Neural Networks 23 (2010) 926–938

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Neocognitron trained with winner-kill-loser rule

Kunihiko Fukushima *

Kansai University, Takatsuki, Osaka, Japan

ARTICLE INFO

Article history:
Received 16 March 2010
Accepted 27 April 2010

Keywords:
Winner-kill-loser
Competitive learning
Visual pattern recognition
Neocognitron
Disinhibited inhibitory surround
Hierarchical network

ABSTRACT

The *neocognitron*, which was proposed by Fukushima (1980), is a hierarchical multi-layered neural network capable of robust visual pattern recognition. It acquires the ability to recognize patterns through learning.

This paper proposes a new rule for competitive learning, named *winner-kill-loser*, and apply it to the neocognitron. The winner-kill-loser rule resembles the winner-take-all rule. Every time when a training stimulus is presented, non-silent cells compete with each other. The winner, however, not only takes all, but also kills losers. In other words, the winner learns the training stimulus, and losers are removed from the network. If all cells are silent, a new cell is generated and it learns the training stimulus. Thus feature-extracting cells gradually come to distribute uniformly in the feature space. The use of winner-kill-loser rule is not limited to the neocognitron. It is useful for various types of competitive learning, in general.

This paper also proposes several improvements made on the neocognitron: such as, disinhibition to the inhibitory surround in the connections to C-cells (or complex cells) from S-cells (or simple cells); and square root shaped saturation in the input-to-output characteristics of C-cells. As a result of these improvements, the recognition rate of the neocognitron has been largely increased.

1. Introduction

The author previously proposed an artificial neural network *neocognitron* for robust visual pattern recognition (Fukushima, 1980, 1988, 2003; Fukushima & Miyake, 1982). Its architecture was initially suggested by neurophysiological findings on the visual systems of mammals. It is a hierarchical multi-layered network similar to the classical hypothesis of Hubel and Wiesel (1962, 1965). It acquires the ability to robustly recognize visual patterns through learning.

The neocognitron consists of layers of S-cells, which resemble simple cells of the visual cortex, and layers of C-cells, which resemble complex cells. These layers of S-cells and C-cells are arranged alternately in a hierarchical manner. In other words, a number of modules, each of which consists of an S-cell layer and a C-cell layer, are connected in a cascade in the network.

Input connections of S-cells are variable and are modified through learning. After the learning, S-cells come to work as feature-extracting cells, and extract local features from stimulus images presented to the input layer (or photoreceptor array).

C-cells, whose input connections are fixed, exhibit an approximate invariance to the position of the stimuli presented within their receptive fields. We can also express this operation that the response of a layer of S-cells is spatially blurred in the succeeding layer of C-cells.

The C-cells in the highest stage work as recognition cells, which indicate the result of pattern recognition. After learning, the neocognitron can recognize input patterns robustly, with little effect from deformation, change in size, or shift in position.

Varieties of modifications, extensions and applications of the neocognitron, as well as varieties of related networks, have been reported so far (Elliffe, Rolls, & Stringer, 2002; Hildebrandt, 1991; LeCun et al., 1989; LeCun, Bottou, Bengio, & Haffner, 1998; Lo et al., 1995; Riesenhuber & Poggio, 1999; Sato, Kuroiwa, Aso, & Miyake, 1999). They are all hierarchical multi-layered networks and have an architecture of *shared connections*, which is sometimes called a *convolutional net*. They also have a mechanism of pooling outputs of feature-extracting cells. The pooling operation can also be interpreted as a blurring operation. In the neocognitron, the pooling operation, which is done by C-cells, is performed by a nonlinear saturation of the weighted sum of the outputs of feature-extracting S-cells. In some networks, the pooling is realized by simply reducing the density of cells in higher layers. In some other networks, it is replaced by a MAX operation.

This paper proposes a new rule for competitive learning, named *winner-kill-loser*, and apply it to the neocognitron. The winner-kill-loser rule resembles the winner-take-all rule. Every time when a training stimulus is presented, non-silent cells compete with each other. The winner, however, not only takes all, but also kills losers. In other words, the winner learns the training stimulus, and losers

* Permanent address: 634-3, Miwa, Machida, Tokyo 195-0054, Japan. Tel.: +81 44 988 5272; fax: +81 44 988 5272.
E-mail address: fukushima@m.ieice.org.
URL: http://www4.ocn.ne.jp/~fuku_k/index-e.html.

0893-6080/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2010.04.008

are removed from the network. Since cells that are silent to the stimulus do not join the competition, they are not categorized as losers and are not removed from the network. They are expected to work for other stimuli. If all cells are silent, a new cell is generated and learns the training stimulus.

During the learning, a number of training stimuli are presented repeatedly to the network. With the winner-kill-loser rule, generation of new cells and removal of redundant cells are repeated in the network. In the areas where feature-extracting cells are missing in the feature space, new cells are generated. In the areas where similar cells exist duplicately, redundant cells are removed. Thus feature-extracting cells gradually come to distribute uniformly in the feature space.

The use of winner-kill-loser rule is not limited to the neocognitron. It is useful for various types of competitive learning, in general.

This paper also proposes several improvements made on the neocognitron: such as, disinhibition to the inhibitory surround in the connections to C-cells from feature-extracting S-cells; and square root shaped saturation in the input-to-output characteristics of C-cells. As a result of these improvements and the new learning rule, the recognition rate of the neocognitron has been largely increased.

The design of the new neocognitron is inspired from various neurophysiological findings. The new neocognitron, however, is not necessarily intended to be a faithful model of the visual system. We aim to obtain useful algorithms for information processing from the biological system.

This paper mainly discusses the modifications given to the neocognitron, and the mechanisms that are in common with conventional neocognitrons are explained in Appendix. A detailed mathematical description of the whole network also appears in Appendix.

2. Architecture of the network

The neocognitron is a hierarchical multi-layered network. It consists of layers of S-cells, which resemble simple cells in the primary visual cortex, and layers of C-cells, which resemble complex cells. These layers of S-cells and C-cells are arranged alternately in a hierarchical manner. In other words, a number of modules, each of which consists of an S-cell layer and a C-cell layer, are connected in a cascade in the network.

S-cells are feature-extracting cells, whose input connections are variable and are modified through learning. C-cells, whose input connections are fixed, exhibit an approximate invariance to the position of the stimuli presented within their receptive fields.

The C-cells in the highest stage work as recognition cells, which indicate the result of the pattern recognition. After learning, the neocognitron can recognize input patterns robustly, with little effect from deformation, change in size, or shift in position.

Fig. 1 shows the architecture of the network that is discussed in this paper. In the figure, U_{S1}, for example, indicates the layer of S-cells of the 1st stage. The network has four stages of S- and C-cell layers.

Each layer of the network is divided into a number of sub-layers, called *cell-planes*, depending on the difference in the features to which cells respond preferentially. Incidentally, a cell-plane is a group of cells that are arranged retinotopically and share the same set of input connections (Fukushima, 1980, 1988). As a result, all cells in a cell-plane have receptive fields of an identical characteristic, but the locations of the receptive fields differ from cell to cell. For example, in layer U_{S1} of edge-extracting cells, cells of the same preferred orientation constitute a cell-plane.

The stimulus pattern is presented to input layer U_0, which consists of a two-dimensional array of photoreceptors. The output

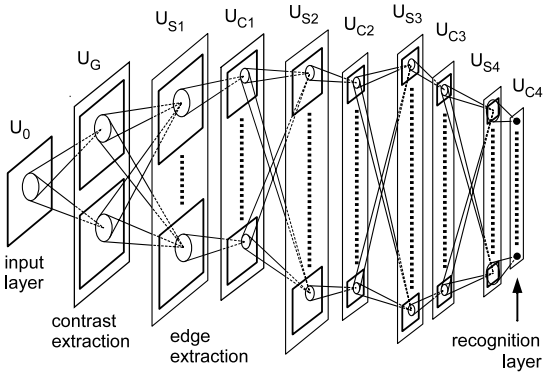


Fig. 1. The architecture of the neocognitron (Fukushima, 2003).

of layer U_0 is fed to contrast-extracting layer U_G, whose cells resemble retinal ganglion cells or lateral geniculate nucleus cells. Layer U_G consists of two cell-planes: one with concentric on-center receptive fields, and the other with off-center receptive fields. The former cells extract positive contrast in brightness, whereas the latter extract negative contrast from the image presented to the input layer.

The output of layer U_G is fed to edge-extracting layer U_{S1}. The S-cells of this layer resemble simple cells in the primary visual cortex, and respond selectively to edges of a particular orientation. Namely, layer U_{S1} consists of K_1 cell-planes, and all cells in the kth cell-plane respond selectively to edges of orientation $2\pi k/K_1$. As a result, the contours of the input image are decomposed into edges of every orientation.

S-cell layers of intermediate stages, U_{S2} and U_{S3}, are trained by *winner-kill-loser* rule, which is discussed below in more detail.

Training method for the S-cell layer of the highest stage, U_{S4}, is almost the same as for the conventional neocognitron (Fukushima, 2003), and is explained in Appendix D.4.

Since the main structure of the new neocognitron is almost the same as that of the conventional neocognitron (Fukushima, 2003), we mainly discuss the difference from the conventional neocognitron in the text, and a detailed explanation of the whole network appears in Appendix.

3. Winner-kill-loser rule

We propose a new learning rule, which we call *winner-kill-loser*. It is a competitive learning method similar to the winner-take-all rule, but the winner, not only takes all, but also kills losers.

We use this learning rule for training U_{S2} and U_{S3} (layers of S-cells of the 2nd and the 3rd stages). In this section, we first discuss the basic idea of the learning rule, which can be applied to pattern recognition systems in general. We then discuss how to apply the rule to the neocognitron, which has architecture of shared connections.

3.1. Feature-extracting S-cells

Before explaining the winner-kill-loser rule, we discuss the behavior of S-cells, which work as feature-extracting cells after having finished learning. Fig. 2 shows an equivalent circuit of an S-cell. As can be seen from the figure, the inhibitory signal works in a divisional manner.

