Report on a research paper

CS152: Neural Networks, Prof. Keller

Anton Bakalov

Paper

Neocognitron of a New Version: Handwritten Digit Recognition

Kunihiko Fukushima*

Tokyo University of Technology
1404-1, Katakura, Hachioji, Tokyo 192-0982, Japan
fukushima@karl.teu.ac.jp

Abstract

The author previously proposed a neural network model neocognitron for robust visual pattern recognition. This paper proposes an improved version of the neocognitron and demonstrates its ability using a large database of handwritten digits (ETL-1).

To improve the recognition rate of the neocognitron, several modifications have been applied: such as, the inhibitory surround in the connections from S-cells to C-cells, contrast-extracting layer between input and edge-extracting layers, self-organization of line-extracting cells, supervised competitive learning at the highest stage, and so on. These modifications allowed the removal of accessory circuits that were appended to the previous versions, resulting in an improvement of recognition rate as well as simplification of the network architecture.

The recognition rate varies depending on the number of training patterns. When we used 3000 digits (300 patterns for each digit) for the learning, for example, the recognition rate was 98.5% for a blind test set (3000 digits), and 100% for the training set.

1 Introduction

The author previously proposed a neural network model *neocognitron* for robust visual pattern recognition [1][2]. This paper proposes an improved version of the neocognitron and demonstrates its ability using a large database of handwritten digits.

The neocognitron was initially proposed as a neural network model of the visual system that has a hierarchical multilayered architecture similar to the classical hypothesis of Hubel and Wiesel [3],[4]. It acquires the ability to recognize robustly visual patterns through learning. 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.

*The author is also with Nippon Institute of Technology

S-cells are feature-extracting cells, whose input connections are variable and are modified through learning. C-cells, whose input connections are fixed and unmodified, 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 indicating 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.

The neocognitron of a new version, which this paper proposes, also has a network architecture similar to that of the conventional neocognitron, but several new ideas have been introduced: such as, the inhibitory surround in the connections from S-cells to C-cells, a contrast-extracting layer followed by an edge-extracting layer, self-organization of line-extracting cells, supervised competitive learning at the highest stage, and so on.

This paper shows that these modifications allow the removal of accessory circuits appended to the neocognitron of recent versions [5], resulting in an improvement of recognition rate as well as simplification of the network architecture.

2 Architecture of the Network

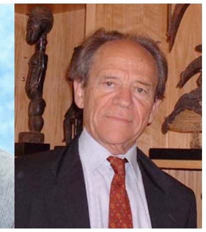
2.1 Outline of the Network

Figure 1 shows the architecture of the proposed model. (See also Fig. 3, which shows a typical response of the network.) As can be seen from the figure, the network has 4 stages of S- and C-cell layers: $U_0 \rightarrow U_G \rightarrow U_{S1} \rightarrow U_{C1} \rightarrow U_{S2} \rightarrow U_{C2} \rightarrow U_{S3} \rightarrow U_{C3} \rightarrow U_{S4} \rightarrow U_{C4}$.

The stimulus pattern is presented to the input layer (photoreceptor layer) U_0.

A layer of contrast-extracting cells (U_G), which correspond to retinal ganglion cells or lateral geniculate nucleus cells, follows layer U_0. The contrast-extracting layer U_G consists of two cell-planes: one cell-plane consisting of cells with concentric on-center receptive fields, and one cell-plane consisting of cells with off-center receptive fields. The former cells extract positive contrast in brightness, whereas the

0-7803-7044-9/01/$10.00 ©2001 IEEE



J. Physiol. (1962), **160**, *pp.* 106–154
With 2 plates and 20 text-figures
Printed in Great Britain

RECEPTIVE FIELDS, BINOCULAR INTERACTION AND FUNCTIONAL ARCHITECTURE IN THE CAT'S VISUAL CORTEX

BY D. H. HUBEL AND T. N. WIESEL

From the Neurophysiology Laboratory, Department of Pharmacology Harvard Medical School, Boston, Massachusetts, U.S.A.

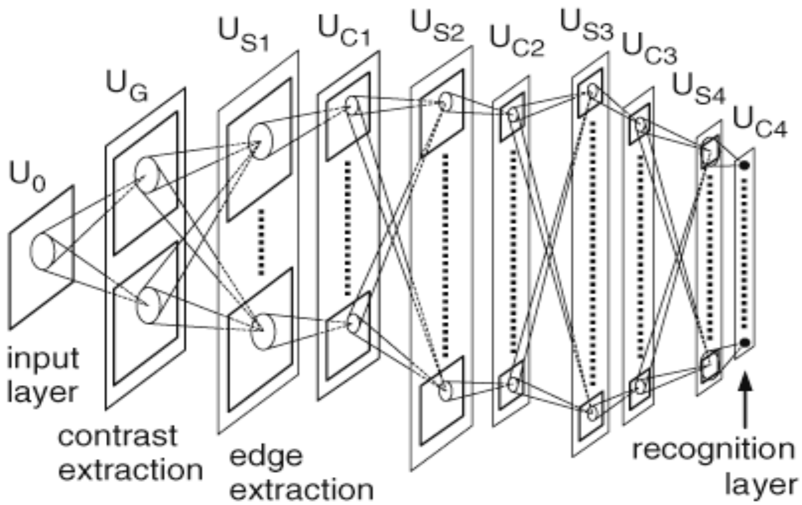
(*Received* 31 *July* 1961)

