive constant which determines the degree of saturation of the output In the computer simulation discussed in Section 6, we chose $\alpha = 0.5$

S-cells and C-cells are excitatory cells, i e , the output terminals of these cells are connected only to excitatory input terminals of other cells

On the other hand, V_S-cells and V_C-cells are inhibitory cells, whose output terminals are connected only to inhibitory input terminals of other cells A V_S-cell has only excitatory input terminals and the output of the cell is proportional to the sum of all the inputs weighted with the interconnecting coefficients That is, a V_S-cell yields an output proportional to the (weighted) arithmetic mean of its inputs

A V_C-cell also has only excitatory input terminals, but its output is proportional to the (weighted) root-mean-square of its input Let $u(1), u(2), \ldots, u(N)$ be the inputs to a V_C-cell and $c(1), c(2), \ldots, c(N)$ be the interconnecting coefficients of its input terminals The output w of this V_C-cell is defined by

$$w = \sqrt{\sum_{v=1}^{N} c(v)\ u^2(v)} \quad (8)$$

3. STRUCTURE OF THE NETWORK

As shown in Fig 2, the neocognitron consists of a cascade connection of a number of modular structures preceded by an input layer U_0 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* In the neocognitron, only the input interconnections to S-cells are variable and modifiable and the input interconnections to other cells are fixed and unmodifiable

The input layer U_0 consists of a photoreceptor array The output of a photoreceptor is denoted by $u_0(\mathbf{n})$ where $\mathbf{n}$ is the two-dimensional co-ordinates indicating the location of the cell

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 a "cell-plane" We will also use the terminology S-plane and C-plane, representing cell-planes consisting of S-cells and C-cells, respectively

It is determined that all the cells in a single cell-plane have input interconnections of the same spatial distribution and only the positions of the preceding cells from which their input interconnections come are shifted in parallel This situation is illustrated in Fig 3 Even in the process of learning, in which the values of the input interconnections of S-cells are varied, the variable interconnections are always modified under this restriction

We will use the notation $u_{Sl}(k_l, \mathbf{n})$ to represent the output of an S-cell in the k_l-th S-plane in the l-th module, and $u_{Cl}(k_l, \mathbf{n})$ to represent the output of a C-cell in the k_l-th C-plane in that module, where $\mathbf{n}$ is the two-dimensional co-ordinates representing the position of these cells' receptive fields on the input layer

Figure 4 is a schematic diagram illustrating the interconnections 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-

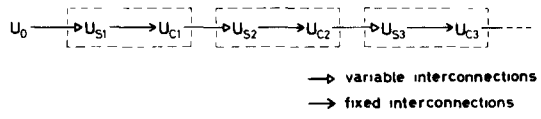

Fig 2 The hierarchical structure of the neocognitron

* S-cells and C-cells are named after simple cells and complex cells in physiological terms, respectively

planes or C-planes are enclosed, represents an S-layer or a C-layer

In Fig 4, for the sake of simplicity, only one cell is shown in each cell-plane Each of these cells receives input interconnections 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 interconnections of the same spatial distribution and only the positions of the preceding cells, to which their input terminals are connected, 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 4, 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 in the size of the receptive fields Hence, the total number of the cells in each cell-plane decreases with the depth of the cell-plane in the network 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

As mentioned in Section 2, the S-cells and C-cells are excitatory cells Although it is not shown in Fig 4, we also have inhibitory cells, namely, V_S-cells in S-layers and V_C-cells in C-layers

Here we will describe the outputs of the cells in the network with numerical expressions

As was discussed in Section 2, S-cells have inhibitory inputs with a shunting mechanism The output of an S-cell of the k_l-th S-plane in the l-th module is given by

$$u_{Sl}(k_l, \mathbf{n}) = r_l \ \varphi\left[\frac{1 + \sum_{k_{l-1}=1}^{K_{l-1}} \sum_{v \in S_l} a_l(k_{l-1}, v, k_l)\ u_{Cl-1}(k_{l-1}, \mathbf{n}+v)}{1 + \frac{r_l}{1+r_l}\ b_l(k_l)\ v_{Cl-1}(\mathbf{n})} - 1\right] \quad (9)^*$$

where $\varphi[\ ]$ is a function defined by equation (2) before In the case of $l = 1$ in equation (9), $u_{Cl-1}(k_{l-1}, \mathbf{n})$ stands for $u_0(\mathbf{n})$ and we have $K_{l-1} = 1$

Here, $a_l(k_{l-1}, v, k_l)$ and $b_l(k_l)$ represent the values of the excitatory and inhibitory variable interconnecting coefficients, respectively As described before, all the S-cells in the same S-plane have an identical set of input interconnections Hence, $a_l(k_{l-1}, v, k_l)$ and $b_l(k_l)$ do not contain any argument representing the position $\mathbf{n}$ of the receptive field of the cell $u_{Sl}(k_l, \mathbf{n})$

Parameter r_l in equation (9) 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 A more detailed discussion on the response of S-cells will be given in Section 4 2

The inhibitory cell $v_{Cl-1}(\mathbf{n})$, which is sending an inhibitory signal to cell $u_{Sl}(k_l, \mathbf{n})$, receives its input interconnections from the same cells as $u_{Sl}(k_l, \mathbf{n})$ does, and yields an output proportional to the weighted root-mean-square of its inputs

$$v_{Cl-1}(\mathbf{n}) = \sqrt{\sum_{k_{l-1}=1}^{K_{l-1}} \sum_{v \in S_l} c_{l-1}(v)\ u_{Cl-1}^2(k_{l-1}, \mathbf{n}+v)} \quad (10)$$

The values of fixed interconnections $c_{l-1}(v)$ are determined so as to decrease monotonically with respect to $|v|$ and to satisfy

$$\sum_{k_{l-1}=1}^{K_{l-1}} \sum_{v \in S_l} c_{l-1}(v) = 1 \quad (11)$$

The size of the connecting area S_l of these cells is set to be small in the first module and to increase with the depth l

The interconnections from S-cells to C-cells are fixed and unmodifiable As illustrated in Fig 4, each C-cell has input interconnections leading from a group of S-cells in the S-plane preceding it (i e , in the S-plane with the same k_l-number as that of the C-cell) This means that all of the S-cells in the C-cell's connecting area extract the same stimulus features but from slightly different positions on the input layer The values of the interconnections are determined in such a way that the C-cell will be activated whenever at least one of these 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 still keep responding as before, because another neighboring S-cell in its connecting area will become active instead of the first In other words, a C-cell responds to the same stimulus feature as the S-cells preceding it, but is less sensitive to a shift in position of the stimulus feature

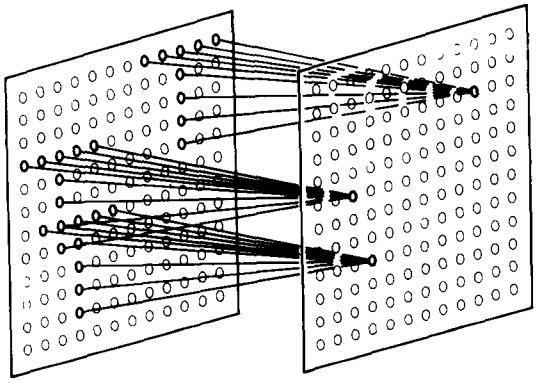

Fig 3 Illustration showing the input interconnections to the cells of an arbitrary cell-plane

