Biol. Cybernetics 36, 193–202 (1980)

Biological Cybernetics

Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position

Kunihiko Fukushima

NHK Broadcasting Science Research Laboratories, Kinuta, Setagaya, Tokyo, Japan

Abstract. A neural network model for a mechanism of visual pattern recognition is proposed in this paper. The network is self-organized by "learning without a teacher", and acquires an ability to recognize stimulus patterns based on the geometrical similarity (Gestalt) of their shapes without affected by their positions. This network is given a nickname "neocognitron". After completion of self-organization, the network has a structure similar to the hierarchy model of the visual nervous system proposed by Hubel and Wiesel. The network consists of an input layer (photoreceptor array) followed by a cascade connection of a number of modular structures, each of which is composed of two layers of cells connected in a cascade. The first layer of each module consists of "S-cells", which show characteristics similar to simple cells or lower order hypercomplex cells, and the second layer consists of "C-cells" similar to complex cells or higher order hypercomplex cells. The afferent synapses to each S-cell have plasticity and are modifiable. The network has an ability of unsupervised learning: We do not need any "teacher" during the process of self-organization, and it is only needed to present a set of stimulus patterns repeatedly to the input layer of the network. The network has been simulated on a digital computer. After repetitive presentation of a set of stimulus patterns, each stimulus pattern has become to elicit an output only from one of the C-cells of the last layer, and conversely, this C-cell has become selectively responsive only to that stimulus pattern. That is, none of the C-cells of the last layer responds to more than one stimulus pattern. The response of the C-cells of the last layer is not affected by the pattern's position at all. Neither is it affected by a small change in shape nor in size of the stimulus pattern.

1. Introduction

The mechanism of pattern recognition in the brain is little known, and it seems to be almost impossible to reveal it only by conventional physiological experiments. 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 an ability of pattern recognition like a human being.

Several models were proposed with this intention (Rosenblatt, 1962; Kabrisky, 1966; Giebel, 1971; Fukushima, 1975). The response of most of these models, however, was severely affected by the shift in position and/or by the distortion in shape of the input patterns. Hence, their ability for pattern recognition was not so high.

In this paper, we propose an improved neural network model. The structure of this network has been suggested by that of the visual nervous system of the vertebrate. This network is self-organized by "learning without a teacher", and acquires an ability to recognize stimulus patterns based on the geometrical similarity (Gestalt) of their shapes without affected by their position nor by small distortion of their shapes.

This network is given a nickname "neocognitron"[1], because it is a further extention of the "cognitron", which also is a self-organizing multilayered neural network model proposed by the author before (Fukushima, 1975). Incidentally, the conventional cognitron also had an ability to recognize patterns, but its response was dependent upon the position of the stimulus patterns. That is, the same patterns which were presented at different positions were taken as different patterns by the conventional cognitron. In the neocognitron proposed here, however, the response of the network is little affected by the position of the stimulus patterns.

1 Preliminary report of the neocognitron already appeared elsewhere (Fukushima, 1979a, b)

0340-1200/80/0036/0193/$02.00

The neocognitron has a multilayered structure, too. It also has an ability of unsupervised learning: We do not need any "teacher" during the process of self-organization, and it is only needed to present a set of stimulus patterns repeatedly to the input layer of the network. After completion of self-organization, the network acquires a structure similar to the hierarchy model of the visual nervous system proposed by Hubel and Wiesel (1962, 1965).

According to the hierarchy model by Hubel and Wiesel, the neural network in the visual cortex has a hierarchy structure: LGB (lateral geniculate body)→simple cells→complex cells→lower order hypercomplex cells→higher order hypercomplex cells. It is also suggested that the neural network between lower order hypercomplex cells and higher order hypercomplex cells has a structure similar to the network between simple cells and complex cells. In this hierarchy, 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.

It is true that the hierarchy model by Hubel and Wiesel does not hold in its original form. In fact, there are several experimental data contradictory to the hierarchy model, such as monosynaptic connections from LGB to complex cells. This would not, however, completely deny the hierarchy model, if we consider that the hierarchy model represents only the main stream of information flow in the visual system. Hence, a structure similar to the hierarchy model is introduced in our model.

Hubel and Wiesel do not tell what kind of cells exist in the stages higher than hypercomplex cells. Some cells in the inferotemporal cortex (i.e. one of the association areas) of the monkey, however, are reported to respond selectively to more specific and more complicated features than hypercomplex cells (for example, triangles, squares, silhouettes of a monkey's hand, etc.), and their responses are scarcely affected by the position or the size of the stimuli (Gross et al., 1972; Sato et al., 1978). These cells might correspond to so-called "grandmother cells".

Suggested by these physiological data, we extend the hierarchy model of Hubel and Wiesel, and hypothesize the existance of a similar hierarchy structure even in the stages higher than hypercomplex cells. In the extended hierarchy model, the cells in the highest stage are supposed to respond only to specific stimulus patterns without affected by the position or the size of the stimuli.

