

第21回IEICE ICT Pioneers Webinar, 2022.1.11

脳に学ぶ視覚情報処理

福島 邦彦

Kunihiko Fukushima

fukushima@m.ieice.org

<http://personalpage.flsi.or.jp/fukushima/>



本日の話題

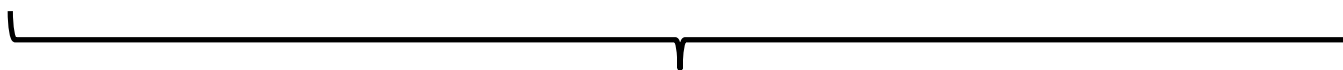
- 1) CNN（畳み込み神経回路）: ネオコグニトロン の着想
- 2) 頑強な 視覚パターン認識 の原理
- 3) 最近のネオコグニトロン: 通常のCNNとの相違点
- 4) ネオコグニトロンから発展したシステム
一部が隠されたパターンの認識
選択的注意機構のモデル, など

Deep CNN: ネオコグニトロンの着想 (Fukushima, 1979)

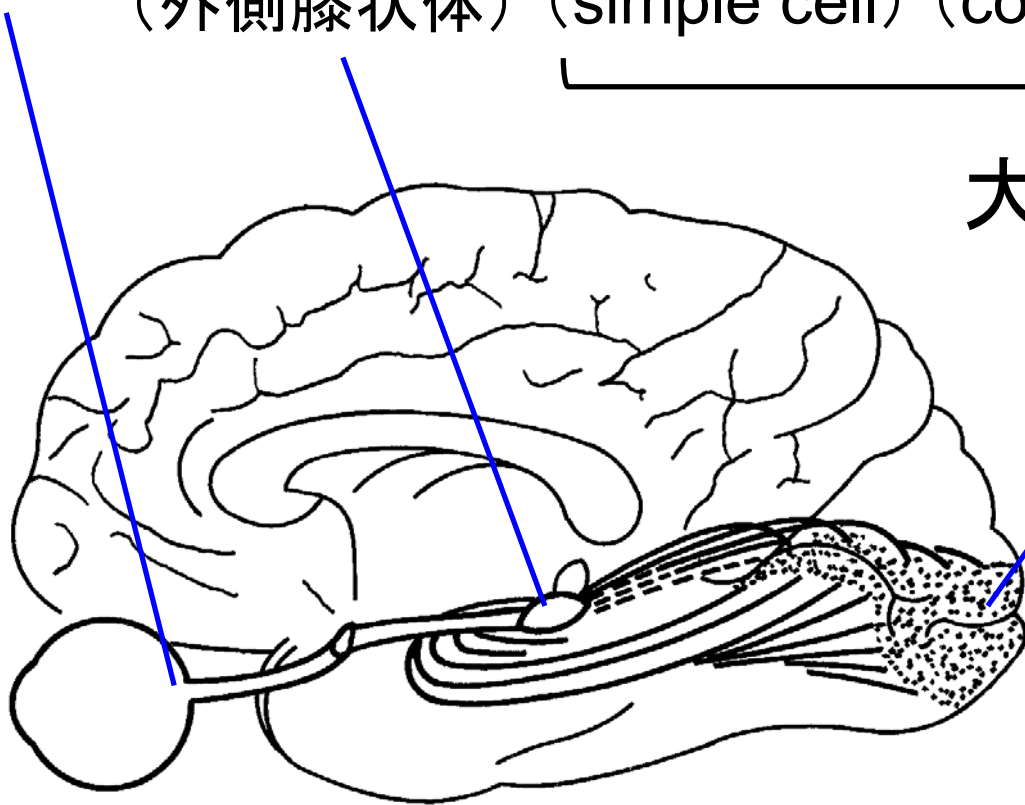
— 生物の脳に学ぶ —

Hubel-Wiesel の階層仮説

網膜 → LGN → 単純型 → 複雑型 → 超複雑型
(外側膝状体) (simple cell) (complex cell) (hyper-complex cell)

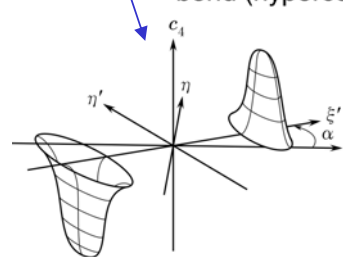
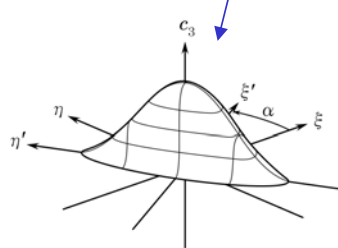
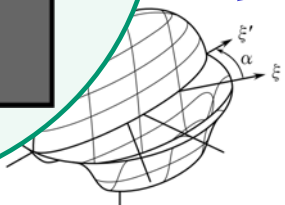
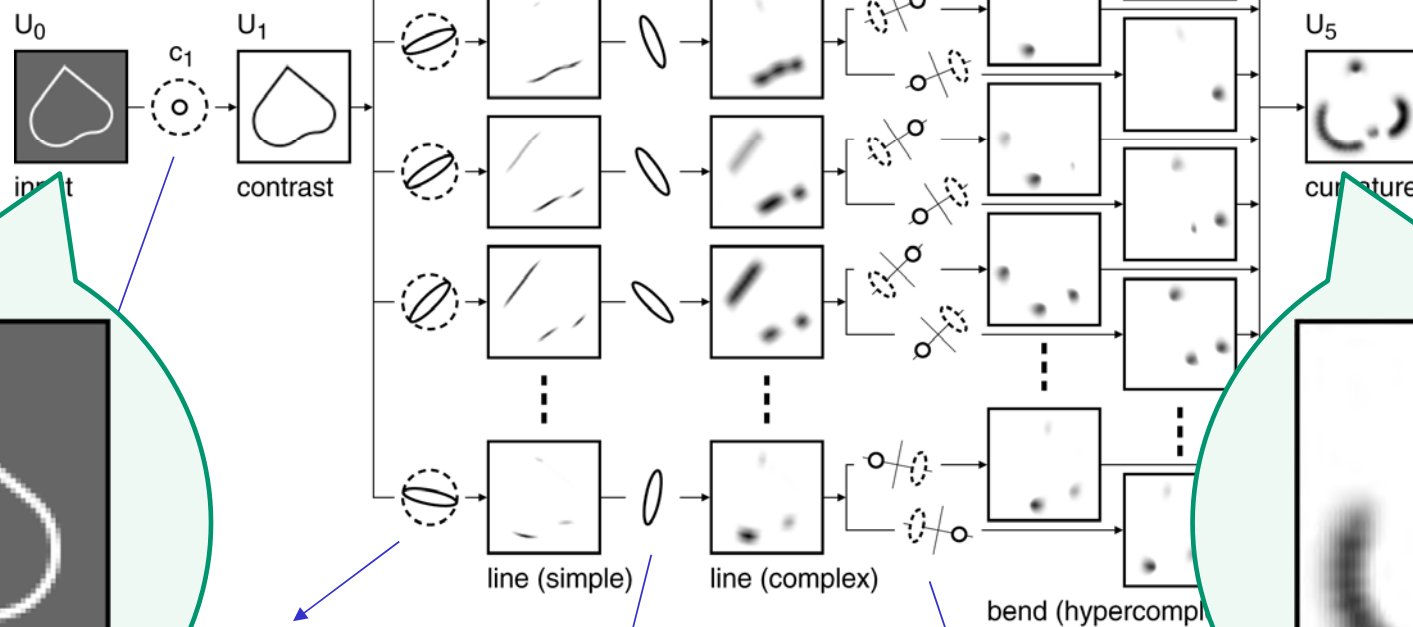
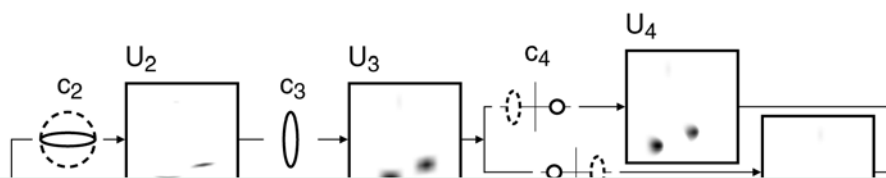


大脳 視覚野(V1)



曲率抽出回路

回路構造 = deep CNN (深層畳み込み神経回路)

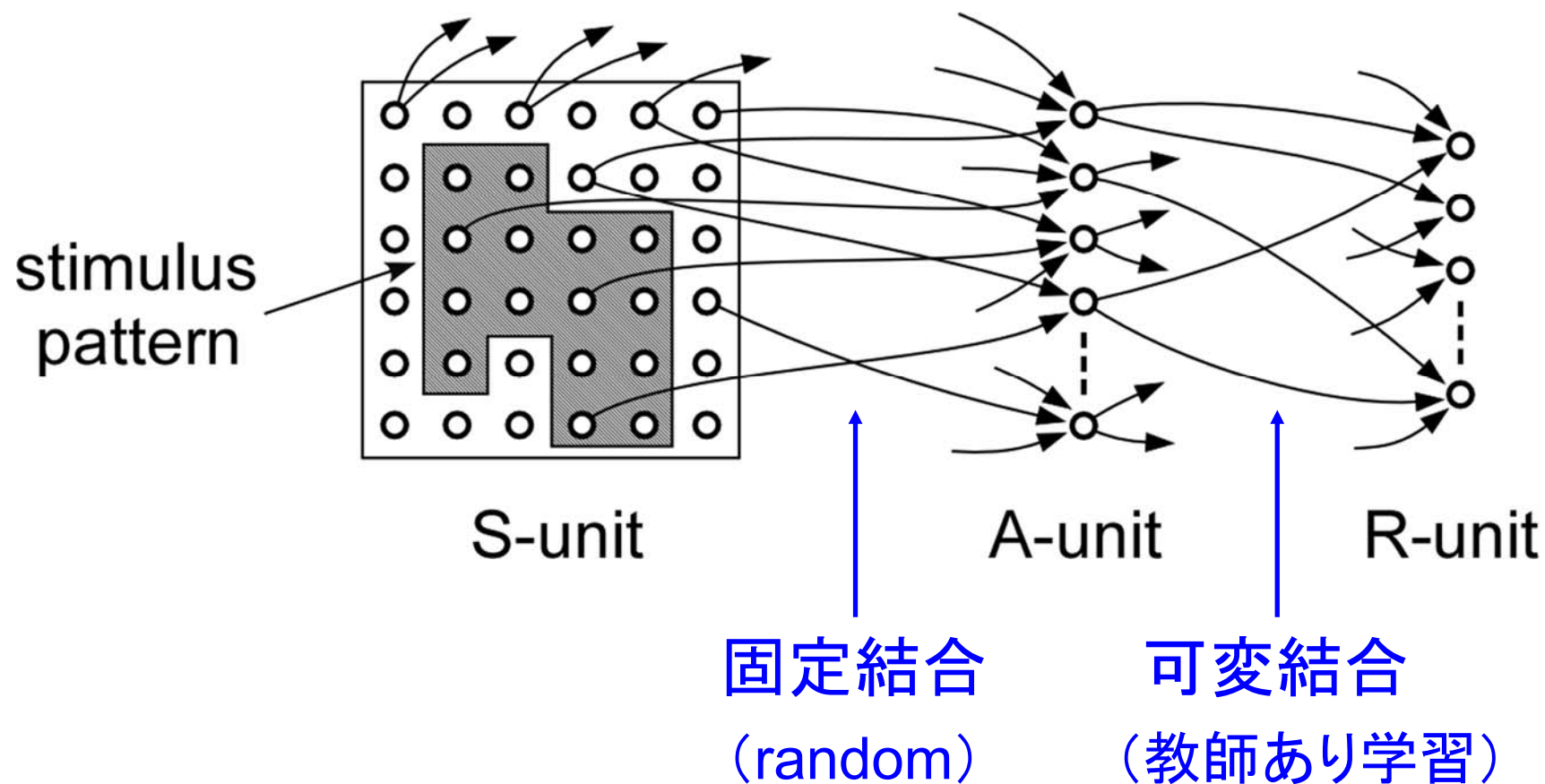


(Fukushima 1971)

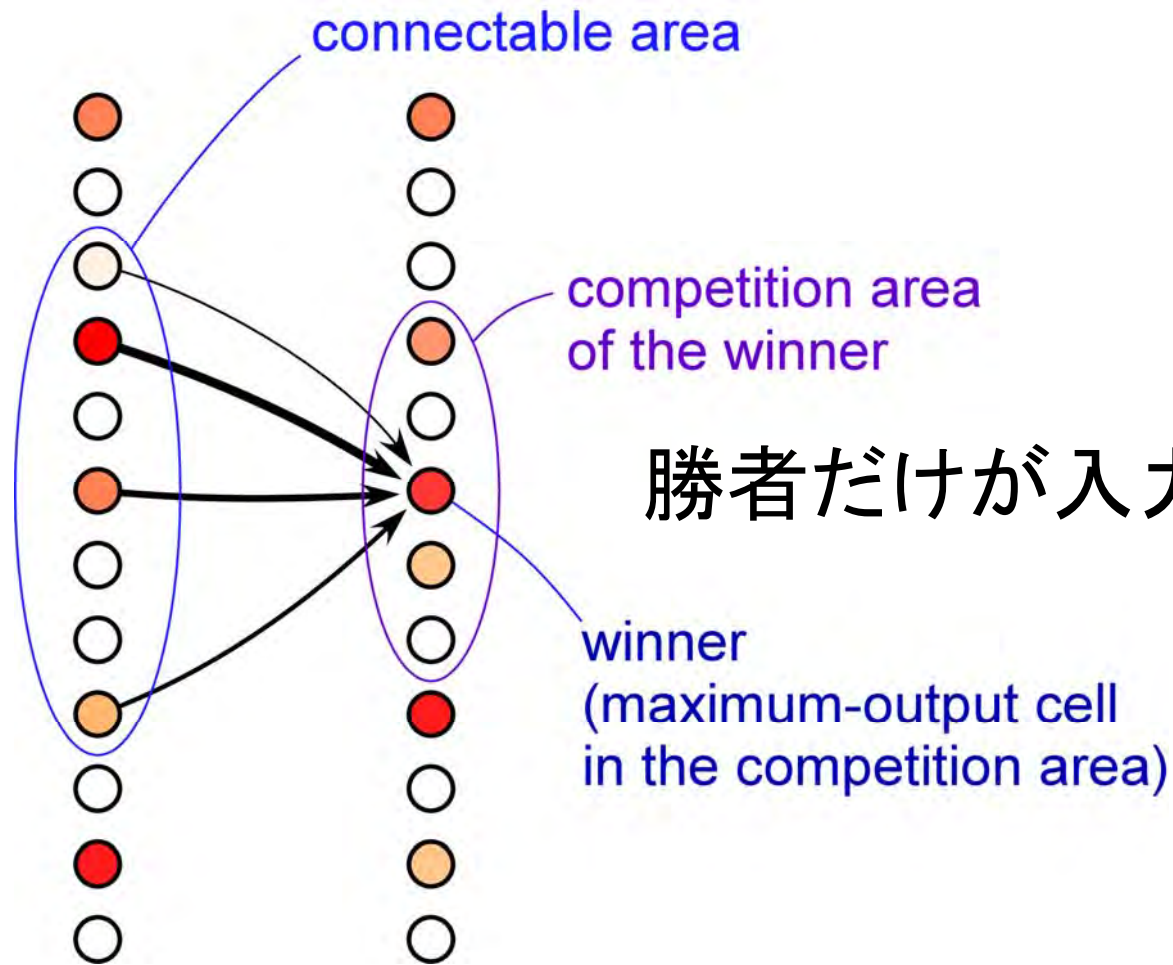
3-layered Perceptron (F. Rosenblatt)

(3層パーセプトロン)

多層回路の学習法は未解決

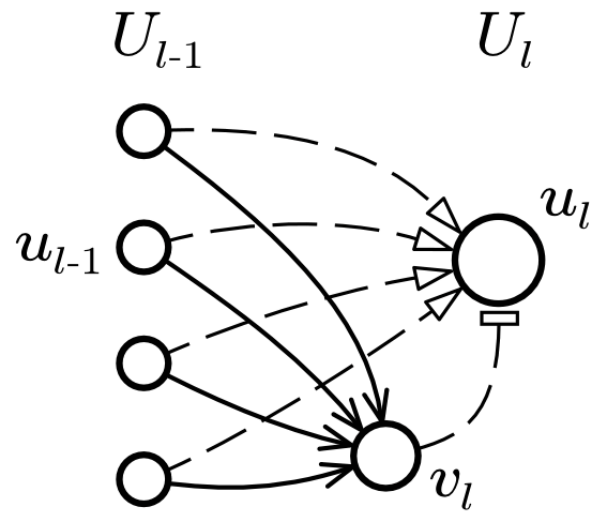


多層回路の学習に 競合学習則 —— コグニトロン(cognitron)

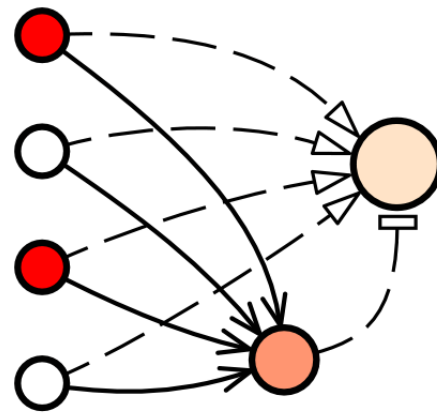


(Fukushima, 1975)

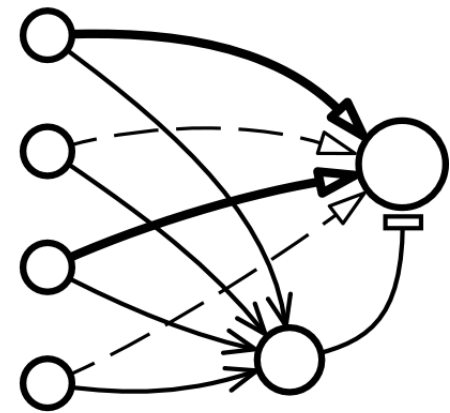
Learning process of the cognitron



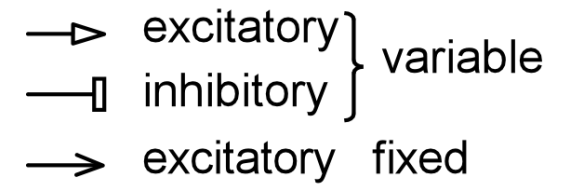
Initial state



Stimulus presentation



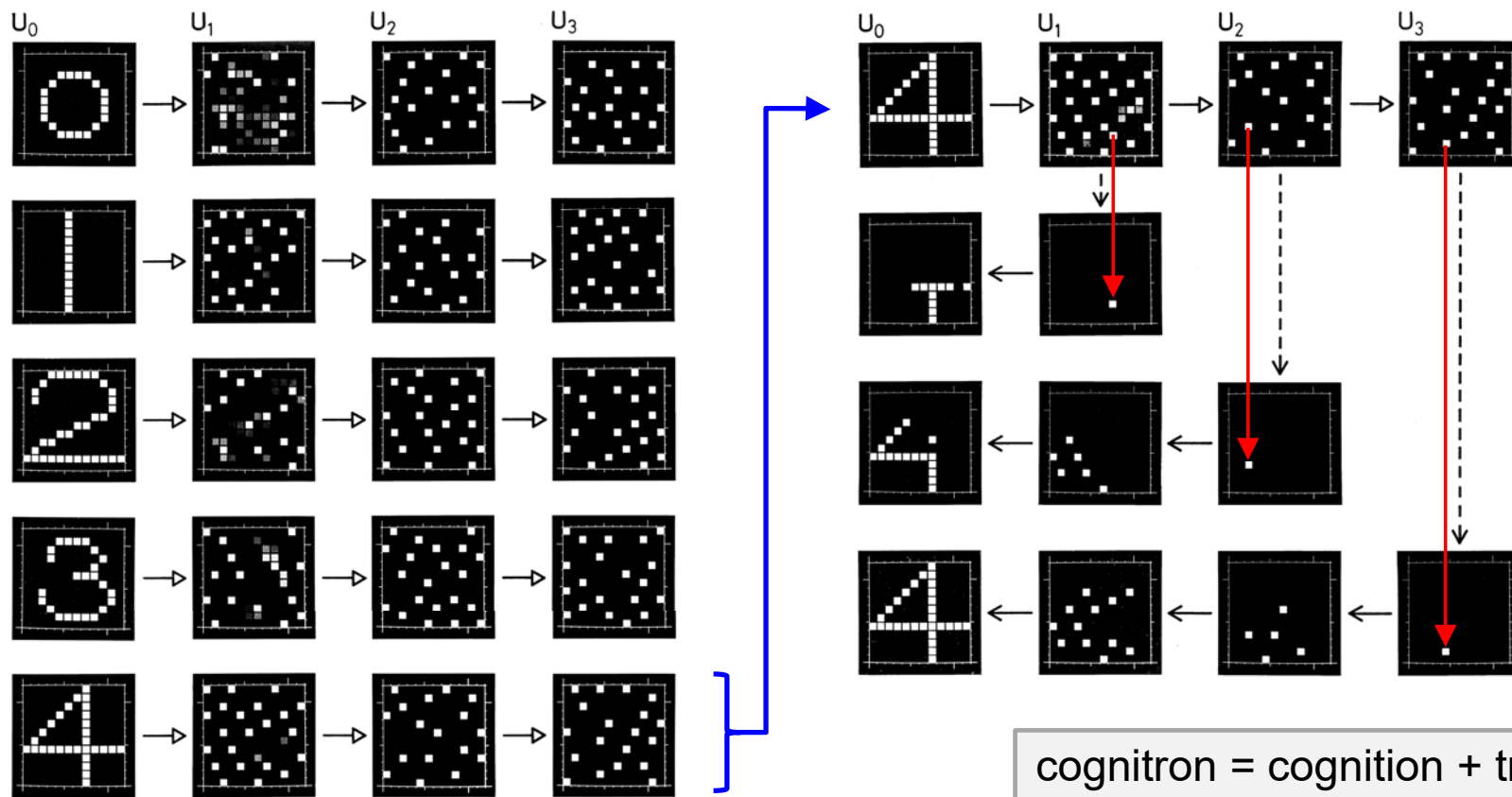
After reinforcement



cognitron

競合学習則 ----- 多層回路の教師なし学習を可能に

competitive learning ----- unsupervised learning for multi-layered network



cognitron = cognition + tron

c.f.: perceptron = perception + tron

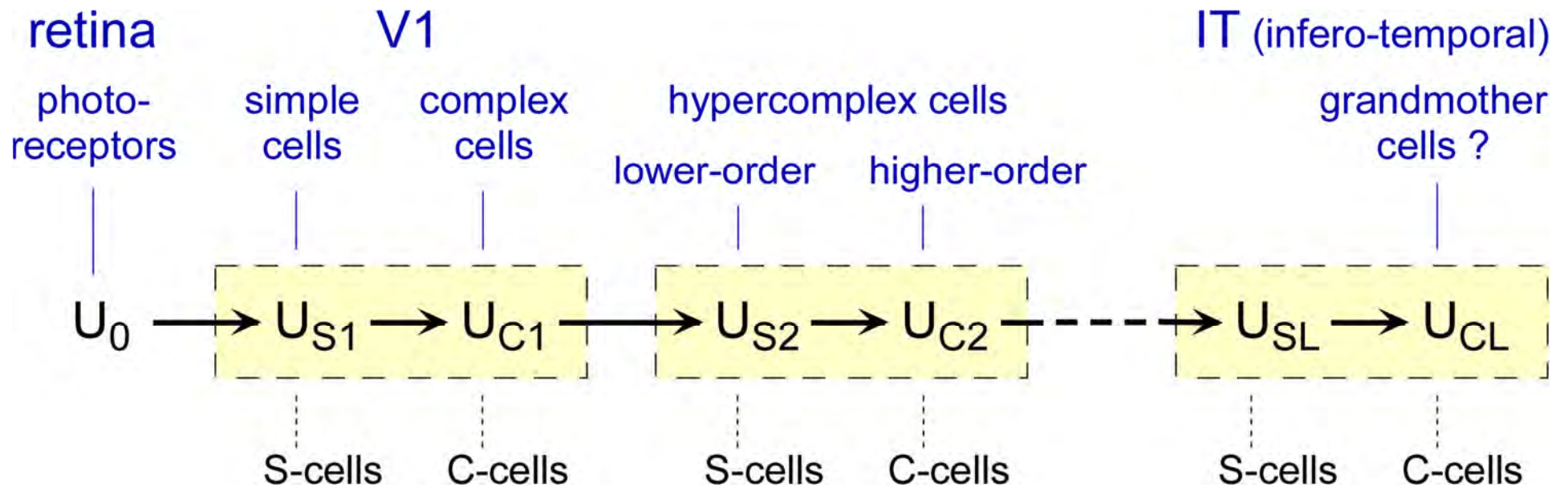
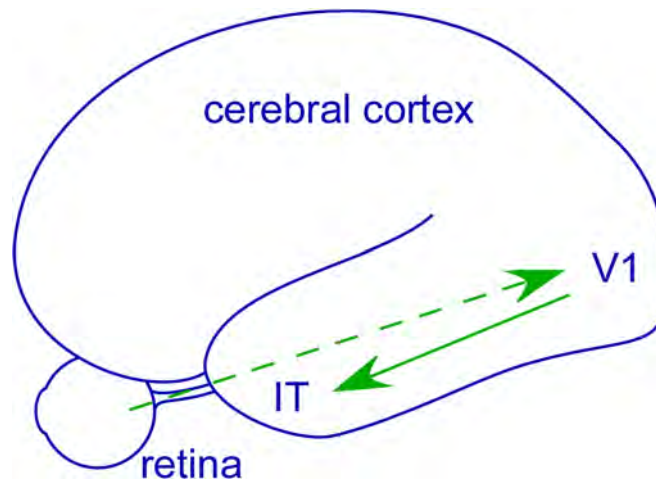
(Fukushima, 1975)

cognitron:

教師なし学習でパターン認識能力を獲得（1975）

次の目標

入力パターンの 変形・位置ずれ・拡大縮小 に対応
CNN(畳み込み神経回路)の学習

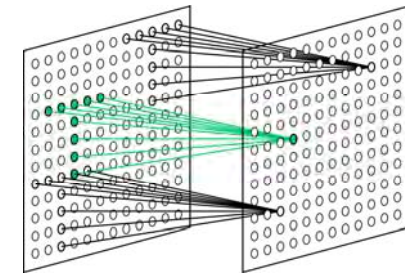
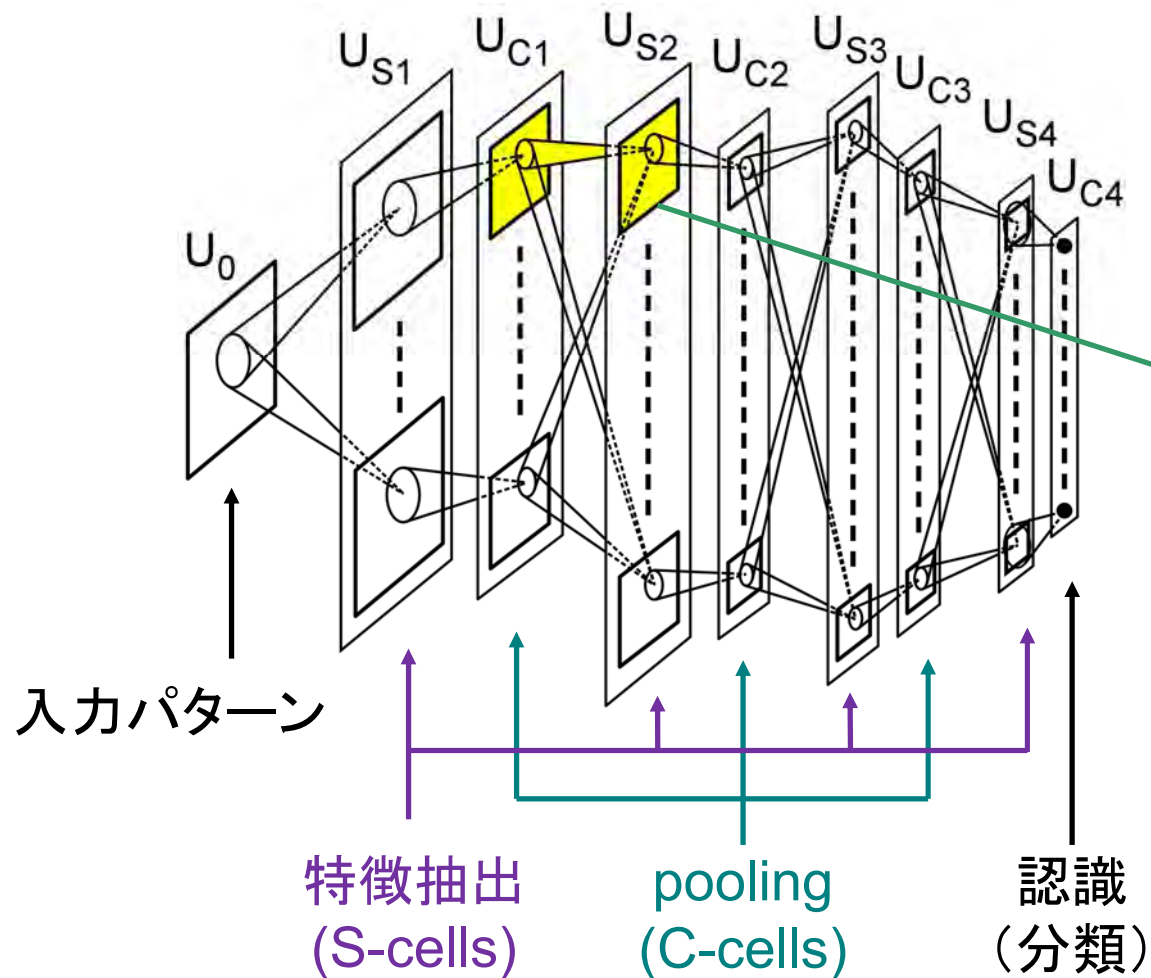


neocognitron

(Fukushima 1979, 1980)

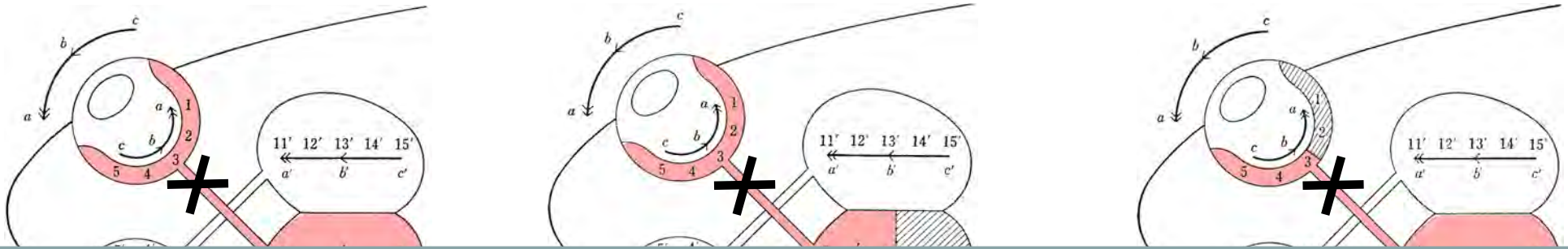
neocognitron = 視覚パターン認識能力を持つ

dCNN(深層畳み込み神経回路)



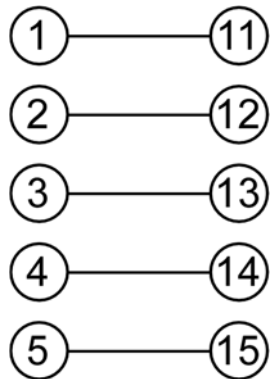
共有結合
= spatial filtering
= **convolution**

網膜 (retina) 視蓋 (optic tectum) 間の結合の再生



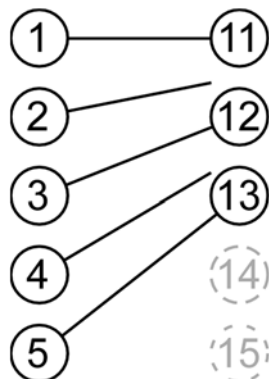
結合発生・再生時にも retinotopy は常に保たれる。

normal



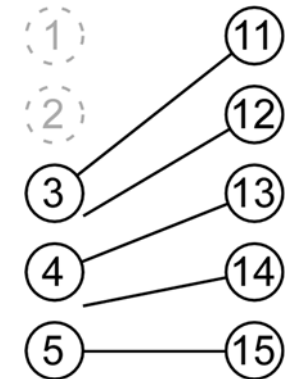
retina tectum

half tectum



retina tectum

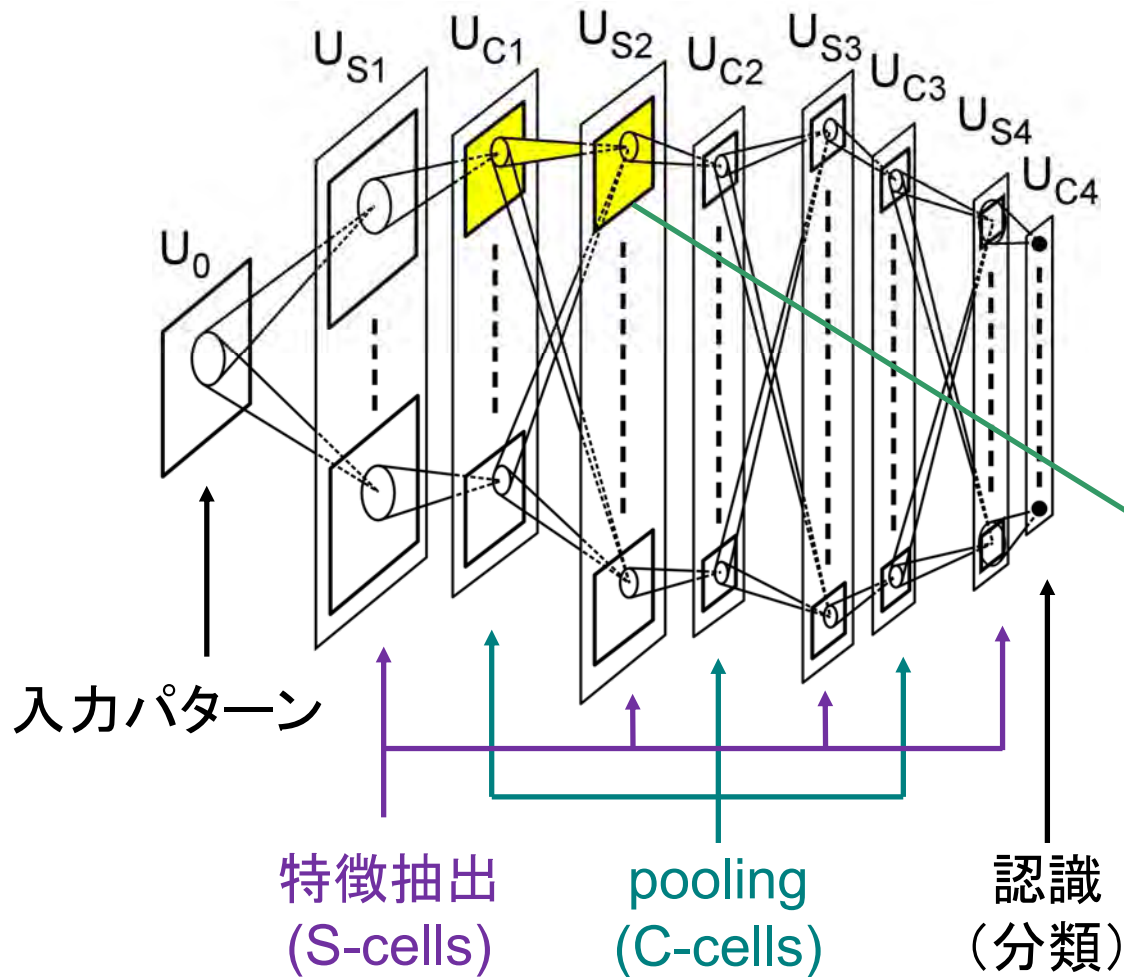
half retina



retina tectum

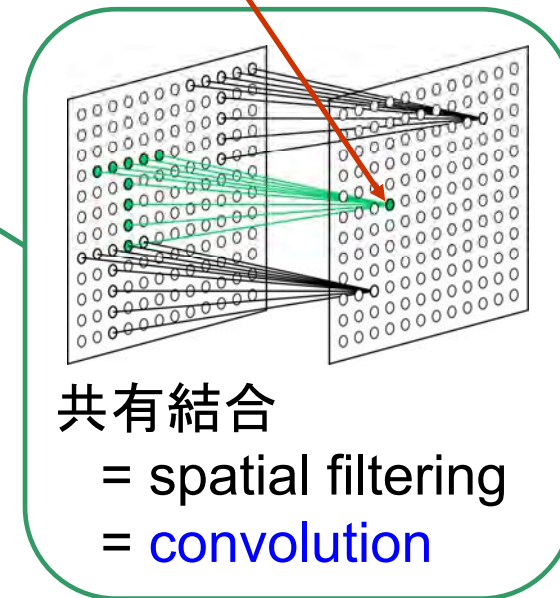
neocognitron = 視覚パターン認識能力を持つ dCNN

(深層畳み込み神経回路)



seed cell (結晶成長の核のような働き)

他の細胞も seed cell に従って
入力結合を変化

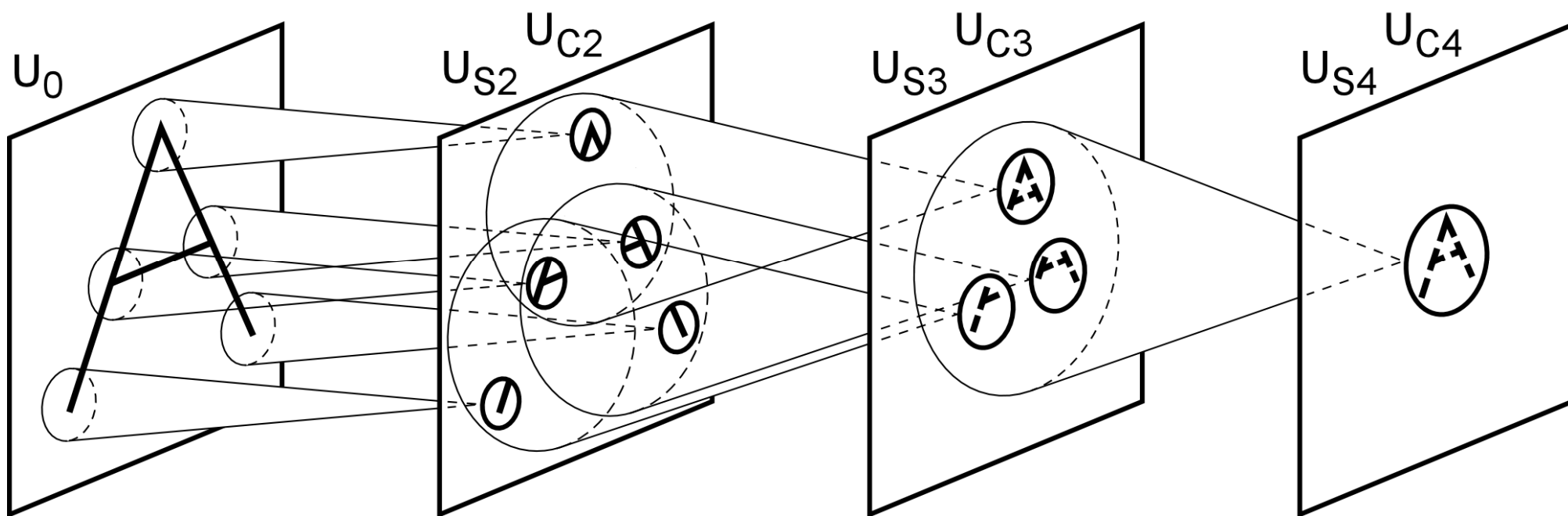


(Fukushima 1979, 1980)

頑強な 視覚パターン認識 の原理

Principles of Robust Visual Pattern Recognition

ネオコグニトロンにおける 階層的情報処理

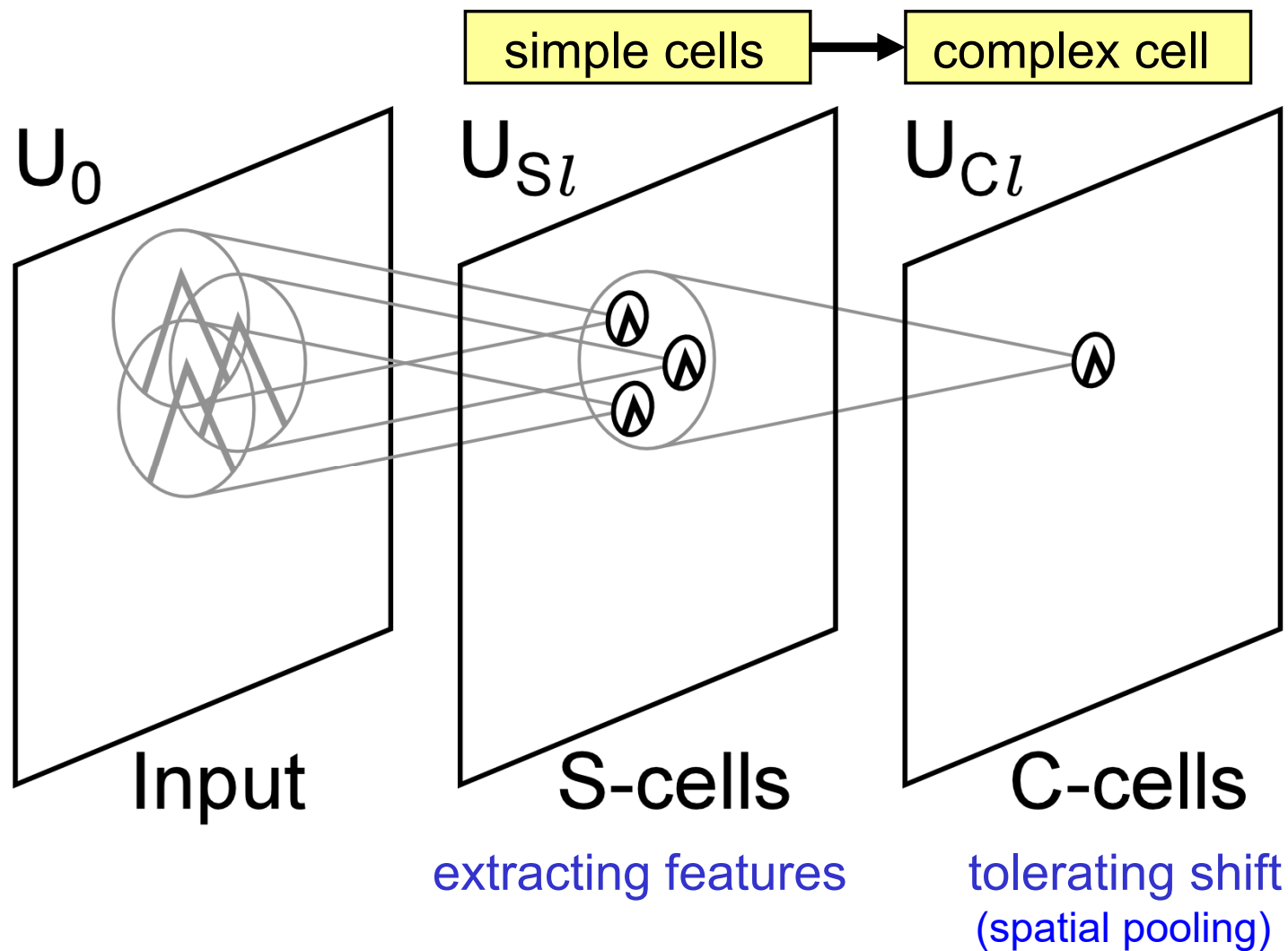


C細胞（複雑型細胞）

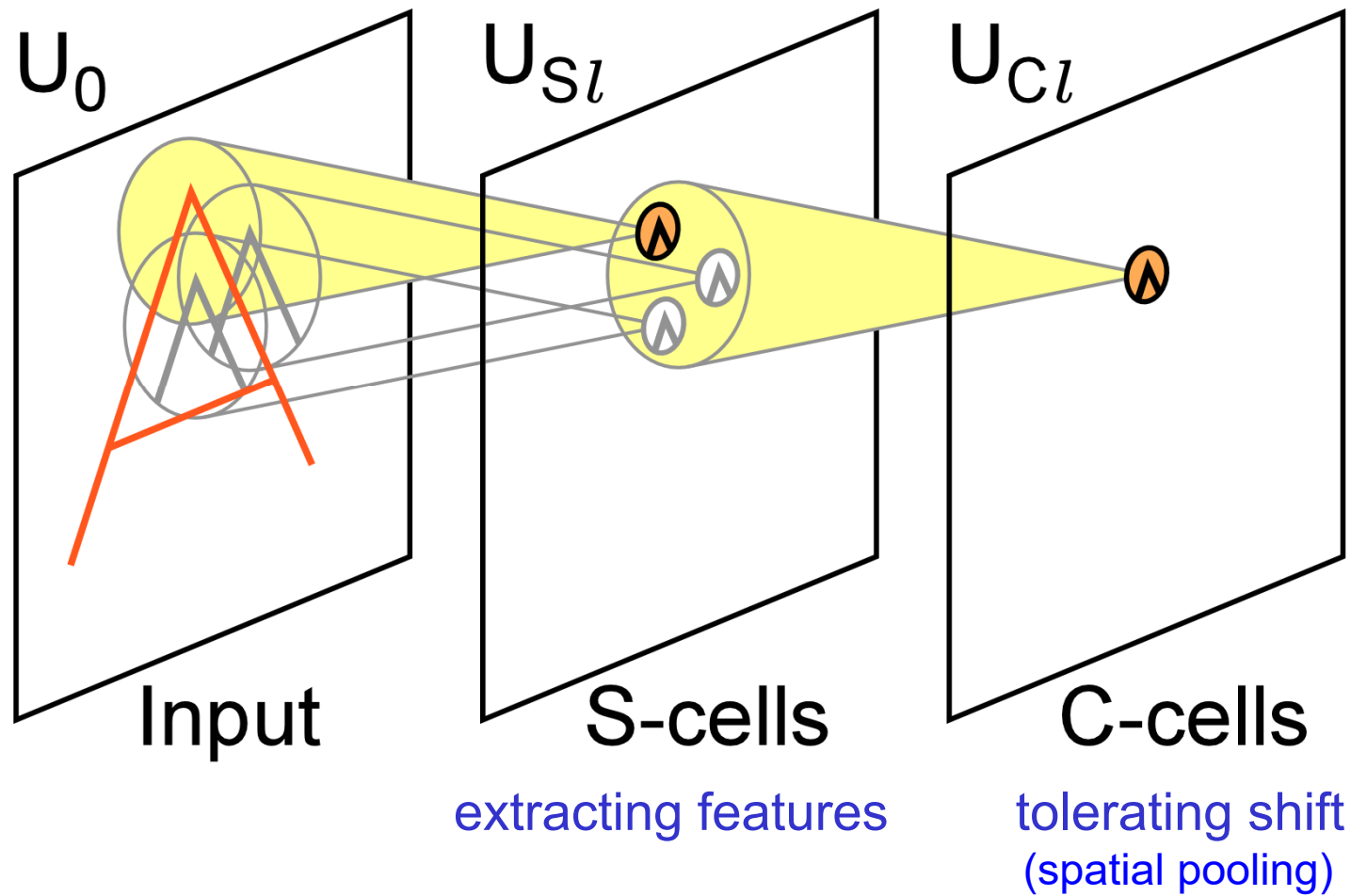
—— ぼかしによって位置ずれを許容 ——

C-cells (complex cells)