To show the essence of the learning algorithm, we extract only the circuit converging to a single S-cell and analyze its behavior. Fig. 3 shows the circuit. The S-cell receives excitatory signals directly from a group of C-cells, which are cells of the preceding layer. It also receives an inhibitory signal through a V-cell, which accompanies the S-cell. The V-cell receives fixed

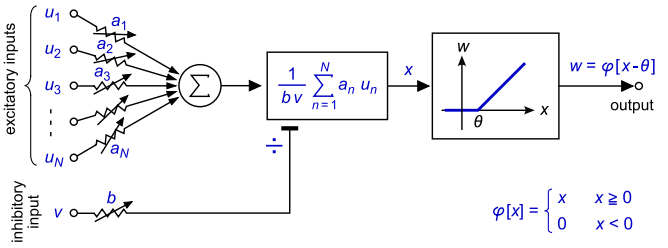


Fig. 2. An equivalent circuit of an S-cell.

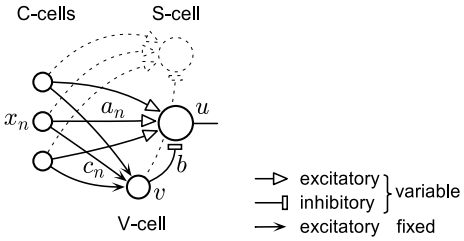


Fig. 3. Input connections converging to a feature-extracting S-cell. Broken thin lines show that, in the network, there are other S-cells that receive signals from the same set of C-cells.

excitatory connections from the same group of C-cells as does the S-cell, and always responds with the average intensity (root-mean-square) of the output of the C-cells.

Let a_n be the strength of the excitatory variable connection to the S-cell from the nth C-cell, whose output is x_n. We also use vector notation $\boldsymbol{x}$ to represent the response of all C-cells, from which the S-cell receive excitatory signals. Namely,

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_n, \ldots). \tag{1}$$

Let b be the strength of the inhibitory variable connection from the inhibitory V-cell, whose output is v.

The output u of the S-cell is given by

$$u = \frac{1}{1-\theta} \cdot \varphi\left[\frac{\sum_n a_n x_n}{b\, v} - \theta\right], \tag{2}$$

where $\varphi[\ ]$ is a function defined by $\varphi[x] = \max(x, 0)$. θ is the threshold of the S-cell ($0 < \theta < 1$). The response of the V-cell is given by

$$v = \sqrt{\sum_n c_n {x_n}^2}, \tag{3}$$

where c_n is the strength of the fixed excitatory connection from the nth C-cell.

We now define *weighted* inner product of arbitrary two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ by

$$(\boldsymbol{x}, \boldsymbol{y}) = \sum_n c_n x_n y_n, \tag{4}$$

where the strength of the input connections to the V-cell, c_n, is used as the weight for the inner product. We also define the norm of a vector, using the *weighted* inner product, by

$$\|\boldsymbol{x}\| = \sqrt{(\boldsymbol{x}, \boldsymbol{x})}. \tag{5}$$

Using this vector notation, the response of the V-cell, which is given by (3), can also be expressed by

$$v = \|\boldsymbol{x}\|. \tag{6}$$

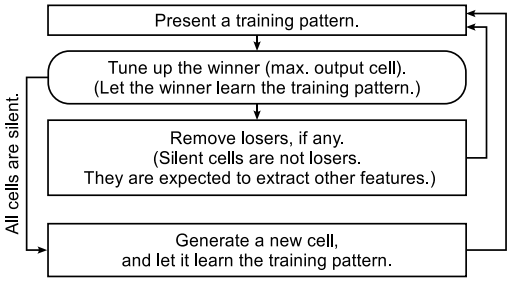


Fig. 4. The process of competitive learning with winner-kill-loser.

3.2. Competitive learning with winner-kill-loser rule

3.2.1. Outline of the learning rule

Training of S-cells progresses through a competitive learning with winner-kill-loser rule. The process of the learning is illustrated in Fig. 4. (See also Fig. 6).

During the learning phase, the output of the C-cells, which is represented by $\boldsymbol{X}$, works as a training vector for the S-cells. When a training vector is presented, the S-cells compete with other S-cells, and the one that yields the largest response to this training vector becomes the winner. The winner learns the training vector, and has its input connections renewed. At the same time, losers are removed from the network. Since cells that are silent to the training vector do not join the competition, they are not categorized as losers and are not removed from the network. They are expected to work for other stimuli. If there is no winner, a new S-cell is generated, and the generated S-cell learns the training vector.

We use dual threshold of S-cells for the learning and the recognition phases (Fukushima & Tanigawa, 1996). Namely, during the competitive learning, the threshold of S-cells is set to a higher value, θ^{L}, than the threshold θ^{R} for the recognition. S-cells join the competition, only when their responses to a training vector are not zero under the high threshold θ^{L} for the learning. This means that, even if an S-cell can yield a non-zero response under the low threshold θ^{R} for the recognition, the S-cell does not join the competition if it is silent under the high threshold θ^{L} for the learning.[1] As mentioned above, silent S-cells, which do not join the competition, do not have their input connections renewed and are not removed, either. Incidentally, the method of dual threshold is useful, not only for the neocognitron, but also for various competitive learning methods in general.

These processes are discussed below in more detail.

3.2.2. Renewing input connections of the winner

Let an S-cell has become a winner for a training vector $\boldsymbol{X}$. The nth excitatory input connection a_n is increased in proportion to the output X_n of the presynaptic C-cell, from which the connection is leading. Namely, the amount of increase is

$$\Delta a_n = c_n X_n, \tag{7}$$

where c_n is the value of the fixed input connection to the inhibitory V-cell.

The S-cell usually becomes a winner several times during the training phase. Let $\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, \ldots, \boldsymbol{X}^{(t)}, \ldots$ be the training vectors that have made the S-cell a winner. We write

$$\boldsymbol{X} = \sum_t \boldsymbol{X}^{(t)}. \tag{8}$$

[1] Here we express the learning process assuming the use of dual threshold for the learning and for the recognition. We can also express this process in another way, assuming that S-cells always take the same threshold value θ^{R} and that only S-cells whose outputs are larger than a certain value Θ join the competition. Both expressions are equivalent, if we take $\Theta = (\theta^{\mathrm{L}} - \theta^{\mathrm{R}})/(1 - \theta^{\mathrm{L}})$.

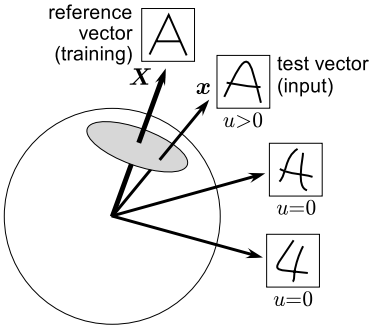


Fig. 5. Response of an S-cell in a multi-dimensional feature space.

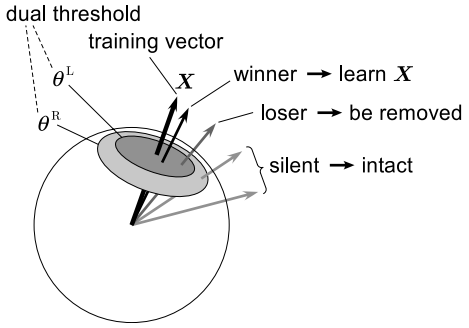


Fig. 6. Competition in the multi-dimensional feature space. The threshold for the learning θ^{L}, which is higher than the threshold for the recognition θ^{R}, is used during the competition. The winner is tuned up by learning the training vector, and losers are removed from the network. Silent cells are kept intact, because they do not join the competition.

After having become winners for these training vectors, the strength of the input connection a_n becomes

$$a_n = \sum_t c_n X_n^{(t)}. \tag{9}$$

The inhibitory input connection b of the winner is renewed depending on the renewed strengths of the excitatory connections a_n. That is,

$$b = \sqrt{\sum_n \frac{{a_n}^2}{c_n}}. \tag{10}$$

Using vector notation and weighted inner product defined by Eq. (4), we have

$$\sum_n a_n x_n = (\boldsymbol{X}, \boldsymbol{x}) \tag{11}$$

and

$$b = \|\boldsymbol{X}\|. \tag{12}$$

Substituting (11), (12) and (6) to (2), we have

$$u = \frac{\varphi[s-\theta]}{1-\theta}, \tag{13}$$

where

$$s = \frac{(\boldsymbol{X}, \boldsymbol{x})}{\|\boldsymbol{X}\| \cdot \|\boldsymbol{x}\|}. \tag{14}$$

In the multi-dimensional feature space, s shows a kind of similarity between $\boldsymbol{x}$ and $\boldsymbol{X}$. We call $\boldsymbol{X}$, which is the sum of the training vectors, the reference vector of the S-cell. We sometimes call $\boldsymbol{x}$, which is the stimulus to the S-cell, the test vector.

Eq. (13) shows that the response of the S-cell takes a maximum value 1 when the test vector is identical to the reference vector, and becomes 0 if the similarity s is less than the threshold θ of the cell. In the multi-dimensional feature space, the area that satisfies $s < \theta$ becomes the tolerance area in feature extraction by the S-cell, and the threshold θ determines the size of the tolerance area. In other words, a non-zero response is elicited from the S-cell, if and only if the test vector $\boldsymbol{x}$ is within a tolerance area around the reference vector $\boldsymbol{X}$ (Fig. 5).

The selectivity of the S-cell to its preferred feature (or the reference vector) can thus be controlled by the threshold θ. A higher value of θ produces a smaller tolerance area. If the threshold is low, the radius of the tolerance area becomes large, and the S-cell responds even to features somewhat deformed from the reference vector. This makes a situation like a population coding of features rather than grandmother cell theory: many S-cells respond to a single feature if the response of an entire layer is observed. This situation of low threshold in the recognition phase usually produces a better recognition rate of the neocognitron.

As mentioned before, S-cells take a higher threshold value θ^{L} during the learning phase, and S-cells join the competition only when their outputs are not zero. This means that, only S-cells whose reference vectors have similarity larger than θ^{L} ($>\theta^{\mathrm{R}}$) join the competition. The training vector is added to the reference vector of the winner, which has the largest similarity to the training vector.

3.2.3. Removing losers from the network

If a training vector elicits non-zero responses from two or more S-cells, it means that preferred features of these cells resemble each other, and that they work redundantly in the network. Hence only the winner has its input connections renewed to fit more to the training vector, and the other cells, namely losers, are removed from the network. This is the basic idea of the winner-kill-loser rule (Fig. 6).