*The notation $b_l(k_l)$ in this paper corresponds to $2b_l(k_l)$ in the previous papers by Fukushima [1-3]

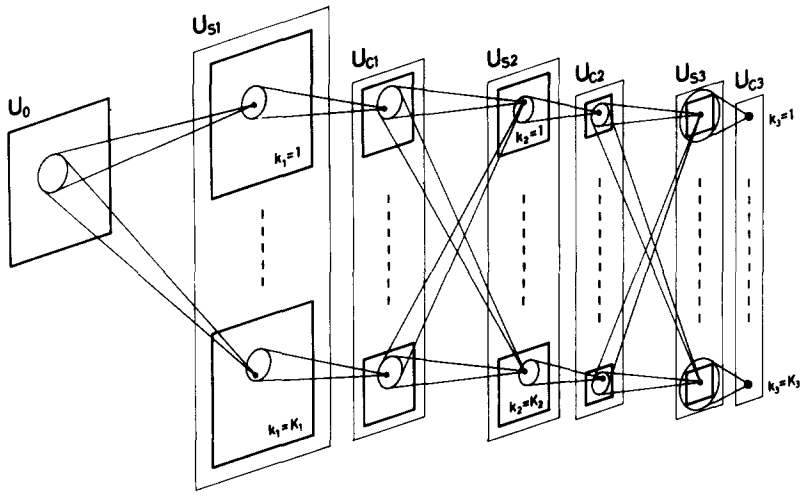

Fig 4 Schematic diagram illustrating the interconnections between layers in the neocognitron

Quantitatively, the output of a C-cell of the k_l-th C-plane in the l-th module is given by

$$u_{Cl}(k_l, \mathbf{n}) = \psi\left[\frac{1 + \sum_{v \in D_l} d_l(v)\ u_{Sl}(k_l, \mathbf{n} + v)}{1 + v_{Sl}(\mathbf{n})} - 1\right], \tag{12}$$

where $\psi[\ ]$ is a function defined by equation (7)

The inhibitory cell $v_{Sl}(\mathbf{n})$, which sends inhibitory signals to this C-cell and makes up the system of lateral inhibition, yields an output proportional to the (weighted) arithmetic mean of its inputs

$$v_{Sl}(\mathbf{n}) = \frac{1}{K_l} \sum_{k_l=1}^{K_l} \sum_{v \in D_l} d_l(v)\ u_{Sl}(k_l, \mathbf{n} + v) \tag{13}$$

In equations (12) and (13), the values of fixed interconnections $d_l(v)$ are determined so as to decrease monotonically with respect to $|v|$, as with $c_l(v)$ The size of the connecting area D_l is set to be small in the first module and to increase with the depth l

4 SELF-ORGANIZATION OF THE NETWORK

4 1 *Reinforcement of variable interconnections*

The self-organization of the neocognitron is performed by means of unsupervised learning, "learning without a teacher" During the process of self-organization, the network is repeatedly presented with a set of stimulus patterns to the input layer, but it does not receive any other information about the category of the stimulus patterns

As was discussed in Section 3, all the S-cells in a single S-plane should always have input interconnections of the same spatial distribution but from different positions In order to modify the variable interconnections always keeping this restriction, we use the following procedure

At first, several "representative" S-cells are chosen from each S-layer every time that a stimulus pattern is presented The representatives are chosen from among those S-cells which have yielded large outputs, but the number of the representatives is so restricted that more than one representative should not be chosen from any single S-plane The detailed procedure for choosing the representatives is given in Section 4 2

For a representative S-cell, only the input interconnections through which non-zero signals are coming are reinforced With this procedure, the representative S-cell becomes selectively responsive only to the stimulus feature which is now presented A detailed discussion on S-cell response will appear in Section 4 3 All the other S-cells in the S-plane, from which the representative is chosen, have their input interconnections reinforced by the same amounts as those for their representative These relations can be quantitatively expressed as follows

Let cell $u_{Sl}(\hat{k}_l, \hat{\mathbf{n}})$ be chosen as a representative The variable interconnections $a_l(k_{l-1}, v, \hat{k}_l)$ and $b_l(\hat{k}_l)$ which are incoming to the S-cells of this S-plane, are reinforced by the amount shown below

$$\Delta a_l(k_{l-1}, v, \hat{k}_l) = q_l\ c_{l-1}(v)\ u_{Cl-1}(k_{l-1}, \hat{\mathbf{n}} + v), \tag{14}$$

$$\Delta b_l(\hat{k}_l) = q_l \cdot v_{Cl-1}(\hat{\mathbf{n}}), \tag{15}$$

where q_l is a positive constant which determines the speed of increment

The cells in the S-plane from which no repre-

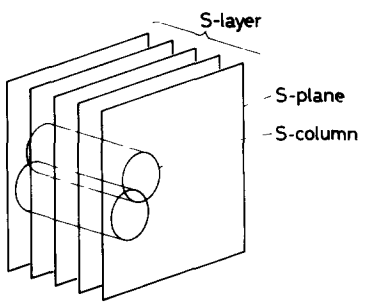

Fig 5 Relation between S-planes and S-columns within an S-layer

sentative is chosen, however, do not have their input interconnections varied at all

The choice of initial values of the variable interconnections has little effect on the performance of the neocognitron, provided that they are small and are determined in such a way that each S-plane has a different set of initial values for its input interconnections In the computer simulation discussed in Section 6, the initial values of the excitatory variable interconnections $a_l(k_{l-1}, v, k_l)$ are set to be small positive values in such a way that each S-cell has a very weak orientation slectivity and that its preferred orientation differs from S-plane to S-plane That is, the initial values of these variable interconnections are given by a function of v, (k_l/K_l) and $|k_{l-1}/K_{l-1} - k_l/K_l|$, but they do not have any randomness The initial values of inhibitory variable interconnections $b_l(k_l)$ are set to be zero

4 2 *Choosing the representatives*

The procedure for choosing the representatives is as follows First, in an S-layer we pick up a group of S-cells whose receptive fields are situated within a small area on the input layer If we arrange the S-planes of an S-layer in the manner shown in Fig 5, such a group of S-cells constitutes a column in an S-layer Accordingly, we call the group an "S-column" An S-column contains S-cells from all the S-planes, i e an S-column contains various kinds of feature-extracting cells in it, but the receptive fields of these cells are situated almost at the same position There are a lot of such S-columns in a single S-layer Since S-columns overlap with each other, there is a possibility that a single S-cell is contained in two or more S-columns

From each S-column, every time that a stimulus pattern is presented, the S-cell which is yielding the largest output is chosen as a candidate for the representative Hence, there is a possibility that a number of candidates appear in a single S-plane If two or more candidates appear in a single S-plane, only the one which is yielding the largest output among them is chosen as the representative from that S-plane In case only one candidate appears in an S-plane, the candidate is unconditionally determined as the representative from that S-plane If no candidate appears in an S-plane, no representative is chosen from that S-plane

