

Increasing robustness against background noise: Visual pattern recognition by a neocognitron

Kunihiko Fukushima*

Fuzzy Logic Systems Institute, Iizuka, Fukuoka, Japan

ARTICLE INFO

Article history:

Received 14 September 2010
Received in revised form 22 February 2011
Accepted 16 March 2011

Keywords:

Visual pattern recognition
Background noise
Neocognitron
Feature extraction
Subtractive inhibition
Root-mean-square

ABSTRACT

The *neocognitron* is a hierarchical multi-layered neural network capable of robust visual pattern recognition. It has been demonstrated that recent versions of the neocognitron exhibit excellent performance for recognizing handwritten digits. When characters are written on a noisy background, however, recognition rate was not always satisfactory. To find out the causes of vulnerability to noise, this paper analyzes the behavior of feature-extracting S-cells. It then proposes the use of subtractive inhibition to S-cells from V-cells, which calculate the average of input signals to the S-cells with a root-mean-square. Together with this, several modifications have also been applied to the neocognitron. Computer simulation shows that the new neocognitron is much more robust against background noise than the conventional ones.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

The author previously proposed an artificial neural network *neocognitron* for robust visual pattern recognition (Fukushima, 1980, 1988, 2003, 2010a; Fukushima & Miyake, 1982). Its architecture was initially suggested by neurophysiological findings on the visual systems of mammals (e.g., Hubel & Wiesel, 1962, 1965). It is a hierarchical multi-layered network and 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.

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 that S-cells' response 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 having finished 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 (Cardoso & Wichert, 2010; Elliffe, Rolls, & Stringer, 2002; Hildebrandt, 1991; LeCun et al., 1989; LeCun, Bottou, Bengio, & Haffner, 1998; Lo et al., 1995; Riesenhuber & Poggio, 1999; Satoh, 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 conventional neocognitrons, 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.

It has been demonstrated that neocognitrons of recent versions exhibit excellent performance for recognizing handwritten digits (Fukushima, 2003, 2010a). Most of the experiments for these neocognitrons have been made using characters written on backgrounds containing little noise.

When characters are written on a background contaminated with noise as shown in Fig. 1, however, the recognition rate of these neocognitrons is not always satisfactory. In Fig. 1(a) and (b), the background noise is a faint image of a different digit. This situation often occurs when we rewrite a character after erasing another character insufficiently. It also occurs when we write characters on a thin paper, through which the printing on the reverse side

* 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.

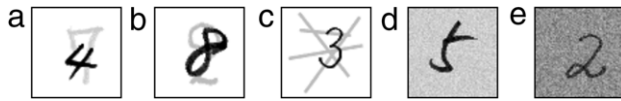


Fig. 1. Input patterns contaminated with background noise.

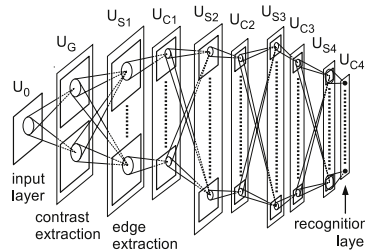


Fig. 2. The architecture of the neocognitron (Fukushima, 2003).

shows through. In Fig. 1(c), line segments are randomly located in the background. This emulates a case when we write characters on a ruled notepaper or fill in a form. In the case shown in Fig. 1(d) and (e), a white noise is superimposed on a character.

To find out the cause of vulnerability to noise, we analyze the behavior of feature-extracting S-cells. We then propose the use of subtractive inhibition to S-cells from V-cells, which calculate the average of input signals to the S-cells with a root-mean-square. Together with this, several modifications have also been applied to C-cells.¹

In Section 5, we test the behavior of the new neocognitron by computer simulation and show that the new neocognitron is much more robust against background noise than the conventional ones.

2. Outline of the network

The neocognitron is a hierarchical multi-layered network. It consists of layers of S-cells, which resemble simple cells in 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.