The neocognitron proposed here has such an extended hierarchy structure. After completion of self-organization, the response of the cells of the deepest layer of our network is dependent only upon the shape of the stimulus pattern, and is not affected by the position where the pattern is presented. That is, the network has an ability of position-invariant pattern-recognition.

In the field of engineering, many methods for pattern recognition have ever been proposed, and several kinds of optical character readers have already been developed. Although such machines are superior to the human being in reading speed, they are far inferior in the ability of correct recognition. Most of the recognition method used for the optical character readers are sensitive to the position of the input pattern, and it is necessary to normalize the position of the input pattern beforehand. It is very difficult to normalize the position, however, if the input pattern is accompanied with some noise or geometrical distortion. So, it has long been desired to find out an algorithm of pattern recognition which can cope with the shift in position of the input pattern. The algorithm proposed in this paper will give a drastic solution also to this problem.

2. Structure of the Network

As shown in Fig. 1, the neocognitron consists of a cascade connection of a number of modular structures preceded by an input layer U_0. Each of the modular structure is composed of two layers of cells connected in a cascade. The first layer of the module consists of "S-cells", which correspond to simple cells or lower order hypercomplex cells according to the classification of Hubel and Wiesel. We call it S-layer and denote the S-layer in the l-th module as U_{Sl}. The second layer of the module consists of "C-cells", which correspond to complex cells or higher order hypercomplex cells. We call it C-layer and denote the C-layer in the l-th module as U_{Cl}. In the neocognitron, only the input synapses to S-cells are supposed to have plasticity and to be modifiable.

The input layer U_0 consists of a photoreceptor array. The output of a photoreceptor is denoted by $u_0(\mathbf{n})$, where $\mathbf{n}=(n_x, n_y)$ is the two-dimensional coordinates indicating the location of the cell.

S-cells or C-cells in a 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 a terminology, S-plane and C-plane representing cell-planes consisting of S-cells and C-cells, respectively.

It is assumed that 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

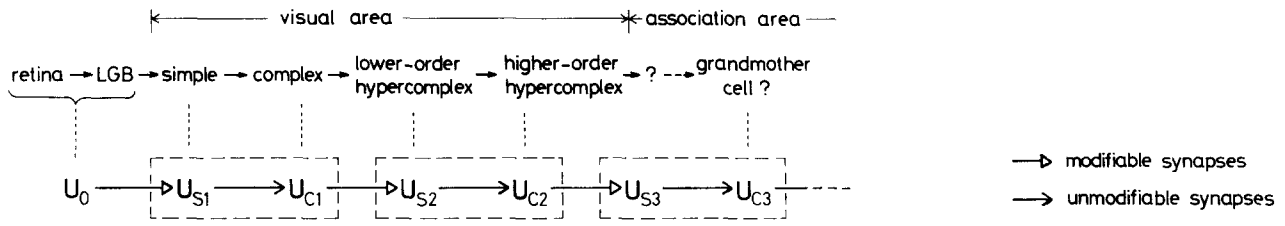

Fig. 1. Correspondence between the hierarchy model by Hubel and Wiesel, and the neural network of the neocognitron

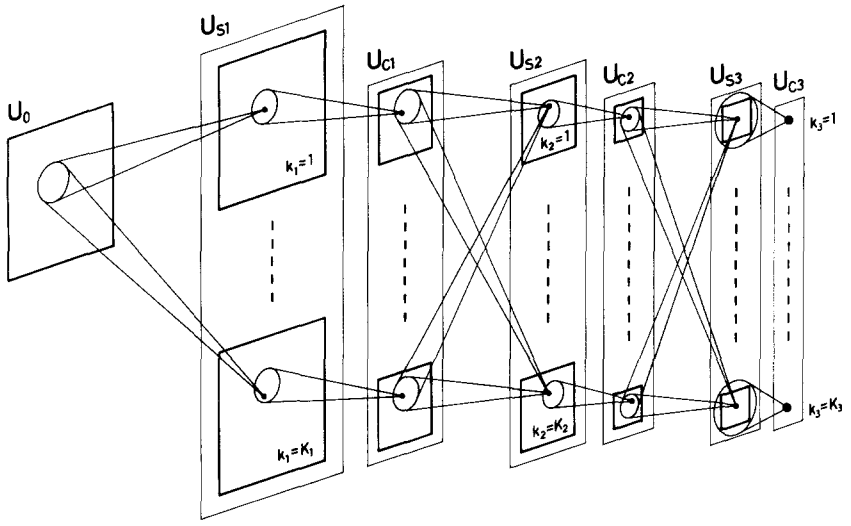

Fig. 2. Schematic diagram illustrating the interconnections between layers in the neocognitron

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.

We will use notations $u_{Sl}(k_l, \mathbf{n})$ 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 cell's receptive fields in the input layer.

Figure 2 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-planes or C-planes are enclosed, represents an S-layer or a C-layer.

In Fig. 2, a cell of each layer receives afferent connections from the cells within the area enclosed by the elipse in its preceding layer. To be exact, as for the S-cells, the elipses in Fig. 2 does not show the *connecting* area but the *connectable* area to the S-cells. That is, all the interconnections coming from the elipses are not always formed, because the synaptic connections incoming to the S-cells have plasticity.