Since silent S-cells (namely, the S-cells whose response to the training vector are zero) do not join the competition, they are not removed, even if they might yield non-zero responses under the lower threshold θ^{R} for the recognition phase. These cells are expected to work for extracting other features.

3.2.4. Generating new cells

If no winner appears for a training vector, a new S-cell is generated. The initial value of the excitatory input connections of the generated S-cell is given by

$$a_n = c_n X_n, \tag{15}$$

which is the same value produced by (7). The initial value of the inhibitory input connection b is determined from the excitatory connections a_n by (10).

3.2.5. Merit of winner-kill-loser

In the learning phase, a number of training vectors are presented sequentially to the network. During this process, generation of new cells and removal of redundant cells are repeated in the network (Fig. 7). In the areas where feature-extracting cells are missing in the multi-dimensional feature space, new cells are generated. In the areas where similar cells exist duplicately, redundant cells are removed. By the repetition of this process, reference vectors of the cells gradually come to distribute uniformly in the multi-dimensional feature space.

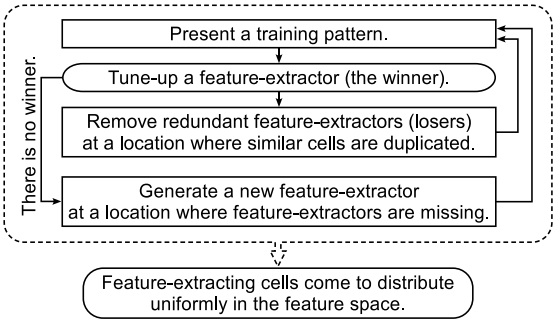


Fig. 7. The merit of winner-kill-loser rule.

3.3. Winner-kill-loser in the neocognitron

To apply this learning rule to the neocognitron, some modifications are required. This is because several S-cells, which have receptive fields at different locations on the input layer U_0, have to be trained simultaneously by a single presentation of a training pattern to the input layer U_0. It is also because cells of the neocognitron have shared connections. This section discusses the basic idea qualitatively, and a more exact mathematical description appears in Appendix.

In the new neocognitron, the winner-kill-loser rule is applied to intermediate layers (U_{S2} and U_{S3}). Let l ($l = 2$ or 3) be the layer that is now being trained. We assume that the training of the preceding stages has already been completely finished. S-cells of U_{Sl} take the threshold θ^{L} for the learning. S-cells of the preceding stages take the threshold θ^{R} for the recognition.

In the neocognitron, each layer of cells is divided into subgroups called *cell-planes*. All cells in a cell-plane share the same set of input connections. This condition of shared connections has to be kept even during the learning phase, when input connections to S-cells are to be renewed. If a winner is chosen from a cell-plane, its input connections are renewed following the responses of the C-cells presynaptic to it. This is the same process discussed in 3.2.2. Since all cells in the cell-plane share the same set of connections, all other cells in the cell-plane come to have the same connections as the winner. This means that the winner works like a seed in crystal growth. Hence we call it a *seed-cell*.

Every time when a training pattern is presented to the input layer U_0, S-cells in layer U_{Sl} compete with each other. Since there are a number of cell-planes, and since there are S-cells that have receptive fields at different locations on U_0, a number of seed-cells are usually chosen at a single presentation of a training pattern. The seed-cells have their input connections renewed, and losers of the competition are removed from the layer.

Only non-silent S-cells join the competition. A non-silent cell means a cell from which non-zero response is elicited by the training pattern.

Renewing input connections of seed-cells and removal of losers are performed step by step, in the order of output values of the winners, which might become seed-cells if some conditions are satisfied (Fig. 8).

At first, the S-cell whose response is the largest in the layer is chosen as a seed-cell. The seed-cell has its input connections renewed depending on the training vector presented to it. Once the connections to the seed-cell are renewed, all cells in the cell-plane, from which the seed-cell is chosen, come to have the same set of input connections as the seed-cell because of the shared connections.

If, there are any other non-silent cells at the location of the seed-cell, they are losers. The cell-planes, to which losers belong, are removed from the layer. It should be noted here that silent cells are not treated as losers, because they do not join the competition.

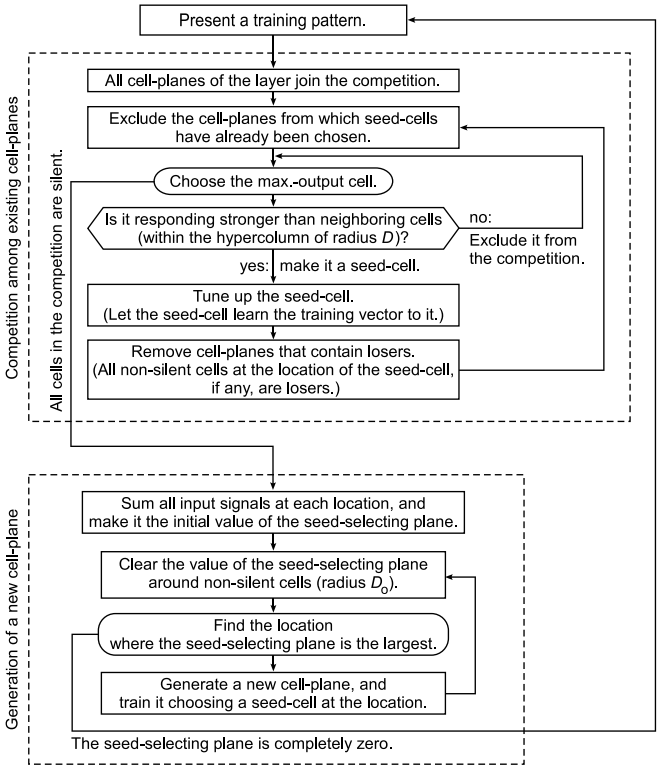


Fig. 8. The winner-kill-loser rule applied to the neocognitron.

Once a seed-cell has been chosen and had its input connections renewed, we repeat the following process to choose other seed-cells. Excluding the cells of the cell-planes from which seed-cells have already been chosen, we search the maximum response cell, and make it a candidate of the next seed-cell. The candidate becomes a seed-cell, if no other cell is responding stronger than it within its competition area of radius D, which has a shape of a hypercolumn (Fukushima, 1980, 1988). If the candidate is defeated by other cells in its competition area, we again search the largest response cell, excluding the defeated candidate, and make it the next candidate. Incidentally, the cells that respond stronger than the candidate, if any, come from cell-planes from which seed-cells have already been chosen.

Every time when a seed-cell has thus been chosen, its input connections are renewed depending on the training vector presented to it. This means that all cells in the cell-plane, from which the seed-cell is chosen, come to have the same set of input connections as the seed-cell.

At the same time, if there are losers at the location of the seed-cell, the cell-planes, to which losers belong, are removed from the layer. The cells in the cell-planes, from which seed-cells have already been chosen, however, do not become losers regardless of the values of their responses, because they do not join the competition any more for this training pattern.

After having finished choosing seed-cells, the renewal of connections of seed-cells, and the removal of cell-planes that include losers, we go to the process of generating new cell-planes.

If there is a location where no S-cell is responding to non-silent input signals, a new cell-plane is generated there. In other words, if there is a location where the response of the C-cell of the preceding layer is not silent (to be more exact, the output of the C-cell is larger than a certain small threshold), and if all S-cells are silent within the hypercolumn of radius D_o ($<D$) around it, a new cell-plane is generated.

To make this process easier, we use a virtual cell-plane called *seed-selecting plane*. The outputs of the C-cells of all cell-planes

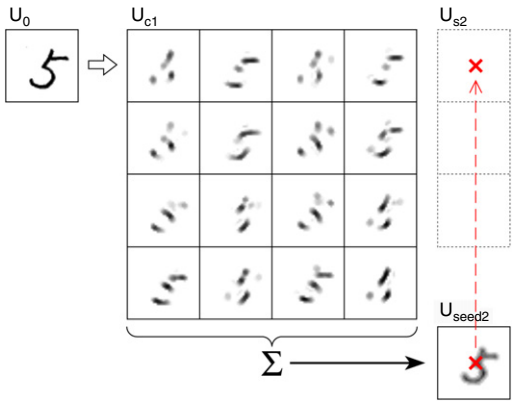


Fig. 9. The initial value of seed-selecting plane $U_{\text{seed}2}$ is calculated from the response of the preceding layer U_{C1}. The location of the maximum output cell of $U_{\text{seed}2}$ indicates the location of the seed-cell of the newly generated cell-plane of U_{S2}.

of the preceding C-cell layer are summed and copied to the seed-selecting plane (Fig. 9). A small threshold value, which is determined as a certain percentage of the maximum output in the seed-selecting plane, is subtracted from the output of each cell in the seed-selecting plane. If the subtracted value is negative, it is replaced with zero. The output of a cell of the seed-selecting plane is inhibited to zero, if any S-cell yields a non-zero response in the hypercolumn of radius D_o around it. We then search the maximum output cell in the seed-selecting plane, and its location is used as the location of the seed-cell for the newly generated cell-plane. The input connections of the newly generated cell-plane are determined by the use of the seed-cell. We then calculate the response of the newly generated cell-planes to the training pattern. The output of a cell of the seed-selecting plane is again inhibited to zero, if there is any non-silent S-cell of the newly generated cell-plane within the area of radius D_o around it. We repeat this process, until all cells of the seed-selecting plane become silent.

We then proceed to the presentation of the next training pattern.

4. C-cell layers

In each stage ($l \leq 3$), except the highest stage, the layer U_{Cl} of C-cells have the same number of cell-planes as the layer U_{Sl} of S-cells. In other words, there is one-to-one correspondence between the cell-planes of the two layers in the same stage.