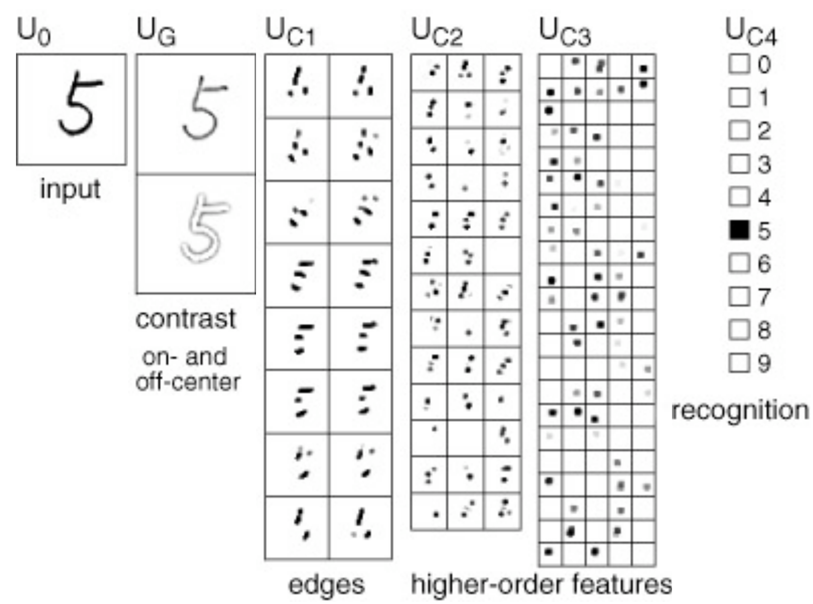
What chiefly distinguishes cerebral cortex from other parts of the central nervous system is the great diversity of its cell types and interconnexions. It would be astonishing if such a structure did not profoundly modify the response patterns of fibres coming into it. In the cat's visual cortex, the receptive field arrangements of single cells suggest that there is indeed a degree of complexity far exceeding anything yet seen at lower levels in the visual system.

In a previous paper we described receptive fields of single cortical cells, observing responses to spots of light shone on one or both retinas (Hubel & Wiesel, 1959). In the present work this method is used to examine receptive fields of a more complex type (Part I) and to make additional observations on binocular interaction (Part II).

This approach is necessary in order to understand the behaviour of individual cells, but it fails to deal with the problem of the relationship of one cell to its neighbours. In the past, the technique of recording evoked slow waves has been used with great success in studies of functional anatomy. It was employed by Talbot & Marshall (1941) and by Thompson, Woolsey & Talbot (1950) for mapping out the visual cortex in the rabbit, cat, and monkey. Daniel & Whitteridge (1959) have recently extended this work in the primate. Most of our present knowledge of retinotopic projections, binocular overlap, and the second visual area is based on these investigations. Yet the method of evoked potentials is valuable mainly for detecting behaviour common to large populations of neighbouring cells; it cannot differentiate functionally between areas of cortex smaller than about 1 mm². To overcome this difficulty a method has in recent years been developed for studying cells separately or in small groups during long micro-electrode penetrations through nervous tissue. Responses are correlated with cell location by reconstructing the electrode tracks from histological material. These techniques have been applied to

Architecture of the network

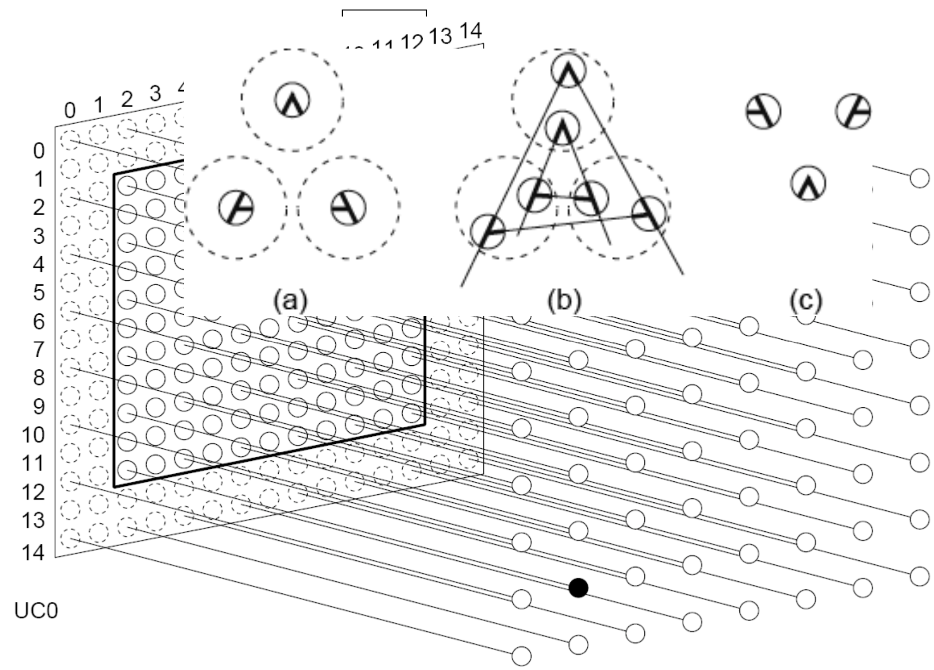




Elements of the Neocognitron

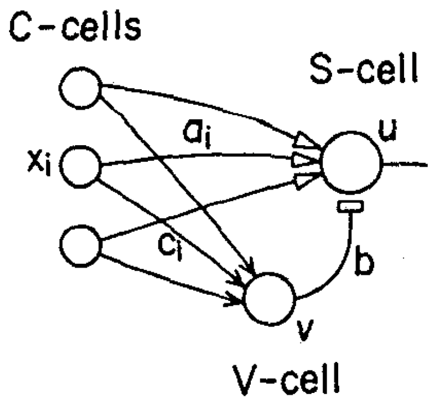
- S-cells and C-sells
  - what are they?
  - layer l
  - cell-plane k
  - location n
  - output notation:

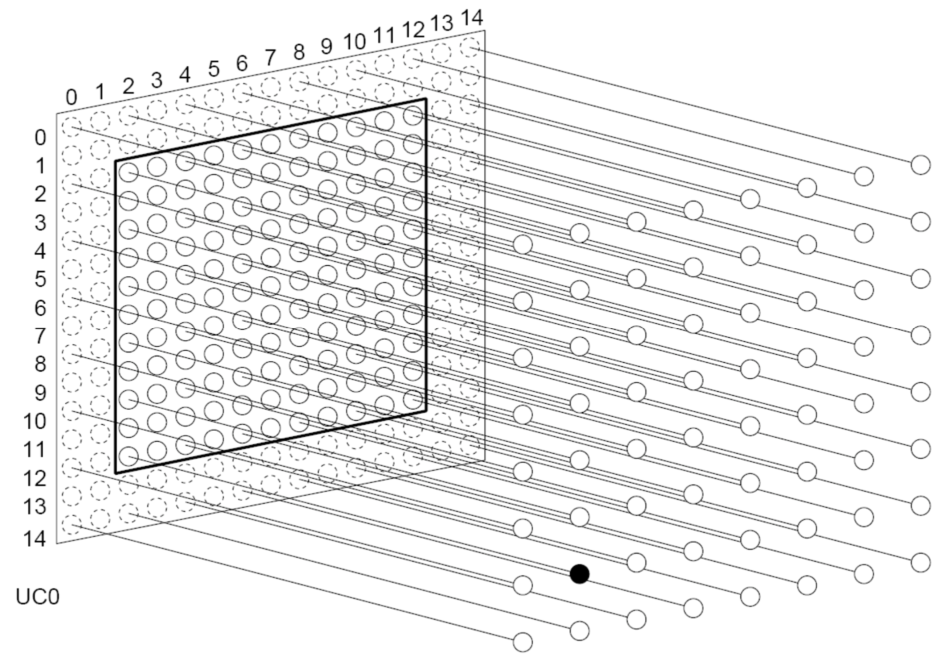
$$u_{S\ell}(\boldsymbol{n}, \boldsymbol{k}) \quad u_{C\ell}(\boldsymbol{n}, \boldsymbol{k})$$



Elements of the Neocognitron (cont.)

- V-cells
  - what are they?



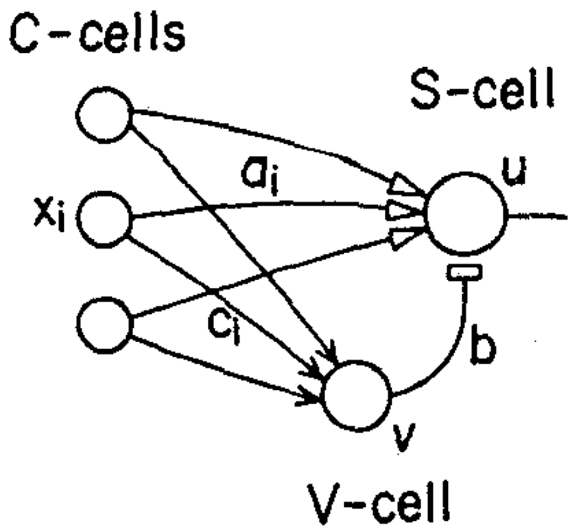


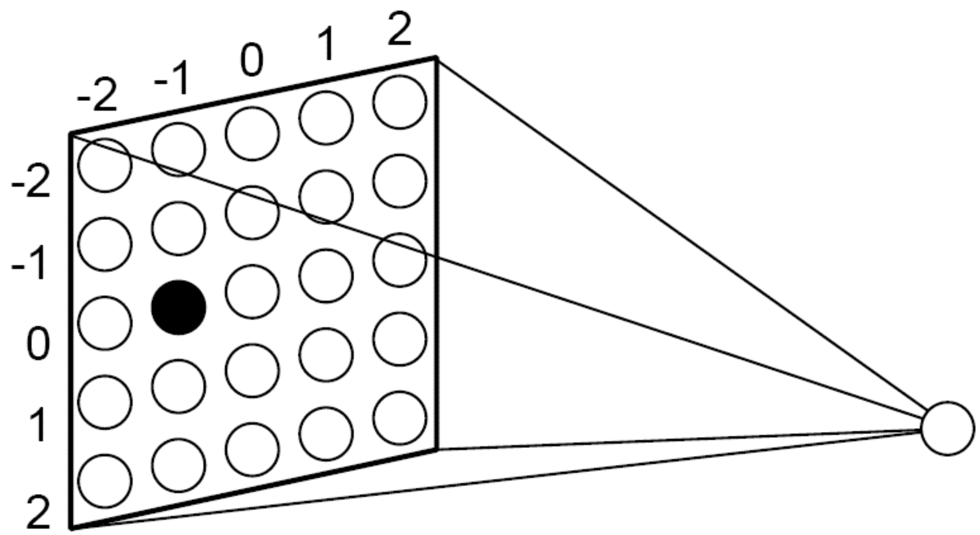
  - layer l
  - location n
  - output notation:

$$u_{V\ell}(\boldsymbol{n})$$

Types of weights

- From C-cells to S-cells $\Delta a_{Sl}(\boldsymbol{\nu}, \kappa, \hat{k}) = q_l \cdot c_{Sl}(\boldsymbol{\nu}) \cdot u_{Cl-1}(\hat{\boldsymbol{n}}+\boldsymbol{\nu}, \kappa)$