In Fig. 2, for the sake of simplicity of the figure, only one cell is shown in each cell-plane. In fact, all the cells in a cell-plane have input synapses of the same spatial distribution as shown in Fig. 3, and only the positions of the presynaptic cells are shifted in parallel from cell to cell.

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. Hence, the total number of the cells in each cell-plane decreases with the depth of the cell-plane in the network. In the last module, the receptive field of each C-cell becomes so large as to cover the whole area of input layer U_0, and each C-plane is so determined as to have only one C-cell.

The S-cells and C-cells are excitatory cells. That is, all the efferent synapses from these cells are excitatory. Although it is not shown in Fig. 2, we also have

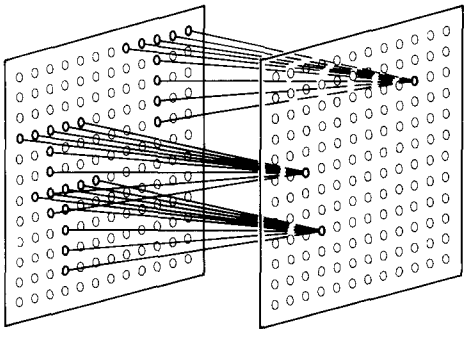

Fig. 3. Illustration showing the input interconnections to the cells within a single cell-plane

inhibitory cells $v_{Sl}(\mathbf{n})$ and $v_{Cl}(\mathbf{n})$ in S-layers and C-layers.

Here, we are going to describe the outputs of the cells in the network with numerical expressions.

All the neural cells employed in this network is of analog type. That is, the inputs and the output of a cell take non-negative analog values proportional to the pulse density (or instantaneous mean frequency) of the firing of the actual biological neurons.

S-cells have shunting-type inhibitory inputs similarly to the cells employed in the conventional cognitron (Fukushima, 1975). The output of an S-cell in the k_l-th S-plane in the l-th module is described below.

$$u_{Sl}(k_l,\mathbf{n})=r_l\cdot\varphi\left[\frac{1+\sum_{k_{l-1}=1}^{K_{l-1}}\sum_{\mathbf{v}\in S_l}a_l(k_{l-1},\mathbf{v},k_l)\cdot u_{Cl-1}(k_{l-1},\mathbf{n}+\mathbf{v})}{1+\frac{2r_l}{1+r_l}\cdot b_l(k_l)\cdot v_{Cl-1}(\mathbf{n})}-1\right], \tag{1}$$

where

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

In case of $l=1$ in (1), $u_{Cl-1}(k_{l-1},\mathbf{n})$ stands for $u_0(\mathbf{n})$, and we have $K_{l-1}=1$.

Here, $a_l(k_{l-1},\mathbf{v},k_l)$ and $b_l(k_l)$ represent the efficiencies of the excitatory and inhibitory synapses, respectively. As was described before, it is assumed that all the S-cells in the same S-plane have identical set of input synapses. Hence, $a_l(k_{l-1},\mathbf{v},k_l)$ 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 (1) prescribes the efficacy of the inhibitory input. The larger the value of r_l is, more selective becomes cell's response to its specific feature (Fukushima, 1978, 1979c). Therefore, the value of r_l should be determined with a compromise between the ability to differentiate similar patterns and the ability to tolerate the distortion of the pattern's shape.

The inhibitory cell $v_{Cl-1}(\mathbf{n})$, which have inhibitory synaptic connections to this S-cell, has an r.m.s.-type (root-mean-square type) input-to-output characteristic. That is,

$$v_{Cl-1}(\mathbf{n})=\sqrt{\sum_{k_{l-1}=1}^{K_{l-1}}\sum_{\mathbf{v}\in S_l}c_{l-1}(\mathbf{v})\cdot u_{Cl-1}^2(k_{l-1},\mathbf{n}+\mathbf{v})}, \tag{3}$$

where $c_{l-1}(\mathbf{v})$ represents the efficiency of the unmodifiable excitatory synapses, and is set to be a monotonically decreasing function of $|\mathbf{v}|$. The employment of r.m.s.-type cells is effective for endowing the network with an ability to make reasonable evaluation of the similarity between the stimulus patterns. Its effectiveness was analytically proved for the conventional cognitron (Fukushima, 1978, 1979c), and the same discussion can be applied also to this network.

As is seen from (1) and (3), the area from which a single cell receives its input, that is, the summation range S_l of $\mathbf{v}$ is determined to be identical for both cells $u_{Sl}(k_l,\mathbf{n})$ and $v_{Cl-1}(\mathbf{n})$.

The size of this range S_l is set to be small for the foremost module ($l=1$) and to become larger and larger for the hinder modules (in accordance with the increase of l).