Basically, a C-cell has fixed excitatory connections from a group of S-cells of the corresponding cell-planes of S-cells. Through these connections, the response of a cell-plane of S-cells is spatially blurred in the succeeding cell-planes of C-cells. The blurring operation is very important for endowing neural networks with an ability to recognize patterns robustly. In this paper, however,

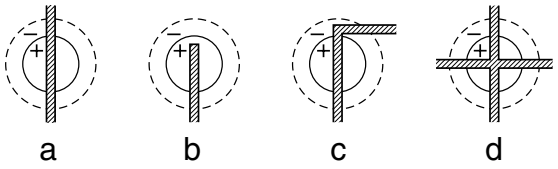


Fig. 11. Various stimuli presented to the center and the surround of the receptive field of a C-cell, whose preferred orientation is vertical. Responses to preferred stimuli presented to the center are modulated by stimuli in the surround.

we do not discuss the importance of the blurring operation any more, because it has already been discussed in many papers so far (e.g., Fukushima, 1989).

Since neighboring C-cells come to show similar responses as a result of the blur, the density of C-cells in each cell-plane is reduced from that of S-cells. The reduction of the density is made by a thinning-out of the cells. In our system, we make 2: 1 thinning-out in both horizontal and vertical directions.

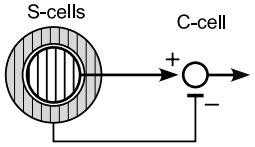
4.1. Inhibitory surround

In layers U_{C1} and U_{C2}, an inhibitory surround is introduced around the excitatory input connections of a C-cell as shown in Fig. 10(a) (Fukushima, 2003). Through the connections, each C-cell receives excitatory and inhibitory signals from S-cells of a preceding cell-plane. These S-cells have the same preferred orientation.

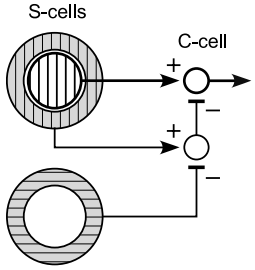
The concentric inhibitory surround endows the C-cells with the characteristics of end-stopped cells, and C-cells behave like hypercomplex cells of the visual cortex (Hubel & Wiesel, 1965). Incidentally, it is known that visual cortical neurons have both a center, or classical receptive field, where stimuli elicit spike responses, and a surround, or extraclassical receptive field, where stimuli modulate responses due to stimulation of the classical receptive field. The presence of a surround stimulus of orientation similar to the cell's preferred orientation suppresses the response to an optimal stimulus within the receptive field center (Jones, Grieve, Wang, & Sillito, 2001; Ozeki et al., 2004; Seriès, Lorenceau, & Frégnac, 2003; Walker, Ohzawa, & Freeman, 2000). Although the mechanism of creating surround suppression might slightly different from that in the brain, C-cells show a similar behavior.

Bend points and end points of lines are important features for pattern recognition. C-cells, whose input connections have inhibitory surrounds, participate in extraction of bend points and end points of lines while they are making a blurring operation. Stimulus like Fig. 11(b) elicits a larger response from the C-cell than the stimulus like Fig. 11(a). In other words, the response from an end of line becomes larger than that from a middle point of the line. Thus, S-cells of the next stage can easily detect end points of lines.

The inhibitory surrounds in the connections also have another benefit. The blurring operation by C-cells, which usually is effective



(a) Inhibitory surround in the connections to a C-cell.



(b) Inhibition from the surround is disinhibited by perpendicularly oriented edges.

Fig. 10. Connections to a C-cell. Preferred orientations of S-cells are indicated by the orientation of the hatches.

for improving robustness against deformation of input patterns, sometimes makes it difficult to detect whether a lump of blurred response is generated by a single feature or by two independent features of the same kind. For example, a single line and a pair of parallel lines of a very narrow separation generate a similar response when they are blurred. The inhibitory surround in the connections to C-cells creates a non-responding zone between the two lumps of blurred responses. This silent zone makes the S-cells of the next stage easily detect the number of original features even after blurring (Fukushima, 2003).

4.2. Disinhibition to inhibitory surround

In the 1st stage of the new neocognitron, connections to a C-cell have also inhibitory surround, but the inhibition from the surround is suppressed by disinhibition from signals of S-cells of orthogonal preferred orientation (Fig. 10(b)). It should be noted here that S-cells of the 1st stage (U_{S1}) have been trained to extract oriented edges.

Incidentally, a number of electrophysiological studies have reported that the stimulation of the surrounding area of classical receptive field facilitates responses to optimal stimuli within the center, when the stimulus to the surround has an orientation significantly different from the cell's preferred orientation. Several models have been proposed to account for the mechanism of facilitation, and some of them hypothesize the mechanism of disinhibition (Dragoi & Sur, 2000; Ozeki et al., 2004; Seriès et al., 2003). The network of the new neocognitron was suggested from these electrophysiological studies.

Suppose a stimulus like Fig. 11(c) be presented to the receptive field of a C-cell, whose preferred orientation is vertical. The inhibitory signal elicited by the vertical line component in the surround is disinhibited by the horizontal line component in the surround. Hence a larger response is elicited from the C-cell by stimulus (c) than (b). Thus, the disinhibition makes bend points of lines, which are an important feature for recognizing patterns, more salient.

Suppose another stimulus like Fig. 11(d) be presented. Since the response of presynaptic S-cells, which also have vertical preferred orientation as the C-cell, are significantly suppressed by the presence of horizontal line component, the excitatory input to the C-cell is also suppressed compared to the input by a stimulus like Fig. 11(a), if the mechanism of disinhibition does not exist. However, the inhibition to the C-cell from vertical line component in the surround can be removed by the disinhibition by the horizontal line component in the surround. Hence the depression of response of S-cells at the crossing point of lines can be recovered in the response of the C-cell. This also helps to improve the recognition rate.

5. Nonlinearity of a C-cell

Since a role of a C-cell is to detect whether any of its presynaptic S-cells is active, it is better to have some saturation in the input-to-output characteristic of the cell. In the new neocognitron, the nonlinearity is determined by a square root function, as shown in Fig. 12.

We now consider a network like Fig. 13. Let ξ_n be the weighted sum of all inputs to the nth C-cell. The output of the C-cell is given by $x_n = \sqrt{\varphi[\xi_n]}$. (See (E.1) in Appendix for more exact expressions). Suppose the level of all input signals to the group of C-cells be increased by a factor of, say β. In other words, vector $\boldsymbol{\xi}$ is increased by a factor of β. As a result, the output of these C-cells, which is shown by vector $\boldsymbol{x}$, increases by a factor of $\sqrt{\beta}$. The response u of the S-cell, however, does not change, because similarity s in (14) is not affected by $\sqrt{\beta}$, the level of $\boldsymbol{x}$.

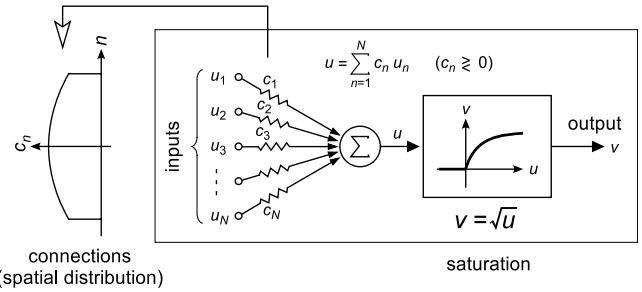


Fig. 12. The saturation in input-to-output characteristic of a C-cell, and the spatial distribution of its input connections. Inhibitory surround in the connections, which is introduced in C-cells of the 1st and the 2nd stages, is abbreviated from this illustration.

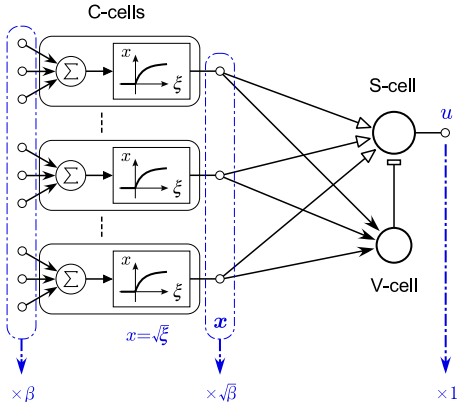


Fig. 13. The effect of amplitude change in the inputs to a group of C-cells, from which an S-cell receives signals. If all input signals to the group of C-cells increase by a factor of β, the outputs of the C-cells, which are shown by vector $\boldsymbol{x}$, increase by a factor of $\sqrt{\beta}$, but the output of the S-cell, which is given by (13) and (14), does not change.

In contrast to this, in the conventional neocognitron (e.g., Fukushima, 2003), the response of the C-cell was given by $x_n = \varphi[\xi_n]/(\alpha + \varphi[\xi_n])$, and the optimal value of α changed depending on the level of $\boldsymbol{\xi}$. In the new neocognitron, a parameter such as α is not required.

6. Computer simulation

We tested the behavior of the neocognitron by computer simulation using handwritten digits (free writing) randomly sampled from the ETL1 database.[2] Incidentally, the ETL1 is a large database of segmented handwritten characters written by about 1400 different writers. We compared recognition rate of the new neocognitron with that of the conventional neocognitron (Fukushima, 2003).

The scale of the network of the new neocognitron and the parameters for the simulation are shown in Appendix.

Under different numbers of training patterns in a training set, we measured the recognition rates. The process of the experiment is as follows:

1. Prepare $m + 1$ sets of 1000 patterns (100 writers × 10 digits) randomly sampled from the ETL1 database.
2. Use m blocks (or sets) for the training, and use the remaining 1 block for measuring the recognition rate.

[2] http://www.is.aist.go.jp/etlcdb/#English.

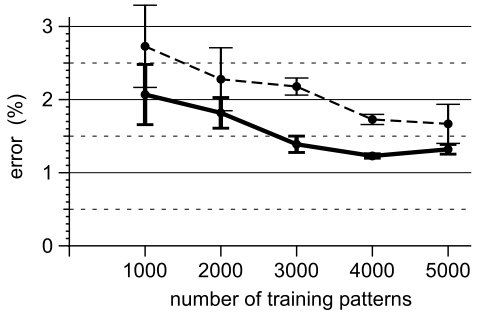


Fig. 14. Error rates and their standard deviations in the recognition by the new (solid line) and the conventional (broken line) neocognitrons trained with different numbers of training patterns. Since there is no rejection in the recognition, recognition rate is equal to (100% − [error rate]).

3. Changing the block that is used for measuring the recognition rate, repeat the experiment $m+1$ times, and calculate the final recognition rate by averaging the recognition rates of the $m+1$ experiments.