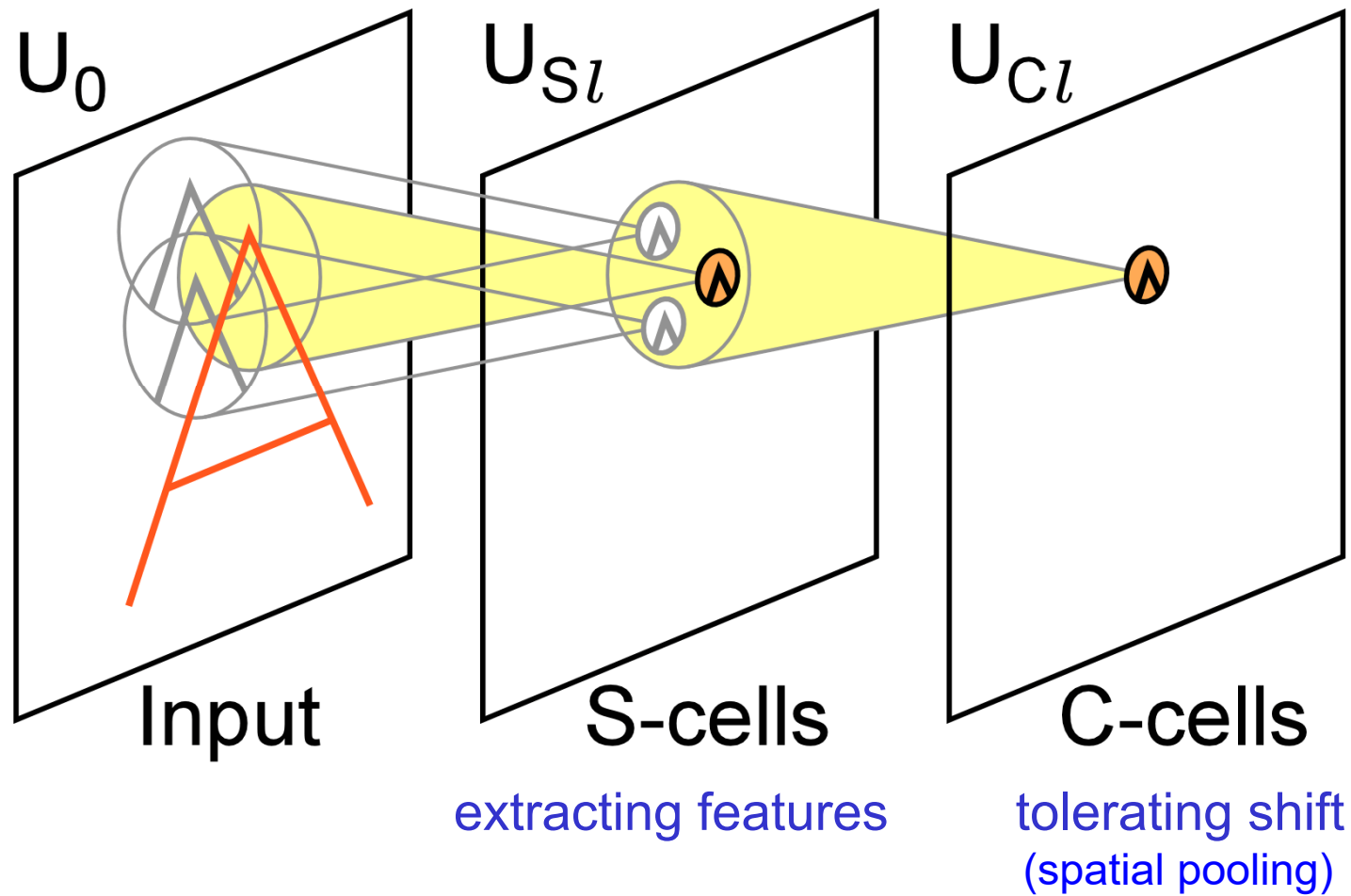
--- tolerating shift by blurring ---



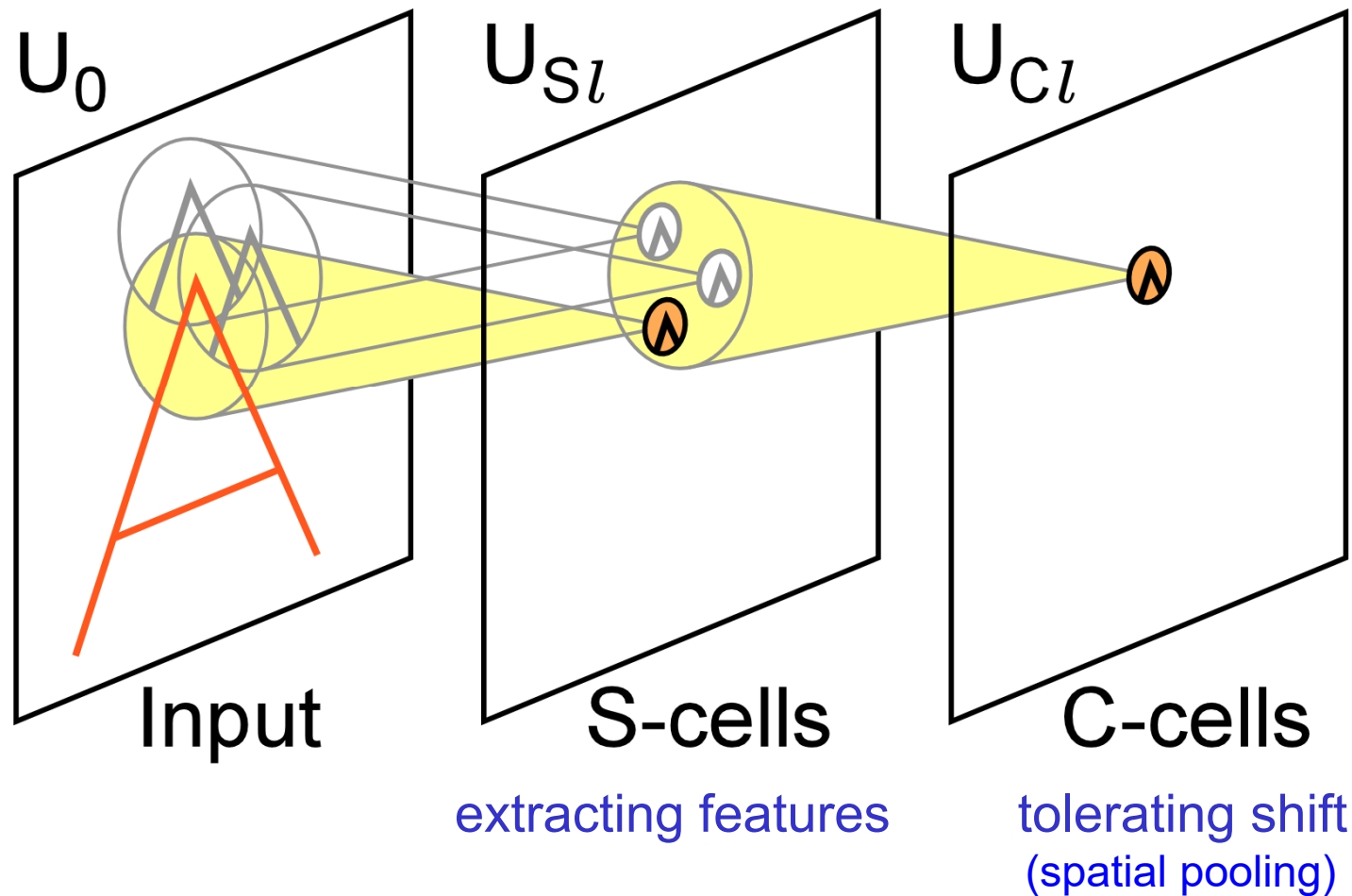
C細胞：局所的特徴の位置ずれを許容



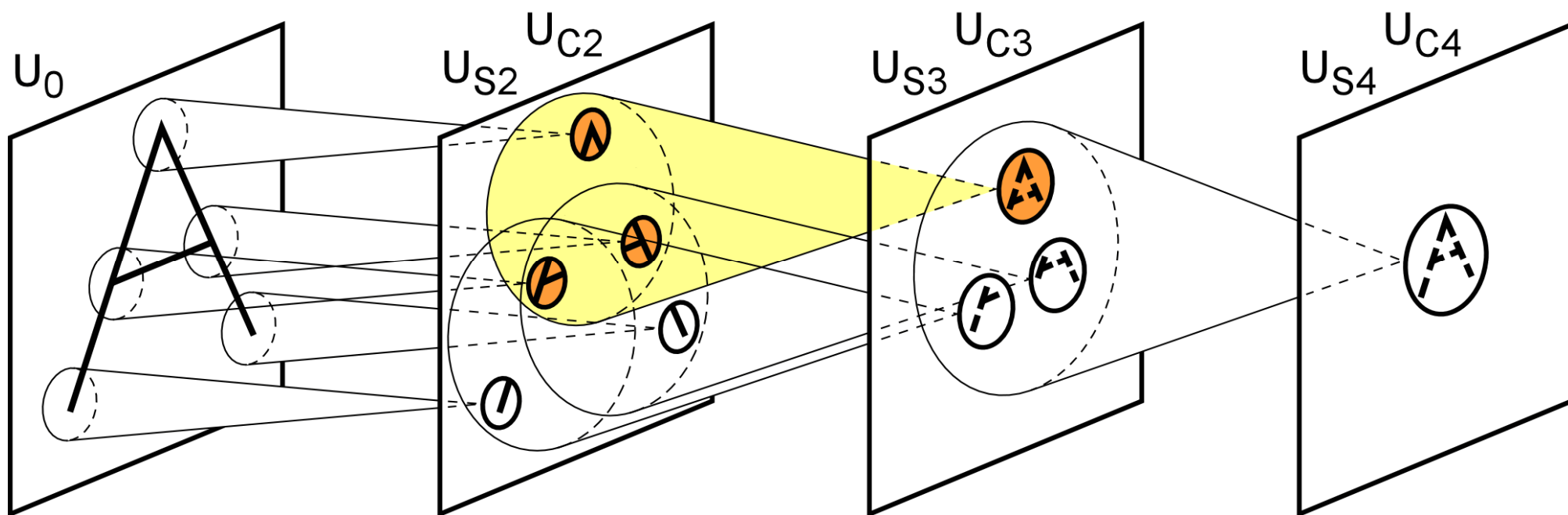
C細胞：局所的特徴の位置ずれを許容

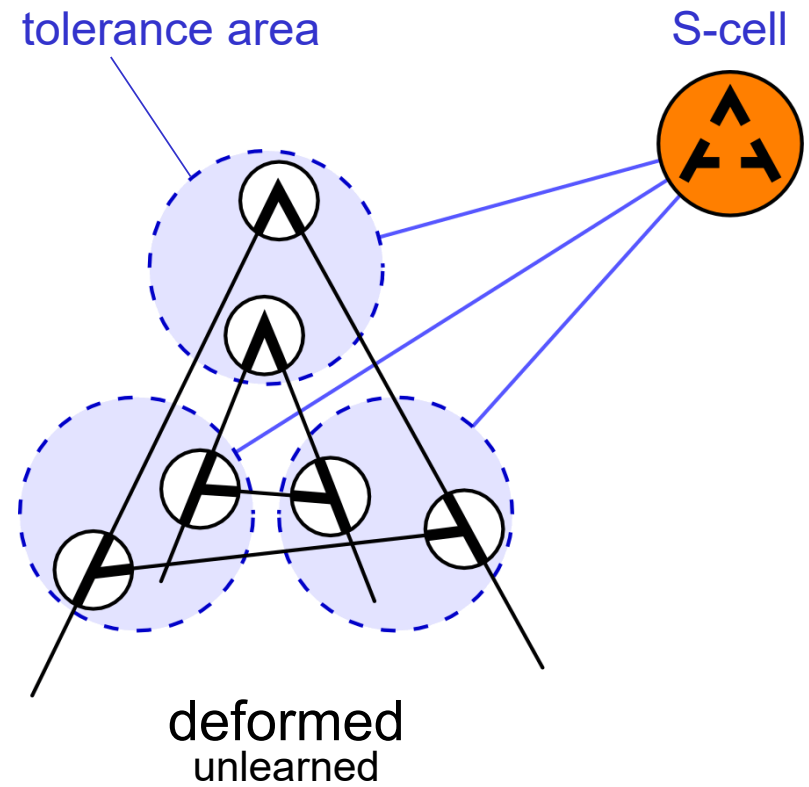
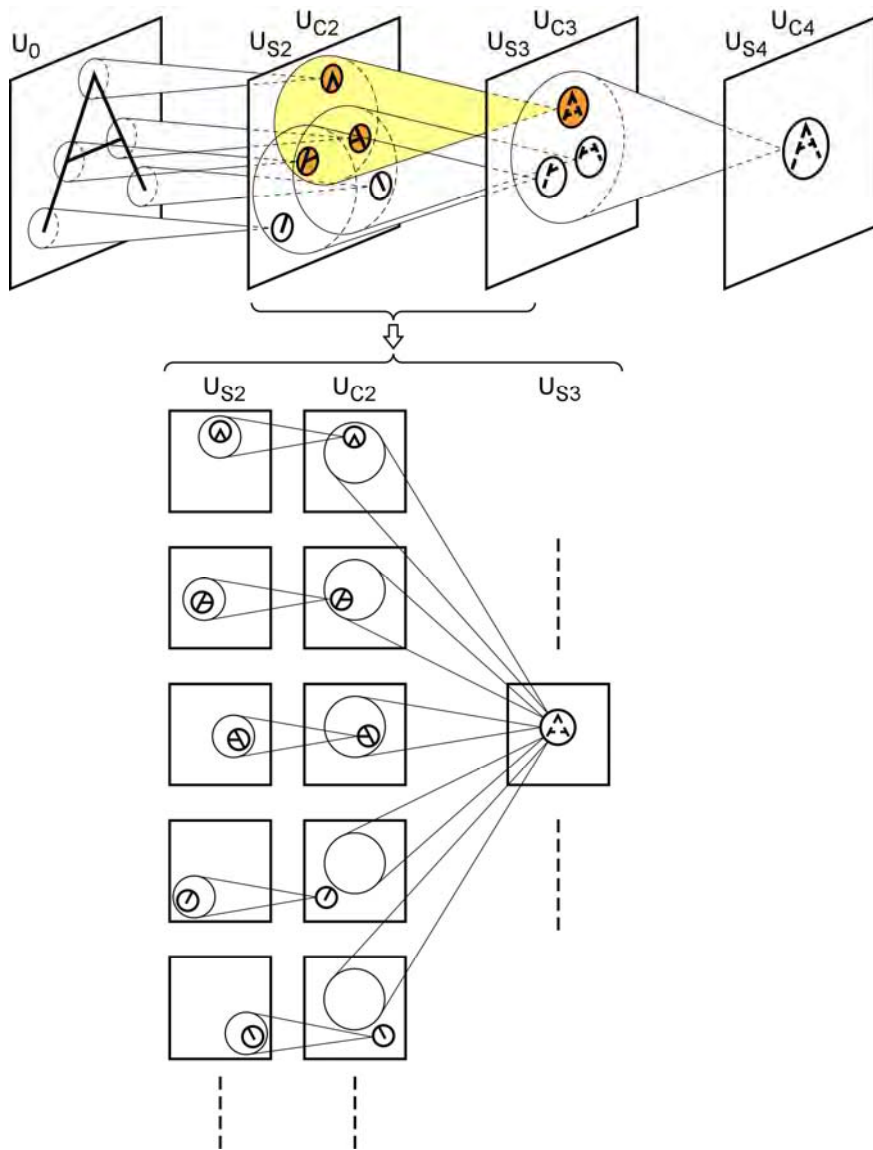


C細胞：局所的特徴の位置ずれを許容



C細胞：局所的特徴の位置ずれを許容

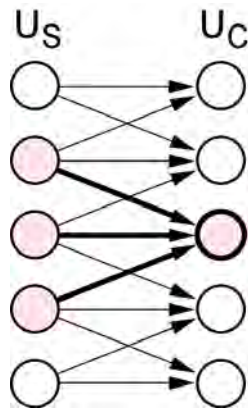




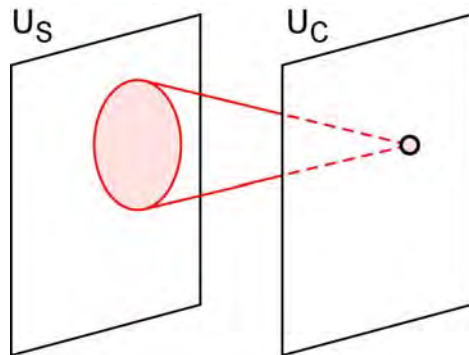
Robust recognition of deformed patterns
by tolerating positional error of local features

C-cells (complex cells) --- tolerating shift by blurring ---

S-cells C-cells

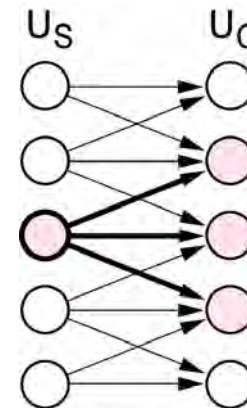


shift tolerance

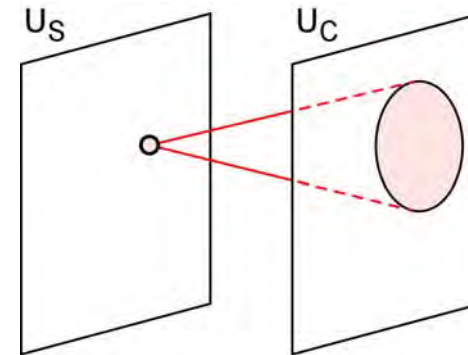


tolerating positional error

S-cells C-cells



spatial blur

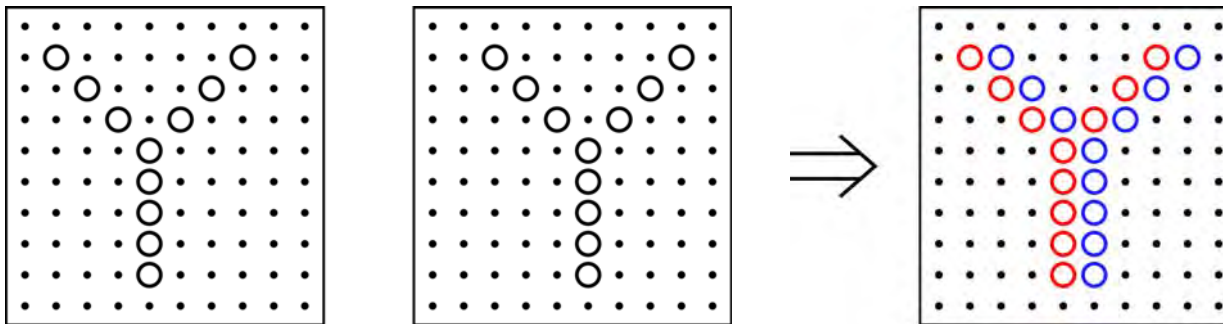


blurring responses

Connections from S- to C-cells --- pooling

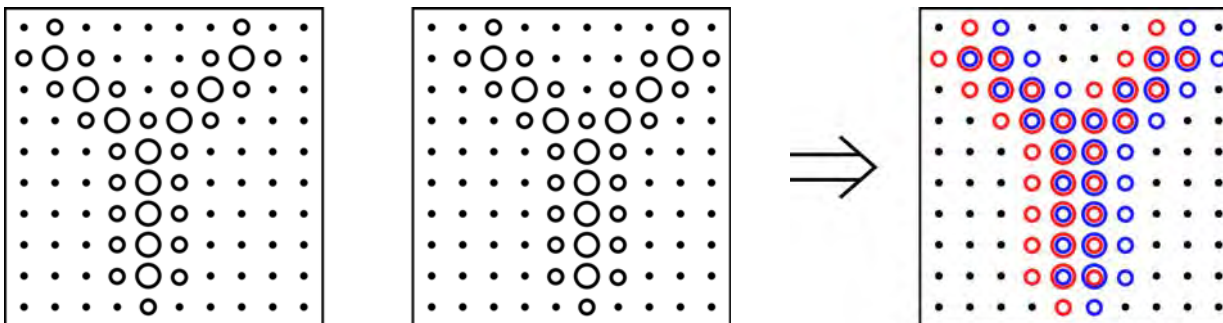
Why blur is useful? → Tolerating shift and deformation



Without blur:



 and  are orthogonal ?!

After blurring operation:



 and  largely overlap each other.

C細胞層の役割:

- ぼかしによって位置ずれを許容
変形に対しても頑強になる
平均化 による pooling, [max-pooling ではない]
- 空間軸に対する粗いサンプリング (= 細胞の 間引き)
細胞密度を減らして計算量を削減
- down-sampling の前に ぼかし
(= 粗いサンプリングの前に low-pass フィルター)
aliasing noise (モアレ, moiré)を防ぐ
- ランダムノイズを平滑化して軽減
平均化操作によって

S細胞 —— 特徴抽出細胞 ——

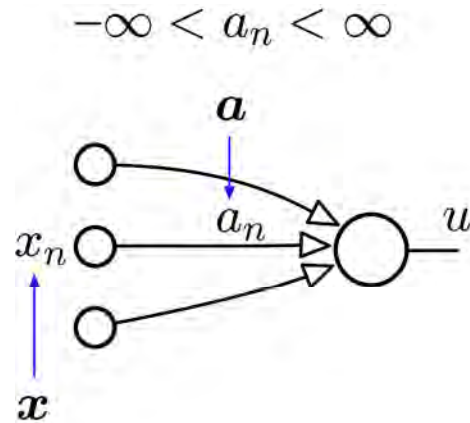
S-cells (simple cells)

--- feature extracting cells ---

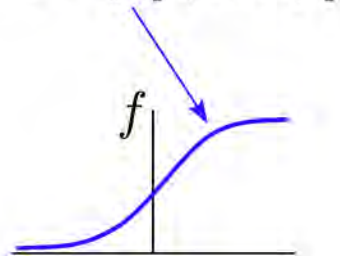
similarity defined by inner product
+
rectified linear

(Fukushima 2011)

Feature-extracting cell (S-cell)

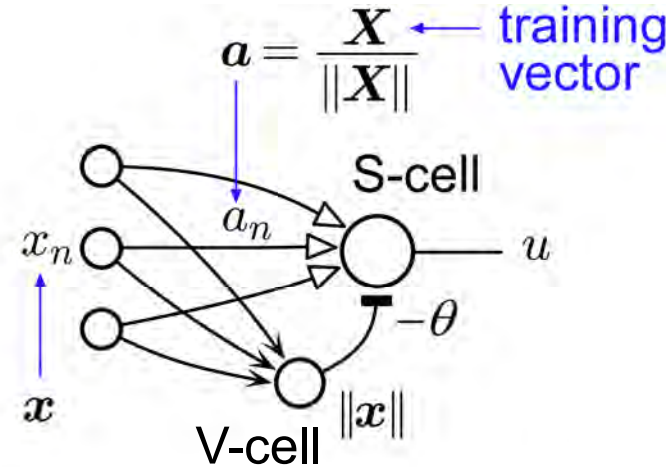


$$u = f [(a, x)]$$

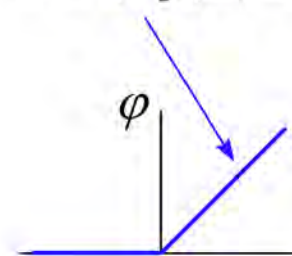


sigmoid

conventional NN



$$u = \varphi [(a, x) - \theta \|x\|]$$



rectified linear (ReLU)

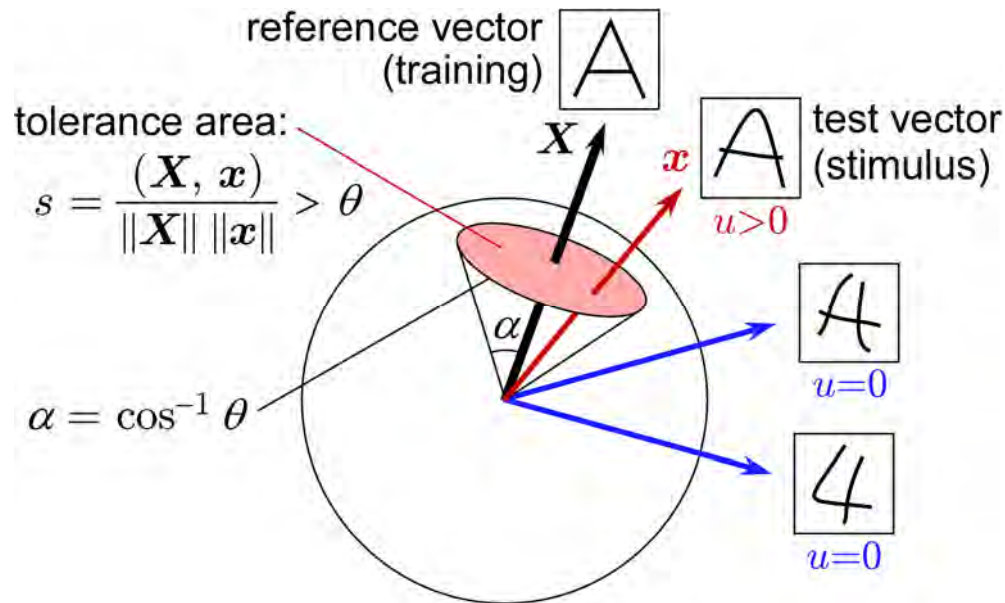
neocognitron

Feature extraction by an S-cell

$$u = \|x\| \frac{\varphi[s - \theta]}{1 - \theta}$$

$$s = \frac{(X, x)}{\|X\| \cdot \|x\|}$$

similarity between
reference vector X
and test vector x



training
vector

X

C-cells

(X_n)

x_n

test vector
 x

$$a = \frac{X}{\|X\|}$$

S-cell

u

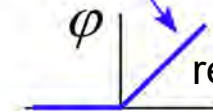
$-\theta$

$\|x\|$

$$u = \frac{1}{1 - \theta} \cdot \varphi[(a, x) - \theta \|x\|]$$

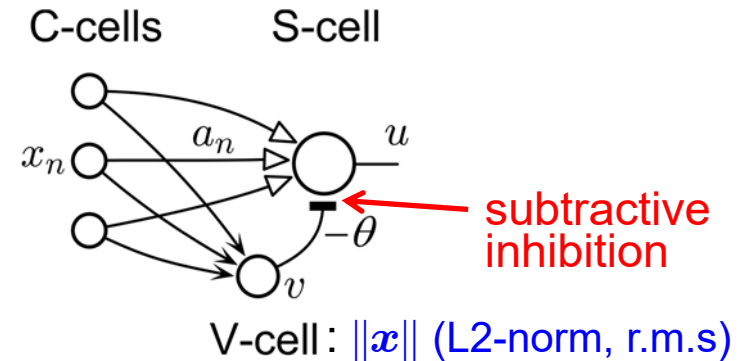
φ

rectified linear (ReLU)