Since the representatives are determined in this manner, each S-plane becomes selectively sensitive to one of the features of the stimulus patterns and there is no possibility of the formation of redundant connections such that two or more S-planes are used for the detection of one and the same feature Incidentally, representatives are chosen from only a small number of S-planes at a time, the rest of the S-planes producing representatives when other stimulus patterns are presented

4 3 *Response of an S-cell*

In this section, we discuss how each S-cell comes 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 $u_{Sl}(k_1, \mathbf{n})$ as a typical example Figure 6 shows the interconnections converging on such a cell For the sake of simplicity, we will omit the suffixes S and $l = 1$ and the arguments k_l and $\mathbf{n}$ and represent the response of this cell simply by u. Similarly, we will use the notation v for the output of the inhibitory cell $v_{C0}(\mathbf{n})$, which sends inhibitory signals to cell u For the other variables, the arguments k_l and $\mathbf{n}$ and suffixes S, C, l and $l-1$ will also be omitted

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

$$p(v) = u_0(\mathbf{n} + v) \tag{16}$$

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

With this notation equations (9) and (10) can be written

$$u = r\ \varphi\left[\frac{1 + \sum_v a(v)\ p(v)}{1 + \frac{r}{1+r}\ b\ v} - 1\right], \tag{17}$$

$$v = \sqrt{\sum_v c(v)\ p^2(v)} \tag{18}$$

When cell u is chosen as a representative, the increments of the variable interconnections are derived from equations (14) and (15), i e.,

$$\Delta a(v) = q\ c(v)\ p(v), \tag{19}$$

$$\Delta b = q\ v \tag{20}$$

Let s be defined by

$$s = \frac{\sum_v a(v)\ p(v)}{b\ v} \tag{21}$$

Then equation (17) reduces, approximately, to

$$u = r\ \varphi\left[\frac{r+1}{r}s - 1\right], \tag{22}$$

provided that $a(v)$ and b are sufficiently large

Let us suppose that cell u has been chosen N-times as a representative for the same stimulus pattern $p(v)$

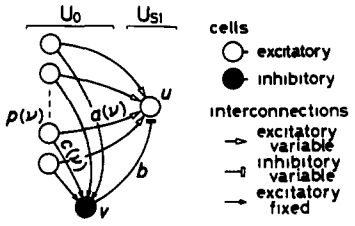

Fig 6 Interconnections converging to an S-cell

* The analysis for the conventional r m s -cognitron[7] can be applied to our neocognitron with a small modification

$= P(v)$ We also suppose that no other cell has been chosen as a representative from the S-plane in which cell u is located Assuming that the initial values of the variable interconnections are small enough to be neglected, we obtain

$$a(v) = Nq \quad c(v) \quad P(v) \tag{23}$$

$$b = Nq \sqrt{\sum_v c(v) \quad P^2(v)} \tag{24}$$

Substituting equations (18), (23) and (24) into equation (21), we obtain

$$s = \frac{\sum_v c(v) \quad P(v) \quad p(v)}{\sqrt{\sum_v c(v) \quad P^2(v)} \quad \sqrt{\sum_v c(v) \quad p^2(v)}}. \tag{25}$$

If we regard $p(v)$ and $P(v)$ as vectors, equation (25) 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 $p(v)$ and $P(v)$ in the multidimensional vector space Therefore, we have $s = 1$ only when $p(v) = P(v)$ and we have $s < 1$ for all patterns such that $p(v) \neq P(v)$ This means that s becomes maximum for the learned pattern and becomes smaller for any other pattern

Consider the case when N is so large that $Nq \gg 1$ In this case, equation (22) holds

When an arbitrary pattern $p(v)$ is presented, if it satisfies $s > r/(r + 1)$, we have $u > 0$ by equation (22) Conversely, for a pattern which makes $s \leqq r/(r + 1)$, cell u does not respond We can interpret this by saying that cell u judges the similarity between patterns $p(v)$ and $P(v)$ using the criterion defined in equation (25), and that it responds only to patterns judged to be similar to $P(v)$ Incidentally, if $p(v) = P(v)$, we have $s = 1$ and consequently $u = 1$

Roughly speaking, during the self-organization process, patterns $p(v)$ which satisfy $s > r/(r + 1)$ are taken to be of the same class as $P(v)$ and patterns $p(v)$ which make $s \leqq r/(r + 1)$ are assumed to be of different classes from $P(v)$

The patterns which are judged to be of different classes will possibly be extracted by the cells in other S-planes

Since the value $r/(r + 1)$ tends to 1 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 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 Appendix we will give some numerical examples and further explain how the selectivity of the cell is controlled by parameter r.

In the above analysis, we supposed that cell u is trained only for one particular pattern $P(v)$ However, this does not necessarily mean that the training pattern sequence should consist of the repetition of only one pattern $P(v)$ Once the self-organization of the network progresses a little, each cell in the network becomes selectively responsive to one particular pattern Hence, even if patterns other than $P(v)$ appear in the training pattern sequence, they usually do not elicit any response from cell u unless they closely resemble $P(v)$ Hence, they do not have any effect on the reinforcement of the input interconnections of u Thus, among many patterns in the training-pattern sequence, only pattern $P(v)$ effectively works for the training of cell u

In the above discussion we also assumed that the stimulus pattern is always presented at the same position in the input layer and consequently that the representative from an arbitrary S-plane, if any, is always the same cell Actually, however, this restriction is not necessary for our discussion We have the same results even if the position of presentation of the stimulus pattern varies each time We can explain this as follows If the stimulus pattern is presented at a different position, another cell in that S-plane will become the representative, but the position of the new representative relative to the stimulus pattern will usually be kept as before On the other hand, according to the algorithm of self-organization discussed in Section 4 1, all the S-cells in that S-plane have their input interconnections modified in the same manner as their representative of the moment Hence, the shift in position of the stimulus pattern does not affect the result of self-organization of the network

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 their connectable area in the preceding layer as its input pattern

4 4 *The temporal progress of self-organization*

In order to help the understanding of the algorithm of self-organization we will give a simple example and explain how the reinforcement of the network progresses "without-a-teacher"

As is seen from the discussions in the above sections, the fundamental rule for self-organization of the neocognitron, as with that of the conventional cognitron, lies in the principle that only maximum-output cells have their input interconnections reinforced We will first explain this principle using a simplified network

Let us consider the network shown in Fig 7[a] This network consists of only nine photoreceptor cells, one inhibitory cell and three S-cells These S-cells are named A, B and C The structure of the interconnections converging to each of the three S-cells are identical to the network of Fig 6 We will explain how the self-organization progresses if three training patterns, say vertical, horizontal and diagonal lines as shown in Fig 7[b] (i)–(iii), are presented by turns The excitatory variable interconnections have very small positive initial values, while the initial values of the

inhibitory interconnections are zero Hence, if the vertical line shown in Fig 7[b] (i) is first presented to the photoreceptor array, all the S-cells yield very small but non-zero positive outputs Among these three S-cells we choose the one which is yielding the maximum output Let this cell be, say, cell A Cell A has its input interconnections reinforced according to equations (19) and (20), i e , the interconnections from cells p_2, p_5, p_8 and v, through which non-zero input signals are coming to cell A, are reinforced It should be noticed that not only the excitatory interconnections but also the inhibitory interconnection is reinforced If the value of parameter q in equations (19) and (20) is large enough, compared to the initial values of the variable interconnections, cell A acquires, after only this single training, selective responsiveness to this vertical line