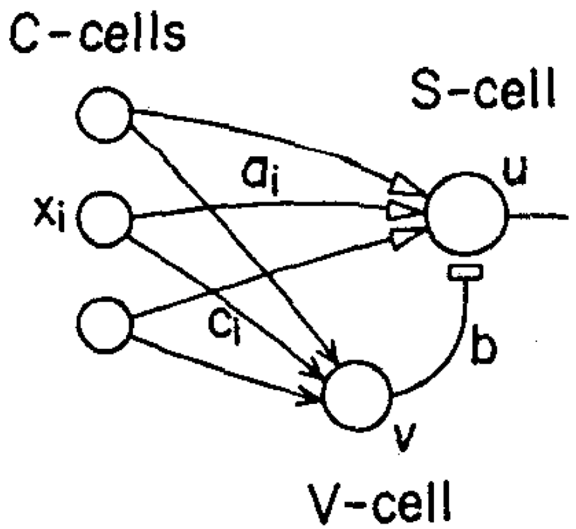


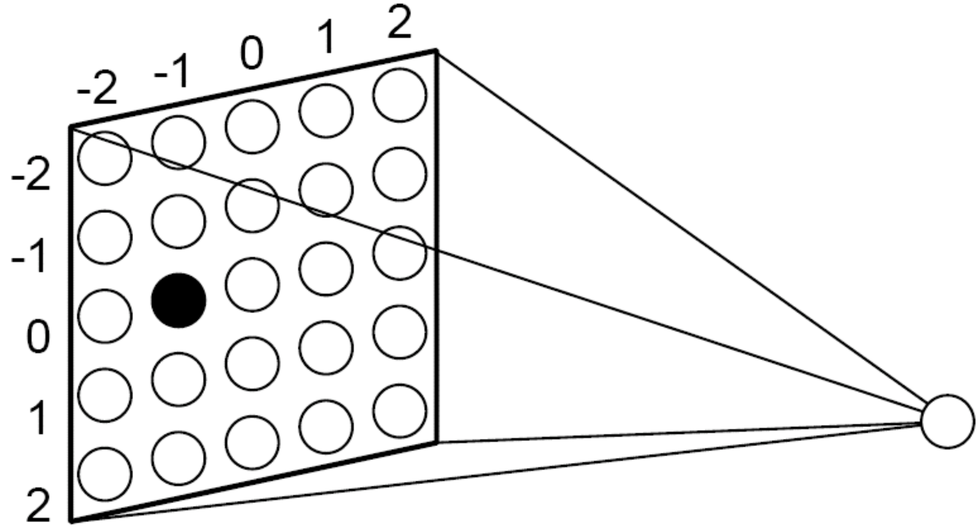


Types of weights (cont.)

- From V-cells to S-cells

$$b_{Sl}(\hat{k}) = \sqrt{\sum_{|\boldsymbol{\nu}|<A_{Sl}} \frac{\{a_{Sl}(\boldsymbol{\nu},\kappa,\hat{k})\}^2}{c_{Sl}(\boldsymbol{\nu})}}$$





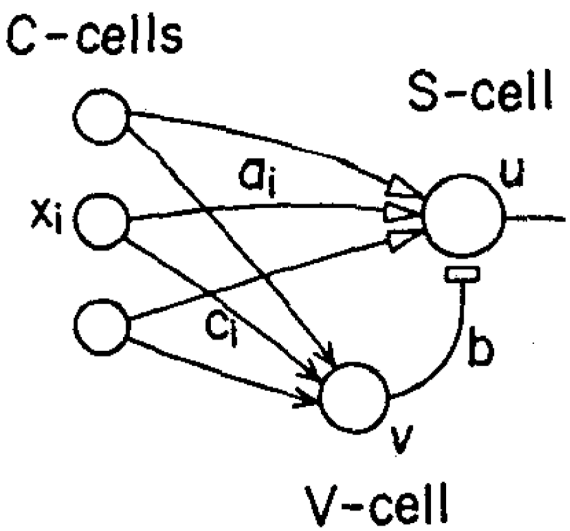
Types of weights (cont.)

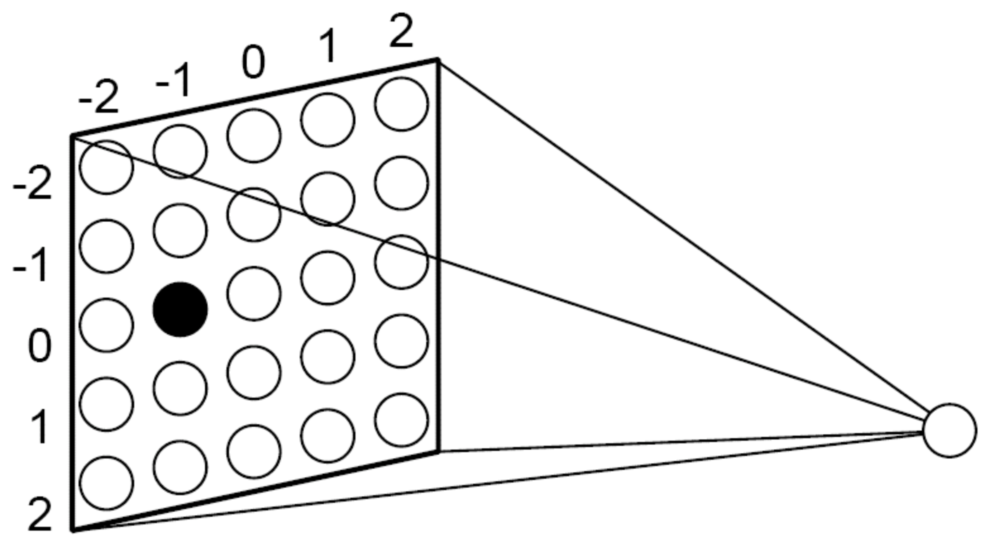
- From C-cells to V-cells
- From S-cells to C-cells