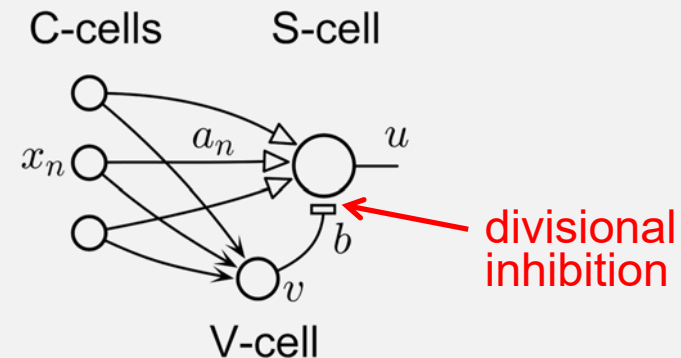
S-cell with **subtractive** inhibition ----- (recent neocognitron)

$$u = \|\mathbf{x}\| \frac{\varphi[s - \theta]}{1 - \theta}$$



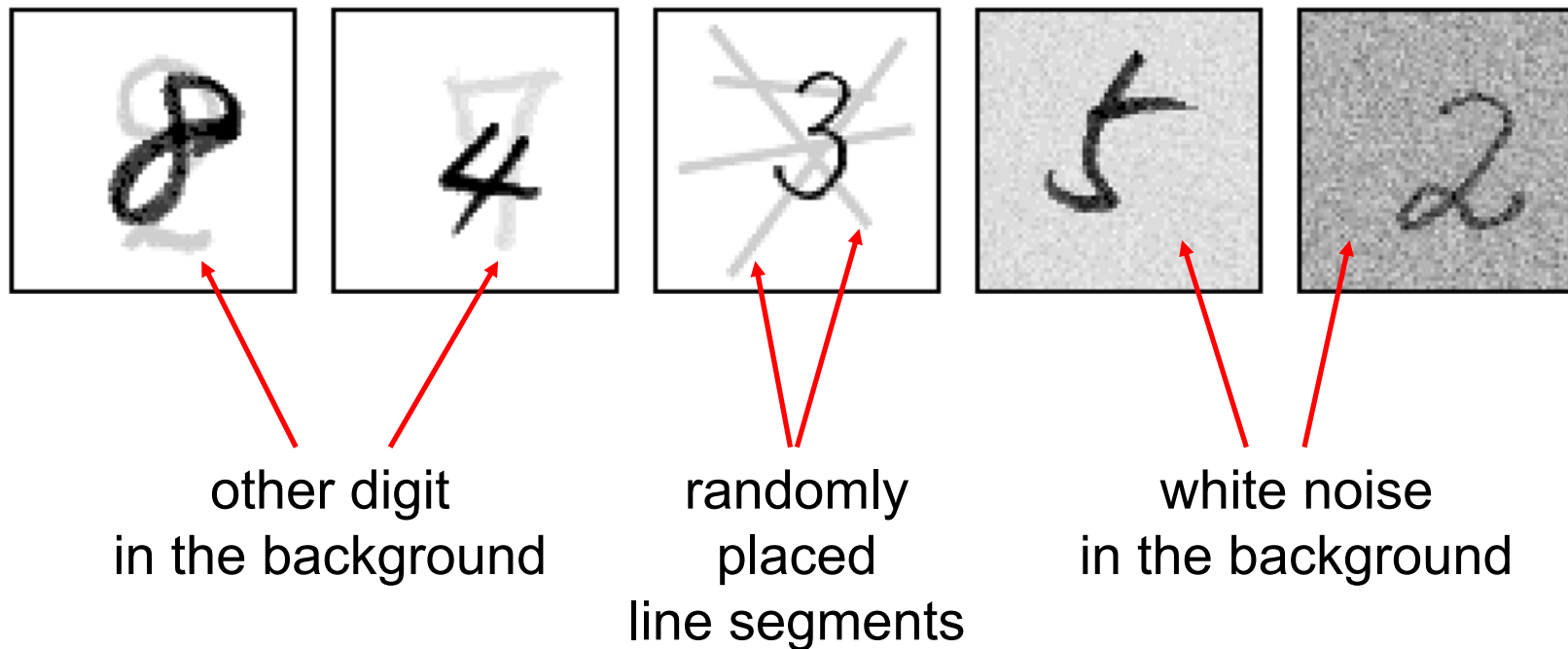
S-cell with **divisional** inhibition ----- (old neocognitron)

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

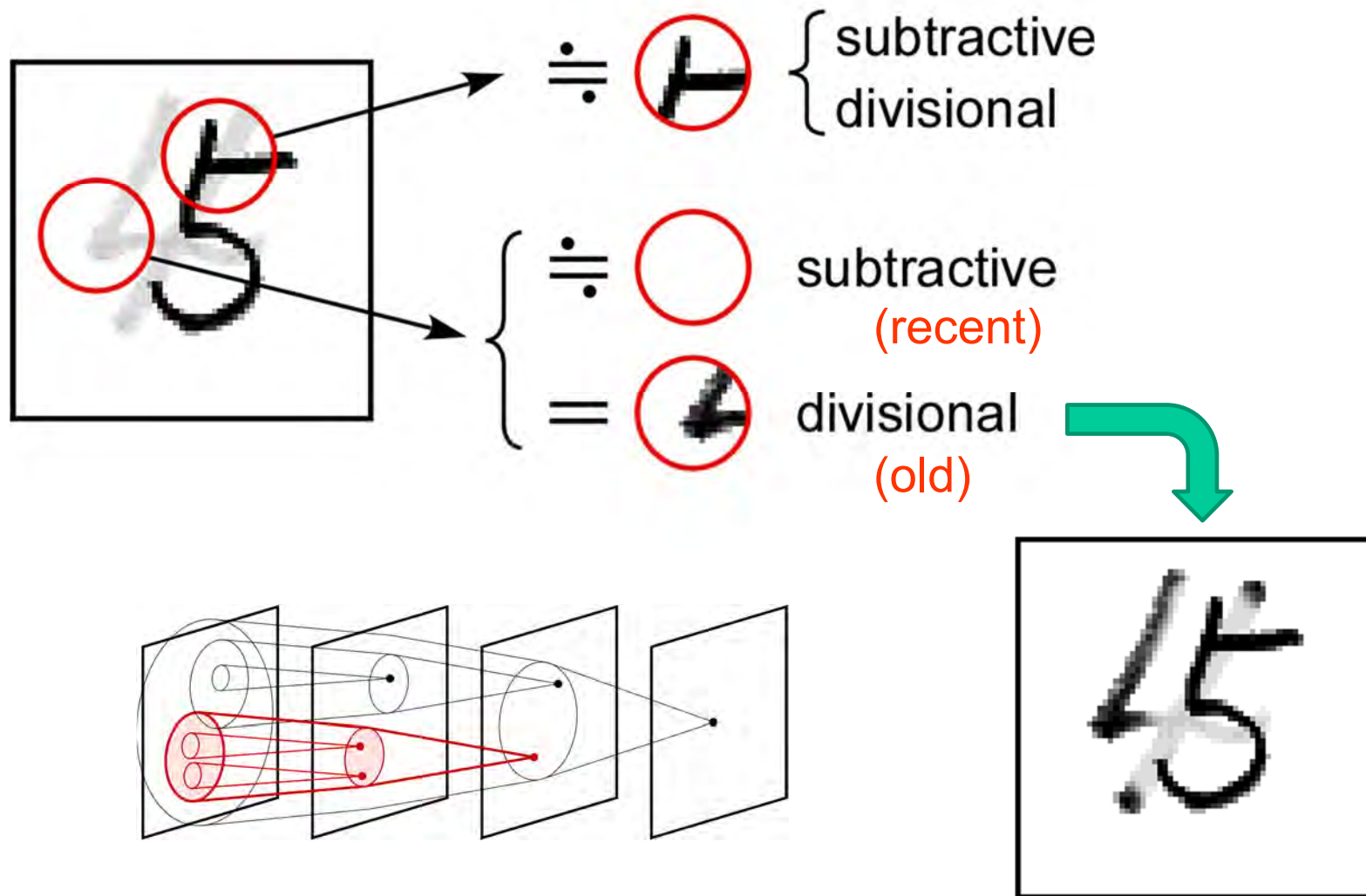


$$s = \frac{(\mathbf{X}, \mathbf{x})}{\|\mathbf{X}\| \cdot \|\mathbf{x}\|} \quad \text{-----} \quad \begin{cases} \text{similarity between} \\ \text{training vector } \mathbf{X} \\ \text{and test vector } \mathbf{x} \end{cases}$$

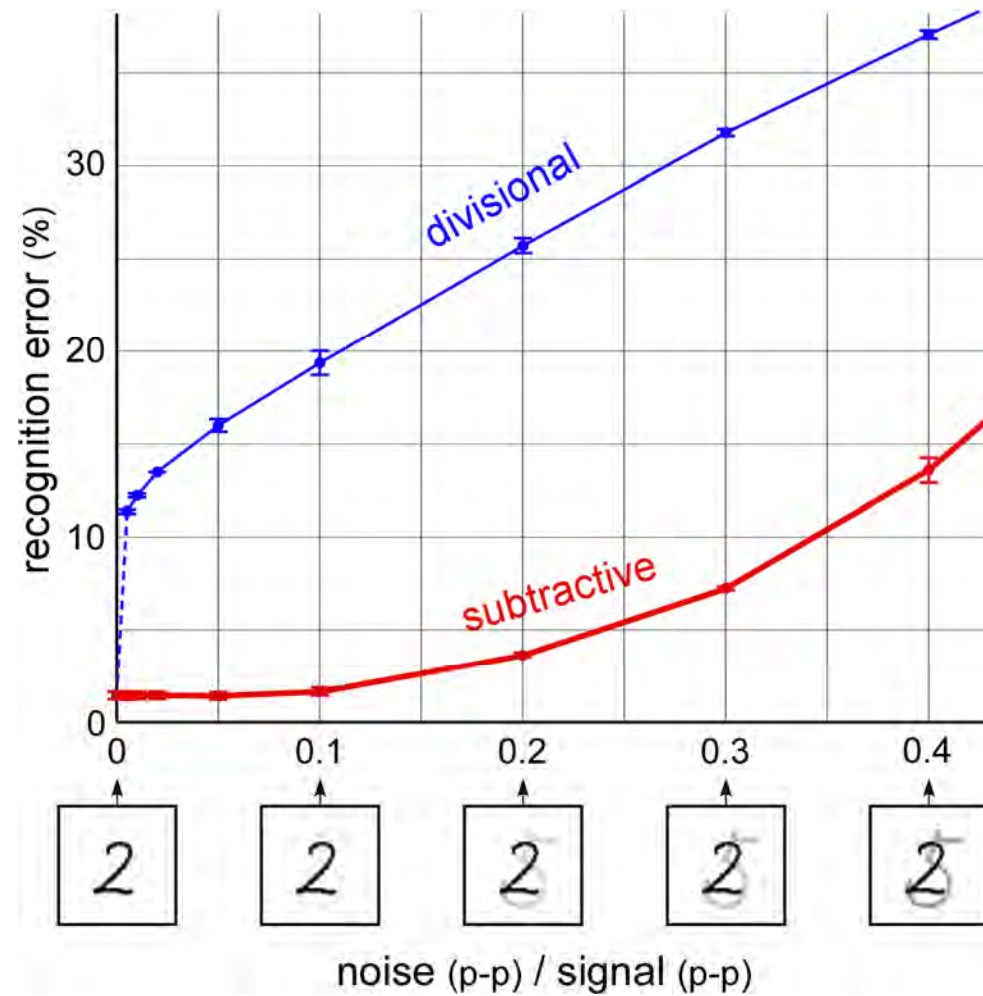
$$\varphi[x] = \max(x, 0): \text{ rectified linear (ReLU)}$$



Examples of patterns with background noise



Difference in the responses of S-cells
between **subtractive** and **divisional** inhibition



Recognition error under background noise

a faint image of a different digit

(rep = 1)

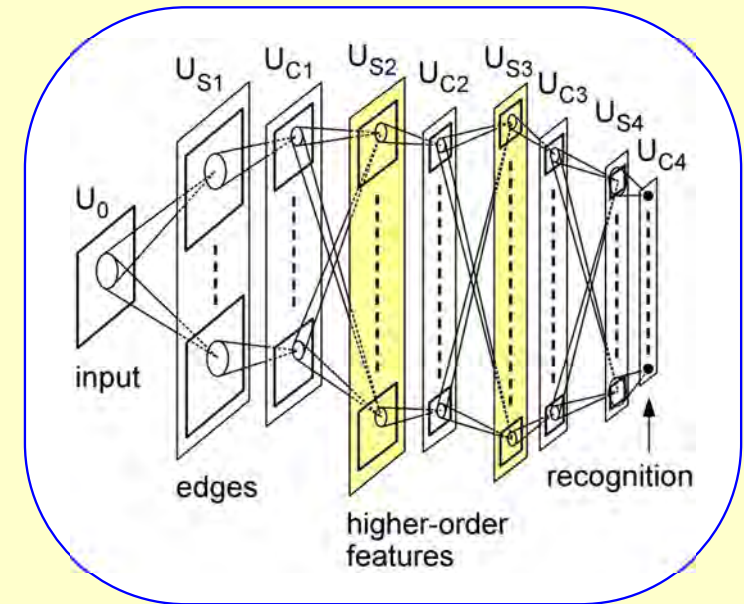
中間層の S細胞の学習

Training S-cells
--- intermediate layers ---

AiS則による 教師なし学習

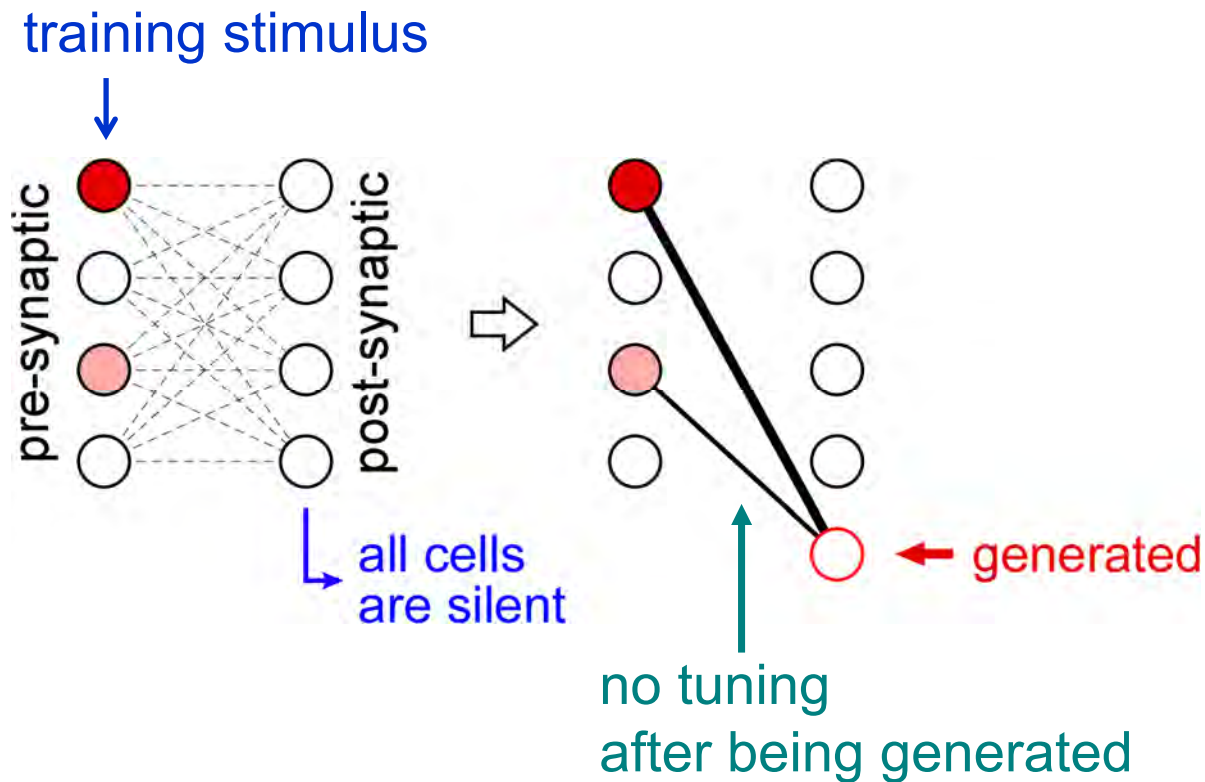
unsupervised learning

AiS (add-if-silent) rule

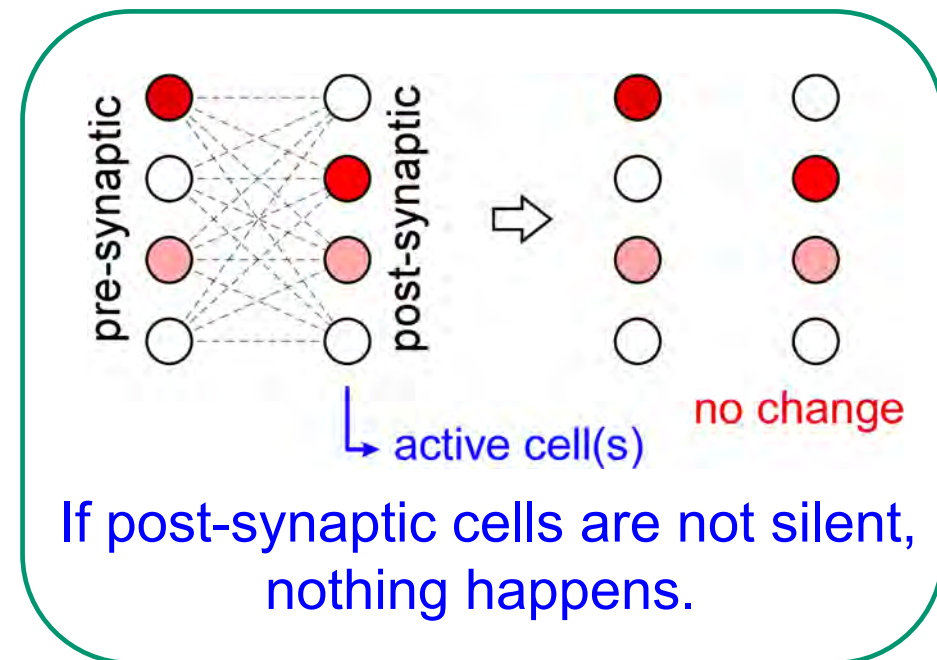


(Fukushima 2013)

Generation of a new S-cell: AiS (add-if-silent rule)

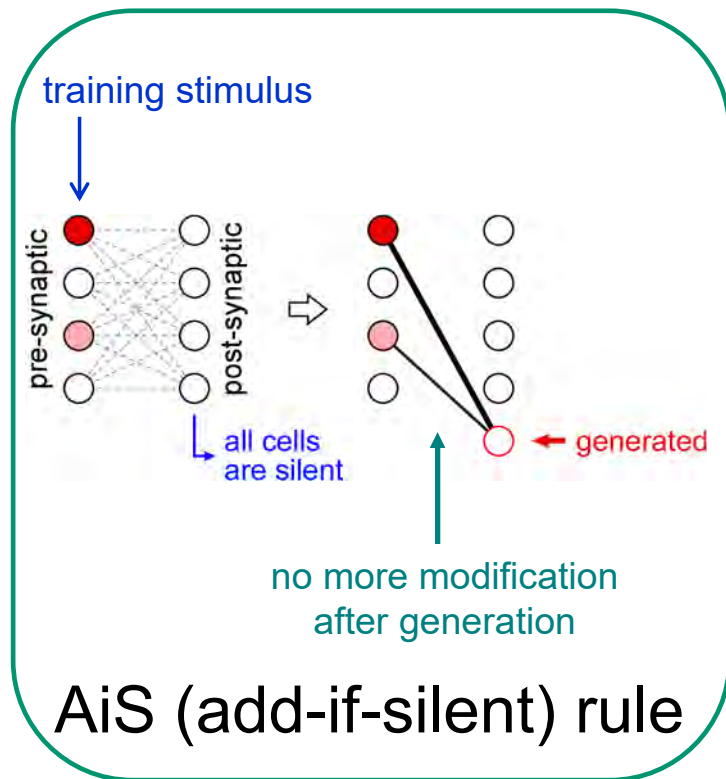


A new cell (cell-plane) is generated if all post-synaptic cells are silent.

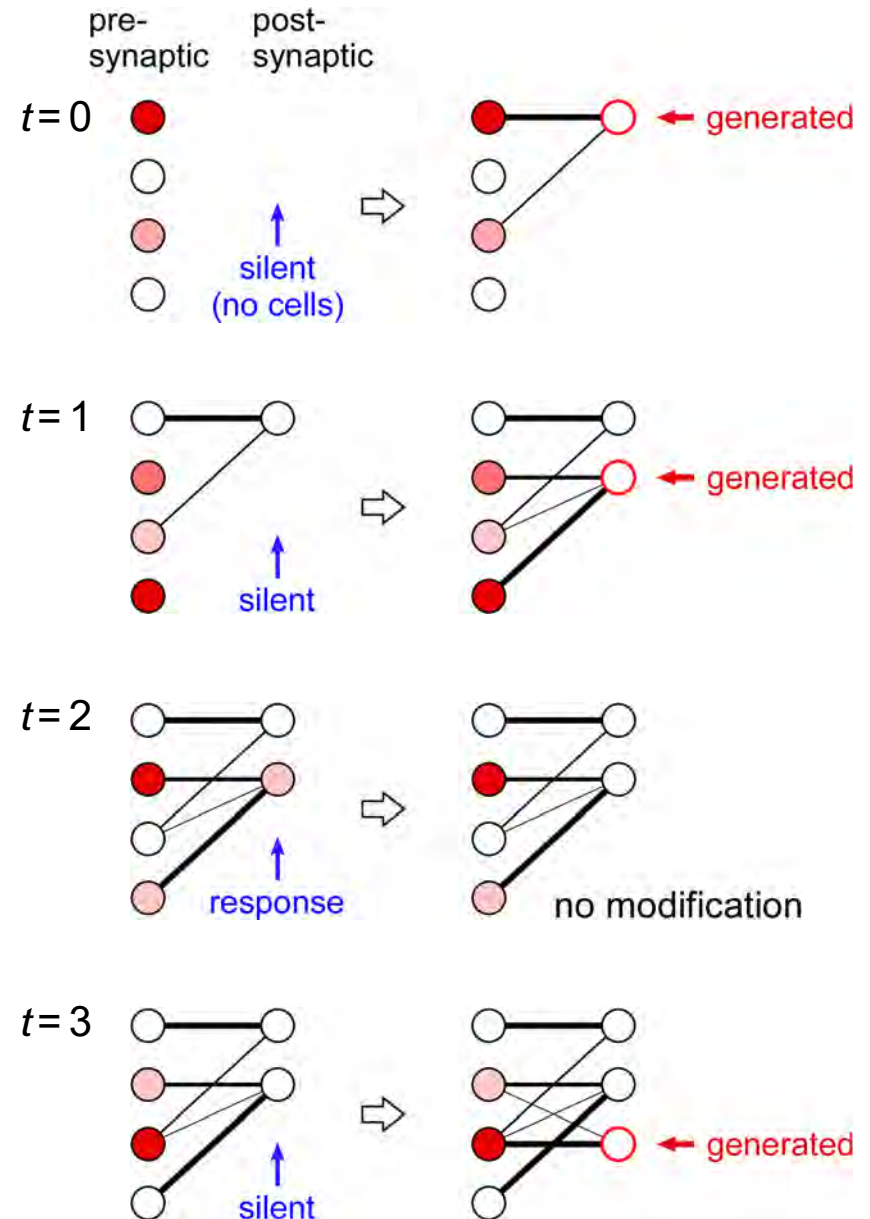


(Fukushima, 2013)

The progress of the learning by the AiS (add-if-silent) rule

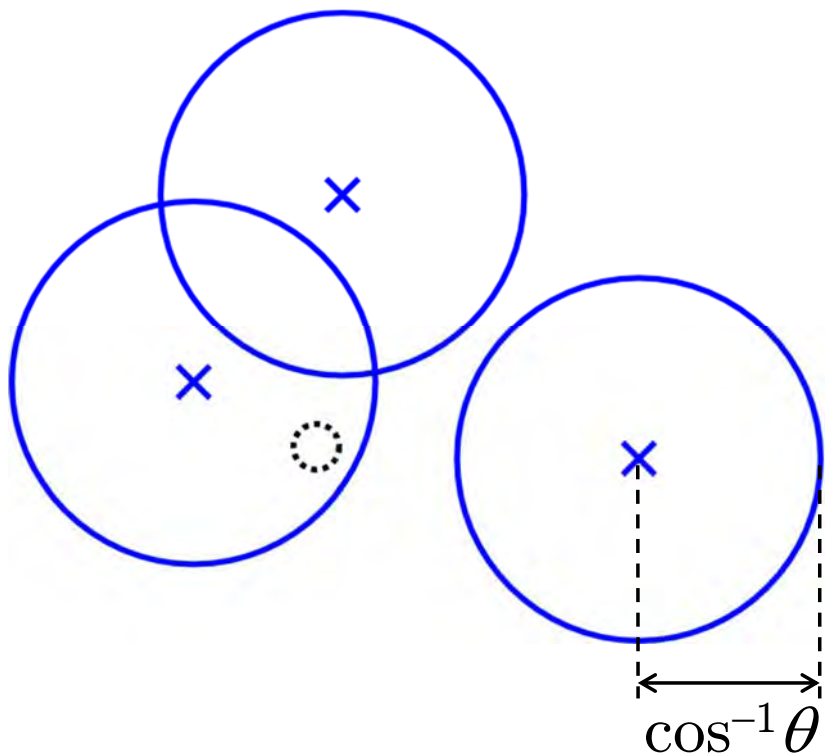


Learning starts from the state where there is no cell.