Next, let a horizontal line shown in Fig 7[b](ii) be presented Since cell A has already come to respond selectively to a vertical line, it does not respond to the horizontal line, provided the value of parameter r is not too small Only cells B and C yield non-zero output this time Let the maximum-output cell among these be, say, cell C This time, cell C has its input interconnections reinforced according to equations (19) and (20) and acquires selective responsiveness to the horizontal line

Let the next training pattern be the diagonal line shown in Fig 7[b](iii) Since cells A and C have already become selectively responsive to vertical and horizontal lines, respectively, the rest of the cells, i e , cell B, yields a non-zero output to the diagonal line Hence, cell B has its input interconnections reinforced and becomes selectively responsive to the diagonal line

If the next training pattern be, say, a horizontal line again, only cell C, which has once been reinforced to the horizontal line, is activated Hence, cell C has its input interconnections reinforced to the horizontal line again and the selectivity to the horizontal line becomes more stable

In the above example, we assumed that the training patterns were presented in the sequence order, vertical, horizontal, diagonal, We obtain the same result for self-organization, even if the identical training pattern appears successively in the training sequence, say vertical, vertical, horizontal, Let cell A be the maximum output cell and let it be chosen to be reinforced on the first presentation of a vertical line It is a matter of course that cell A again yields the maximum output among the three when the vertical line is again presented as the second training pattern Hence, the same cell A is again reinforced to the vertical line In other words, there is no possibility that two or more cells could be reinforced to the same training pattern, even if the same training pattern is successively presented to the network

If the number of different training patterns is four or more, none of the S-cells yield a positive output to the fourth training pattern, because cells A, B and C have already been reinforced and have acquired selective responsiveness to the first three training patterns

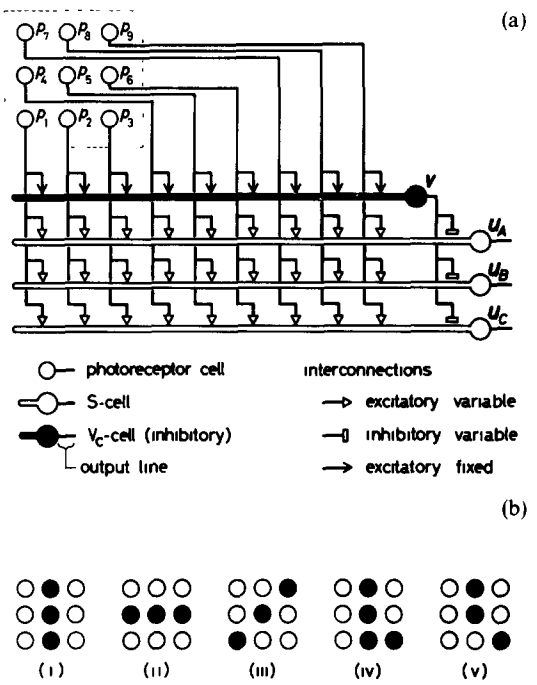

Fig 7 Illustrations for the progress of self-organization (a), Simplified network explaining the algorithm of self-organization (b), Some examples of the stimulus pattern used to train the network of (a)

Consequently, no S-cell can be reinforced to the fourth training pattern in a network which has only three S-cells Hence, in designing a network such as Fig 7[a] we have to prepare a greater number of S-cells than the number of different training patterns to be recognized In designing the whole system of the neocognitron, the number of S-planes in layer U_{S1} should be larger than the number of different stimulus features contained in the stimulus patterns to be recognized This will be further discussed in Section 5 In the computer simulation discussed in Section 6, the number of S-planes in layer U_{S1} is 24

As can be seen from the above examples, once an S-cell has been trained to one particular pattern the cell usually does not respond to any other patterns which are judged to be different from the former one by the criterion defined in Section 4 3 and the Appendix Consequently, the cell usually cannot be reinforced to a new pattern and will preserve the same response characteristics In a special situation, however, it is possible to make the cell come to respond to a completely new pattern This can be performed if we change its response characteristics little by little as shown in the next example

Suppose cell A is first reinforced to pattern (i) in Fig 7[b] It is assumed that the value of parameter r is chosen in such a way that cell A responds also to pattern (iv) but not to pattern (v) (Further discussions on the choice of the value of r appear in the Appendix) Hence, if training pattern (iv) is presented next, cell A yields a small but non-zero output If no other cell has yet been trained to pattern (iv), cell A is probably the

maximum-output cell and consequently is reinforced to pattern (iv) If we repeat the presentation of pattern (iv) many times, the response of cell A to pattern (iv) gradually increases and becomes larger than that to pattern (i) When the response to pattern (iv) becomes large enough, after repeated reinforcement, cell A now comes to yield a non-zero output to pattern (v) because pattern (v) has only one missing element compared to pattern (iv) (See Appendix) Hence, cell A can now be reinforced to pattern (v), to which cell A did not respond at the beginning Thus, we can gradually change the response characteristics of a cell so as to make it respond to a considerably different pattern, if the speed of variation of the training pattern is slow enough This situation resembles the phenomenon which we experience in recognition of a human face the face of a person gradually changes with time but we scarcely perceive the change if we see him every day

We have explained how the self-organization is performed by means of the principle that only maximum output cells have their input interconnections reinforced In the neocognitron, this principle is applied only to the "representative cells" and the other cells mimic their representatives Let us compare the network of Fig 7 and that of the neocognitron The photoreceptor cells in Fig 7 correspond to cells of layer U_0 in the neocognitron and cells A, B and C correspond to S-cells of layer U_{S1}, which have receptive fields at the same position but belong to different S-planes Actually, in the neocognitron, each S-plane has many S-cells in parallel, each of which should have input interconnections of the same spatial distribution as those of cell A, B or C

Concerning the network of Fig 7, we have just explained the case in which the positions of the receptive fields of cells A, B and C coincide exactly and the training patterns are always presented at the same position in these receptive fields In the real situation of self-organization, however, the position of training-pattern presentation is not fixed Nevertheless, the result of self-organization is not affected at all Assume the case where a training pattern which can elicit a maximum response from cell A, is not presented to the receptive field of cell A but to the receptive field of, say, cell A', which belongs to the same S-plane as cell A In this case, cell A' yields a larger output than cell A and cell A' is chosen as the representative instead of cell A Since all the cells in the S-plane in which cells A and A' are situated have their input interconnections reinforced in the same manner as their representative, we have the same result of reinforcement without it being affected by whether the training pattern is presented to the receptive field of cell A or A'

In the above explanation, we assumed that each training pattern contains only one stimulus feature However, each training pattern usually contains many different stimulus features in it Hence, in order to perform an effective self-organization, it is necessary to choose numbers of representatives from each S-layer at a time Therefore, the technique discussed in Section 4 2 is employed to choose the representative, i e , each S-plane is divided into many subgroups and maximum-output cells are individually chosen from these subgroups as candidates for the representatives and the representatives are chosen from these candidates