After completion of self-organization, the procedure of which will be discussed in the next chapter, a number of feature extracting cells of the same function are formed in parallel within each S-plane, and only the positions of their receptive fields are different to each other. Hence, if a stimulus pattern which elicits a response from an S-cell is shifted in parallel in its position on the input layer, another S-cell in the same S-plane will respond instead of the first cell.

The synaptic connections from S-layers to C-layers are fixed and unmodifiable. As is illustrated in Fig. 2, a C-cell have synaptic connections from a group of S-cells in its corresponding S-plane (i.e. the preceding S-plane with the same k_l-number as that of the C-cell). The efficiencies of these synaptic connections are so determined that the C-cell will respond strongly whenever at least one S-cell in its connecting area yields a large output. Hence, even if a stimulus pattern which has elicited a large response from a C-cell is shifted a little in position, the C-cell will keep responding as before, because another presynaptic S-cell will become to respond instead.

Quantitatively, C-cells have shunting-type inhibitory inputs similarly as S-cells, but their outputs show a saturation characteristic. The output of a C-cell in the k_l-th C-plane in the l-th module is given by the equation below.

$$u_{Cl}(k_l,\mathbf{n})=\psi\left[\frac{1+\sum_{\mathbf{v}\in D_l}d_l(\mathbf{v})\cdot u_{Sl}(k_l,\mathbf{n}+\mathbf{v})}{1+v_{Sl}(\mathbf{n})}-1\right], \tag{4}$$

where

$$\psi[x]=\varphi[x/(\alpha+x)]. \tag{5}$$

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:

$$v_{Sl}(\mathbf{n})=\frac{1}{K_l}\sum_{k_l=1}^{K_l}\sum_{\mathbf{v}\in D_l}d_l(\mathbf{v})\cdot u_{Sl}(k_l,\mathbf{n}+\mathbf{v}). \tag{6}$$

In (4) and (6), the efficiency of the unmodifiable excitatory synapse $d_l(\mathbf{v})$ is set to be a monotonically decreasing function of $|\mathbf{v}|$ in the same way as $c_l(\mathbf{v})$, and the connecting area D_l is small in the foremost module and becomes larger and larger for the hinder modules. The parameter α in (5) is a positive constant which specifies the degree of saturation of C-cells.

3. Self-organization of the Network

The self-organization of the neocognitron is performed by means of "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 stimulus patterns.

As was discussed in Chap. 2, one of the basic hypotheses employed in the neocognitron is the assumption that all the S-cells in the same S-plane have input synapses of the same spatial distribution, and that only the positions of the presynaptic cells shift in parallel in accordance with the shift in position of individual S-cells' receptive fields.

It is not known whether modifiable synapses in the real nervous system are actually self-organized always keeping such conditions. Even if it is assumed to be true, neither do we know by what mechanism such a self-organization goes on. The correctness of this hypothesis, however, is suggested, for example, from the fact that orderly synaptic connections are formed between retina and optic tectum not only in the initial development in the embryo but also in regeneration in the adult amphibian or fish: In regeneration after removal of half of the tectum, the whole retina come to make a compressed orderly projection upon the remaining half tectum (e.g. review article by Meyer and Sperry, 1974).

In order to make self-organization under the conditions mentioned above, the modifiable synapses are reinforced by the following procedures.

At first, several "representative" S-cells are selected from each S-layer every time when a stimulus pattern is presented. The representative is selected among the S-cells which have yielded large outputs, but the number of the representatives is so restricted that more than one representative are not selected from any single S-plane. The detailed procedure for selecting the representatives is given later on.

The input synapses to a representative S-cell are reinforced in the same manner as in the case of r.m.s.-type cognitron[2] (Fukushima, 1978, 1979c). All the other S-cells in the S-plane, from which the representative is selected, have their input synapses 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 selected as a representative. The modifiable synapses $a_l(k_{l-1}, \mathbf{v}, \hat{k}_l)$ and $b_l(\hat{k}_l)$, which are afferent to the S-cells of the $\hat{k}_l$-th S-plane, are reinforced by the amount shown below:

$$\Delta a_l(k_{l-1}, \mathbf{v}, \hat{k}_l) = q_l \cdot c_{l-1}(\mathbf{v}) \cdot u_{Cl-1}(k_{l-1}, \hat{\mathbf{n}} + \mathbf{v}), \tag{7}$$

$$\Delta b_l(\hat{k}_l) = (q_l/2) \cdot v_{Cl-1}(\hat{\mathbf{n}}), \tag{8}$$

where q_l is a positive constant prescribing the speed of reinforcement.

The cells in the S-plane from which no representative is selected, however, do not have their input synapses reinforced at all.

In the initial state, the modifiable excitatory synapses $a_l(k_{l-1}, \mathbf{v}, k_l)$ are set to have small positive values such that the S-cells show very weak orientation selectivity, and that the preferred orientation of the S-cells differ from S-plane to S-plane. That is, the initial values of these modifiable synapses are given by a function of $\mathbf{v}$, (k_l/K_l) and $|k_{l-1}/K_{l-1} - k_l/K_l|$, but they don't have any randomness. The initial values of modifiable inhibitory synapses $b_l(k_l)$ are set to be zero.