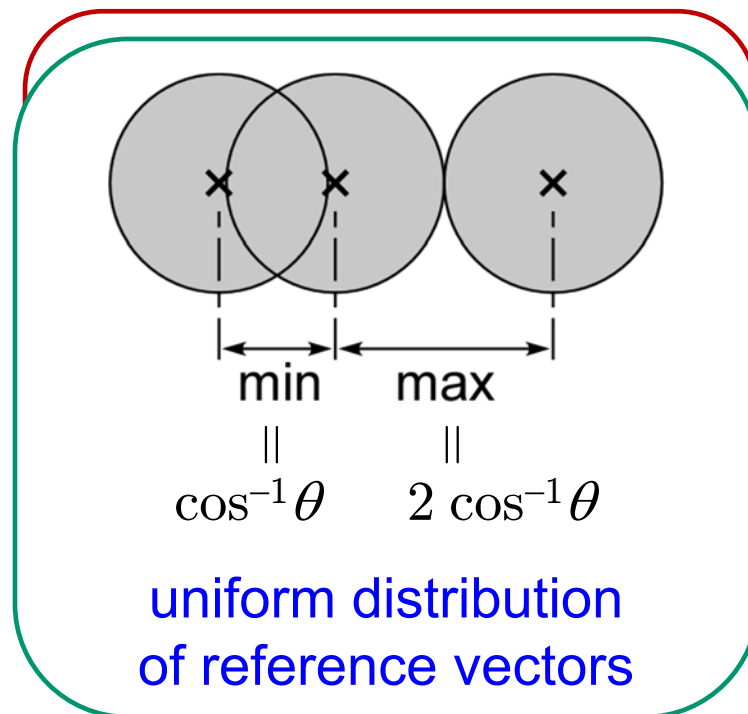
AiS (add-if-silent): a new cell is generated, if all cells are silent.

In the vector space:



- : training vector
- × : silent
- : active response
- ◆ : generated cell

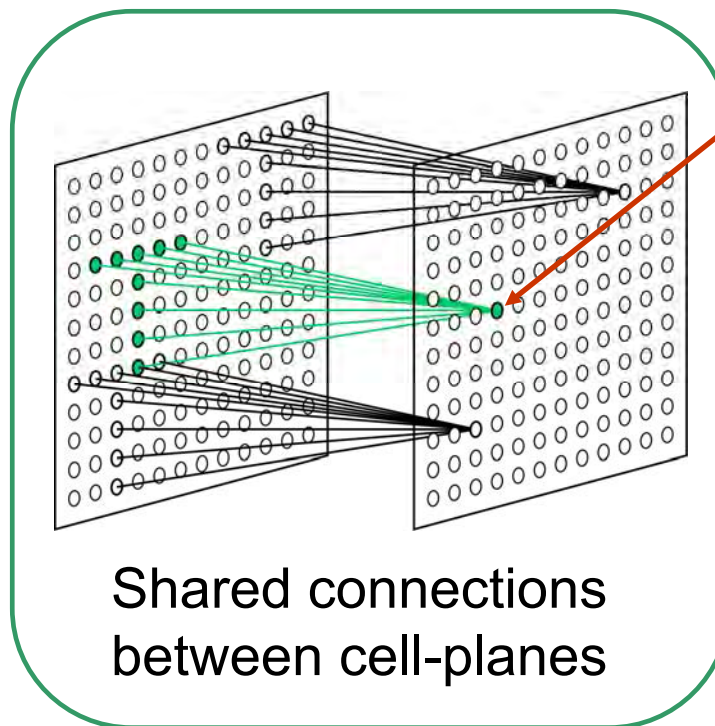
Threshold θ determines
the minimum distance between reference vectors



AiS (add-if-silent) with feedback

Training S-cells
--- intermediate layers ---

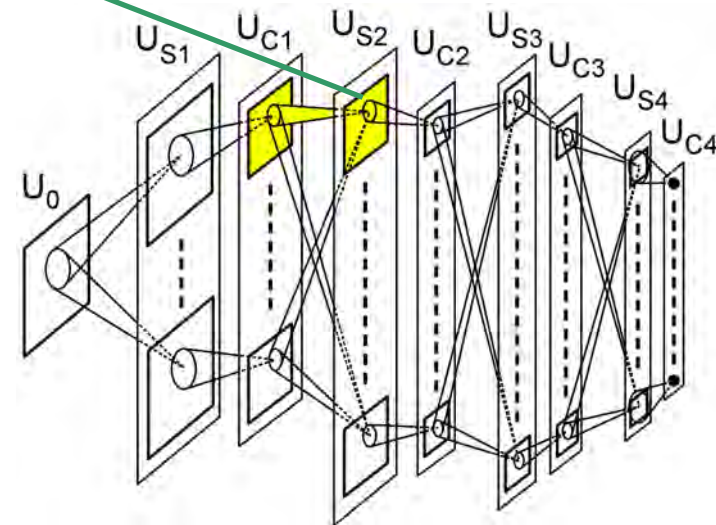
When applying the **add-if-silent** to the neocognitron, condition of **shared connections** should always be kept even during the learning.

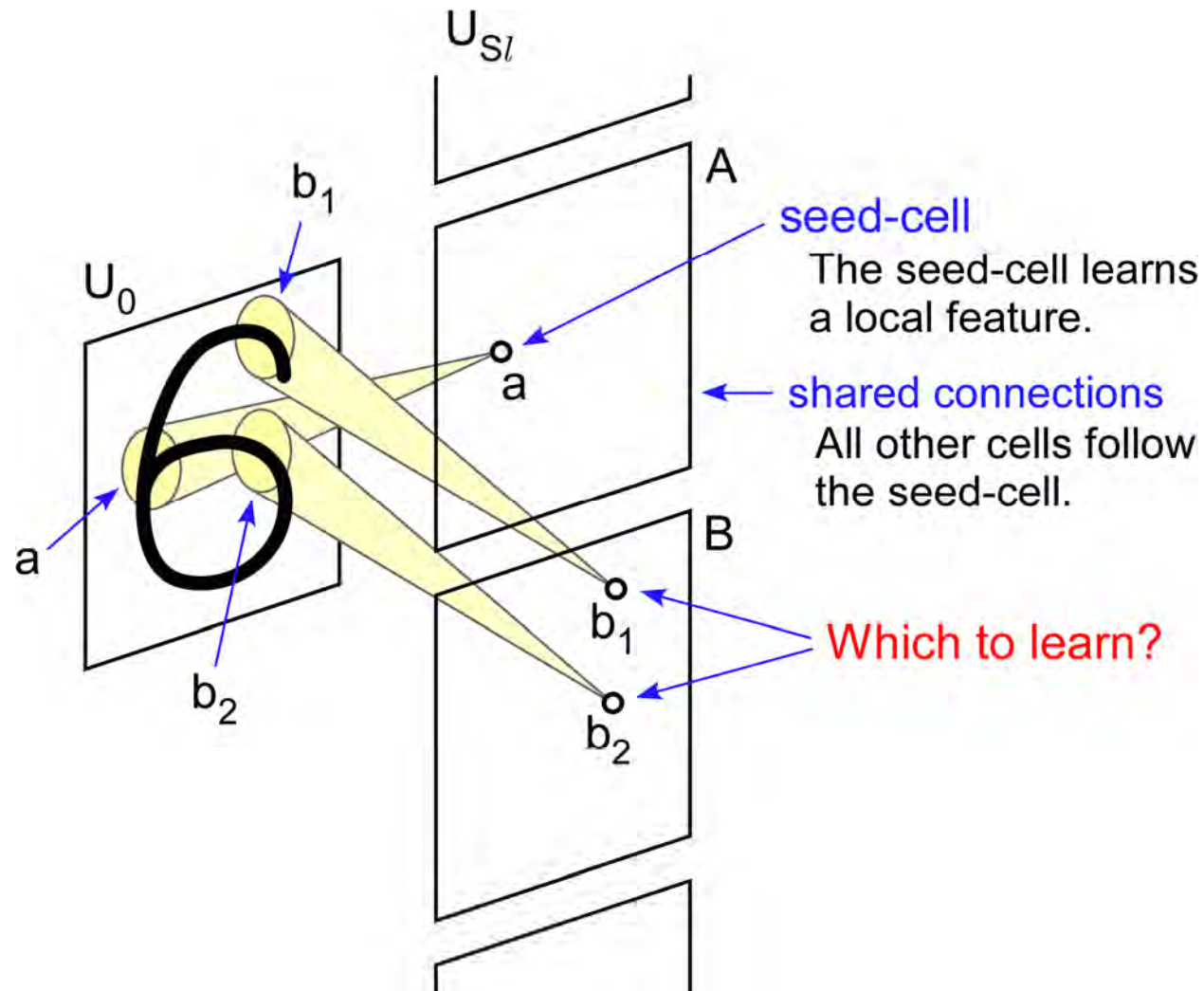


seed cell

(like a seed in crystal growth)

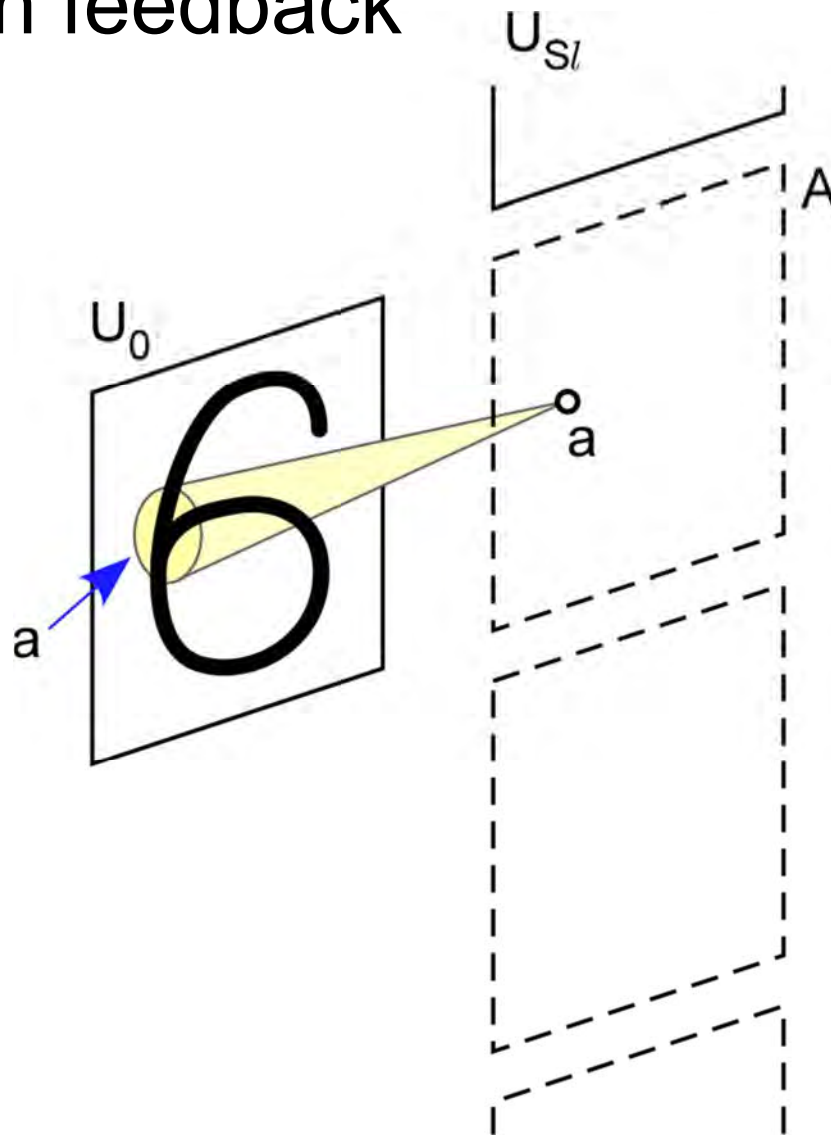
Following the seed cell,
all other cells change their connections.





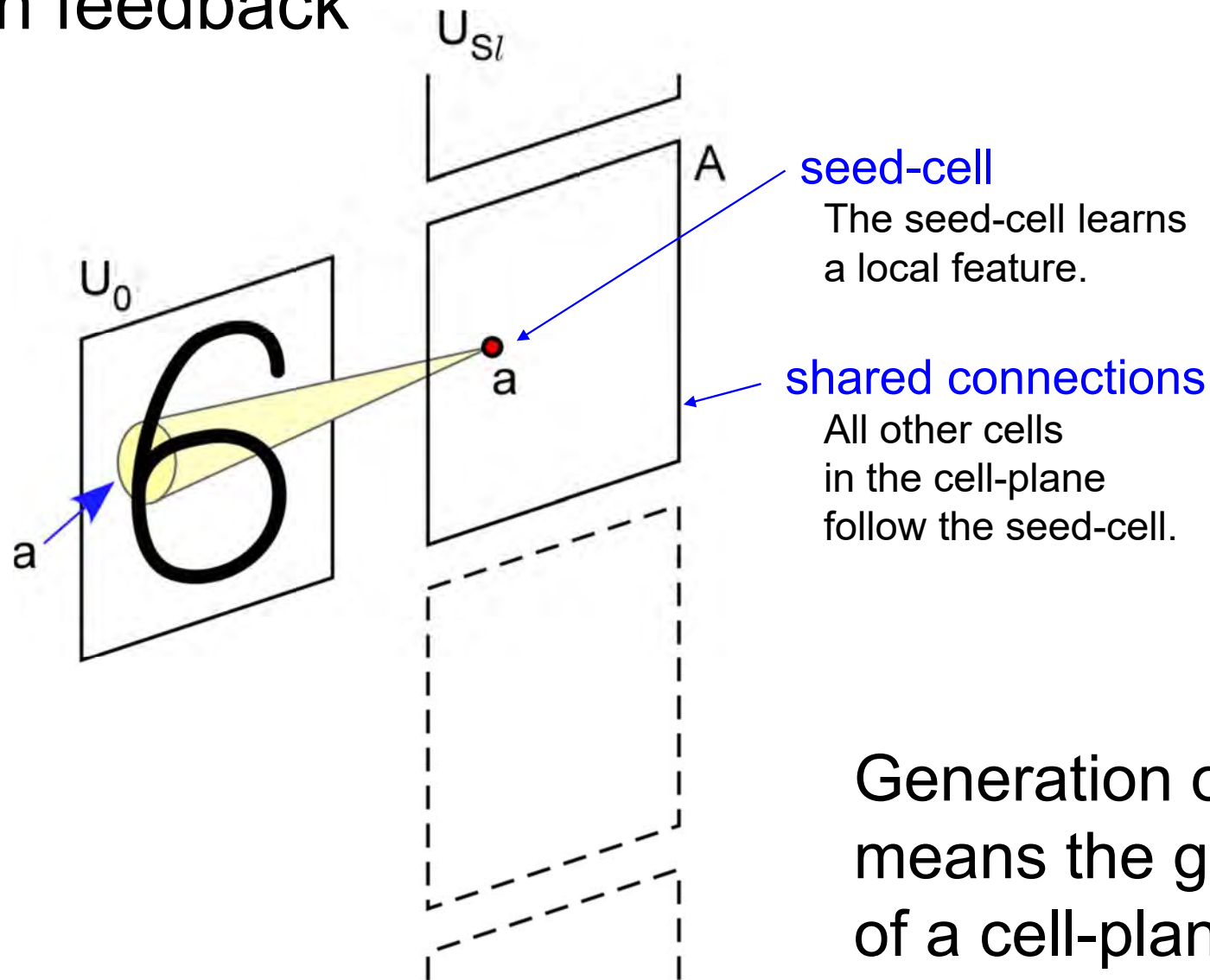
At which retinotopic location should the seed-cell be located?

AiS with feedback

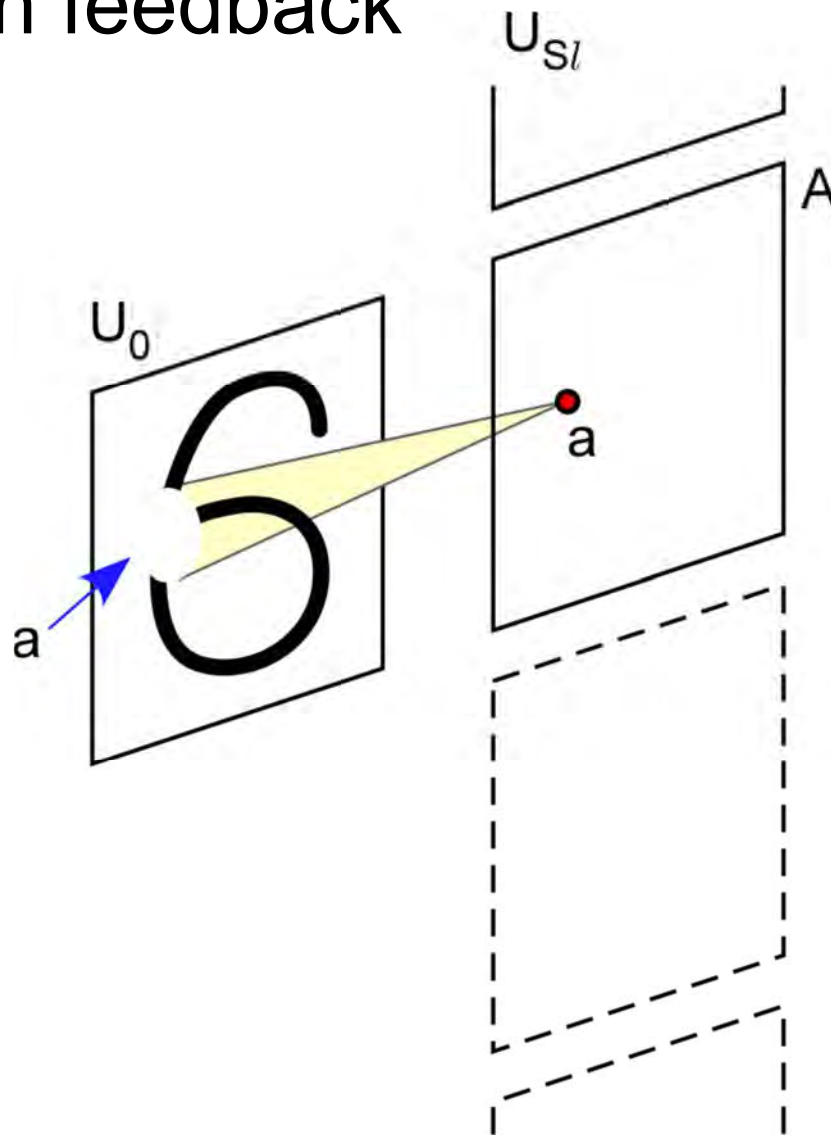


Search the location
of a feature.
Generate a cell
that extract the feature.

AiS with feedback

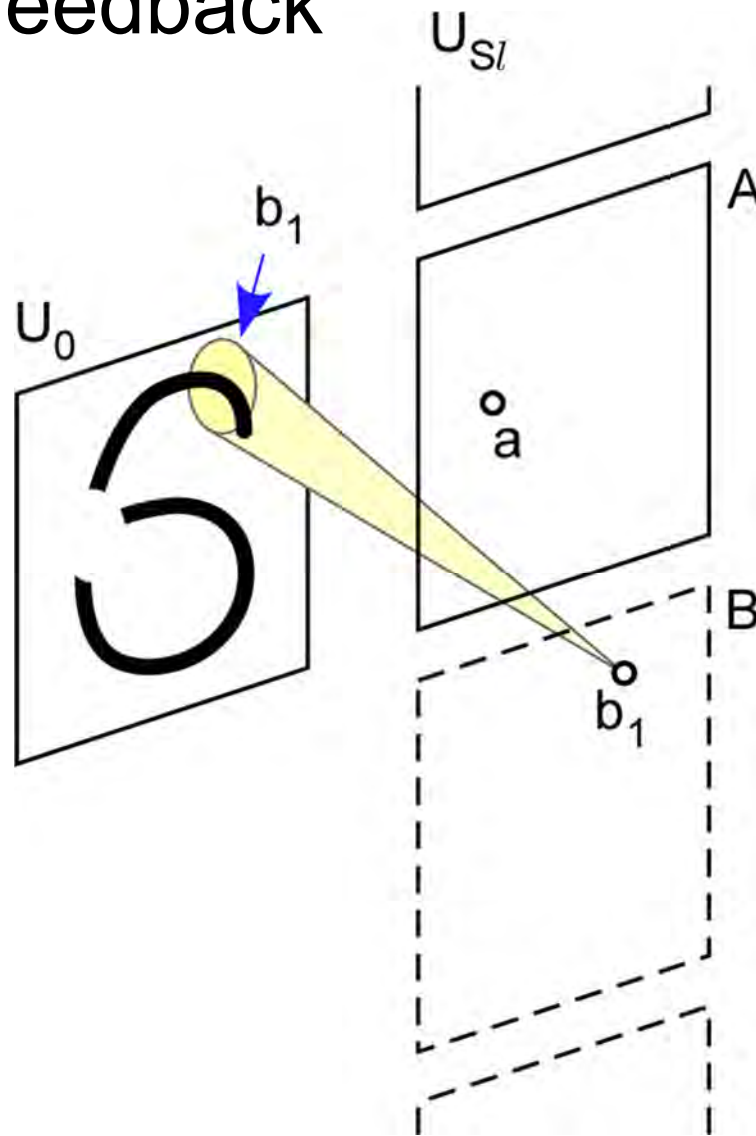


AiS with feedback



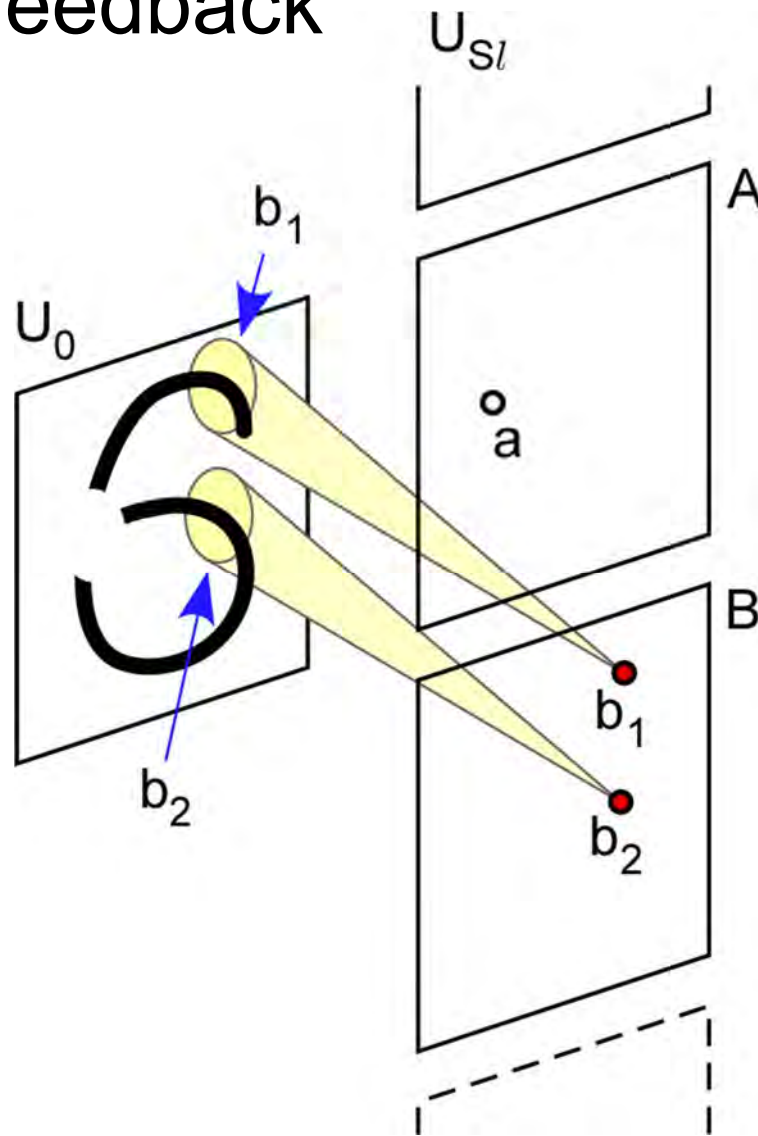
Negative feedback
from active cells inhibits
the presynaptic activities.