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. 2 shows the architecture of the network that is discussed in this paper. In the figure, U_{S_l} , for example, indicates the layer of S-cells of the l th 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 feature 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). 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.

The stimulus pattern is presented to the input layer (photoreceptor layer) U_0 . A layer of contrast-extracting cells (U_C), which

correspond to retinal ganglion cells or lateral geniculate nucleus cells, follows layer U_0 . It consists of two cell-planes: one consisting of cells with concentric on-center receptive fields, and one consisting of cells with off-center receptive fields. The former cells extract positive contrast in brightness, whereas the latter extract negative contrast from the images presented to U_0 . At the same time, the dc component of spatial frequency (namely, the averaged gray level) of the input pattern is eliminated, because the input connections to a single cell of layer U_C are designed in such a way that their total sum is equal to zero (see Appendix B for more detail). Hence the difference in the gray level of the background between Fig. 1(d) and (e), for example, is removed in the output of U_C . The output of U_C is sent to U_{S1} .

The S-cells of U_{S1} resemble simple cells in the primary visual cortex, and respond selectively to edges of a particular orientation. To be more specific, layer U_{S1} consists of $K_1 = 16$ cell-planes, and all cells in the k th cell-plane respond preferentially to edges of orientation $2\pi k/K_1 = k \times 22.5^\circ$. As a result, the contours of the input image are decomposed into edges of every orientation.

The input connections of S-cells of higher stages are variable and are modified through learning. After having finished learning, S-cells come to work as feature-extracting cells. In higher stages, they extract more global features.

In each stage of the hierarchical network, the output of layer U_{S_l} is fed to layer U_{C_l} . C-cells, whose input connections are fixed, exhibit an approximate invariance to the position of the stimuli presented within their receptive fields. In other words, a blurred version of the response of U_{S_l} appears in U_{C_l} . The blurring operation is essential for endowing the neocognitron with an ability to recognize patterns robustly, with little effect from deformation, change in size, or shift in position of input patterns. The C-cells in the highest stage work as recognition cells, which indicate the result of the pattern recognition.

3. Feature-extracting S-cells

3.1. Response of an S-cell

3.1.1. Several types of inhibitory mechanisms

An S-cell receives both excitatory and inhibitory input signals. The excitatory signals come from C-cells of the preceding stage, and the inhibitory signal comes from a V-cell, which accompanies the S-cell. (We discuss this later in more detail. See Fig. 4.) We now compare several types of inhibitory mechanisms and discuss how they affect the behavior of the S-cell.

In the original neocognitron (Fukushima, 1980, 2003), the inhibitory signal from the V-cell works in a shunting manner. Fig. 3 shows an equivalent circuit of an S-cell. Let e be the weighted sum of all excitatory inputs, and let h be the inhibitory input. Since h works on e in a shunting manner, the output of the S-cell is determined by nonlinear function $(1 + e)/(1 + h) - 1$. If the value of the nonlinear function is negative, however, the output of the S-cell becomes zero.

Incidentally, cells with shunting inhibition were used already in the *cognitron* (Fukushima, 1975), which is an earlier model before the neocognitron. The nonlinear function $(1 + e)/(1 + h) - 1$ was suggested from the behavior of the membrane potential of a biological neuron (Eccles, 1964). It is assumed that inhibitory input works to shunt the membrane potential toward an equilibrium potential of the IPSP (inhibitory post-synaptic potential), which is more negative than the resting potential of the membrane. On the other hand, excitatory inputs raise the membrane potential toward the equilibrium potential of the EPSP (excitatory post-synaptic potential), which is much higher than the resting potential. It is also known that the threshold at which the neuron fires is much lower

¹ A preliminary short report on these modifications has appeared in Fukushima (2010b).

Download English Version:

<https://daneshyari.com/en/article/10326502>

Download Persian Version:

<https://daneshyari.com/article/10326502>

[Daneshyari.com](https://daneshyari.com)