$$c_\ell(\boldsymbol{\nu}) = \gamma_\ell{}^{|\boldsymbol{\nu}|}$$

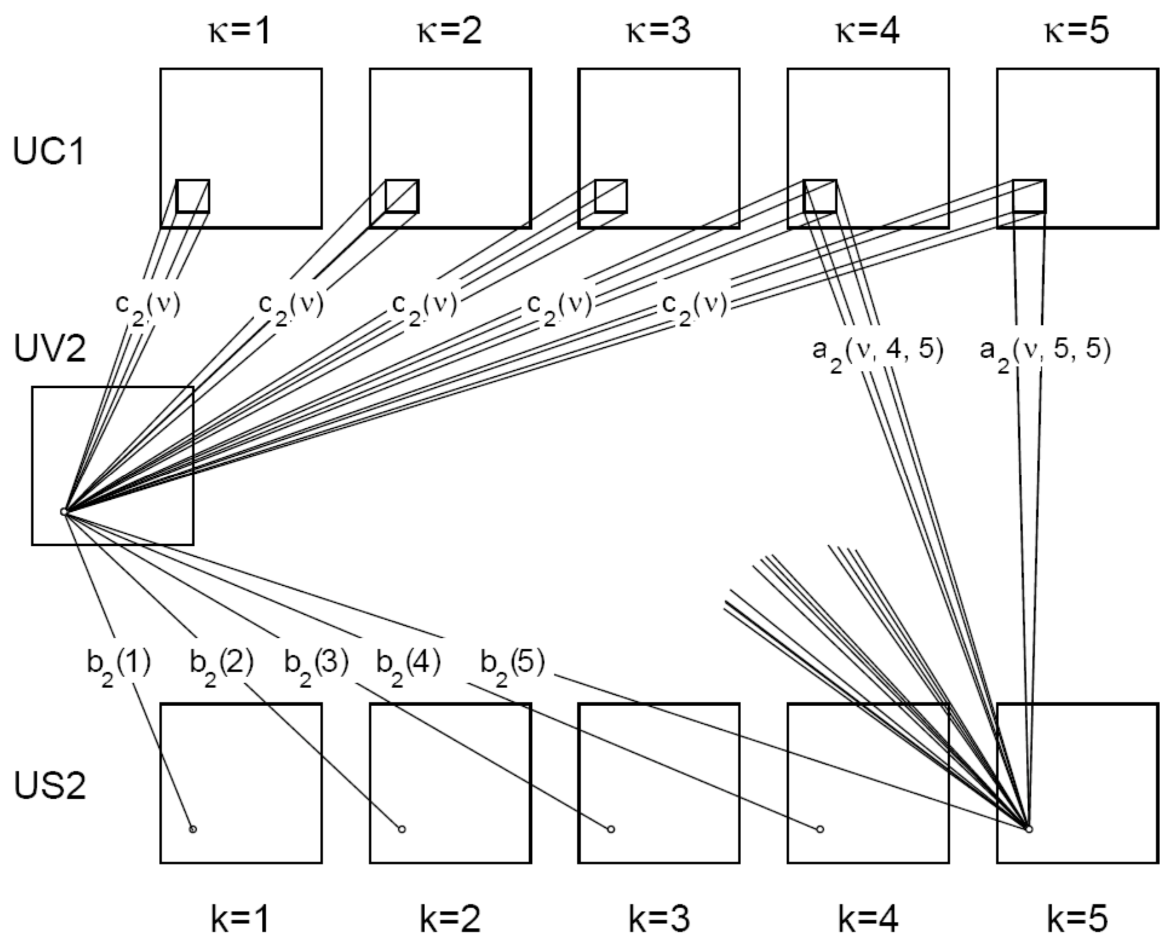
$$d_\ell(\boldsymbol{\nu}) = \overline{\delta}_\ell \cdot \delta_\ell{}^{|\boldsymbol{\nu}|}$$

where $0 < \gamma_\ell, \delta_\ell \leq 1$ and $0 < \overline{\delta}_\ell$





All weights


κ=1
κ=2
κ=3
κ=4
κ=5
UC1
c2(ν)
c2(ν)
c2(ν)
c2(ν)
c2(ν)
UV2
a2(ν, 4, 5)
a2(ν, 5, 5)
b2(1)
b2(2)
b2(3)
b2(4)
b2(5)
US2
k=1
k=2
k=3
k=4
k=5