The procedure for selecting the representatives is given below. It resembles, in some sense, to the procedure with which the reinforced cells are selected in the conventional cognitron (Fukushima, 1975).

At first, in an S-layer, we watch 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 a manner shown in Fig. 4, the group of S-cells constitute a column in an S-layer. Accordingly, we call the group as an "S-column". An S-column contains S-cells from all the S-planes. That is, 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. Hence, the idea of S-columns defined here closely resembles that of "hypercolumns" proposed by Hubel and Wiesel (1977). There are a lot of such S-columns in a single S-layer. Since S-columns have overlapping with one another, there is a possibility that a single S-cell is contained in two or more S-columns.

From each S-column, every time when a stimulus pattern is presented, the S-cell which is yielding the largest output is chosen as a candidate for the representatives. 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 selected as the representative from that S-plane. In

2 Qualitatively, the procedure of self-organization for r.m.s.-type cognitron is the same as that for the conventional cognitron (Fukushima, 1975)

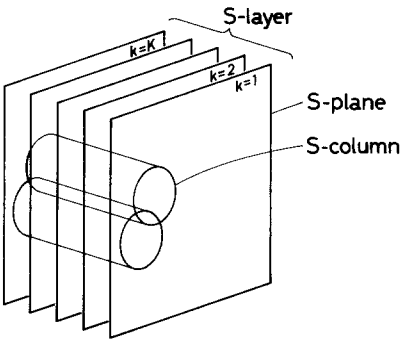

Fig. 4. Relation between S-planes and S-columns within an S-layer

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 selected 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 not a possibility of formation of redundant connections such that two or more S-planes are used for detection of one and the same feature. Incidentally, representatives are selected only from a small number of S-planes at a time, and the rest of the S-planes are to send representatives for other stimulus patterns.

As is seen from these discussions, if we consider that a single S-plane in the neocognitron corresponds to a single excitatory cell in the conventional cognitron (Fukushima, 1975), the procedures of reinforcement in the both systems are analogous to each other.

4. Rough Sketches of the Working of the Network

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

At first, let us assume that the neocognitron has been self-organized with repeated presentations of stimulus patterns like "A", "B", "C" and so on. In the state when the self-organization has been completed, various feature-extracting cells are formed in the network as shown in Fig. 5. (It should be noted that Fig. 5 shows only an example. It does not mean that exactly the same feature extractors as shown in this figure are always formed in this network.)

Here, if pattern "A" is presented to the input layer U_0, the cells in the network yield outputs as shown in

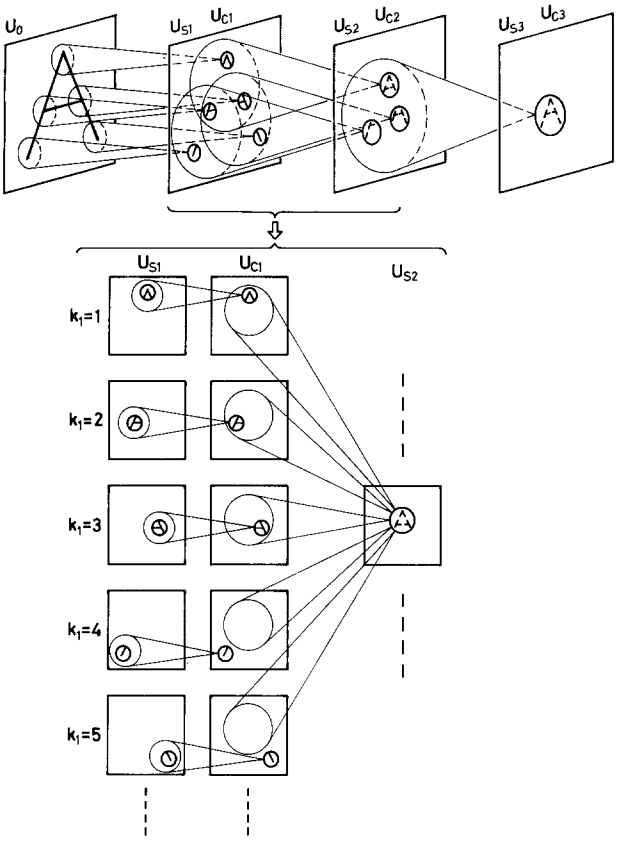

Fig. 5. An example of the interconnections between cells and the response of the cells after completion of self-organization

Fig. 5. For instance, S-plane with $k_1=1$ in layer U_{S1} consists of a two-dimensional array of S-cells which extract Λ-shaped features. Since the stimulus pattern "A" contains Λ-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. 5.