When the reinforcement of the input interconnections of layer U_{S1} progresses to a certain extent and the response of layer U_{S1} to each training pattern becomes large and stable, the output from layer U_{C1}, which is directly driven by the output from layer U_{S1}, now begins to trigger the reinforcement of the variable interconnections between $U_{C1} \rightarrow U_{S2}$ Incidentally, during the initial state of reinforcement when the response characteristics of layer U_{S1} are still unstable and fluctuating, the output of layer U_{S1} is kept very small because the input interconnections to the cells of this layer are still very small and, consequently, the reinforcement of the interconnections between $U_{C1} \rightarrow U_{S2}$ scarcely progresses On the other hand, when the response characteristics of layer U_{S1} become stable after a certain degree of progress in learning, the peaks of the response of layer U_{S1} become large and approach to value 1 0 and, consequently, the reinforcement of the interconnections between $U_{C1} \rightarrow U_{S2}$ begins to make rapid progress

Thus, it is important in the self-organization of a hierarchical multilayered network to retard the reinforcement of the interconnections to the succeeding layers until the response characteristics of the preceding layers become stable What would occur if the network were not designed in such a way? If some cells of layer U_{S1} happen to yield strange outputs during the initial stage of training, some cells of layer U_{S2} will be reinforced so as to detect this strange response If this strange response does not appear in layer U_{S1} afterwards, these U_{S2}-cells which detect this strange response become useless and wasteful In the neocognitron, however, this defect is avoided, because the progress of self-organization of each layer in the network is optimally controlled by the employment of analog cells with shunting inhibition and by the optimum choice of parameter q

It can be seen from the above discussions that parameter q, which determines the speed of reinforcement, should be chosen as follows When we want to reinforce a network which has only one layer of S-cells, we can choose a large value of q, by which self-organization of the network can be completed with only a single presentation of each training pattern On the other hand, in the neocognitron which has many stages of modifiable layers (c f in the computer simulation in Section 6, we consider the network with three modifiable layers), the value of q should be small enough to retard the build up of the outputs of the cells of each layer until the response characteristics of all the cells in the layer become stable Furthermore, during the stage of self-organization all the training patterns should be evenly presented with a certain number of repetitions, otherwise a decrease in value of q

becomes of little avail Thus, self-organization of the neocognitron progresses step by step from the front layer, in the order U_{S1}, U_{S2}, U_{S3}

5 A ROUGH SKETCH OF THE WORKING OF THE NEOCOGNITRON

In order to help with the understanding of the principles by which the neocognitron performs pattern recognition, we will make a rough sketch of the working of the network in the state after completion of self-organization The description in this Section, however, is not so strict, because the purpose of this Section is only to show an outline of the working of the network

First, let us assume that the neocognitron has been self-organized with repeated presentations of a set of stimulus patterns such as "A", "B", "C" and so on In the state when self-organization has been completed, various feature-extracting cells are formed in the network, as shown in Fig 8 (It should be noted that Fig 8 shows only an example It does not mean that exactly the same feature extractors as shown in this figure are always formed in the network)

If pattern "A" is presented to the input layer U_0, the cells in the network yield outputs as shown in Fig 8 For instance, the S-plane with $k_1 = 1$ in layer U_{S1} consists of a two-dimensional array of S-cells which extract $\wedge$-shaped features Since the stimulus pattern "A" contains a $\wedge$-shaped feature at the top, an S-cell near the top of this S-plane yields a large output as shown in the enlarged illustration in the lower part of Fig 8

A C-cell in the succeeding C-plane (i e , the C-plane in layer U_{C1} with $k_1 = 1$) has interconnections from a group of S-cells in this S-plane For example, the C-cell shown in Fig 8 has interconnections from the S-cells situated within the thin-lined circle and it is activated whenever at least one of these S-cells yields a large output Hence, the C-cell responds to a $\wedge$-shaped feature situated in a certain area in the input layer and its response is less affected by the shift in position of the stimulus pattern than that of the preceding S-cells Since this C-plane consists of an array of such C-cells, several C-cells which are situated near the top of this C-plane respond to the $\wedge$-shaped feature of the stimulus pattern "A" In layer U_{C1}, besides this C-plane, we also have C-planes which extract features with shapes like $\nearrow$, $\nwarrow$ and so on

In the next module, each S-cell receives signals from all the C-planes of layer U_{C1} For example, the S-cell of layer U_{S2} shown in Fig 8 receives signals from C-cells within the thin-lined circles in layer U_{C1} Its input interconnections have been reinforced in such a way that this S-cell responds only when $\wedge$-shaped, $\nearrow$-shaped and $\nwarrow$-shaped features are presented in its receptive field with configuration like $\nearrow\!\!\wedge\!\!\nwarrow$ Hence, pattern "A" elicits a large response from this S-cell, which is situated a little above the center of this S-plane Even if the positional relation of these three features is changed a little, this cell will still keep

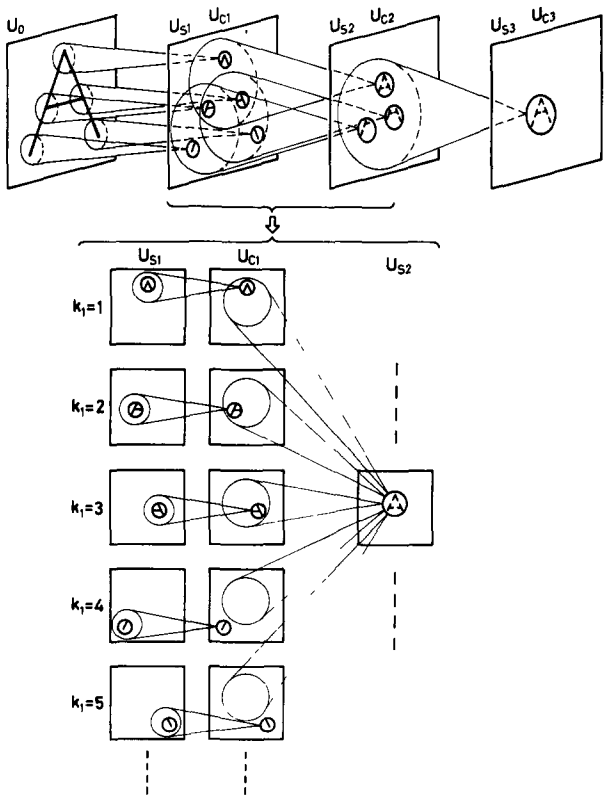

Fig 8 An example of the interconnections between cells and the response of the cells after completion of the self-organization

responding, because the preceding C-cells are not so sensitive to the positional error of these features However, if the positional relation of these three features is changed beyond some allowance, this S-cell stops responding This S-cell also checks the condition that other features such as ends-of-lines, which are to be extracted in S-planes in layer U_{S1} with $k_1 = 4, 5$ and so on, are not presented in its receptive field The inhibitory V_C-cell, which makes inhibitory interconnections to this S-cell of layer U_{S2}, plays an important role in checking the absence of such irrelevant features