AiS with feedback



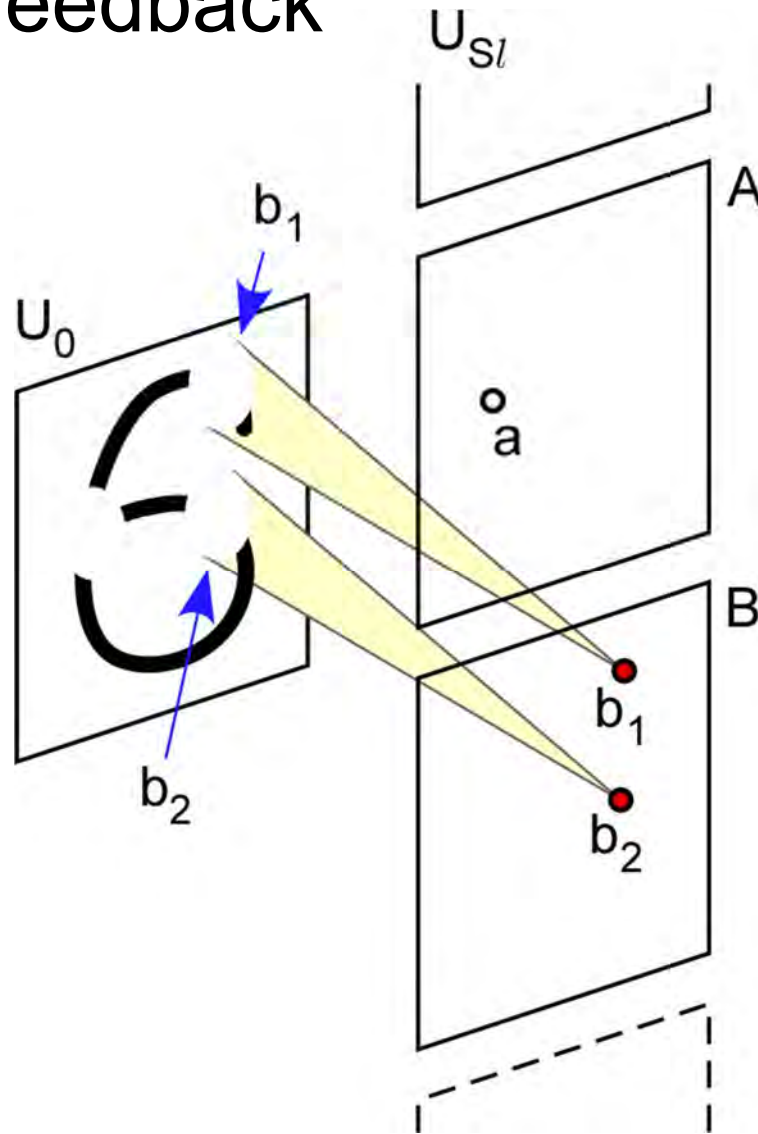
Search the location
of another feature,
which has not been
inhibited yet.
Generate a cell
that extract the feature.

AiS with feedback



Cells in the cell-plane that learned feature “b1” respond also to a similar feature “b2”.

AiS with feedback



Negative feedback from active cells inhibits the presynaptic activities at both “b1” and “b2”.

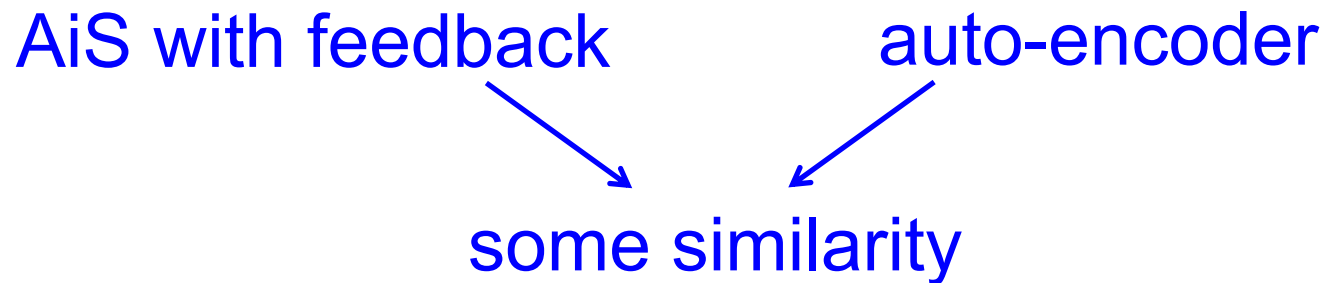
Generation of redundant cell-planes can be avoided.

AiS with feedback:

negative feedback は、
どこに seed-cell を置くべきかを決めるためだけに用いる。

seed-cell は、negative feedback によって抑制される以前の
前シナプス細胞の反応強度を学習する。

フィードバック誤差が無くなるまで 細胞発生を続ける。



各層の細胞数

最近のネオコグニトロン:

AiS = Add-if-Silent 則

細胞が無い状態から学習を開始.

学習時に, 必要に応じて細胞を発生.

細胞間の結合は固定 (細胞発生時に決定し, tuning なし).

通常の deep CNN:

細胞数は固定: 回路の設計時に定める.

学習によって変化せず.

学習によって変わるのは, 細胞間の結合強度.

中間層の学習

	通常の deep CNN	ネオコグニトロン AiS = Add-if-Silent 則
細胞数	固定 回路の設計時に定める 学習によって変化せず	可変 必要に応じて細胞発生 学習開始時は 細胞なし
結合強度	可変 学習時に変更 (tuning)	固定 細胞発生時に決定 (no tuning)

単純なAIS則でなぜ高い認識率？

入力パターンの最終的な認識(分類)は
中間層ではなく 最深層で行なわれる

中間層の役割は、入力パターンを
1個の細胞の反応ではなく
細胞集団の反応(population coding)
によって、正確に表現することである

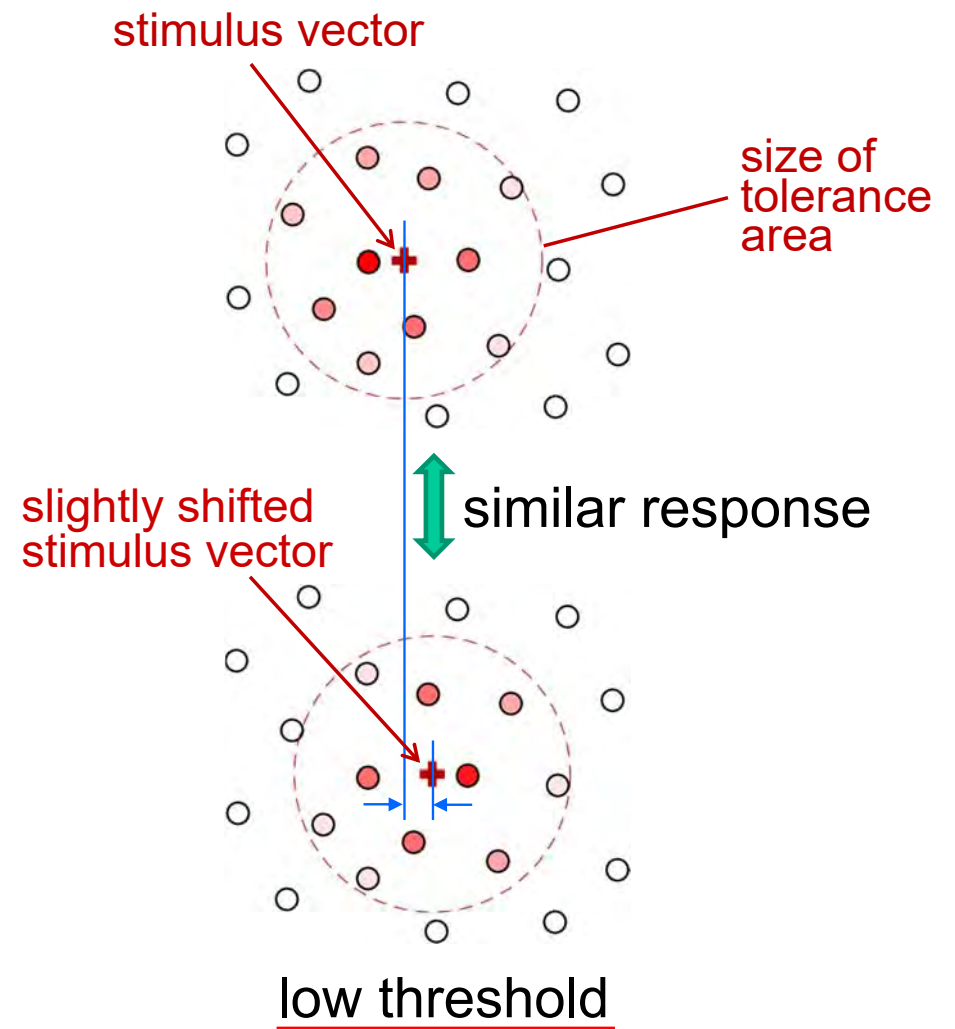
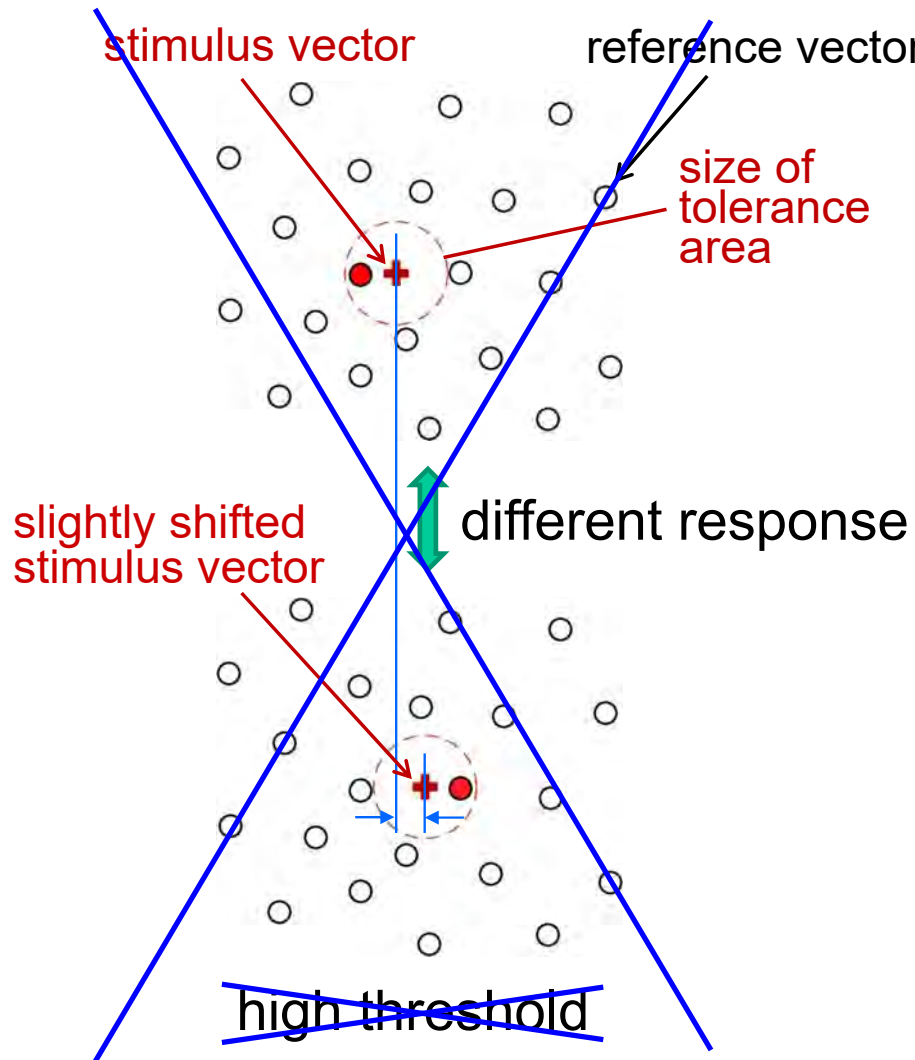
→ 個々の細胞の反応を
特定の入力刺激に対応させることは
必ずしも重要ではない

S細胞に 二重しきい値

(学習時と認識時とで異なるしきい値)

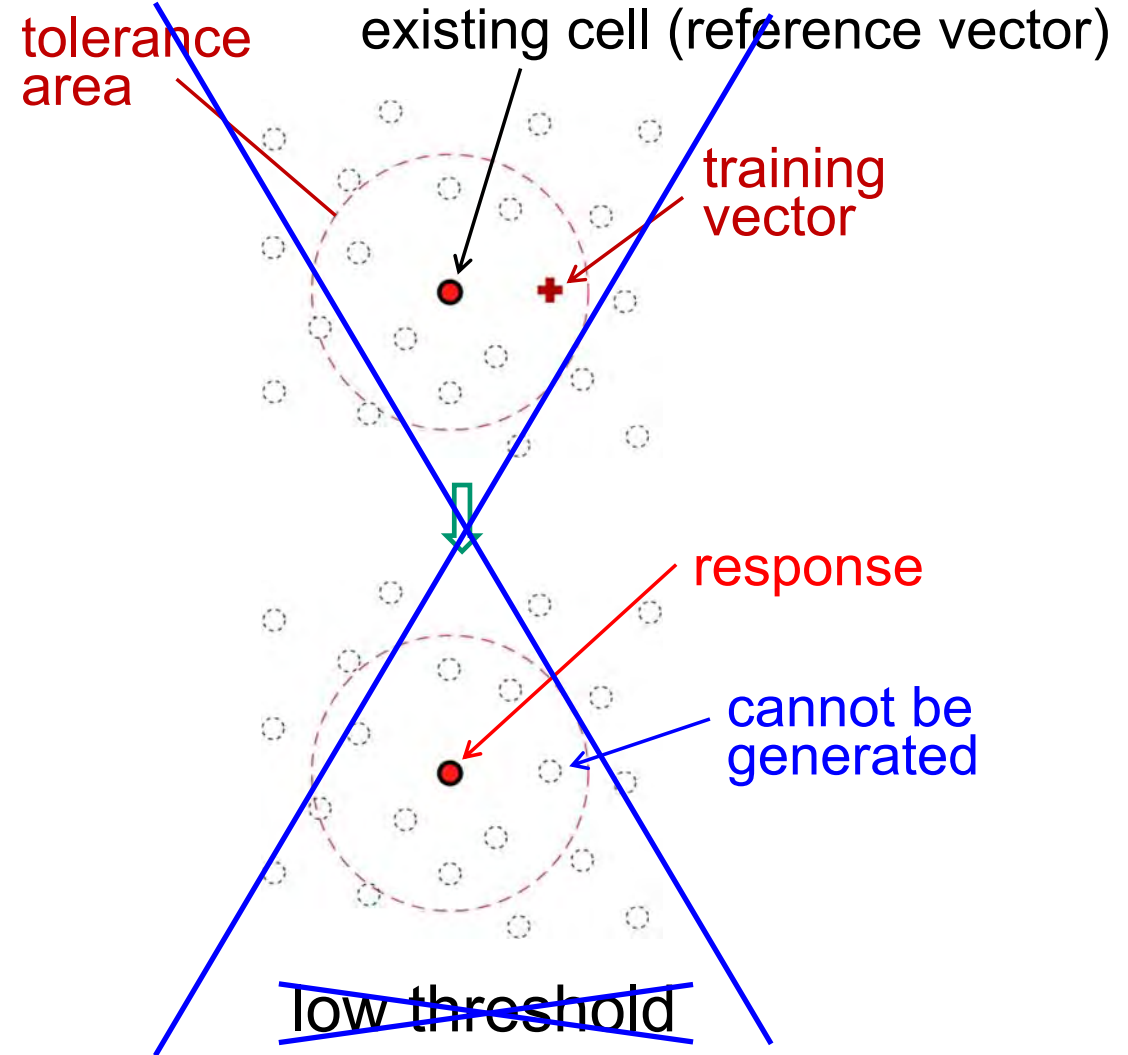
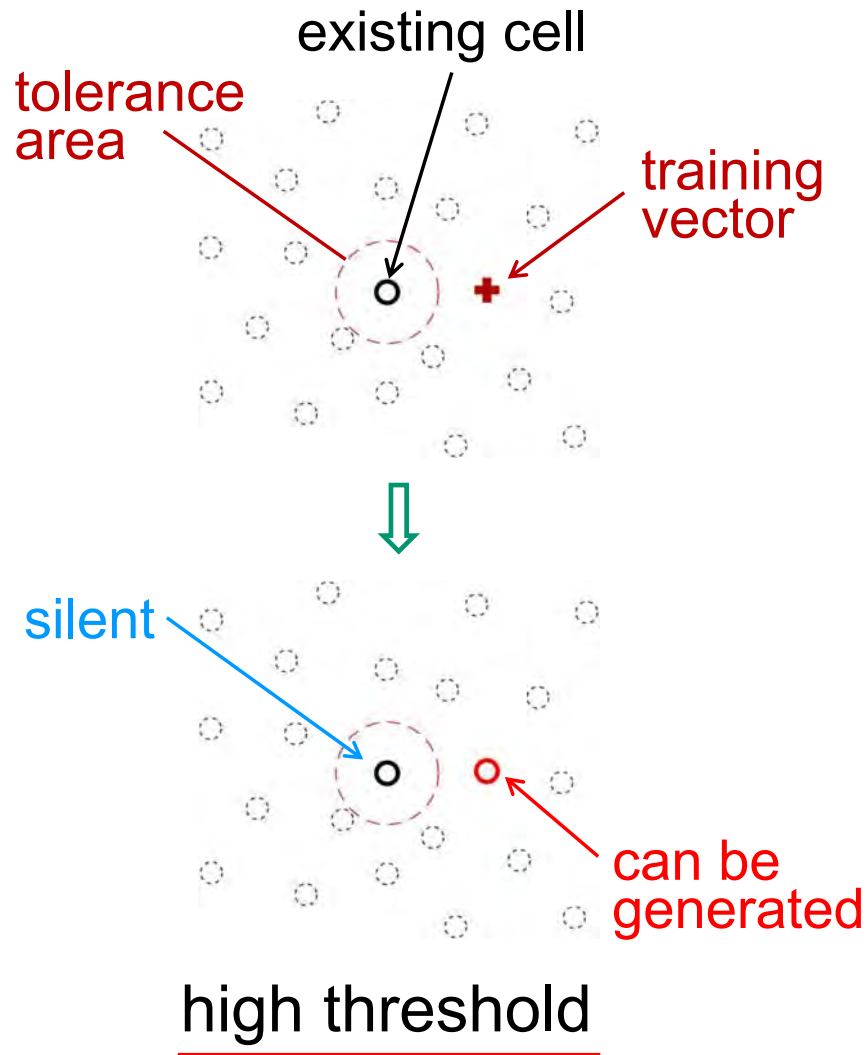
Dual threshold for S-cells
of intermediate layers

(Fukushima & Tanigawa, 1996)



Recognition phase

(Response of S-cells in the multi-dimensional feature space)



Learning phase

(Response of S-cells in the multi-dimensional feature space)

S細胞に 二重しきい値

- 学習時 → 高い しきい値 (θ^L)

おばあさん細胞的反応 (各特徴に1個の細胞)

⇒ 特徴空間内に 細胞が一様に分布

θ^L は細胞面 (= 抽出する特徴) の数を決定する

- 認識時 → 低い しきい値 (θ^R)

特徴空間での ぼかし

sparse population coding (各特徴に多数の細胞が反応)

⇒ 変形したパターンも頑強に認識

θ^R は反応の sparseness を決定する

(Fukushima & Tanigawa, 1996)

C細胞

----- 網膜座標での ぼかし (= pooling)
(spatial pooling)

S細胞（認識時：低い しきい値）

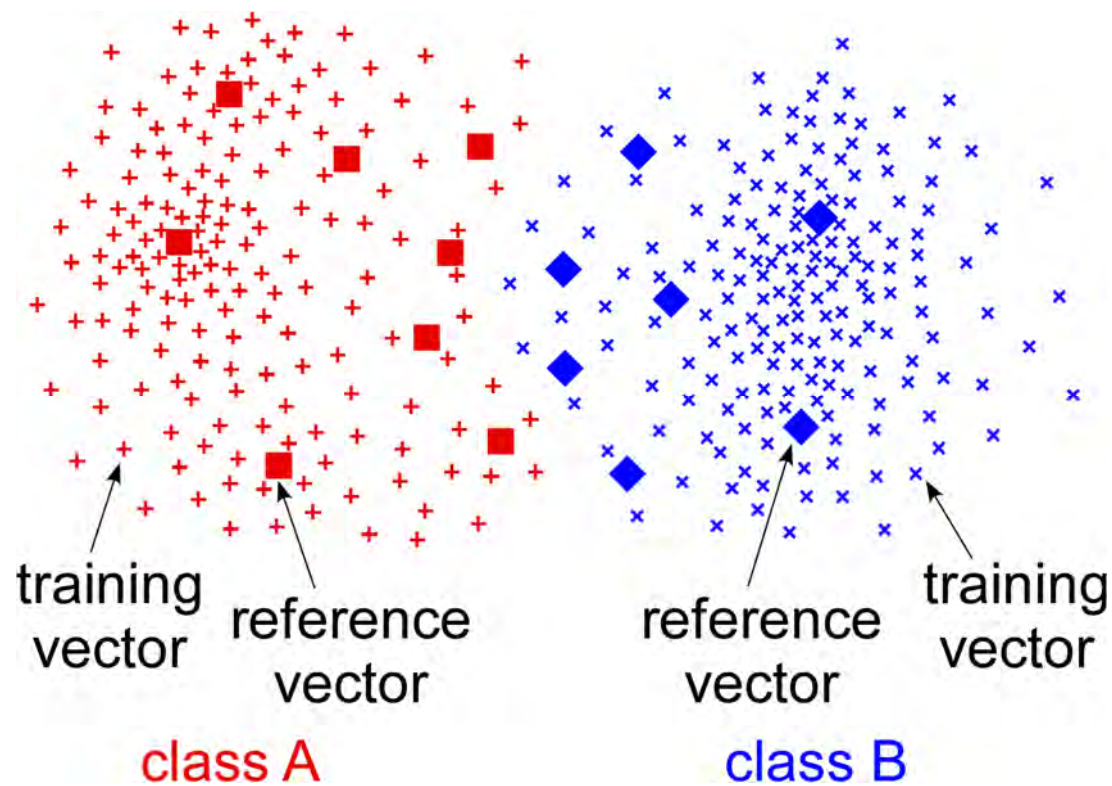
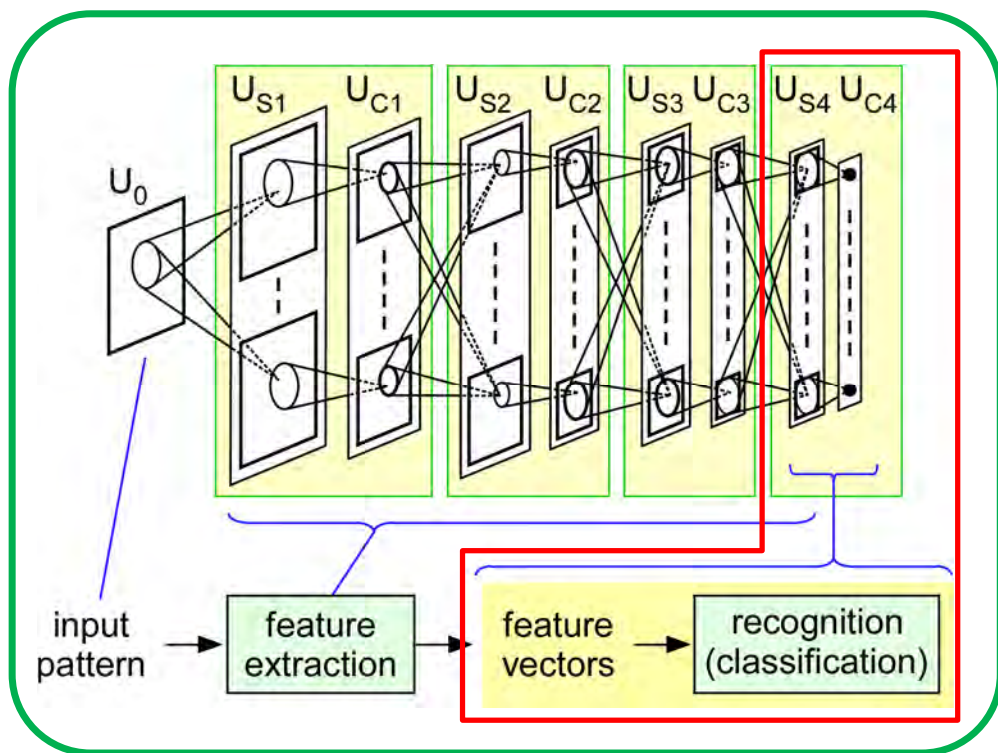
----- 特徴空間での ぼかし (= pooling)