A C-cell in the succeeding C-plane (i.e. C-plane in layer U_{C1} with $k_1=1$) has synaptic connections from a group of S-cells in this S-plane. For example, the C-cell shown in Fig. 5 has synaptic connections from the S-cells situated within the thin-lined circle, and it responds whenever at least one of these S-cells yields a large output. Hence, the C-cell responds to a Λ-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 presynaptic 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 Λ-shaped feature contained in 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 shown in Fig. 5 receives signals from C-cells within the thin-lined circles in layer U_{C1}. Its input synapses have been reinforced in such a way that this S-cell responds only when $\wedge$-shaped, $/$-shaped and $\backslash$-shaped features are presented in its receptive field with configuration like $\overset{\wedge}{/\backslash}$. Hence, pattern "A" elicits a large response from this S-cell, which is situated a little above the center of this S-plane. If positional relation of these three features are 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 with $k_1=4$, 5 and so on, are not presented in its receptive field. The inhibitory cell v_{C1}, which makes inhibitory synaptic connection to this S-cell, 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 becomes to have 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 of the input pattern. Thus, one C-cell in the last layer U_{C3} yields a large response only when, say, pattern "A" is presented to the input layer, regardless of the pattern's position. Although only one cell which responds to pattern "A" is drawn in Fig. 5, cells which respond to other patterns, such as "B", "C" and so on, have been formed in parallel in the last layer.

From these discussions, it might be felt as if an enormously large number of feature-extracting cell-planes become necessary with the increase in the number of input patterns to be recognized. However, it is not the case. With the 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 patterns. Hence, each cell-plane, especially the one near the input layer, will generally be used in common for the 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 the increase in the number of patterns to be recognized.

Viewed from another angle, this procedure for pattern recognition can be interpreted as identical in its principle to the information processing mentioned below.

That is, in the neocognitron, the input pattern is compared with learned standard patterns, which have been recorded beforehand in the network in the form of spatial distribution of the synaptic connections. This comparison is not made by a direct pattern matching in a wide visual field, but by piecewise pattern matchings in a number of small visual fields. Only when the difference between both patterns does not exceed a certain limit in any of the small visual fields, the neocognitron judges that these patterns coincide with each other.

Such comparison in small visual fields is not performed in a single stage, but similar processes are repeatedly applied in a cascade. That is, the output from one stage is used as the input to the next stage. In the comparison in each of these stages, the allowance for the shift in pattern's position is increased little by little. The size of the visual field (or the size of the receptive fields) in which the input pattern is compared with standard patterns, becomes larger in a higher stage. In the last stage, the visual field is large enough to observe the whole information of the input pattern simultaneously.

Even if the input pattern does not match with a learned standard pattern in all parts of the large visual field simultaneously, it does not immediately mean that these patterns are of different categories. Suppose that the upper part of the input pattern matches with that of the standard pattern situated at a certain location, and that, at the same time, the lower part of this input pattern matches with that of the same standard pattern situated at another location. Since the pattern matching in the first stage is tested in parallel in a number of small visual fields, these two patterns are still regarded as the same by the neocognitron. Thus, the neocognitron is able to make a correct pattern recognition even if input patterns have some distortion in shape.

5. Computer Simulation

The neural network proposed here has been simulated on a digital computer. In the computer simulation, we consider a seven layered network: $U_0 \rightarrow U_{S1} \rightarrow U_{C1} \rightarrow U_{S2} \rightarrow U_{C2} \rightarrow U_{S3} \rightarrow U_{C3}$. That is, the network has three stages of modular structures preceded by an input layer. The number of cell-planes K_l in each layer is 24 for all the layers except U_0. The numbers of excitatory cells in these seven layers are: 16×16 in U_0, $16 \times 16 \times 24$ in U_{S1}, $10 \times 10 \times 24$ in U_{C1}, $8 \times 8 \times 24$ in U_{S2}, $6 \times 6 \times 24$ in U_{C2}, $2 \times 2 \times 24$ in U_{S3}, and 24 in U_{C3}. In the last layer U_{C3}, each of the 24 cell-planes contains only one excitatory cell (i.e. C-cell).

The number of cells contained in the connectable area S_l is always 5×5 for every S-layer. Hence, the number of input synapses[3] to each S-cell is 5×5 in layer U_{S1} and $5 \times 5 \times 24$ in layers U_{S2} and U_{S3}, because

3 It does not necessarily mean that all of these input synapses are always fully reinforced. In usual situations, only some of these input synapses are reinforced, and the rest of them remains in small values

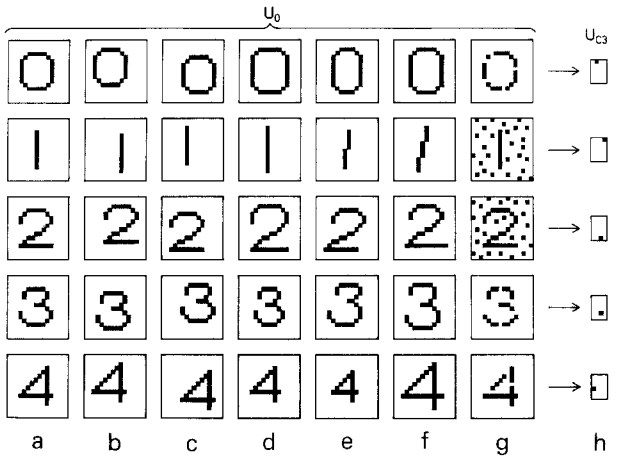

Fig. 6. Some examples of distorted stimulus patterns which the neocognitron has correctly recognized, and the response of the final layer of the network