We measured the final recognition rates under different numbers of training patterns, namely, 1000, 2000, …, 5000 training patterns, where $m = 1, 2, \ldots, 5$, respectively.

We repeated the above experiment three times for each number of training patterns, using pattern sets obtained by different random samplings. Fig. 14 shows the average and the standard deviation of the error rates of the three experiments. Incidentally, the recognition rate is equal to (100% − [error rate]), because we do not accept rejection in the recognition in these experiments.

From this figure, we can see a superiority of the new neocognitron over the conventional one.

7. Discussions

This paper proposed the winner-kill-loser rule for competitive learning, and applied it to the neocognitron. The winner-kill-loser rule resembles the winner-take-all rule, but the winner, not only takes all, but also kills losers. In the areas where similar cells exist duplicately in the feature space, redundant cells, which become losers, can be removed. Silent cells are not categorized as losers and are kept intact. If all cells are silent, a new cell is generated and learns the training stimulus. Through a repeated presentation of training stimuli during the learning phase, feature-extracting cells gradually come to distribute uniformly in the feature space.

A uniform distribution of feature extractors in a feature space is desirable for pattern recognition by a multi-layered network. If a larger number of feature-extracting cells come to respond to a particular feature than other, the feature comes to affect the result of pattern recognition much stronger. This relatively reduces the effect of other features, and is not desirable for pattern recognition. The undesirable non-uniformity can be reduced by the winner-kill-loser rule.

The use of dual threshold for feature-extracting cells is important when training networks with the winner-kill-loser rule. Namely, a higher threshold is used for the learning phase than for the recognition phase. Since all non-silent cells except the winner are removed from the network, each training stimulus comes to elicit a response from only one cell. If the higher threshold for the learning is still used during the recognition phase when the learning has been finished, feature-extracting cells in the network will behave like grandmother cells. Hence the robustness of the network against deformations is lost. When a stimulus that is slightly deformed from a training stimulus is presented to the network, we will have a situation where either no cell responds or completely different cells might respond. This means that a slight change in the stimulus elicits a completely different response from the network, and that the network loses an ability to generalize stimuli.

Since we actually use a lower threshold for the recognition phase, a number of feature-extracting cells come to respond to each stimulus, and we can have a situation like a population coding. This largely increases the robustness of the network. The use of winner-kill-loser rule together with dual threshold thus helps homogenizing the distribution of feature extractors in the feature space, and, at the same time, producing a situation like a population coding in the recognition phase.

Although this paper mainly discussed the application of the winner-kill-loser rule to the neocognitron, the use of winner-kill-loser rule is not limited to the neocognitron. It will be useful for various types of competitive learning, in general.

This paper also proposed several modifications made on the neocognitron. We already knew the effect of inhibitory surround in the connections to C-cells (Fukushima, 2003). Adding disinhibition to the inhibitory surround for U_{C1}, we had a further increase in the recognition rate. The square root shaped saturation in the input-to-output characteristics of C-cells is also effective for making the network robust against changes in intensity of input signals.

In the computer simulation, we searched the optimal thresholds that produce the best recognition rate. Since there are a large number of combinations in the threshold values of three layers (θ_1^{R}, θ_2^{L}, θ_2^{R}, θ_3^{L} and θ_3^{R}), however, a complete search for all combinations has not been finished yet. We showed here a result with a set of threshold values that seems to be nearly optimal.

Although we have not mentioned explicitly in the text, the number of parameters that have to be determined when designing a neocognitron has been largely decreased compared with that in the conventional neocognitron.[3] The characteristic of an S-cell, which is expressed by (2) has also been simplified from the original one.

Acknowledgements

The author thanks Prof. Isao Hayashi (Kansai University), Dr. Hayaru Shouno (University of Electro-Communications) and Dr. Masayuki Kikuchi (Tokyo University of Technology) for critical discussions, helpful comments and suggestions. The author also thanks Mr. Yuki Makino (Tokyo University of Technology) for measuring recognition rates of varieties of neocognitrons by computer simulation. This work was partially supported by Strategic Project to Support the Formation of Research Bases at Private Universities: Matching Fund Subsidy from MEXT, 2008–2012.

Appendix A. Scale of the network

The new neocognitron discussed in this paper is a four staged network as was shown in Fig. 1.

Although there are a number of cell-planes in each layer, the number of connections converging to a cell and the radius of their spatial spread are the same for all cell-planes of a layer. Fig. A.15 shows a one-dimensional cross-section of connections between cell-planes.

The figure also shows the total number of cells (not counting inhibitory V-cells) in each layer. The number of cells in each cell-plane is pre-determined for all layers when designing the network.

[3] For example, parameters such as q_l, θ_4^{R}, θ_4^{L}, and many others, which were used in the conventional neocognitron (Fukushima, 2003), are not required in the new neocognitron. Parameter α mentioned in Section 5, and several parameters related to seed-selecting planes, can also be omitted.

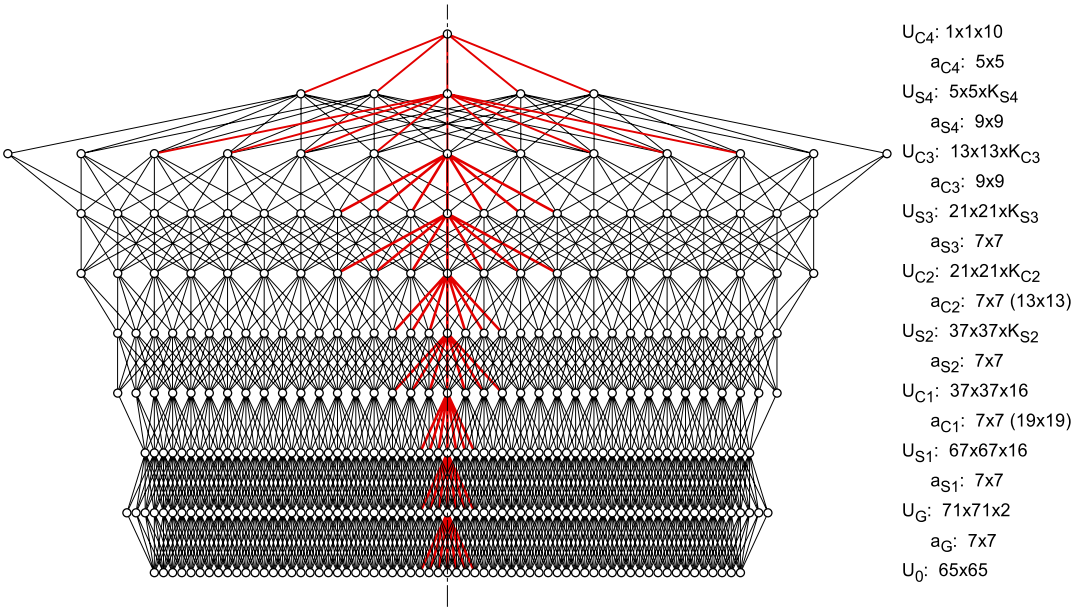


Fig. A.15. Arrangement of cells and connections in the network. A one-dimensional cross-section of connections between cell-planes is drawn. Since the spatial spread of connections converging to a cell is not square but actually is circular, only approximate numbers of connections are shown in the figure. Only positive connections are drawn for a_{C1} and a_{C2}. To clearly show the input connections converging to a single cell, a cell is arbitrarily chosen from each layer, and its input connections are drawn with heavy lines.

Table A.1
Numbers of cell-planes generated in each layer by the learning with 3000 training patterns. The table shows the minimum, maximum and average numbers of K_{Sl} generated in the 12 trials.

	Min	Max	Average
K_{S2}	24	43	34.6
K_{S3}	100	175	136.7
K_{S4}	100	138	116.7

Table A.2
Numbers of cell-planes generated in each layer by the learning with 4000 training patterns. The table shows the minimum, maximum and average numbers of K_{Sl} generated in the 15 trials.

	Min	Max	Average
K_{S2}	26	36	32.6
K_{S3}	109	149	129.1
K_{S4}	114	157	134.3

The reduction in density of cells in cell-planes, namely the thinning-out of the cells, is made from a S-cell layer to the C-cell layer of the same stage. The ratio of the thinning-out from U_{Sl} to U_{Cl} is 2: 1 (in both horizontal and vertical directions) in all stages except U_{C4}.

Layer U_{S1} consists of edge-extracting cells. The number of its cell-planes is $K_{S1} = 16$, which is determined when designing the network (see Appendix D.2).

For other stages higher than it ($l \geq 2$), the number of cell-planes in each S-cell layer (K_{Sl}) is determined automatically by the learning, depending on the training set, which is produced by random sampling from the database.

In each stage except the highest one, the same number of cell-planes is generated in the C-cell layer as in the S-cell layer. Namely, $K_{Cl} = K_{Sl}$ ($l \leq 3$). The recognition layer U_{C4}, however, has only $K_{C4} = 10$ cell-planes, which correspond to the 10 digits to be recognized, and each cell-plane contains only one C-cell.

Tables A.1 and A.2 show the numbers of cell-planes of U_{Sl} that were actually generated in the simulation discussed in Section 6. Changing the size of the training set, which is given by 1000 m, we tested to train the network $3(m+1)$ times under different training sets of the same size. These tables show the minimum, maximum and average numbers of K_{Sl} generated in the $3(m+1)$ trials, when trained with 3000 (Table A.1, $m = 3$) and 4000 (Table A.2, $m = 4$) training patterns, respectively.

The sizes and the shapes of connections converging to single cells (namely, radiuses A_G, A_{Sl}, A_{Cl}, etc.) are shown below in connections with mathematical equations. In the equations, the unit of length is different from layer to layer: namely, the pitch of cells in the cell-plane of the preceding layer is taken as the unit of length. [4]

Appendix B. Extraction of brightness contrast

Contrast-extracting cell of layer U_G have concentric on- and off-center receptive fields.