最深層の細胞発生に

mWTA (margined WTA)

(Add-if-Error, not by Add-if-Silent)

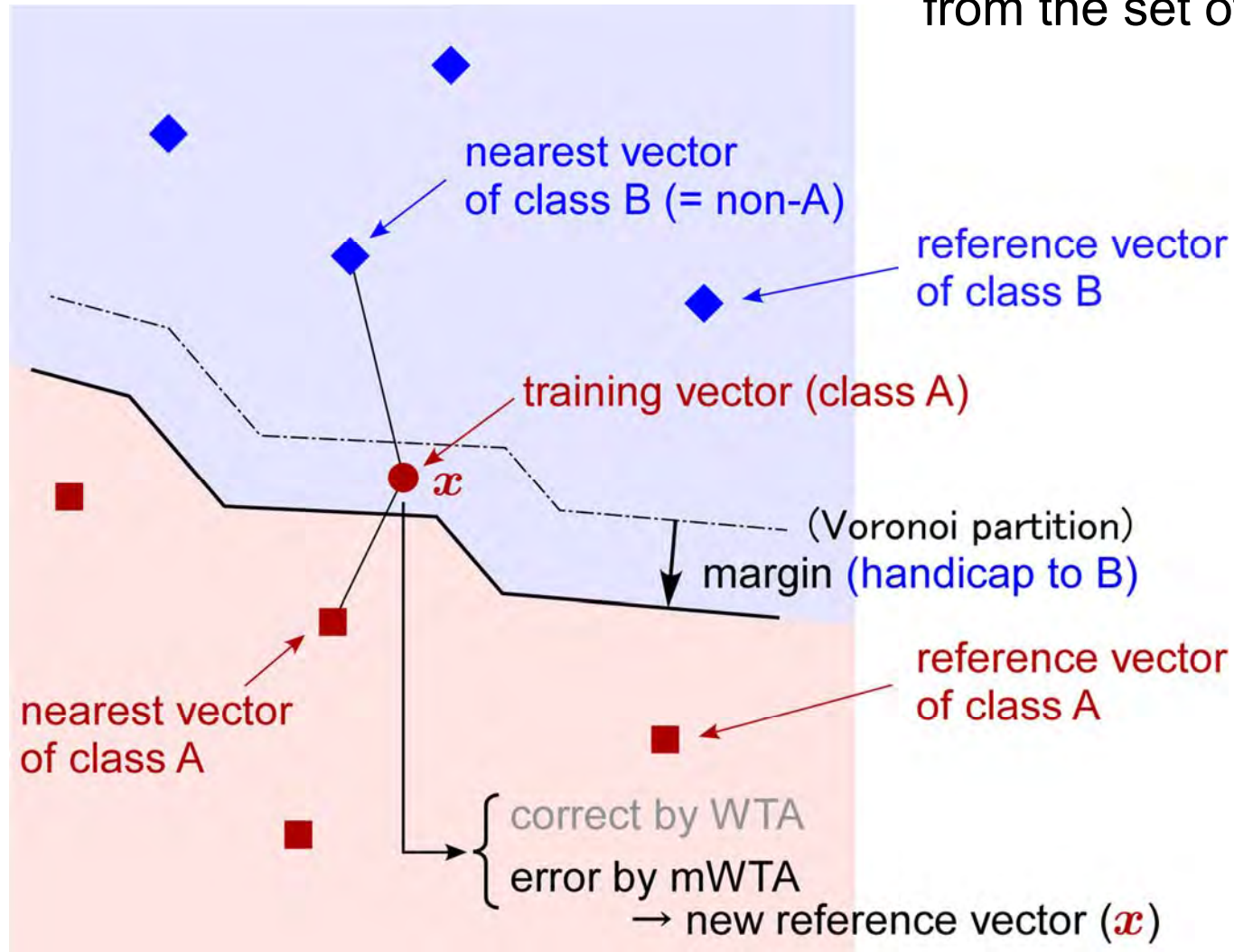
(Fukushima 2018)



最深層では、

学習ベクトルを、少数の**参照ベクトル**で表わす。
(labeled **reference vectors**)

mWTA (margin by WTA) for choosing reference vectors
from the set of training vectors

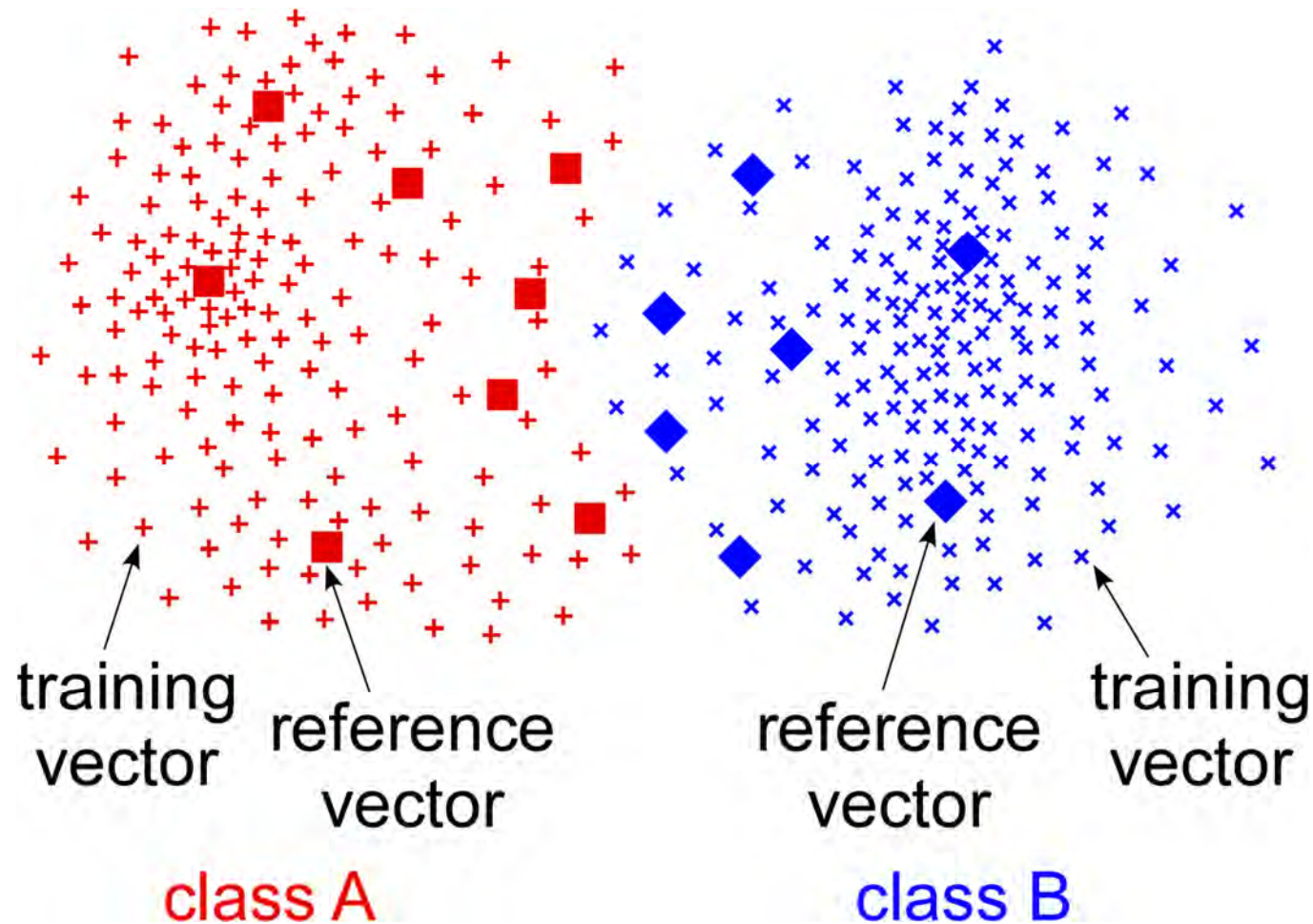


IntVec (Interpolating-Vector) 内挿ベクトル法

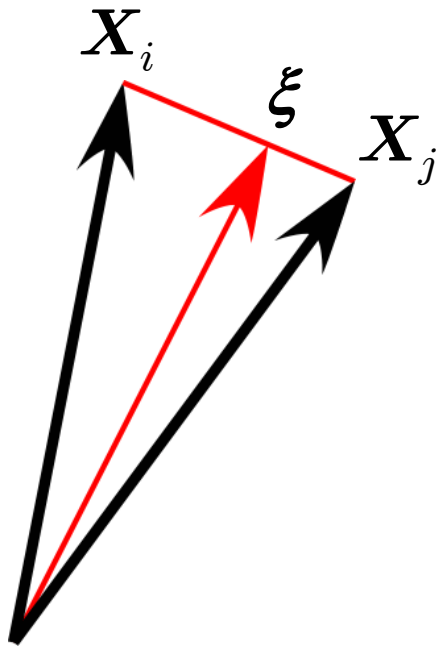
認識時 最深層に適用

(Fukushima 2007)

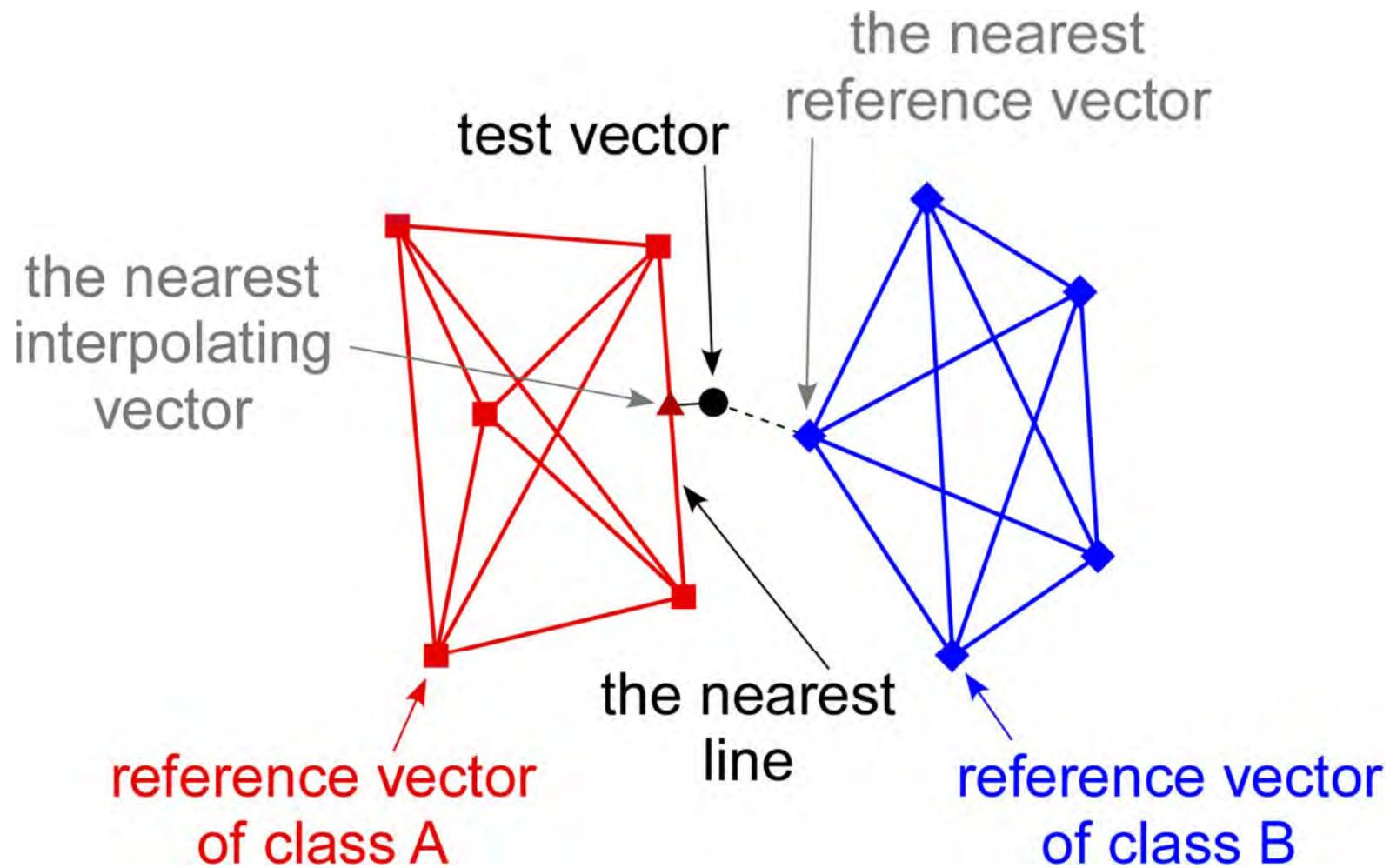
Training vectors are represented by
a smaller number of labeled **reference vectors**.



Interpolating vectors (内挿ベクトル) between two reference vectors of the same label

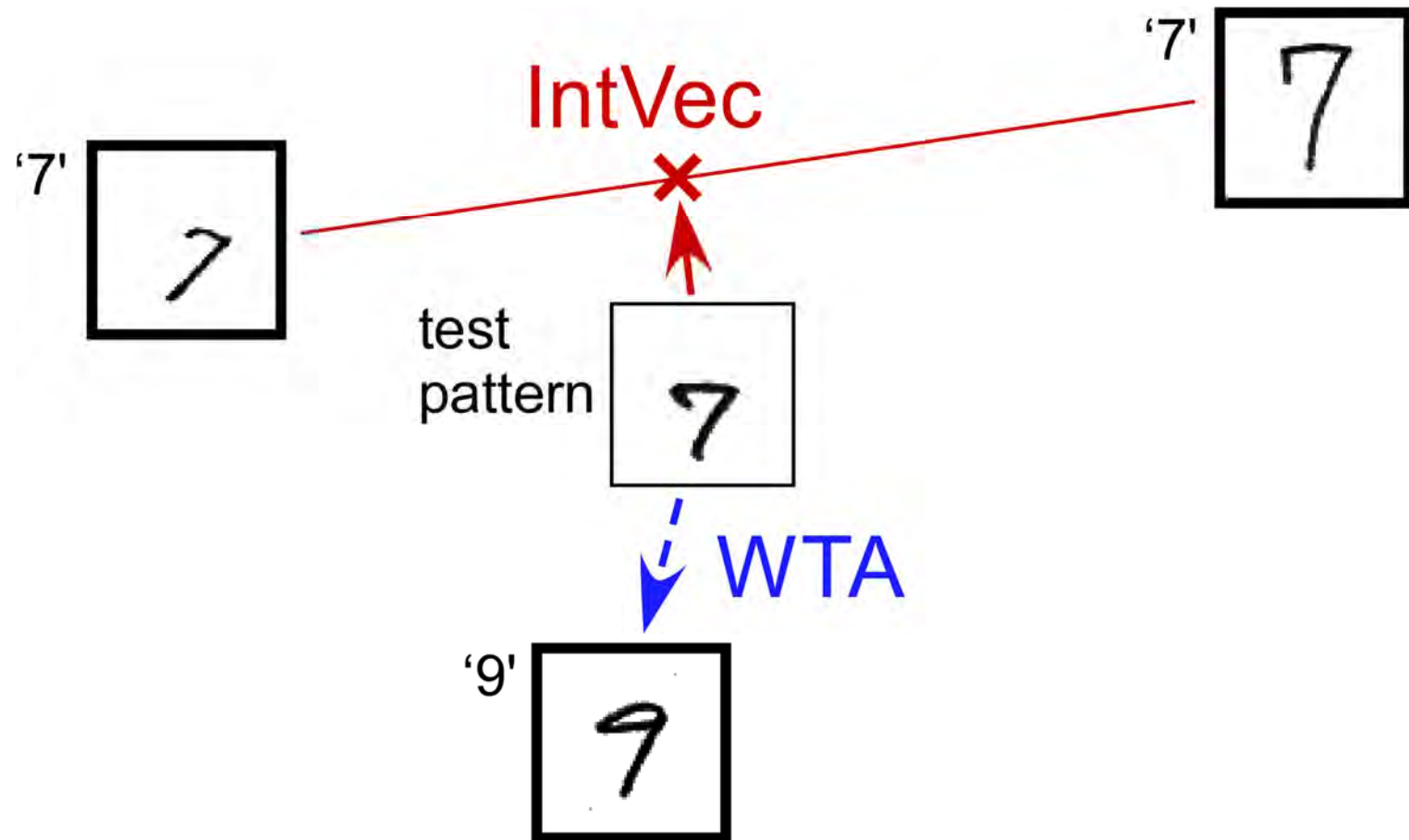


$$\xi = p_i \frac{\mathbf{X}_i}{\|\mathbf{X}_i\|} + p_j \frac{\mathbf{X}_j}{\|\mathbf{X}_j\|}$$
$$(p_i + p_j = 1)$$

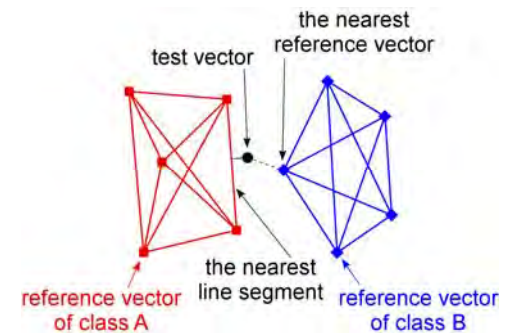
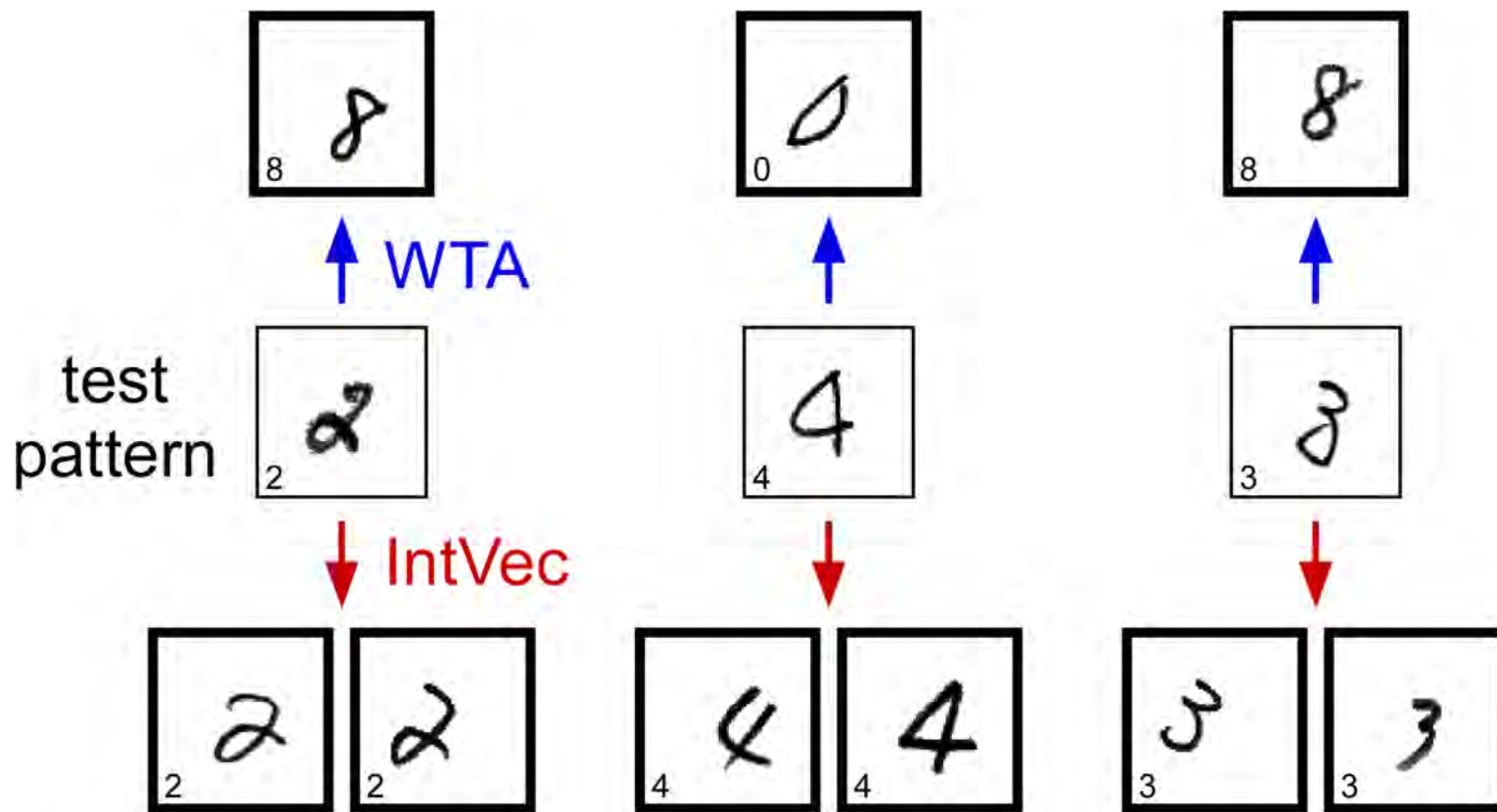


Find the nearest **line**

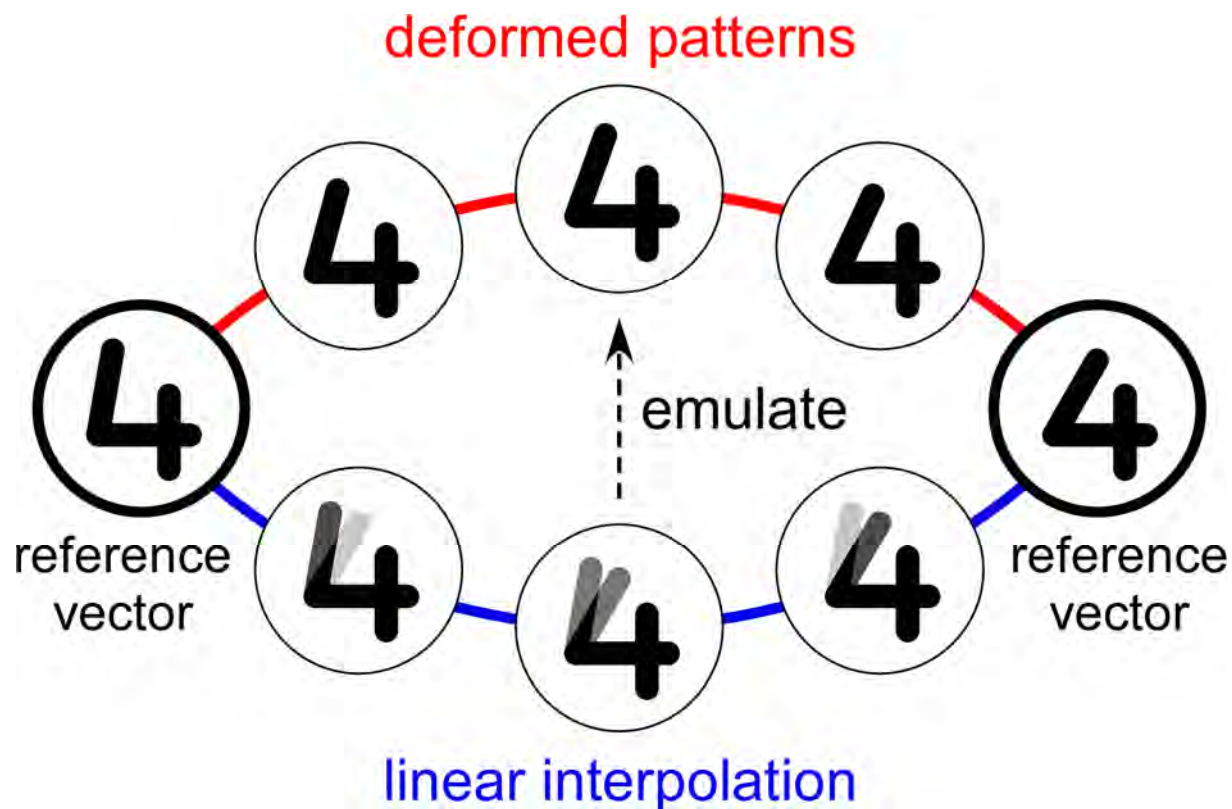
Recognition by IntVec (interpolating-vector)



Examples of patterns that were recognized,
erroneously by the WTA,
but **correctly by the IntVec** (interpolating-vector, Int-2).