Since operations of this kind are repeatedly applied through a cascade connection of modular structures of S- and C-layers, each individual cell in the network comes to have a wider receptive field in accordance with the increased number of modules before it and, at the same time, becomes more tolerant of shift in position and distortion in shape of the input pattern Thus, in the deepest layer, each C-cell has a receptive field large enough to cover the whole input layer and is selectively responsive only to one of the stimulus patterns, say to "A", but its response is not affected by shift in position or even by considerable distortion in shape of the stimulus pattern Although only one cell which responds to pattern "A" is drawn in Fig 8, cells which respond to other patterns, such as "B", "C" and so on, have also been formed in parallel in the deepest layer

From these discussions, it might be felt that an enormously large number of feature-extracting cell-planes becomes necessary with an increase in the number of input patterns to be recognized However, this is not the case With an increase in the number of input patterns, it becomes more and more probable that one and the same feature is contained in common in more than two different kinds of stimulus pattern Hence, each cell-plane, especially the one near the input layer, will generally be used in common for feature extraction not from only one pattern but from numerous kinds of patterns. Therefore, the required number of cell-planes does not increase so much, in spite of an increase in the number of patterns to be recognized

We can summarize the discussion in this Section as follows The stimulus pattern is first observed within a narrow range by each of the 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 slightly larger range and higher-order features are extracted Operations of this kind are repeatedly applied through a cascade 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, plays an important role in endowing the network with an ability to recognize even distorted patterns

6 COMPUTER SIMULATION

6 1 *Parameters for the simulation*

The network proposed here was simulated on a digital computer In the computer simulation we considered a seven layered network $\cdot U_0 \rightarrow U_{S1} \rightarrow U_{C1} \rightarrow U_{S2} \rightarrow U_{C2} \rightarrow U_{S3} \rightarrow U_{C3}$ In other words, the network has three stages of modular structures preceded by an input layer The number of cell-planes K_l is equally 24 for all layers, U_{S1}–U_{C3} The other parameters are listed in Table 1

As can be seen from Table 1, the total number of C-cells in the deepest layer U_{C3} is 24, because each C-plane has only one C-cell

The number of cells contained in a connectable area S_l is always 5 × 5 for every S-layer Hence, the number of excitatory input interconnections* to each S-cell is 5 × 5 in layer U_{S1} and 5 × 5 × 24 in layers U_{S2} and U_{S3}, because layers U_{S2} and U_{S3} are preceded by C-layers consisting of 24 cell-planes each Although the number of cells contained in S_l is the same for every S-layer, the size of S_l, which is projected to and observed at layer U_0, increases with the depth of the S-layer, because of the decrease in the density of the cells in a cell-plane

6.2. *Recognition of five numerals*

During the stage of learning in the first experiment, five training patterns, "0", "1", "2", "3" and "4" (shown

* It does not necessarily mean that all of these input interconnections actually grow up and become effective In usual situations, only a part of these interconnections come to have large values, and the rest of them remain with small values

Table 1 Parameters for the simulation

Layer	No of exc cells	No of exc input interconnections per cell	Size of an S-column (No of cells per S-column)	r_l	q_l
U_0	16 × 16				
U_{S1}	16 × 16 × 24	5 × 5	5 × 5 × 24	4 0	1
U_{C1}	10 × 10 × 24	5 × 5			
U_{S2}	8 × 8 × 24	5 × 5 × 24	5 × 5 × 24	1 5	16
U_{C2}	6 × 6 × 24	5 × 5			
U_{S3}	2 × 2 × 24	5 × 5 × 24	2 × 2 × 24	1 5	16
U_{C3}	24	2 × 2			

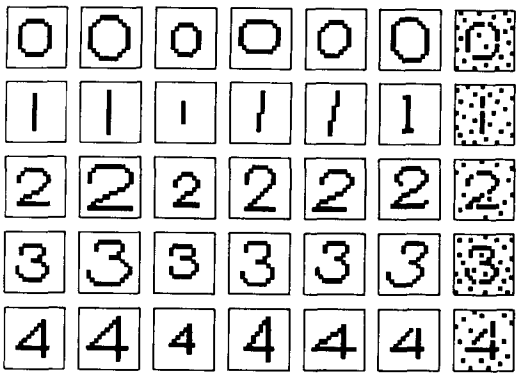

Fig 9 Some examples of the stimulus patterns which the neocognitron recognized correctly The neocognitron was first trained with the patterns shown in the leftmost column

in the leftmost column in Fig 9) were presented repeatedly to the input layer U_0 The positions of these training patterns were randomly shifted at every presentation It does not matter, of course, if the training patterns are presented always at the same position On the contrary, self-organization generally becomes easier if the position of pattern presentation is stationary rather than if it is shifted at random Thus, experimental results under more difficult conditions are shown here

The learning was performed without-a-teacher and the neocognitron did not receive any information about the categories of the training patterns

After repeated presentations of these five training patterns, the neocognitron gradually acquired the ability to classify these patterns by itself, according to the difference in their shape In this simulation, each of the five training patterns was presented 20 times By that time, self-organization of the network was almost completed

Figure 10 collectively shows how the individual cells in the network came to respond to the five training patterns In each of the pictures in Fig 10, out of the seven cell-layers constituting the neocognitron we display only four layers, namely the input layer U_0 and C-cell-layers U_{C1}, U_{C2} and U_{C3}, arranging them vertically in order In the deepest layer U_{C3}, which is displayed in the bottom row of each picture, it is seen that a different cell responds to each stimulus pattern "0", "1", "2", "3" and "4" This means that the neocognitron correctly recognizes these five stimulus patterns

We will see, next, how the response of the neocognitron is changed by a shift in the position of presentation of the stimulus patterns Figure 11 shows how the individual cells of layers U_0, U_{C1}, U_{C2} and U_{C3} respond to stimulus pattern "2" presented at four different positions As can be seen in this figure, the responses of the cells in the intermediate layers, especially the ones near the input layer, vary with a shift in position of the stimulus pattern However, the deeper the layer is, i e , the lower in each of the four pictures the layer is, the smaller is the variation in response of the cells in it Thus, the cells of the deepest layer U_{C3} are not affected at all by a shift in position of the stimulus pattern This means that the neocognitron correctly recognizes the stimulus patterns irrespective of their positions