Outputs of cells

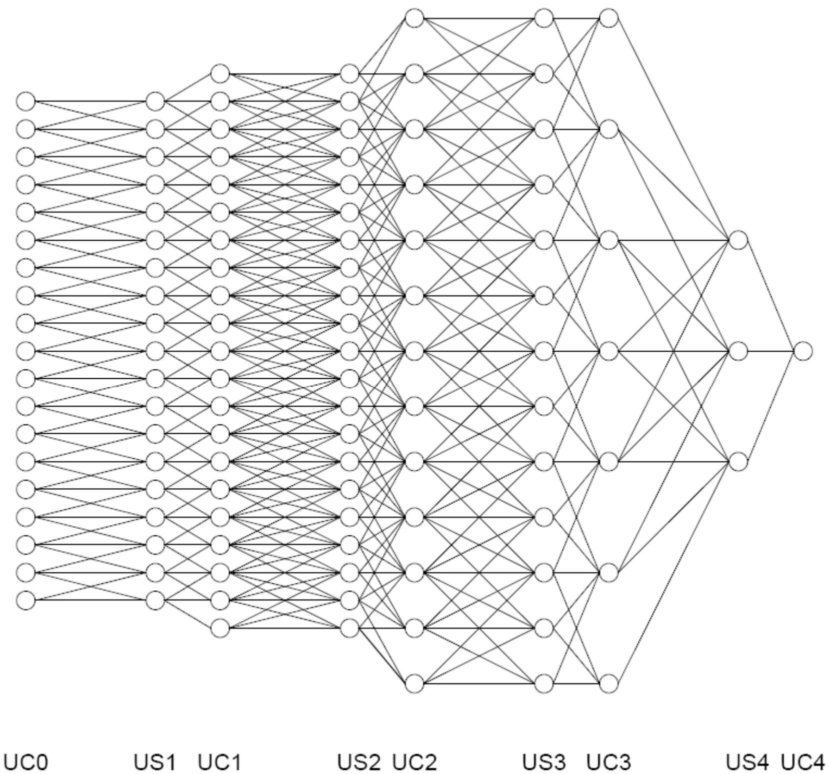
- S-cell:

$$u_{S\ell}(\boldsymbol{n},k) \overset{\text{def}}{=} r_\ell \times$$

$$\varphi\left[\frac{1+\displaystyle\sum_{\kappa=1}^{K_{C\ell-1}}\sum_{\boldsymbol{\nu}\in A_\ell} a_\ell(\boldsymbol{\nu},\kappa,k)\cdot u_{C\ell-1}(\boldsymbol{n}+\boldsymbol{\nu},\kappa)}{1+\dfrac{r_\ell}{r_\ell+1}\cdot b_\ell(k)\cdot u_{V\ell}(\boldsymbol{n})}-1\right]$$

where

$$\varphi(x) \overset{\text{def}}{=} \begin{cases} 0 & \text{if } x<0 \\ x & \text{if } 0\le x \end{cases}$$



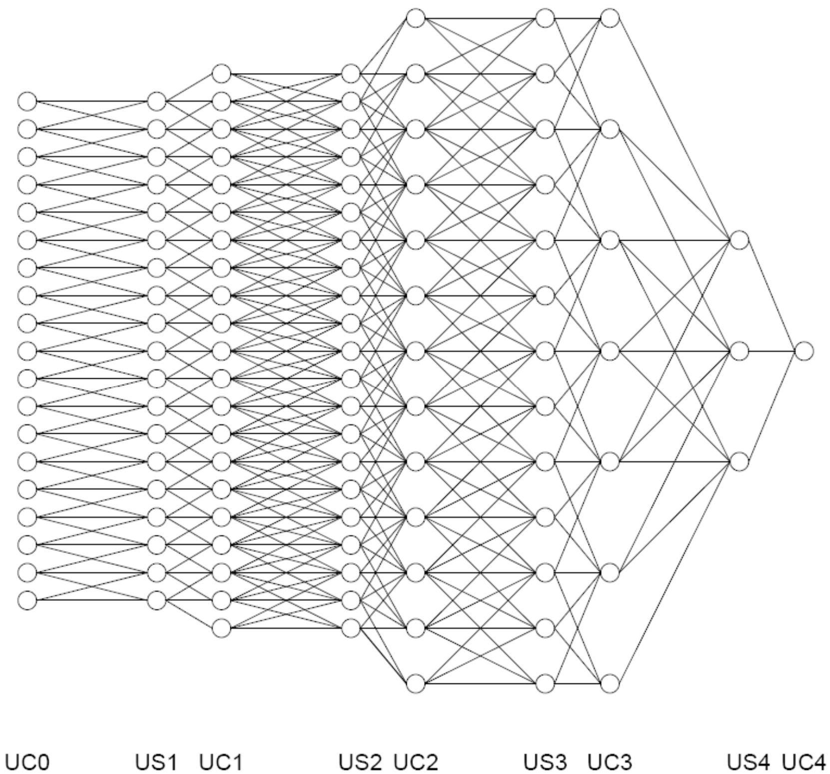
Outputs of cells (cont.)

- C-cell:

$$u_{Cl}(\boldsymbol{n},k)=\psi\Big[\sum_{|\boldsymbol{\nu}|<A_{Cl}} a_{Cl}(\boldsymbol{\nu})\cdot u_{Sl}(\boldsymbol{n}+\boldsymbol{\nu},k)\Big]$$