Let the output of a photoreceptor cell of input layer U_0 be $u_0(\boldsymbol{n})$, where $\boldsymbol{n}$ represent the location of the cell. The output of a contrast-extracting cell of layer U_G, whose receptive field center is located at $\boldsymbol{n}$, is given by

$$u_G(\boldsymbol{n}, k) = \varphi\left[(-1)^k \sum_{|\boldsymbol{v}|<A_G} a_G(\boldsymbol{v}) \cdot u_0(\boldsymbol{n}+\boldsymbol{v})\right], \quad (k = 1, 2), \quad \text{(B.1)}$$

where $\varphi[\ ]$ is a function defined by $\varphi[x] = \max(x, 0)$. Parameter $a_G(\boldsymbol{\xi})$ represents the strength of fixed connections to the cell and takes the shape of a Mexican hat. Layer U_G has two cell-planes: one consisting of on-center cells ($k = 2$) and one of off-center cells ($k = 1$). A_G denotes the radius of summation range of $\boldsymbol{v}$, that is, the size of spatial spread of the input connections to a cell.

The input connections to a single cell are designed in such a way that their total sum is equal to zero. In other words, the connection

[4] It should be noted here that, if two layers have the same radius based on the pitch of the cells in their respective layers, the actual size of the connections measured with the scale of the input layer is larger in the higher layer, because the density of cells is lower in the higher layer.

$a_G(\boldsymbol{v})$ is designed so as to satisfy

$$\sum_{|\boldsymbol{v}|<A_G} a_G(\boldsymbol{v}) = 0. \tag{B.2}$$

This means that the dc component of spatial frequency of the input pattern is eliminated in the contrast-extracting layer U_G. As a result, the output from layer U_G is zero in the area where the brightness of the input pattern is flat.

In the computer simulation, the shape of $a_G(\boldsymbol{\xi})$ is determined by a DoG (difference of Gaussian) function, but its value is truncated to zero in the outside of the connecting area of radius A_G. To be more specific, we first define a truncated DoG function $a'_G(\boldsymbol{\xi})$ by

$$a'_G(\boldsymbol{v}) = \begin{cases} \dfrac{1}{A_{Go}^2}\gamma_G^{(|\boldsymbol{v}|^2/A_{Go}^2)} - \dfrac{1}{A_G^2}\gamma_G^{(|\boldsymbol{v}|^2/A_G^2)} & |\boldsymbol{v}| \le A_G \\ 0 & |\boldsymbol{v}| > A_G. \end{cases} \tag{B.3}$$

The radius of input connections $a_G(\boldsymbol{\xi})$ is $A_G = 3.3$, and the parameter determining the positive center is $A_{Go} = 1.5$. The parameter that determines the deviation of the Gaussian function is $\gamma_G = 0.05$. As a result, the positive center of the Mexican hat becomes about 1.2 in radius. The function $a_G(\boldsymbol{\xi})$ is then modified from $a'_G(\boldsymbol{\xi})$ by

$$a_G(\boldsymbol{v}) = \begin{cases} a'_G(\boldsymbol{v}) & \text{if } a'_G(\boldsymbol{v}) \ge 0 \\ \beta_G \cdot a'_G(\boldsymbol{v}) & \text{if } a'_G(\boldsymbol{v}) < 0 \end{cases} \tag{B.4}$$

where positive constant β_G is adjusted so as to satisfy (B.2).

Appendix C. Response of S-cell layers

Let $u_{Sl}(\boldsymbol{n}, k)$ and $u_{Cl}(\boldsymbol{n}, k)$ be the output of S-cells and C-cells of the kth cell-plane of the lth stage, respectively, where $\boldsymbol{n}$ represents the location of the receptive field center of the cells. Layer U_{Sl} contains not only S-cells but also V-cells, whose output is represented by $v_l(\boldsymbol{n})$. The outputs of S-cells and V-cells are given by

$$u_{Sl}(\boldsymbol{n}, k) = \frac{1}{1-\theta_l} \times \varphi\left[\frac{\sum_{\kappa=1}^{K_{Cl-1}} \sum_{|\boldsymbol{v}|<A_{Sl}} a_{Sl}(\boldsymbol{v}, \kappa, k) \cdot u_{Cl-1}(\boldsymbol{n}+\boldsymbol{v}, \kappa)}{b_{Sl}(k) \cdot v_l(\boldsymbol{n})} - \theta_l\right], \tag{C.1}$$

where

$$v_l(\boldsymbol{n}) = \sqrt{\sum_{\kappa=1}^{K_{Cl-1}} \sum_{|\boldsymbol{v}|<A_{Sl}} c_{Sl}(\boldsymbol{v}) \cdot \{u_{Cl-1}(\boldsymbol{n}+\boldsymbol{v}, \kappa)\}^2}. \tag{C.2}$$

If $l = 1$ in (C.1) and (C.2), $u_{Cl-1}(\boldsymbol{n}, k)$ stands for $u_G(\boldsymbol{n}, k)$, and we have $K_{Cl-1} = 2$.

Parameter $a_{Sl}(\boldsymbol{v}, \kappa, k)$ (≥ 0) is the strength of variable excitatory connection coming from C-cell $u_{Cl-1}(\boldsymbol{n}+\boldsymbol{v}, \kappa)$ of the preceding stage. It should be noted here that all cells in a cell-plane share the same set of input connections, hence $a_{Sl}(\boldsymbol{v}, \kappa, k)$ is independent of $\boldsymbol{n}$. A_{Sl} denotes the radius of summation range of $\boldsymbol{v}$, that is, the size of spatial spread of input connections to a particular S-cell. Parameter $b_{Sl}(k)$ (>0) is the strength of variable inhibitory connection coming from the V-cell.

Parameter $c_{Sl}(\boldsymbol{v})$ represents the strength of the fixed excitatory connections to the V-cell, and is a monotonically decreasing function of $|\boldsymbol{v}|$. It is also used as a weighting function for training connections $a_{Sl}(\boldsymbol{v}, \kappa, k)$, as is shown in (D.1) below.

In the computer simulation, the shape of $c_{Sl}(\boldsymbol{v})$ is determined by a Gaussian function, but its value is truncated to zero in the outside of the connecting area of radius A_{Sl}. Namely,

$$c_{Sl}(\boldsymbol{v}) = \begin{cases} \gamma_{Sl}^{(|\boldsymbol{v}|^2/A_{Sl}^2)} & |\boldsymbol{v}| \le A_{Sl} \\ 0 & |\boldsymbol{v}| > A_{Sl}. \end{cases} \tag{C.3}$$

The radius of input connections to S- and V-cells is: $A_{S1} = A_{S2} = A_{S3} = 3.5$ and $A_{S4} = 4.5$. The parameter that determines the deviation of the Gaussian function is $\gamma_{Sl} = 0.7$ for $l = 1, 2, 3, 4$.

The positive constant θ_l is the threshold of the S-cell and determines the selectivity in extracting features (See 3.2.2).

In the computer simulation, thresholds θ_l^R for the recognition phase are: $\theta_1^R = 0.50, \theta_2^R = 0.50, \theta_3^R = 0.51$ and $\theta_4^R = 0.0$. Threshold θ_l^L for the learning phase are: $\theta_2^L = 0.70$, $\theta_3^L = 0.69$ and $\theta_4^L = 0.0$.

Appendix D. Training S-cell layers

In the hierarchical network of the neocognitron, training (or learning) is performed from lower stages to higher stages: after the training of a lower stage has been completely finished, the training of the succeeding stage begins. The same set of training patterns is used for the training of all stages except layer U_{S1}.

D.1. Renewing connections

Every time when a training pattern is presented to the input layer, a small number of *seed-cells* are selected. The method of selecting seed-cells, which is different from layer to layer, is discussed later.

Although the method for selecting seed-cells during learning is slightly different between layers, the rule for renewing variable connections $a_{Sl}(\boldsymbol{v}, \kappa, k)$ and $b_{Sl}(k)$ is the same for all layers, once the seed-cells have been determined. The connections are renewed depending on the responses of the presynaptic cells (namely, the C-cells of the preceding stage).

Let cell $u_{Sl}(\hat{\boldsymbol{n}}, \hat{k})$ be selected as a seed-cell at a certain time, the variable connections $a_{Sl}(\boldsymbol{v}, \kappa, \hat{k})$ to this seed-cell, and consequently to all the S-cells in the same cell-plane as the seed-cell, are increased by the following amount:

$$\Delta a_{Sl}(\boldsymbol{v}, \kappa, \hat{k}) = c_{Sl}(\boldsymbol{v}) \cdot u_{Cl-1}(\hat{\boldsymbol{n}}+\boldsymbol{v}, \kappa). \tag{D.1}$$

The inhibitory connection $b_{Sl}(\hat{k})$ is determined directly from the values of the excitatory connections $a_l(\boldsymbol{v}, \kappa, \hat{k})$. That is,

$$b_{Sl}(\hat{k}) = \sqrt{\sum_{\kappa=1}^{K_{Cl-1}} \sum_{|\boldsymbol{v}|<A_{Sl}} \frac{\left\{a_{Sl}(\boldsymbol{v}, \kappa, \hat{k})\right\}^2}{c_{Sl}(\boldsymbol{v})}}. \tag{D.2}$$

Once the input connections to a seed-cell have been renewed, all cells in the cell-plane come to have the same set of input connections as the seed-cell, because all cells in the cell-plane share the same set of input connections.

D.2. Edge-extracting layer

Layer U_{S1}, namely, the S-cell layer of the 1st stage, is an edge-extracting layer. It has $K_{S1} = 16$ cell-planes, each of which consists of edge-extracting cells of a particular preferred orientation. Preferred orientations of the cell-planes, namely, the orientations of the training patterns, are chosen at an interval of 22.5°.

The S-cells of this layer are trained with supervised learning (Fukushima, 1988). To train a cell-plane, the "teacher" presents a training pattern, namely a straight edge of a particular orientation, to the input layer U_0. The teacher then points out the location of the feature, which, in this particular case, can be an arbitrary point on the edge. The cell whose receptive field center coincides with the location of the feature takes the place of the seed-cell of the cell-plane, and the process of strengthening connections occurs automatically. It should be noted here that the process of supervised learning is identical to that of the unsupervised learning except the process of choosing seed-cells.

D.3. Competitive learning for intermediate layers

As was discussed in 3.3, we use winner-kill-loser rule to train intermediate layers of S-cells, U_{S2} and U_{S3}. The training (or learning) of layer U_{Sl} ($l = 2$ or 3) begins after the training of the preceding stages has been completely finished. The same set of training patterns is used repeatedly for training these layers.

