Neural Networks 37 (2013) 103–119

Contents lists available at SciVerse ScienceDirect

Neural Networks

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

# Artificial vision by multi-layered neural networks: Neocognitron and its advances

Kunihiko Fukushima *

*Fuzzy Logic Systems Institute, Iizuka, Fukuoka, Japan*

ARTICLE INFO

*Keywords:*
Artificial vision
Neocognitron
Hierarchical network
Bottom-up and top-down
Modeling neural networks

ABSTRACT

The *neocognitron* is a neural network model proposed by Fukushima (1980). Its architecture was suggested by neurophysiological findings on the visual systems of mammals. It is a hierarchical multi-layered network. It acquires the ability to robustly recognize visual patterns through learning. Although the neocognitron has a long history, modifications of the network to improve its performance are still going on. For example, a recent neocognitron uses a new learning rule, named *add-if-silent*, which makes the learning process much simpler and more stable. Nevertheless, a high recognition rate can be kept with a smaller scale of the network. Referring to the history of the neocognitron, this paper discusses recent advances in the neocognitron. We also show that various new functions can be realized by, for example, introducing top-down connections to the neocognitron: mechanism of selective attention, recognition and completion of partly occluded patterns, restoring occluded contours, and so on.

## 1. Introduction

In the visual systems of mammals, visual scenes are analyzed in parallel by separate channels. Loosely speaking, information concerning object shape is mainly analyzed through the temporal pathway in the cerebrum, while information concerning visual motion and location is mainly analyzed through the occipito-parietal pathway. The neocognitron is an artificial neural network, whose architecture was initially suggested from neurophysiological findings on the temporal pathway: retina → LGN → area V1 (primary visual cortex) → area V2 → area V4 → IT (inferotemporal cortex).

In area V1, cells respond selectively to local features of a visual pattern, such as lines or edges in particular orientations (Hubel & Wiesel, 1962, 1965). In areas V2 and V4, cells exist that respond selectively to complex visual features (e.g., Ito and Komatsu (2004), von der Hydt, Peterhans, and Baumgartner (1984) and Desimone and Schein (1987)). In the inferotemporal cortex, cells exist that respond selectively to more complex features, or even to human faces (e.g., Fujita, Tanaka, Ito, and Cheng (1992), Bruce, Desimone, and Gross (1981) and Yamane, Kaji, and Kawano (1988)). Thus, the visual system seems to have a hierarchical architecture, in which simple features are first extracted from a stimulus pattern, and then integrated into more complicated ones. In this hierarchy, a cell in a higher stage generally has a larger receptive field, and is more insensitive to the location of the stimulus. This kind of physiological evidence suggested the network architecture of the neocognitron.

In the 1960s, Hubel and Wiesel classified cells in the visual cortex into simple, complex and hypercomplex cells. They hypothesized that visual information is processed hierarchically through simple cells → complex cells → lower-order hypercomplex cells → higher-order hypercomplex cells (Hubel & Wiesel, 1962, 1965). They suggested that, in this hierarchy, the relation between lower-order hypercomplex cells to higher-order hypercomplex cells resembles that between simple cells to complex cells. Although classifying hypercomplex cells into lower-order and higher-order is not popular among neurophysiologists recently, it is this hypothesis that suggested the original architecture of the neocognitron model when it was first proposed by Fukushima (1980).

In the neocognitron, there are two major types of cells, namely *S-cells* and *C-cells*. *S*-cells, which are named after simple cells, correspond to simple cells or lower-order hypercomplex cells. Similarly, *C*-cells, which are named after complex cells, correspond to complex cells or higher-order hypercomplex cells. As shown in Fig. 1, the neocognitron consists of cascaded connection of a number of modules, each of which consists of a layer of *S*-cells followed by a layer of *C*-cells.

Although the neocognitron has a long history, modifications of the network to improve its performance are still going on. Referring to the history of the neocognitron, this paper discusses recent advances in the neocognitron.

Sections 2 and 3 discuss the basic architecture of the neocognitron and the principles for robust recognition of visual patterns. Section 4 discusses the mechanism of feature extraction by *S*-cells, comparing several learning rules adopted in the neocognitron of recent versions. Among them, a new learning rule named *add-if-silent* makes the learning process much simpler and more stable.

---

* Correspondence to: 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://www.www4.ocn.ne.jp/~fuku_k/index-e.html.

0893-6080/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2012.09.016

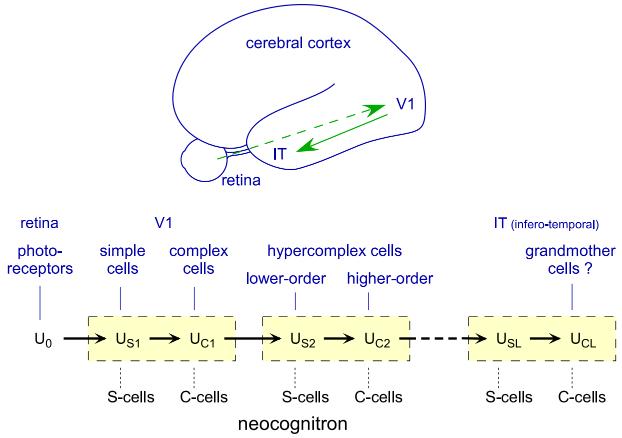

**Fig. 1.** Relation between the architecture of the neocognitron and the classical hypothesis of Hubel and Wiesel.
*Source:* (modified from Fukushima (1980)).