The neocognitron recognizes the stimulus patterns

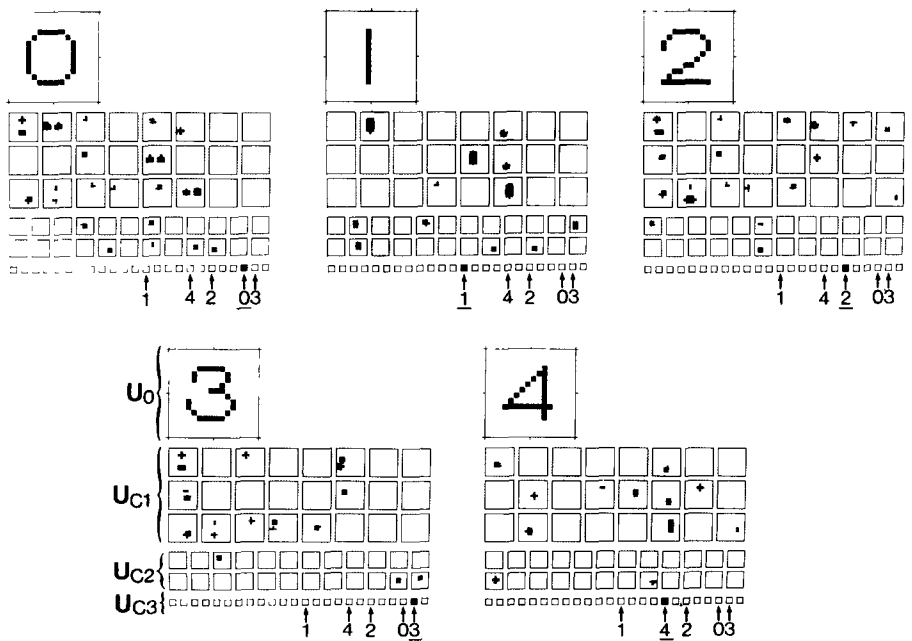

Fig 10 Response of the cells of layers U_0, U_{C1}, U_{C2} and U_{C3} to each of the five stimulus patterns

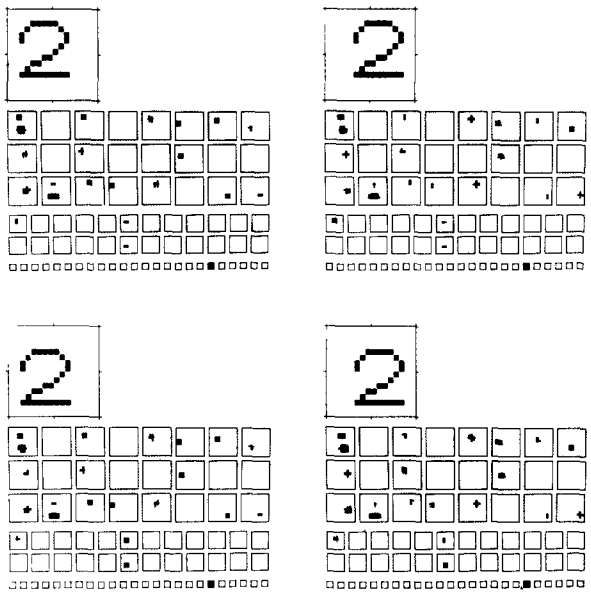

Fig 11 Response of the cells of layers U_0, U_{C1}, U_{C2} and U_{C3} to the same stimulus pattern "2" presented at four different positions

correctly, even if the stimulus patterns are distorted in shape or contaminated with noise Figure 9 shows some examples of distorted stimulus patterns which the neocognitron recognized correctly All the patterns in each row elicited the same response from the cells of the deepest layer U_{C3} Even though a stimulus pattern was increased or diminished in size or was skewed in shape, the response of the cells of the deepest layer was not affected Similarly, the stimulus patterns were correctly recognized even if they were stained with noise or had some parts missing

As an example of the feature extracting cells formed in this network, Fig 12 displays the receptive fields of the cells of each of the 24 S-planes of layer U_{S1} In this figure, the excitatory parts of the receptive fields are indicated by the darkness of the cells in the picture The arrangement of these 24 receptive fields in Fig 12 is the same as that of the C-cells of layer U_{C1} in Figs 10 and 11 Hence, in each display of Figs 10 and 11 it is understood, for example, that the upper left C-plane of layer U_{C1} extracts the horizontal line components from the stimulus pattern

6 3 *Some other experiments*

In order to check whether the neocognitron can acquire the ability for correct pattern recognition, even for a set of stimulus patterns resembling each other, another experiment was carried out In this experiment, the neocognitron was trained using four stimulus patterns, "X", "Y", "T" and "Z" These four patterns resemble each other in shape the upper parts of "X" and "Y" have an identical shape and the diagonal lines in "Z" and "X" have an identical inclination, and so on After repetitive presentation of these resembling patterns, the neocognitron acquired the ability to discriminate between them correctly

In a third experiment, the number of the training patterns was increased and ten different patterns, "0", "1", "2", , "9", were presented during the process of self-organization Even in the case of ten training patterns, it *is* possible to self-organize the neocognitron so as to recognize these ten patterns correctly, provided that various parameters in the network are properly adjusted and that the training patterns are skillfully presented during the process of self-organization In this case, however, a small deviation of the values of the parameters or a small change in the sequence of pattern presentation has a critical influence upon the ability of the self-organized network This means that the number of cell-planes in the network (i e, 24 cell-planes in each layer) is not sufficient for the recognition of ten different patterns If the number of cell-planes was further increased, the neocognitron would steadily, correctly recognize these ten patterns, or even a much greater number of patterns A computer simulation for the case of more than 24 cell-planes in each layer, however, has not yet been carried out, because of the lack of memory capacity on our computer

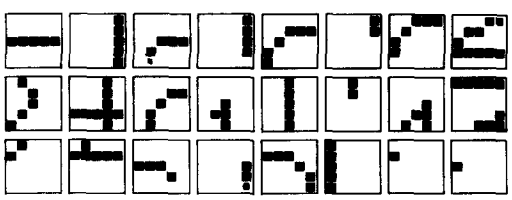

Fig 12 Receptive fields of the cells of each of the 24 S-planes of layer U_{S1}, which has finished learning

7 DISCUSSION

The "neocognitron" proposed in this paper has an

ability to recognize stimulus patterns without being affected by shifts in position or by considerable distortions in shape of the stimulus patterns It also has a function of self-organization, which progresses by means of "learning-without-a-teacher" If a set of stimulus patterns is repeatedly presented to it, it gradually acquires the ability to recognize these patterns It is not necessary to give any information about the categories to which these stimulus patterns should be classified The performance of the neocognitron was demonstrated by computer simulation

One of the greatest, and long-standing, difficulties in designing a pattern-recognizing machine was the problem of how to cope with shifts in position and distortions in shape of the input patterns The neocognitron proposed in this paper gives an important solution to this difficulty We would be able to vastly improve the performance of pattern recognizers if we introduced this algorithm into the design of the machines The same principle could also be applied to auditory information processing, such as speech recognition, if the spatial pattern (the envelope of the vibration) generated on the basilar membrane in the cochlea is considered as the input signal to the network

In this paper, we have demonstrated that the neocognitron can be realized with a multilayered network When we want to build a neocognitron with hardware, however, the multilayered network is not the only method for realization There are a variety of methods for constructing hardware working with this same algorithm For instance, the use of digital filters combined with television-like scanning is one of the easily realizable methods

SUMMARY

Suggested by the structure of the visual nervous system, a new algorithm is proposed for pattern recognition This algorithm can be realized with a multilayered network consisting of cells of analog type The network, which is called a "neocognitron", is self-organized by unsupervised learning only repetitive presentations of a set of stimulus patterns are necessary for the self-organization of the network and no information about the categories to which these patterns should be classified is needed The neocognitron by itself acquires the ability to recognize stimulus patterns according to the differences in their shapes any patterns which we human beings judge to be alike are also judged to be of the same category by the neocognitron The neocognitron recognizes stimulus patterns correctly without being affected by shifts in position or even by considerable distortions in shape of the stimulus patterns