The method of dual threshold (Fukushima & Tanigawa, 1996) is used for the learning of these layers. To be more specific, when training the lth layer U_{Sl}, a higher threshold value θ_l^{L} is used for calculating the response of U_{Sl}, and a lower threshold value θ_l^{R}, which is to be used for recognition, is used for calculating the response of the preceding layers (U_{Sl-1} and layers lower than it).

Every time when a training pattern is presented, the learning processes are repeated: (1) A seed-cell is chosen by a competition among cells that have already been generated. (2) The seed-cell has its input connections renewed. (3) Cell-planes that contain losers of the competition are removed from the layer. (4) A new cell-plane is generated, if there is a place where no cell is responding despite of non-zero stimulus. When the repetition of these processes has been finished for a training pattern, we proceed to the presentation of the next training pattern.

In the computer simulation, each training pattern of the training set was repeatedly presented three times. In other words, we made three rounds of presentation of the training set. Although it produces a slightly better recognition rate than one round of presentation, the difference is not so large. One round of presentation could be enough, for the economy of the training time.

D.3.1. Competition among cells that have already been generated

Since there are a number of cell-planes, a number of seed-cells are usually chosen at a single presentation of a training pattern. Renewing connections of seed-cells and removal of losers are performed step by step, in the order of output values of the winners, which become seed-cells.

At first, the S-cell whose response is the largest in the layer is chosen as a seed-cell. Namely,

$$u_{Sl}(\hat{\boldsymbol{n}}, \hat{\kappa}) = \max_{\boldsymbol{n},\kappa} \{u_{Sl}(\boldsymbol{n}, \kappa)\}, \tag{D.3}$$

where $\hat{\boldsymbol{n}}$ and $\hat{\kappa}$ are the location and the sequence number of the cell-plane of the seed-cell, respectively. The seed-cell has its input connections renewed by (D.1). Once the connections to the seed-cell are renewed, all cells in the cell-plane, from which the seed-cell is chosen, come to have the same set of input connections as the seed-cell because of the shared connections. Now the response of the cell-plane with the renewed connections is recalculated, because it affects the process of choosing other seed-cells by (D.5) below.

If there are any other non-silent cells at the location $\hat{\boldsymbol{n}}$ of the seed-cell, they are losers. Namely, all cells $u_{Sl}(\hat{\boldsymbol{n}}, \kappa)$ that satisfy

$$u_{Sl}(\hat{\boldsymbol{n}}, \hat{\kappa}) > u_{Sl}(\hat{\boldsymbol{n}}, \kappa) > 0, \quad \kappa \neq \hat{\kappa} \tag{D.4}$$

are losers. The cell-planes, to which losers belong, are removed from the layer. It should be noted here that silent cells are not treated as losers, because they do not join the competition.

Once a seed-cell has been chosen and had its input connections renewed, we repeat the following process to choose other seed-cells. The next seed-cell is chosen from the remaining cell-planes, from which seed-cells have not been chosen yet for this training pattern. Incidentally, the cell-planes that contain the losers have already been removed from the layer by this time.

Excluding the cells of the cell-planes from which seed-cells have already been selected, we search the maximum response cell, and make it a candidate of the next seed-cell. The candidate becomes a seed-cell, if no other cell is responding stronger than it within its competition area of radius D_l, which has a shape of a hypercolumn (Fukushima, 1980, 1988). Namely, the candidate cell $u_{Sl}(\hat{\boldsymbol{n}}, \hat{\kappa})$ becomes a seed-cell, if

$$u_{Sl}(\hat{\boldsymbol{n}}, \hat{\kappa}) > u_{Sl}(\hat{\boldsymbol{n}} + \boldsymbol{v}, \kappa), \quad \text{for all } |\boldsymbol{v}| < D_l \text{ and } \kappa \neq \hat{\kappa} \tag{D.5}$$

is satisfied. The seed-cell has its input connections renewed by (D.1), and the response of the cell-plane is recalculated. In the computer simulation, the radius of the hypercolumn is $D_l = A_{Sl}$ for both $l = 2$ and 3, where $A_{Sl} = 3.5$ is the size of the radius of input connections to an S-cell (See Appendix C).

If the candidate is defeated by other cells in its competition area, we again search the largest response cell, excluding the defeated candidate, and make it the next candidate. Incidentally, the cells that respond stronger than the candidate, if any, come from cell-planes from which seed-cells have already been chosen.

The losers are chosen again by (D.4). If there are losers at the location of the seed-cell, the cell-planes, to which losers belong, are removed from the layer. The cells in the cell-planes, from which seed-cells have already been chosen, however, do not become losers regardless of the values of their responses, because they do not join the competition any more for this training pattern.

The process of choosing seed-cells is continued, while any seed-cell can be chosen by this process. After all winners have their input connections renewed and all losers have been removed, the process of generating cell-planes begins.

D.3.2. Generating new cell-planes

If there is any location where C-cells of the preceding stage are active but all S-cells are silent around there (within D_o from the location), a new cell-plane is generated. At each presentation of a training pattern, generation of new cell-planes is repeated until all locations where C-cells are active are covered with active S-cells.

Specifically, we first prepare a virtual cell-plane called a *seed-selecting plane* for the layer. Every time when a training pattern is presented, the seed-selecting plane is set to an initial value of

$$u_{\mathrm{seed}\,l}(\boldsymbol{n}) = \varphi\left[\sum_{\kappa=1}^{K_{Cl-1}} u_{Cl-1}(\boldsymbol{n}, \kappa) - \theta_{\mathrm{seed}\,l}\right]. \tag{D.6}$$

Threshold $\theta_{\mathrm{seed}\,l}$ in this equation is determined by

$$\theta_{\mathrm{seed}\,l} = \epsilon \cdot \max_{\boldsymbol{n}}\left[\sum_{\kappa=1}^{K_{Cl-1}} u_{Cl-1}(\boldsymbol{n}, \kappa)\right], \tag{D.7}$$

where ϵ is a small positive constant. In the computer simulation, we chose $\epsilon = 0.3$.

The value of the seed-selecting plane at $\boldsymbol{n}$ is suppressed to zero if any S-cell is active within a radius of D_{ol} ($<D_l$) around $\boldsymbol{n}$. Namely,

$$u_{\mathrm{seed}\,l}(\boldsymbol{n}) = 0, \quad \text{if } u_{Sl}(\boldsymbol{n} + \boldsymbol{v}, \kappa) > 0 \quad \text{for any } |\boldsymbol{v}| \leq D_{ol} \text{ and } 1 \leq \kappa \leq K_{Sl}, \tag{D.8}$$

where K_{Sl} is the number of the cell-planes that have been generated by that time. In the computer simulation, we chose $D_{ol} = 2.5$ for both $l = 2$ and 3.

If the value of the seed-selecting plane is not completely zero, we generate a new cell-plane. We then search the location $\hat{\boldsymbol{n}}$ at which $u_{\mathrm{seed}\,l}(\boldsymbol{n})$ is maximum, and make it the location of the seed-cell of the newly generated cell-plane. The input connections of the seed-cell is set to the values given by

$$a_{Sl}(\boldsymbol{v}, \kappa, \hat{k}) = c_{Sl}(\boldsymbol{v}) \cdot u_{Cl-1}(\hat{\boldsymbol{n}} + \boldsymbol{v}, \kappa), \tag{D.9}$$

and (D.2), where $\hat{k}$ is the sequence number of the newly generated cell-plane.

The value of the seed-selecting plane is suppressed again by the response of the new cell-plane through (D.8), and search the location of a seed-cell to generate and train another cell-plane. The process of generating new cell-plane is repeated until the value of the seed-selecting plane becomes completely zero. We then proceed to the presentation of the next training pattern.

D.4. Training S-cells of the highest stage

Training method for the highest stage of the network is almost the same as that for the conventional neocognitron (Fukushima, 2003), except the threshold value of S-cells, which is now set to zero for both recognition and learning phases.

S-cells of the highest stage (U_{S4}) are trained using a supervised competitive learning, and the class names of the training patterns are also utilized for the learning. When the network learns varieties of deformed training patterns through competitive learning, more than one cell-plane for one class is usually generated in U_{S4}. Therefore, when each cell-plane first learns a training pattern, the class name of the training pattern is assigned to the cell-plane. Thus, each cell-plane of U_{S4} has a label indicating one of the 10 digits.

Every time a training pattern is presented, competition occurs among all S-cells in the layer. (In other words, the competition area for layer U_{S4} is large enough to cover all cells of the layer.) If the winner of the competition has the same label as the training pattern, the winner becomes the seed-cell and learns the training pattern in the same way as the seed-cells of the lower stages. If the winner has a wrong label (or if all S-cells are silent), however, a new cell-plane is generated, learns the training pattern and is put a label of the class name of the training pattern.

During the recognition phase, the label of the maximum output S-cell of U_{S4} determines the final result of recognition. We can also express this process of recognition as follows. Recognition layer U_{C4} has 10 C-cells corresponding to the 10 digits to be recognized. Every time a new cell-plane is generated in layer U_{S4} in the learning phase, excitatory connections are created from all S-cells of the cell-plane to the C-cell of that class name. Competition among S-cells occurs also in the recognition phase, and only one maximum output S-cell within the whole layer U_{S4} can transmit its output to U_{C4}.

Since the threshold (θ_4) of U_{S4} is chosen to be zero, any input pattern usually elicits responses from several S-cells. Hence the process of finding the largest output S-cell is equivalent to the process of finding the nearest reference vector in the multi-dimensional feature space. Each reference vector has its own territory determined by the Voronoi partition of the feature space. The recognition process in the highest stage resembles the vector quantization (Gray, 1984; Kohonen, 1990) in this sense.

Training vectors that are misclassified in the learning phase usually come from near class borders. Suppose a particular training vector is misclassified in the learning. The reference vector of the winner, which caused a wrong recognition for this training vector, is not renewed this time. A new cell-plane is generated instead, and the misclassified training vector is adopted as the reference vector of the new cell-plane. Generation of a new reference vector causes a shift of decision borders in the feature space, and some of the training vectors, which have been recognized correctly before, might now be misclassified and additional reference vectors have to be generated again to readjust the borders. Hence a repeated presentation of the same set of training vectors is required before the learning converges. Thus, the decision borders are gradually adjusted to fit the real borders between classes.