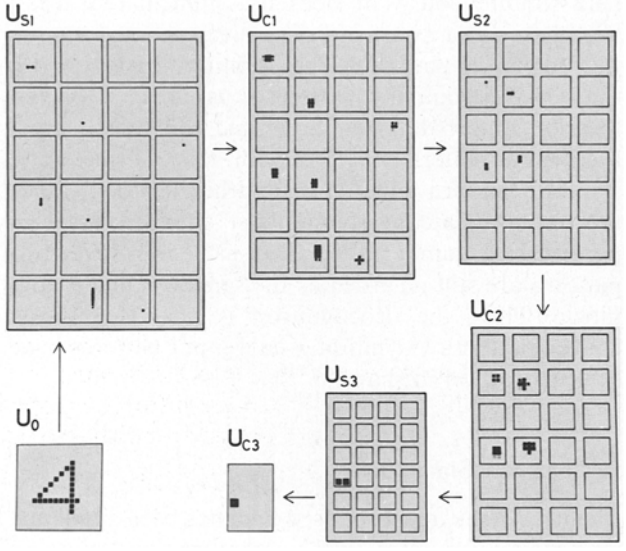

Fig. 7. A display of an example of the response of all the individual cells in the neocognitron

layers U_{S2} and U_{S3} are preceded by C-layers consisting of 24 cell-planes. 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 for the hinder layers because of decrease in density of the cells in a cell-plane.

The number of excitatory input synapses to each C-cell is 5×5 in layers U_{C1} and U_{C2}, and is 2×2 in layer U_{C3}. Every S-column has a size such that it contains $5\times 5\times 24$ cells for layers U_{S1} and U_{S2}, and $2\times 2\times 24$ cells for layer U_{S3}. That is, it contains 5×5, 5×5, and 2×2 cells from each S-plane, in layers U_{S1}, U_{S2}, and U_{S3}, respectively.

Parameter r_l, which prescribe the efficacy of inhibitory input to an S-cell, is set such that $r_1=4.0$ and $r_2=r_3=1.5$. The efficiency of unmodifiable excitatory synapses $c_{l-1}(\mathbf{v})$ is determined so as to satisfy the equation

$$\sum_{k_{l-1}=1}^{K_{l-1}} \sum_{\mathbf{v}\in S_l} c_{l-1}(\mathbf{v})=1\,. \tag{9}$$

The parameter q_l, which prescribe the speed of reinforcement, is adjusted such that $q_1=1.0$ and $q_2=q_3=16.0$. The parameter α, which specifies the degree of saturation, is set to be $\alpha=0.5$.

In order to self-organize the network, we have presented five stimulus patterns "0", "1", "2", "3", and "4", which are shown in Fig. 6 (a) (the leftmost column in Fig. 6), repeatedly to the input layer U_0. The positions of presentation of these stimulus patterns have been randomly shifted at every presentation[4].

Each of the five stimulus patterns has been presented 20 times to the network. By that time, self-organization of the network has almost been completed.

Each stimulus pattern has become to elicit an output only from one of the C-cells of layer U_{C3}, and conversely, this C-cell has become selectively responsive only to that stimulus pattern. That is, none of the C-cells of layer U_{C3} responds to more than one stimulus pattern. It has also been confirmed that the response of cells of layer U_{C3} is not affected by the shift in position of the stimulus pattern at all. Neither is it affected by a slight change of the shape or the size of the stimulus pattern.

Figure 6 shows some examples of distorted stimulus patterns which the neocognitron has correctly recognized. All the stimulus patterns (a)~(g) in each row of Fig. 6 have elicited the same response to C-cells of layer U_{C3} as shown in (h) (i.e. the rightmost patterns in each row). That is, the neocognitron has correctly recognized these patterns without affected by shift in position like (a)~(c), nor by distortion in shape or size like (d)~(f), nor by some insufficiency of the patterns or some noise like (g).

Figure 7 displays how individual cells in the neocognitron have responded to stimulus pattern "4". Thin-lined squares in the figure stand for individual cell-planes (except in layer U_{C3} in which each cell-plane contains only one cell). The magnitude of the output of each individual cell is indicated by the darkness of each small square in the figure. (The size of the square does not have a special meaning here.)

4 It does not matter, of course, even if the patterns are presented always at the same position. On the contrary, the self-organization generally becomes easier if the position of pattern presentation is stationary than it is shifted at random. Thus, the experimental result under more difficult condition is shown here

In order to check whether the neocognitron can acquire the ability of correct pattern recognition even for a set of stimulus patterns resembling each other, another experiment has been made. In this experiment, the neocognitron has been self-organized using four stimulus patterns "X", "Y", "T", and "Z". These four patterns resemble each other in shape: For instance, 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 has also acquired the ability to discriminate them correctly.