where

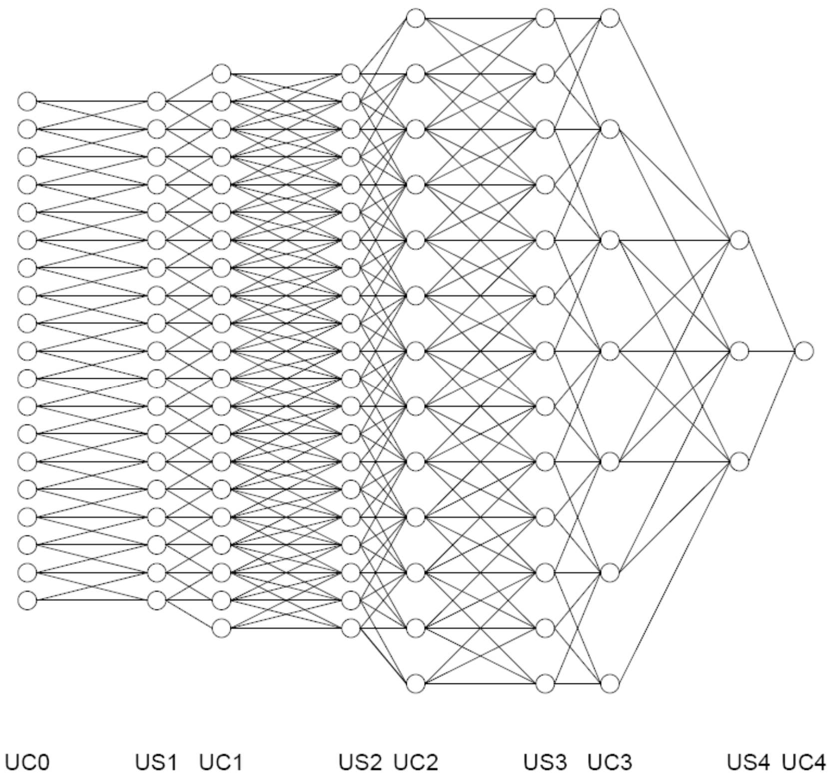
$$\psi[x]=\varphi[x]/(1+\varphi[x])$$



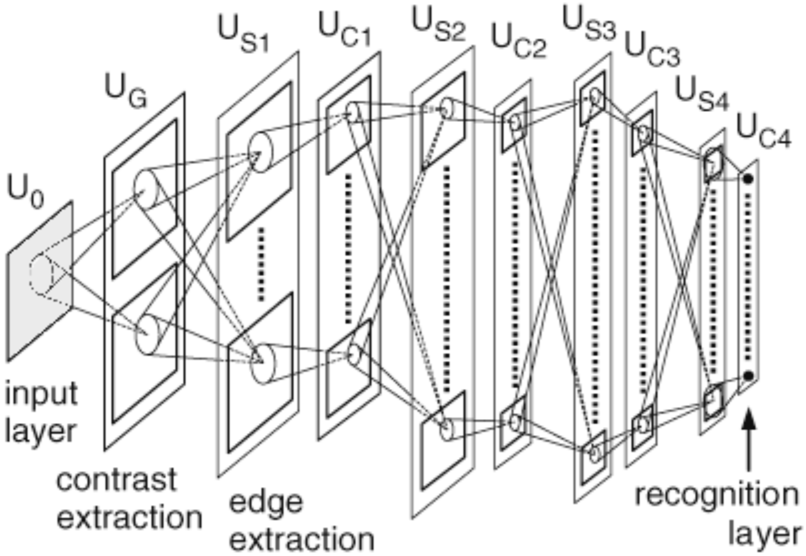
Outputs of cells (cont.)

- V-cell:

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



The big picture – Input layer



- Representation of input:

000010000

000110000

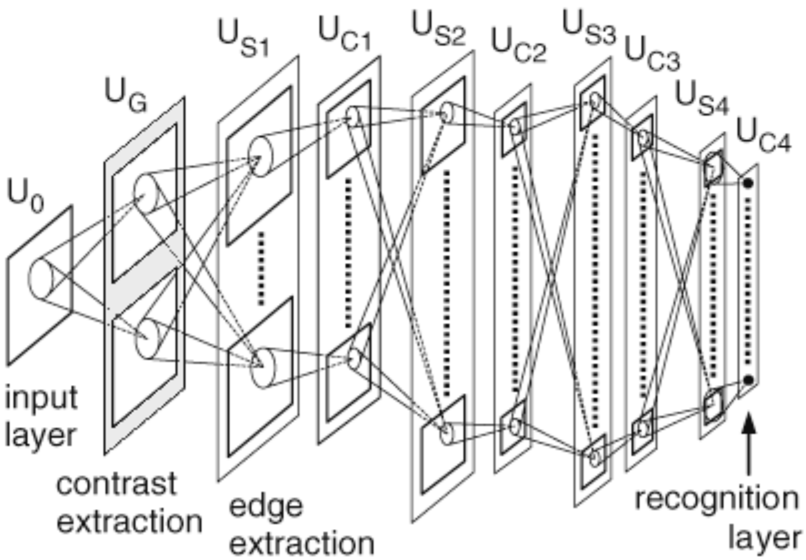
000010000

000010000

000010000

000111000

The big picture – Contrast Extract.