Training vectors that are located near the central area of its class is usually recognized correctly, they can be represented by a small number of reference vectors. Hence reference vectors come to distribute more densely near the class borders, while the density of the reference vectors is much lower in the central area of its class.

The same training set are presented repeatedly until all patterns in the training set come to be recognized correctly. To be more specific, every time when each round of presentation has finished in the learning phase, the number of newly generated cell-planes during that round, which is equal to the number of erroneously recognized patterns, is counted. If it reaches zero, the learning process ends. Although a repeated presentation of a training pattern set is required before the learning converges, the required number of repetition is not so large, say around 4 or 5, in usual cases.

This does not always guarantee, however, that all training patterns will be recognized correctly after finishing the learning, because reference vectors drift slightly even during the last round of presentation of the training set, where each training vector is summed to the reference vector of the winner. Erroneous recognition of training patterns, however, occurs very seldom after finishing the learning. In usual situations, the recognition rate for the training set is 100%.

Appendix E. C-cell layers

In each stage ($l \leq 3$), except the highest stage, the layer of C-cells have the same number of cell-planes as the layer of S-cells. There is one-to-one correspondence between the cell-planes of the S- and C-cell layers in the same stage. Each C-cell of a cell-plane basically receives signals from S-cells of the corresponding cell-plane.

Since C-cells of the 1st stage (U_{C1}) is slightly different from those of other stages, we first discuss the responses of C-cells of the 2nd and the 3rd stages (U_{C2} and U_{C3}).

The response of C-cells of the 2nd and the 3rd stages can be expressed by

$$u_{Cl}(\boldsymbol{n}, k) = \psi\left[\sum_{|\boldsymbol{v}|<A_{Cl}} a_{Cl}(\boldsymbol{v}) \cdot u_{Sl}(\boldsymbol{n} + \boldsymbol{v}, k)\right], \quad (l = 2, 3), \tag{E.1}$$

where function $\psi[x]$ determines the saturation of a C-cell:

$$\psi[x] = \sqrt{\varphi[x]}. \tag{E.2}$$

For the 2nd stage ($l = 2$) (and also for the 1st stage discussed later), the connections $a_{Cl}(\boldsymbol{v})$ has an excitatory center $a^{+}_{Cl}(\boldsymbol{v})$ and an inhibitory surround $a^{-}_{Cl}(\boldsymbol{v})$.

$$a_{Cl}(\boldsymbol{v}) = a^{+}_{Cl}(\boldsymbol{v}) - a^{-}_{Cl}(\boldsymbol{v}) \quad (l = 1, 2), \tag{E.3}$$

where $a^{+}_{Cl}(\boldsymbol{v}) > 0$ in a disk $|\boldsymbol{v}| \leq A_{\mathrm{Co}\,l}$, and $a^{-}_{Cl}(\boldsymbol{v}) > 0$ in an annulus $A_{\mathrm{Co}\,l} < |\boldsymbol{v}| \leq A_{Cl}$.

In the computer simulation, we chose

$$a^{+}_{Cl}(\boldsymbol{v}) = \begin{cases} \gamma_{Cl}^{(|\boldsymbol{v}|^2/A^2_{\mathrm{Co}\,l})} & |\boldsymbol{v}| \le A_{\mathrm{Co}\,l} \quad (l = 1, 2, 3), \\ 0 & \text{otherwise} \end{cases} \tag{E.4}$$

$$a^{-}_{Cl}(\boldsymbol{v}) = \begin{cases} \beta_{Cl} \cdot \gamma_{Cl}^{(|\boldsymbol{v}|^2/A^2_{\mathrm{Co}\,l})} & A_{\mathrm{Co}\,l} < |\boldsymbol{v}| \le A_{Cl} \quad (l = 1, 2), \\ 0 & \text{otherwise} \end{cases} \tag{E.5}$$

where positive constant β_{Cl} is determined so as to satisfy

$$\sum_{|\boldsymbol{v}|<A_{Cl}} a_{Cl}(\boldsymbol{v}) = 0, \quad (l = 1, 2). \tag{E.6}$$

The radiuses of the excitatory center and the inhibitory surround are: $A_{\mathrm{Co}\,1} = 3.5, A_{\mathrm{C}1} = 9.5, A_{\mathrm{Co}\,2} = 3.5, A_{\mathrm{C}2} = 6.9, A_{\mathrm{Co}\,3} = A_{\mathrm{C}3} = 4.9$. The parameter that determines the spatial decay is $\gamma_{Cl} = 0.7$ for $l = 1, 2, 3$.

Different from the 1st and the 2nd stages, the connections to a C-cell of the 3rd stage lack an inhibitory surround and have only an excitatory center. Namely,

$$a_{\mathrm{C}3}(\boldsymbol{v}) = a^{+}_{\mathrm{C}3}(\boldsymbol{v}). \tag{E.7}$$

Connections to a C-cell of the 1st stage also has inhibitory surround, but the signals from the inhibitory surround is suppressed by disinhibition from signals of C-cells of orthogonal preferred orientation (Fig. 10(b)). Mathematically,

$$\begin{aligned} u_{\mathrm{C}1}(\boldsymbol{n}, k) = \psi \Bigg[& \sum_{|\boldsymbol{v}|<A_{\mathrm{C}1}} a^{+}_{\mathrm{C}1}(\boldsymbol{v}) \cdot u_{\mathrm{S}1}(\boldsymbol{n}+\boldsymbol{v}, k) \\ & - \varphi \Bigg[\sum_{|\boldsymbol{v}|<A_{\mathrm{C}1}} a^{-}_{\mathrm{C}1}(\boldsymbol{v}) \cdot \Bigg\{ u_{\mathrm{S}1}(\boldsymbol{n}+\boldsymbol{v}, k) \\ & - \frac{u_{\mathrm{S}1}(\boldsymbol{n}+\boldsymbol{v}, k^{+}) + u_{\mathrm{S}1}(\boldsymbol{n}+\boldsymbol{v}, k^{-})}{2} \Bigg\} \Bigg] \Bigg], \end{aligned} \tag{E.8}$$

where k^{+} and k^{+} represent the sequence numbers of the cell-planes whose preferred orientations are perpendicular to that of the kth cell-plane. Namely, $k^{+} = k + K_{\mathrm{C}1}/4$ and $k^{-} = k - K_{\mathrm{C}1}/4$.[5] This does not necessarily mean, however, that the disinhibition comes only from perpendicular edges but also from edges of somewhat different orientations, because the orientation selectivity of S-cells are adjusted broad enough so that a single S-cell responds also to edged of orientations slightly different from its preferred orientation.

[5] To be more exact, $(k + K_{\mathrm{C}1}/4)$ and $(k - K_{\mathrm{C}1}/4)$ here actually means $(k - 1 + K_{\mathrm{C}1}/4 \bmod K_{\mathrm{C}1}) + 1$ and $(k - 1 + 3K_{\mathrm{C}1}/4 \bmod K_{\mathrm{C}1}) + 1$, respectively, because k^{+} and k^{-} must be numbers in the ranges $1 \le k^{+} \le K_{\mathrm{C}1}$ and $1 \le k^{-} \le K_{\mathrm{C}1}$.

References

Dragoi, V., & Sur, M. (2000). Dynamic properties of recurrent inhibition in primary visual cortex: contrast and orientation dependence of contextual effects. *Journal of Neurophysiology*, *83*, 1019–1030.

Elliffe, M. C. M., Rolls, E. T., & Stringer, S. M. (2002). Invariant recognition of feature combinations in the visual system. *Biological Cybernetics*, *86*, 59–71.

Fukushima, K. (1980). Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*(4), 193–202.

Fukushima, K. (1988). Neocognitron: a hierarchical neural network capable of visual pattern recognition. *Neural Networks*, *1*(2), 119–130.

Fukushima, K. (1989). Analysis of the process of visual pattern recognition by the neocognitron. *Neural Networks*, *2*(6), 413–420.

Fukushima, K. (2003). Neocognitron for handwritten digit recognition. *Neurocomputing*, *51*, 161–180.

Fukushima, K., & Miyake, S. (1982). Neocognitron: a new algorithm for pattern recognition tolerant of deformations and shifts in position. *Pattern Recognition*, *15*(6), 455–469.

Fukushima, K., & Tanigawa, M. (1996). Use of different thresholds in learning and recognition. *Neurocomputing*, *11*(1), 1–17.

Gray, R. M. (1984). Vector quantization. *IEEE ASSP Magazine*, *1*(2), 4–29.

Hildebrandt, T. H. (1991). Optimal training of thresholded linear correlation classifiers. *IEEE Transactions on Neural Networks*, *2*(6), 577–588.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology (London)*, *106*(1), 106–154.

Hubel, D. H., & Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology*, *28*(2), 229–289.

Jones, H. E., Grieve, K. L., Wang, W., & Sillito, A. M. (2001). Surround suppression in primate V1. *Journal of Neurophysiology*, *86*, 2011–2028.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, *78*(9), 1464–1480.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., & Hubbard, W. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, *1*, 541–551.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324.

Lo, S. B., Chan, H., Lin, J., Li, H., Freedman, M. T., & Mun, S. K. (1995). Artificial convolution neural network for medical image pattern recognition. *Neural Networks*, *8*(7–8), 1201–1214.

Ozeki, H., Sadakane, O., Akasaki, T., Naito, T., Shimegi, S., & Sato, H. (2004). Relationship between excitation and inhibition underlying size tuning and contextual response modulation in the cat primary visual cortex. *Journal of Neuroscience*, *24*(6), 1428–1438.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*(11), 1019–1025.

Sato, S., Kuroiwa, J., Aso, H., & Miyake, S. (1999). Recognition of rotated patterns using a neocognitron. In L. C. Jain, & B. Lazzerini (Eds.), *Knowledge based intelligent techniques in character recognition* (pp. 49–64). CRC Press.

Seriès, P., Lorenceau, J., & Frégnac, Y. (2003). The 'silent' surround of V1 receptive fields: theory and experiments. *Journal of Physiology (Paris)*, *97*, 453–474.

Walker, G. A., Ohzawa, I., & Freeman, R. D. (2000). Suppression outside the classical cortical receptive field. *Visual Neuroscience*, *17*, 369–379.