The neocognitron is a multilayered network with a hierarchical structure Specifically, it consists of an input layer (photoreceptor array) followed by a cascaded connection of a number of modular structures, each of which is composed of a layer of "S-cells" followed by a layer of "C-cells"

S-cells have variable input interconnections With the progress of learning, these variable interconnections are modified and the S-cells develop into feature extractors

Each C-cell has excitatory fixed interconnections leading from a group of S-cells All the S-cells in the group extract the same feature but from different positions on the input layer and the C-cell is activated if at least one of these S-cells is active Hence, the C-cell responds to the same feature as the S-cells preceding it but is less sensitive to a shift in position of the feature

In the network consisting of a cascaded connection of these modules, the stimulus pattern is first observed within a narrow range by each of the 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 of these stages of feature extraction, a small amount of positional error is tolerated The operation by which positional errors are tolerated little by little, not at a single stage, plays an important role in endowing the network with an ability to recognize even distorted patterns

When a stimulus pattern is presented at a different position or is distorted in shape, the responses of the cells in intermediate layers, especially the ones near the input layer, vary with the shift in position or the deformation of the stimulus pattern However, the deeper the layer is, the smaller become the variations in the response of the cells Thus, the response of the cells in the deepest layer is entirely unaffected by shifts in position or by deformations of the stimulus pattern

The ability of the neocognitron was demonstrated by a computer simulation

REFERENCES

1 K Fukushima, Neural network model for a mechanism of pattern recognition unaffected by shift in position—neocognitron—, *Trans Inst electronics commun Engrs Japan* **62-A,** 658–665 (1979) (In Japanese)
2 K Fukushima, Neocognitron a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, *Biol Cybernet* **36,** 193–202 (1980)
3 K Fukushima and S Miyake, Neocognitron self-organizing network capable of position-invariant recognition of patterns, *Proc 5th Int Conf Pattern Recognition,* Vol 1, pp 459–461 (1980)
4 F Rosenblatt, *Principles of Neurodynamics* Spartan Books, Washington, DC (1962)
5 M Minsky and S Papert, Perceptrons, *An Introduction to Computational Geometry* MIT Press, Cambridge, Mass (1969)
6 K Fukushima, Cognitron a self-organizing multilayered neural network, *Biol Cybernet* **20,** 121–136 (1975)
7 K Fukushima, Cognitron a self-organizing multilayered neural network model, NHK Technical Monograph No 30 (1981)

8 K Fukushima, Visual feature extraction by a multi-layered network of analog threshold elements, *IEEE Trans Syst Sci Cybernet* **SSC-5**, 322–333 (1969)

APPENDIX

Numerical examples for selectivity control

Using a concrete example, we will further explain how the selectivity of the S-cell shown in Fig 6 is controlled by parameter r We use the same notation as in Section 4 3 For the sake of simplicity, we will consider the case where the value of $c(v)$ is chosen to be a constant and to be independent of v It is also assumed that both $P(v)$ and $p(v)$ consist of elements whose values are either 1 or 0 Let N be the number of active elements (i e , the elements of value 1) in the training pattern $P(v)$ Let us make an elementwise comparison between the stimulus pattern $p(v)$ and the training pattern $P(v)$ Let n_1 be the number of elements which have value zero in $P(v)$ but are active in $p(v)$ Similarly, let n_0 be the number of elements which are active in $P(v)$ but have value zero in $p(v)$ Under these conditions, equation (25) reduces to the following equation

$$s = \frac{1 - \frac{n_0}{N}}{\sqrt{1 + \frac{n_1 - n_0}{N}}} \tag{26}$$

As was discussed in Section 4 3, iff parameter r is in the range $s > r/(r+1)$, i e , in the range

$$\frac{1 - \frac{n_0}{N}}{\sqrt{1 + \frac{n_1 - n_0}{N}}} > \frac{r}{r+1}, \tag{27}$$

the S-cell, which has been reinforced to the training pattern $P(v)$, yields a positive output also to pattern $p(v)$

We will give some numerical examples First, let us consider the case which appeared in the example in Section 4 4 Let the training pattern $P(v)$ be the one shown in Fig 7[b](i) We want to design a network in such a way that the S-cell which has been reinforced to $P(v)$ responds also to pattern (iv) of Fig 7[b] but not to pattern (v) Since pattern (i) has three active elements, we have $N = 3$ Comparing pattern (i) with (iv) and (v), we have $(n_1, n_0) = (1, 0)$ and $(1, 1)$, respectively Hence, in order to make equation (27) hold for $(n_1, n_0) = (1, 0)$ but not for $(1, 1)$, parameter r should be in the interval $2\,00 < r < 6\,46$ When parameter r is in this interval, we can easily see that the cell which is reinforced to pattern (iv) responds also to pattern (v) in this case we have $N = 4$ and $(n_1, n_0) = (0, 1)$ and equation (27) is satisfied when parameter r is in the range $r < 6\,46$

We will give another numerical example Consider an S-cell of layer U_{S1}, which appeared in the computer simulation of Section 6 The S-cell has a receptive field consisting of 5×5 photoreceptor cells Let the training pattern be, say, a vertical line of length 5 in the 5×5 array In this case, we have $N = 5$ The pairs of n_1 and n_0 which satisfy equation (27) can easily be calculated If we take $r = 4\,0$, such pairs are only $(n_1, n_0) = (1, 0), (0, 1)$ and $(2, 0)$ Incidentally, for these pairs, we have $s = 0\,91, 0\,89$ and $0\,85$, respectively This means that a vertical line which has one or two excessive elements or has one missing element is judged to be of the same class as $P(v)$ On the other hand, a vertical line with two missing elements, for example, a vertical line consisting of only three active elements, is judged to be of a different class from the training pattern $P(v)$ Incidentally, for a vertical line consisting of only three active elements, we have $s = 0\,77$

If we want to increase the selectivity and to design a network in such a way that the S-cell should yield output only when $(n_1, n_0) = (1, 0)$ and $(0, 1)$, parameter r should be in the interval $5\,46 < r < 8\,47$

About the Author—KUNIHIKO FUKUSHIMA received his B Sc degree in Electronics in 1958 and Ph D degree in Electrical Engineering in 1966, both from Kyoto University, Kyoto, Japan Since 1958 he has been working for the NHK (Japan Broadcasting Corporation) and he is presently a Senior Research Scientist of the NHK Broadcasting Science Research Laboratories His research interests are in the discovering of the algorithm for information processing in the brain, and he is engaged in the synthesis of neural network models for the mechanisms of visual pattern recognition, memory, learning and self-organization in the brain He is the author of the books, written in Japanese, "Physiology and Bionics of the Visual Systems" published by the Institute of Electronics and Communication Engineers, Japan in 1976 and "Neural Networks and Self-Organization" published by Kyoritsu Shuppan in 1979, among others

About the Author—SEI MIYAKE received his B Sc degree in Electronics in 1970 from Waseda University, Tokyo, Japan Since 1970 he has been working for the NHK (Japan Broadcasting Corporation) and he is currently a Research Scientist of the NHK Broadcasting Science Research Laboratories His major interest is in the mechanisms of pattern recognition in the brain