Nevertheless, a high recognition rate can be kept with a smaller scale of the network. Section 5 discusses the blurring operation by *C*-cells. Section 6 discusses the process of pattern classification at the highest stages of the network. We show that the method of *interpolating-vector* can greatly increase the recognition rate.

Section 7 discusses several networks extended from the neocognitron. We show that various new functions can be realized by, for example, introducing top-down connections to the neocognitron: mechanisms of selective attention, recognition and completion of partly occluded patterns, restoring occluded contours, and so on.

Incidentally, varieties of modifications, extensions and applications of the neocognitron, as well as varieties of related networks, have also been reported so far by several groups other than the author's (e.g., LeCun, Bottou, Bengio, and Haffner (1998), Mutch and Lowe (2008), Riesenhuber and Poggio (1999), Satoh, Kuroiwa, Aso, and Miyake (1999) and Serre, Oliva, and Poggio (2007)). 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 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.

## 2. Outline of the network

The neocognitron is a multi-layered network, which consists of layers of *S*-cells and *C*-cells. These layers of *S*-cells and *C*-cells are arranged alternately in a hierarchical manner.

*S*-cells work as feature-extracting cells. Their input connections are variable and are modified through learning. After learning, each *S*-cell comes to respond selectively to a particular visual feature presented in its receptive field. The features extracted by *S*-cells are determined during learning. Generally speaking, *local* features, such as edges or lines in particular orientations, are extracted in lower stages. More *global* features, such as parts of learned patterns, are extracted in higher stages.

*C*-cells are inserted in the network to allow for positional errors in the features of the stimulus. The input connections of *C*-cells, which come from *S*-cells of the preceding layer, are fixed and invariable. Each *C*-cell receives excitatory input connections from a group of *S*-cells that extract the same feature, but from slightly different locations. The *C*-cell responds if at least one of these *S*-cells yields an output. Even if the stimulus feature shifts and another *S*-cell comes to respond instead of the first one, the same *C*-cell keeps responding. Thus, the *C*-cell's response is less sensitive to a shift in location of the input pattern. We can also express that *C*-cells make a blurring operation, because the response of a layer of *S*-cells is spatially blurred in the response of the succeeding layer of *C*-cells.

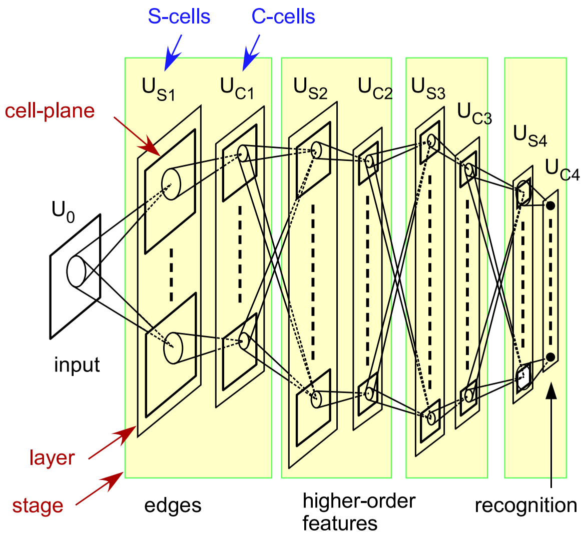

**Fig. 2.** A typical architecture of the neocognitron. The neocognitron consists of a number of *stages* of modules connected in a cascade in a hierarchical manner. Each stage consists of a *layer* of *S*-cells followed by a layer of *C*-cells. Each layer is divided into a number of sub-layers, called *cell-planes*, depending on the feature to which cells respond preferentially.

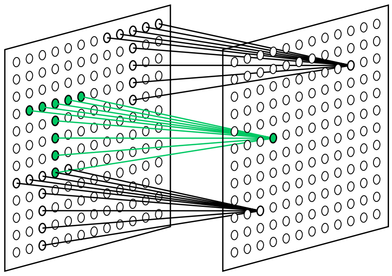

**Fig. 3.** An illustration of shared connections between two cell-planes. All cells in a cell-plane share the same set of input connections (Fukushima, 1980).

There are several versions of the neocognitron, which have slightly different architectures. Fig. 2 shows a typical architecture of the network. The hierarchical network has a number of *stages* of modules, each of which consists of a layer of *S*-cells followed by a layer of *C*-cells. Here we use notation like $U_{Sl}$, for example, to indicate the layer of *S*-cells of the *l*th stage.

There are retinotopically ordered connections between cells of adjoining layers. Each cell receives input connections that lead from cells situated in a limited area on the preceding layer. Since cells in higher stages come to have larger receptive fields, the density of cells in each layer is designed to decrease with the order of the stage.

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. In Fig. 2, each rectangle drawn with thick lines represents a cell-plane. Incidentally, a cell-plane is a group of cells that are arranged retinotopically and share the same set of input connections (Fukushima, 1980). Namely, all cells in a cell-plane share the same set of input connections, as illustrated in Fig. 3. In other words, the connections to a cell-plane have a translational symmetry. As a result, all cells in a cell-plane have identical receptive fields but at different locations. The modification of variable