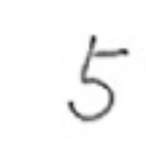
- Output of a contrast extracting cell:

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

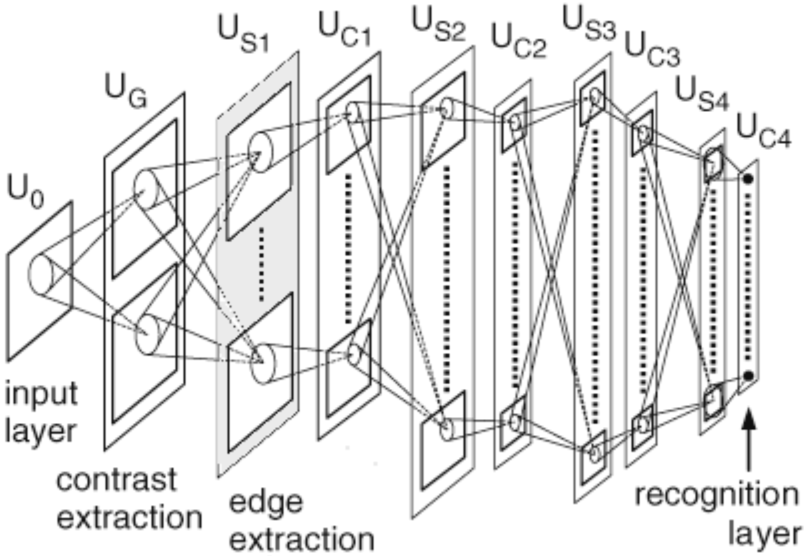
where $\varphi[\ ]$ is a function defined by $\varphi[x] = \max(x, 0)$

- k = 1

- k = 2



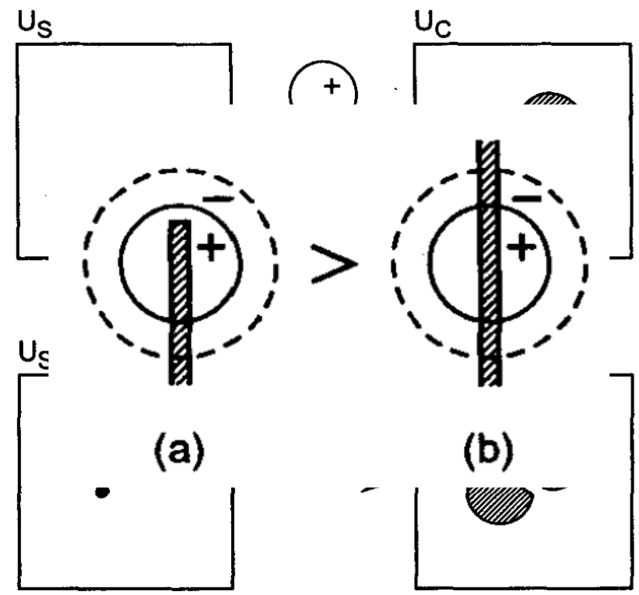
The big picture – Edge Extraction



- Layer Us1
- 16 cell-planes
- Supervised learning

The big picture – Inhibitory Surround

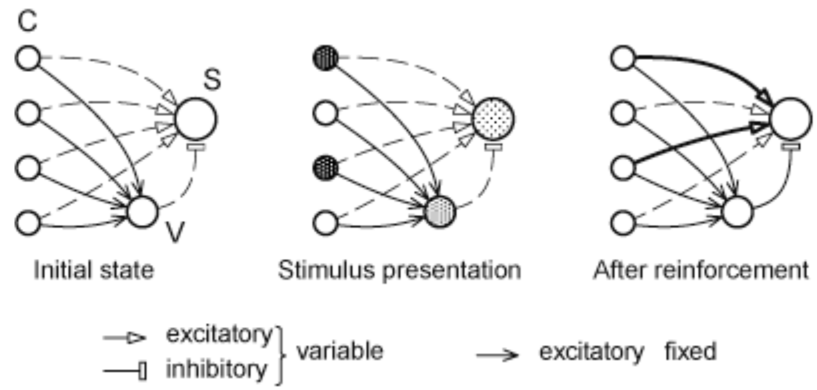
- Creates a non-responding zone
- Endows C-cells with the characteristics of end-stopped cells
- Participates in extraction of bend points and end points
- Threshold adjustment
- Introduced in Uc1, Uc2, not Uc3. Why?



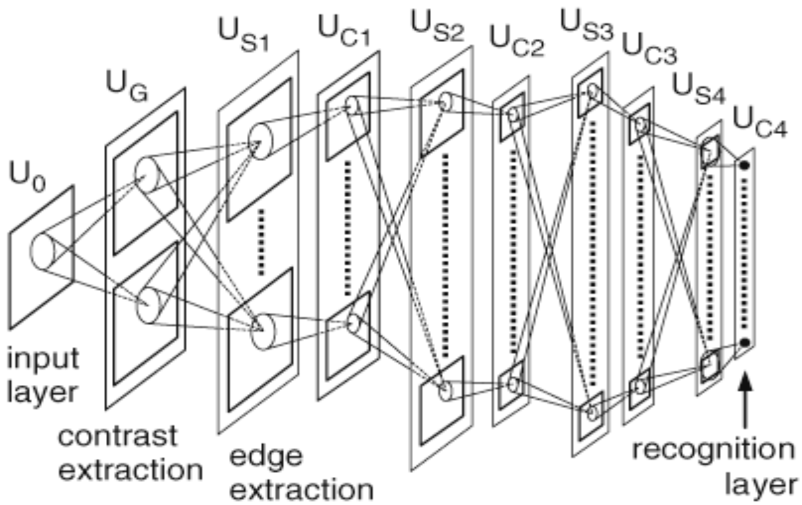
(b) Inhibitory surround (proposed)

The big picture – Us2, Us3

- Self organized using unsupervised competitive learning
- Selectivity of feature-extracting cells



The big picture – Us4



- Supervised competitive learning
- Presenting a training pattern
- Recognition phase

References

- *Kunihiko Fukushima* “Neocognitron of a New Version: Handwritten Digit Recognition”
- *D. R. Lovell, T. Downs and A. C. Tsoi* “An Evaluation of The Neocognitron”
- *D. H. Hubel and T. N. Wiesel* “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex”

Questions