In a third experiment, the number of stimulus patterns has been increased, and ten different patterns "0", "1", "2", ..., "9" have been presented during the process of self-organization. Even in the case of ten stimulus 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 stimulus 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 of the way of pattern presentation, has critically influenced upon the ability of the self-organized network. This would mean that the number of cell-planes in the network (that is, 24 cell-planes in each layer) is not sufficient enough for the recognition of ten different patterns. If the number of cell-planes is further increased, it is presumed that the neocognitron would steadily make correct recognition of these ten patterns, or even much more number of patterns. The computer simulation for the case of more than 24 cell-planes in each layer, however, has not been made yet, because of the lack of memory capacity of our computer.

6. Conclusion

The "neocognitron" proposed in this paper has an ability to recognize stimulus patterns without affected by shift in position nor by a small distortion 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 are repeatedly presented to it, it gradually acquires the ability to recognize these patterns. It is not necessary to give any instructions about the categories to which the stimulus patterns should belong. The performance of the neocognitron has been demonstrated by computer simulation.

The author does not advocate that the neocognitron is a complete model for the mechanism of pattern recognition in the brain, but he proposes it as a working hypothesis for some neural mechanisms of visual pattern recognition.

As was stated in Chap. 1, the hierarchy model of the visual nervous system proposed by Hubel and Wiesel is not considered to be entirely correct. It is a future problem to modify the structure of the neocognitron lest it should be contradictory to the structure of the visual system which is now being revealed.

It is conjectured that, in the human brain, the process of recognizing familiar patterns such as alphabets of our native language differs from that of recognizing unfamiliar patterns such as foreign alphabets which we have just begun to learn. The neocognitron probably presents a neural network model corresponding to the former case, in which we recognize patterns intuitively and immediately. It would be another future problem to model the neural mechanism which works in deciphering illegible letters.

The algorithm of information processing proposed in this paper is of great use not only as an inference upon the mechanism of the brain but also to the field of engineering. One of the largest and long-standing difficulties in designing a pattern-recognizing machine has been the problem how to cope with the shift in position and the distortion in shape of the input patterns. The neocognitron proposed in this paper gives a drastic solution to this difficulty. We would be able to extremely improve the performance of pattern recognizers if we introduce this algorithm in the design of the machines. The same principle can 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.

References

Fukushima, K.: Cognitron: a self-organizing multilayered neural network. Biol. Cybernetics **20**, 121–136 (1975)

Fukushima, K.: Improvement in pattern-selectivity of a cognitron (in Japanese). Pap. Tech. Group MBE**78**-27, IECE Japan (1978)

Fukushima, K.: Self-organization of a neural network which gives position-invariant response (in Japanese). Pap. Tech. Group MBE **78**-109, IECE Japan (1979a)

Fukushima, K.: Self-organization of a neural network which gives position-invariant response. In: Proceedings of the Sixth International Joint Conference on Artificial Intelligence. Tokyo, August 20–23, 1979, pp. 291–293 (1979b)

Fukushima, K.: Improvement in pattern-selectivity of a cognitron (in Japanese). Trans. IECE Japan (A), J **62**-A, 650–657 (1979c)

Giebel, H.: Feature extraction and recognition of handwritten characters by homogeneous layers. In: Pattern recognition in biological and technical systems. Grüsser, O.-J., Klinke, R. (eds.), pp. 162–169. Berlin, Heidelberg, New York: Springer 1971

Gross, C.G., Rocha-Miranda, C.E., Bender, D.B.: Visual properties of neurons in inferotemporal cortex of the macaque. J. Neurophysiol. **35**, 96–111 (1972)

Hubel, D.H., Wiesel, T.N.: Receptive fields, binocular interaction and functional architecture in cat's visual cortex. J. Physiol. (London) **160**, 106–154 (1962)

Hubel, D.H., Wiesel, T.N.: Receptive fields and functional architecture in two nonstriate visual area (18 and 19) of the cat. J. Neurophysiol. **28**, 229–289 (1965)

Hubel, D.H., Wiesel, T.N.: Functional architecture of macaque monkey visual cortex. Proc. R. Soc. London, Ser. B **198**, 1–59 (1977)

Kabrisky, M.: A proposed model for visual information processing in the human brain. Urbana, London: Univ. of Illinois Press 1966

Meyer, R.L., Sperry, R.W.: Explanatory models for neuroplasticity in retinotectral connections. In: Plasticity and function in the central nervous system. Stein, D.G., Rosen, J.J., Butters, N. (eds.), pp. 45–63. New York, San Francisco, London: Academic Press 1974

Rosenblatt, F.: Principles of neurodynamics. Washington, D.C.: Spartan Books 1962

Sato, T., Kawamura, T., Iwai, E.: Responsiveness of neurons to visual patterns in inferotemporal cortex of behaving monkeys. J. Physiol. Soc. Jpn. **40**, 285–286 (1978)

Received: October 28, 1979

Dr. Kunihiko Fukushima
NHK Broadcasting Science Research Laboratories
1-10-11, Kinuta, Setagaya
Tokyo 157
Japan