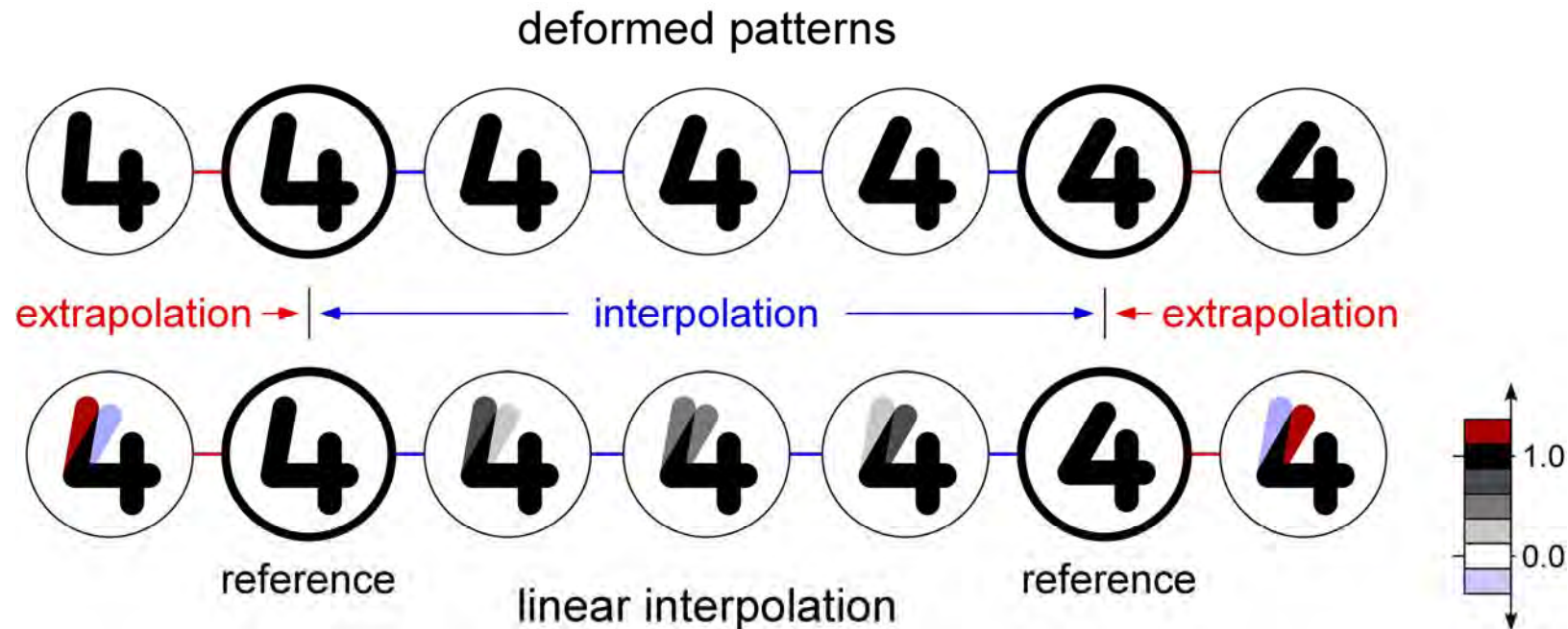


Recognition by Int-Vec (interpolating-vector)



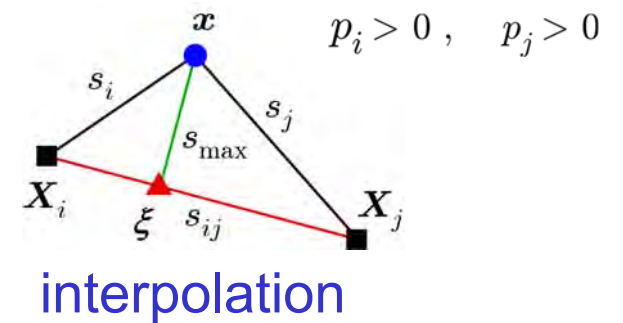
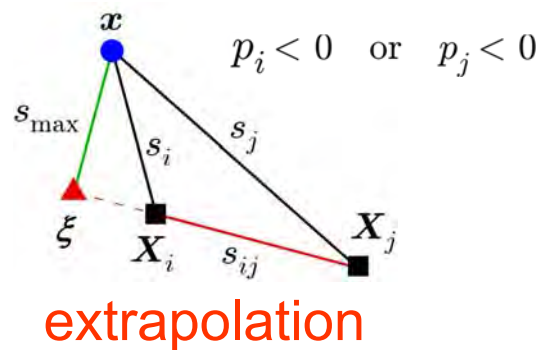
This operation is applied,
~~not directly to input images, but~~ to extracted features.

Extrapolation, as well as interpolation



$$\xi = p_i \frac{X_i}{\|X_i\|} + p_j \frac{X_j}{\|X_j\|}$$

$$(p_i + p_j = 1)$$



内挿ベクトル法(IntVec)はなぜ強力か？

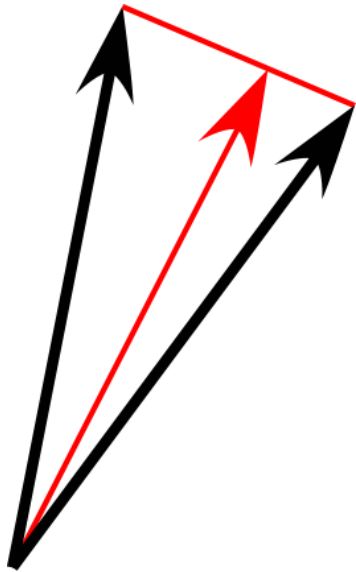
一般に CNN（畳込み神経回路）の認識率は、
学習パターン数を増やすことによって向上する。

十分多数の学習パターンが得られないときには、
人工的に作った学習パターンで不足を補っても、
認識率を向上させることができる。

----- data augmentation（データ拡張） -----

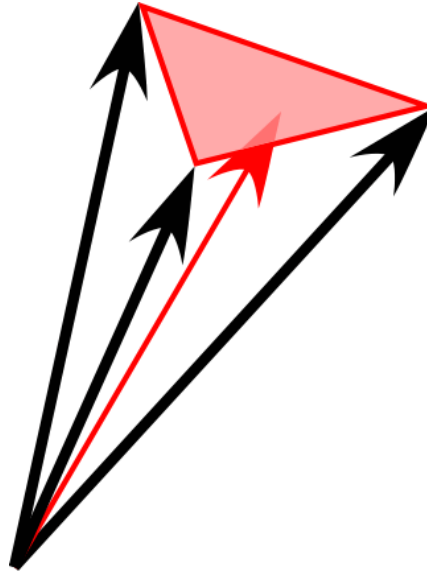
内挿ベクトル法(IntVec)では data augmentation を、
学習時ではなく、認識時に行なっている。
比較すべき特徴の数を、学習パターン数を増やさずに、
実質的に増やしている。

Int-Vec (interpolating-vector) from **three** or **four** reference vectors



Search the nearest
line.

Int-2 (interpolating-vector
from 2 reference vectors)



Search the nearest
plane.

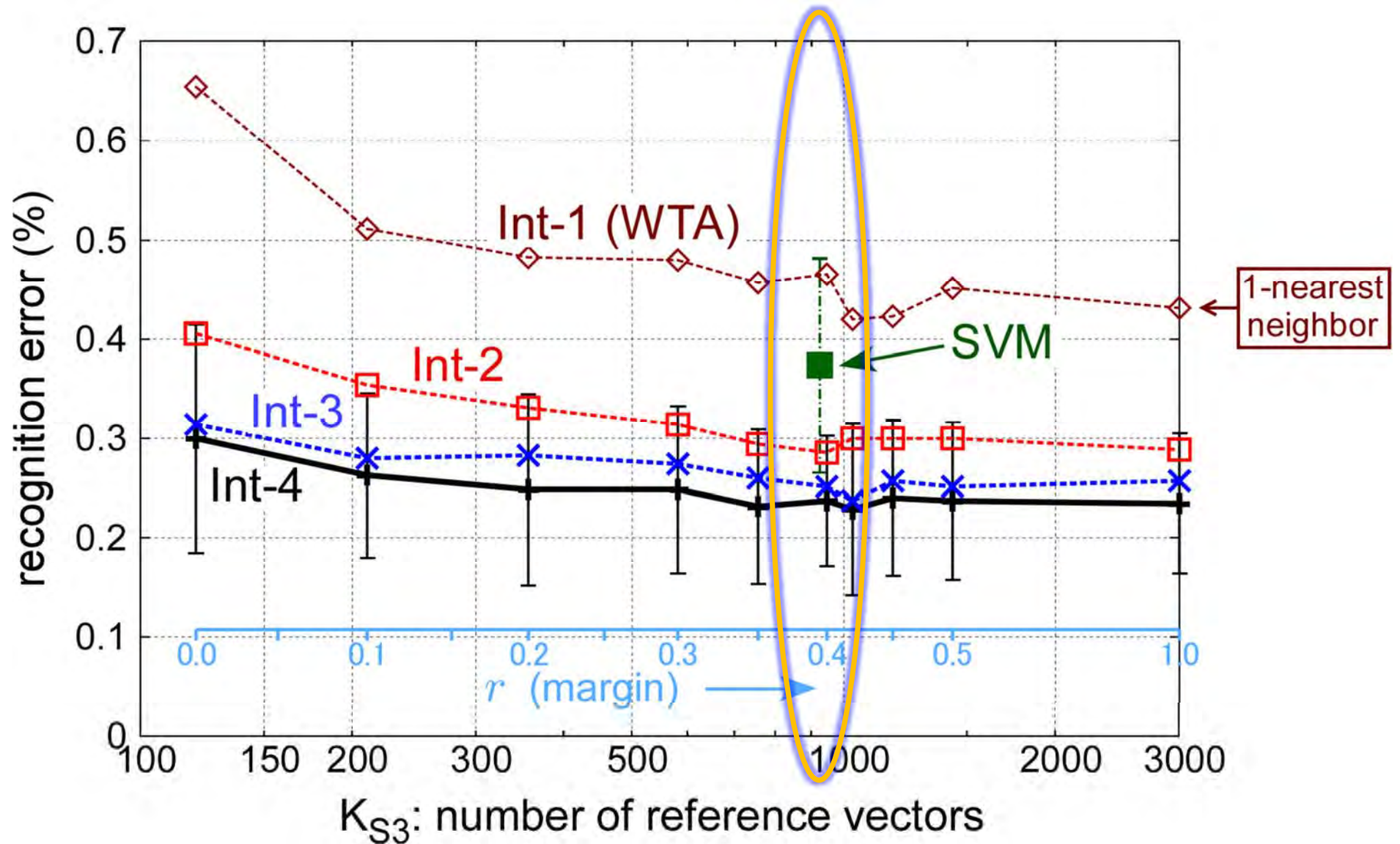
Int-3 (interpolating-vector
from 3 reference vectors)

hyper-plane
made of
4 reference vectors

Search the nearest
hyper-plane.

Int-4 (interpolating-vector
from 4 reference vectors)

neocognitron trained by mWTA with IntVec



IntVec の 演算コスト削減

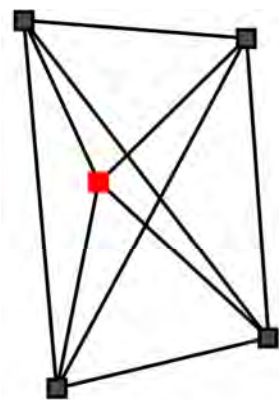
(内挿ベクトル法)

Reducing computational cost
of the Interpolating-Vector

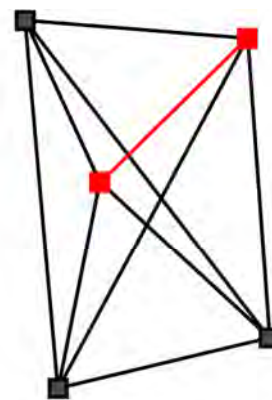
(IntVec \rightarrow WTA)

(Fukushima 2019)

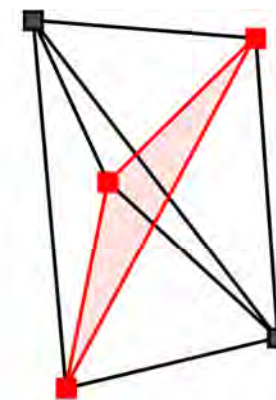
参照ベクトルの数 と 超平面の数



Int-1 (WTA)



Int-2



Int-3

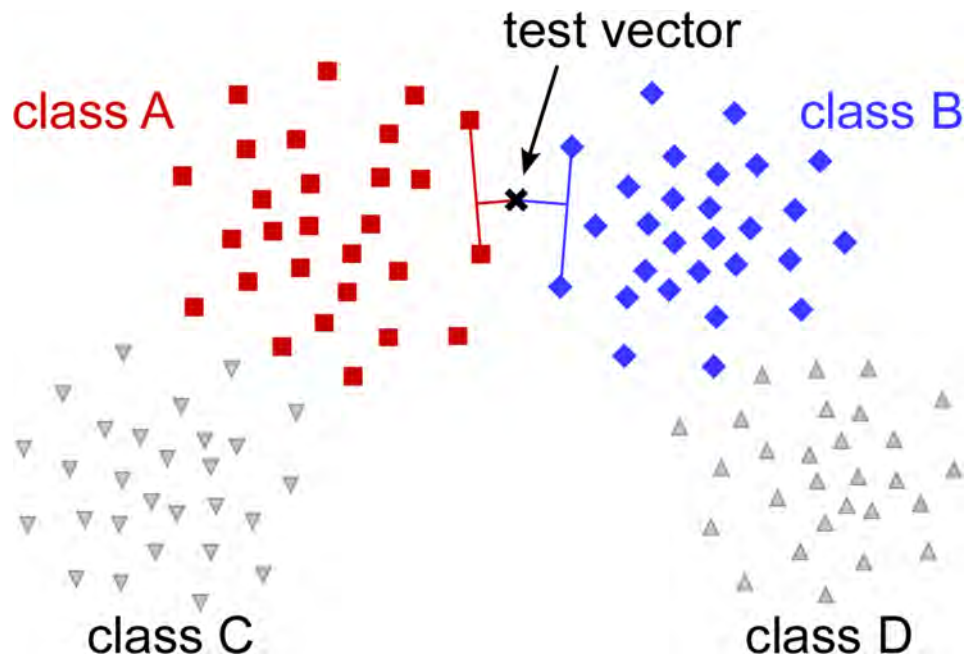


Int-4

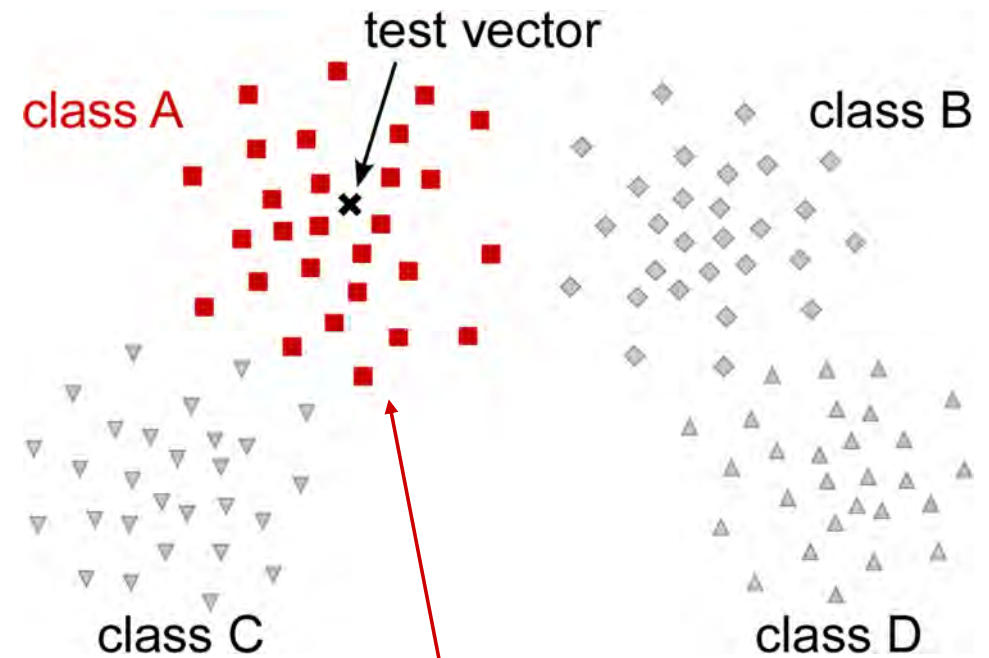
Int- m	number of hyperplanes
Int-1 (WTA)	K
Int-2	$K(K-1)/2 \doteq K^2/2$
Int-3	$K(K-1)(K-2)/6 \doteq K^3/6$
Int-4	$K(K-1)(K-2)(K-3)/24 \doteq K^4/24$

K : number of reference vectors of a class

IntVec (内挿ベクトル法) の演算量削減



WTA で 完全敗者
IntVec の演算省略
(class C 及び D)

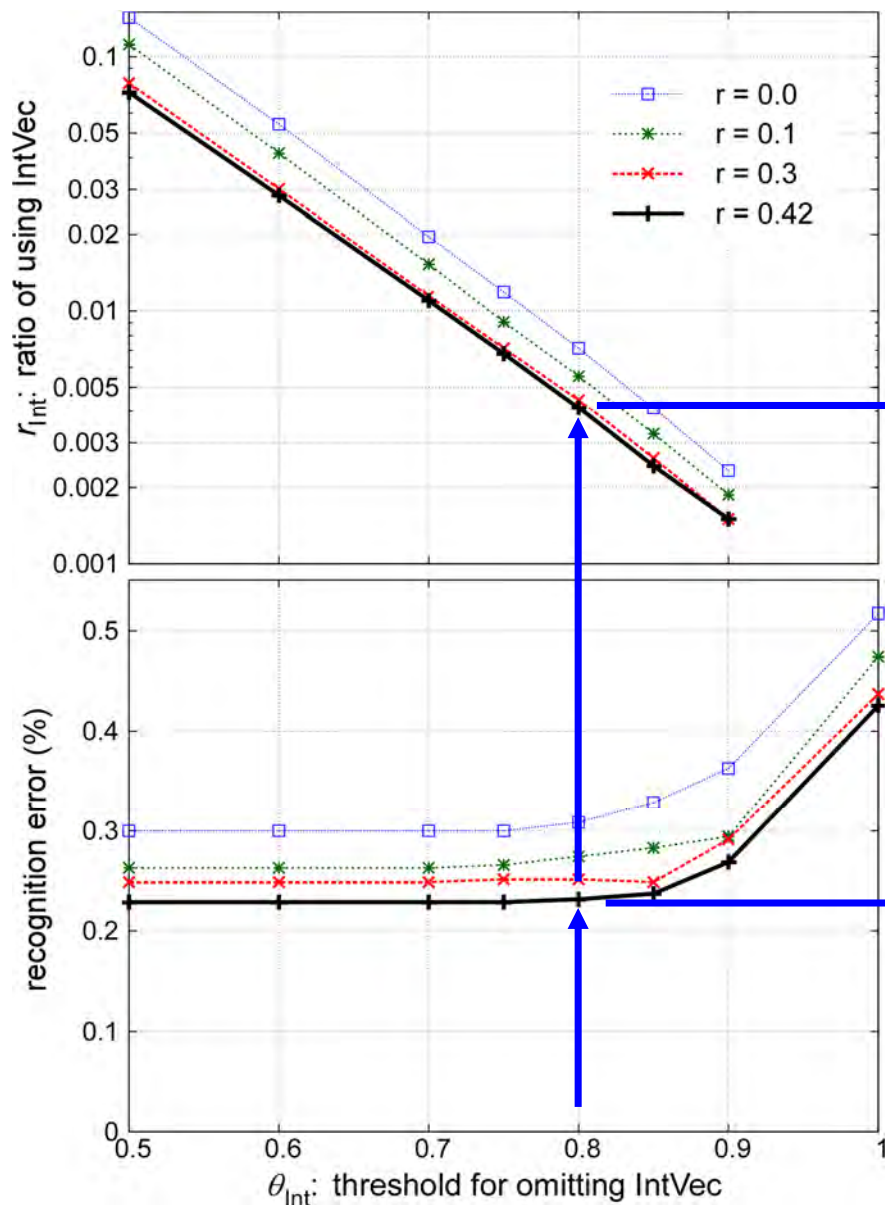


WTA で 一人勝ち
全class に対して
IntVec の演算不要
(class A に対しても)

(Fukushima 2019)

演算コスト

認識誤差



IntVec (Int-4)
with reduced computational cost

--- Omitting Calculation of IntVec ---

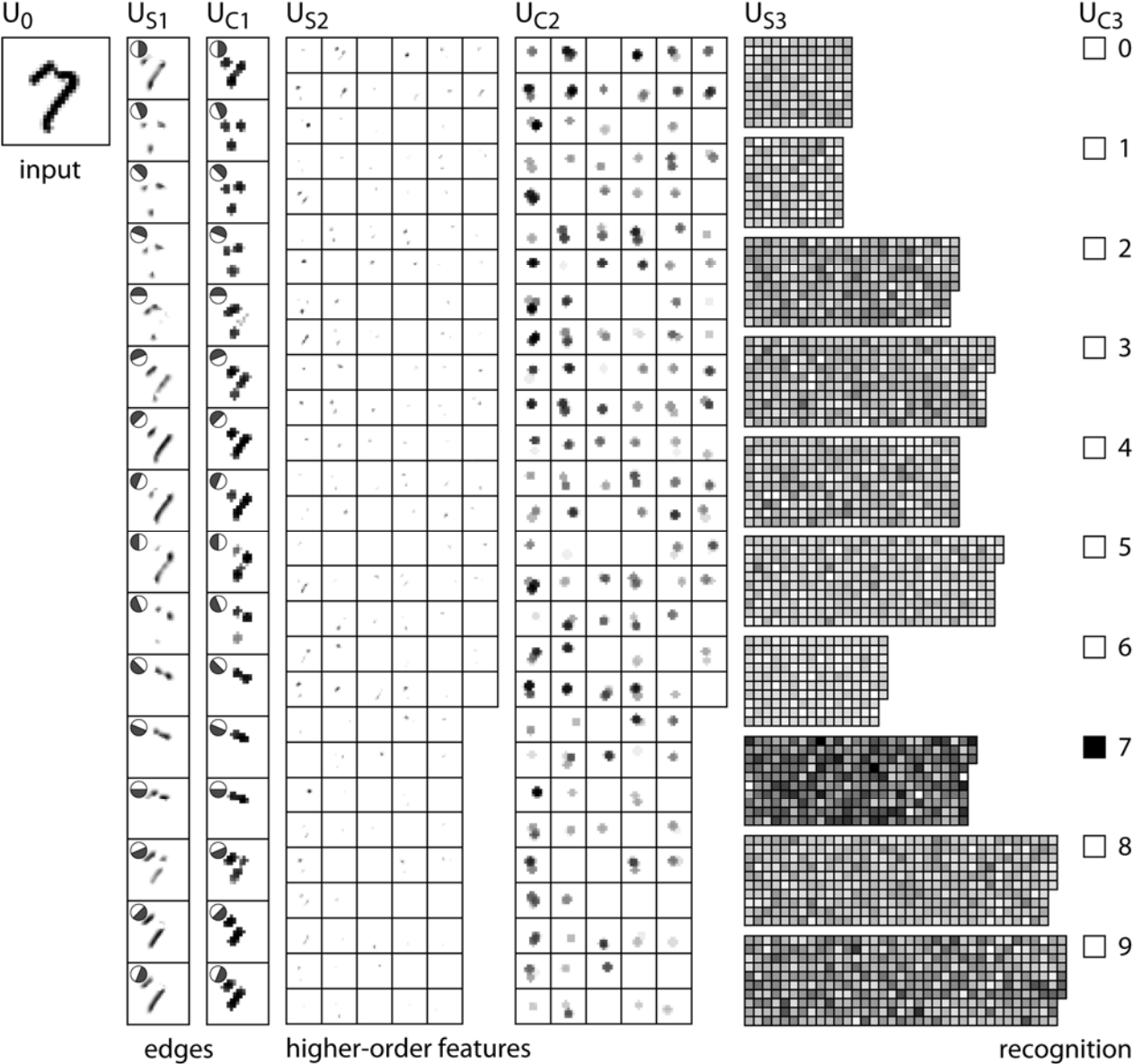
演算コスト: ほぼ 0.4% に減少

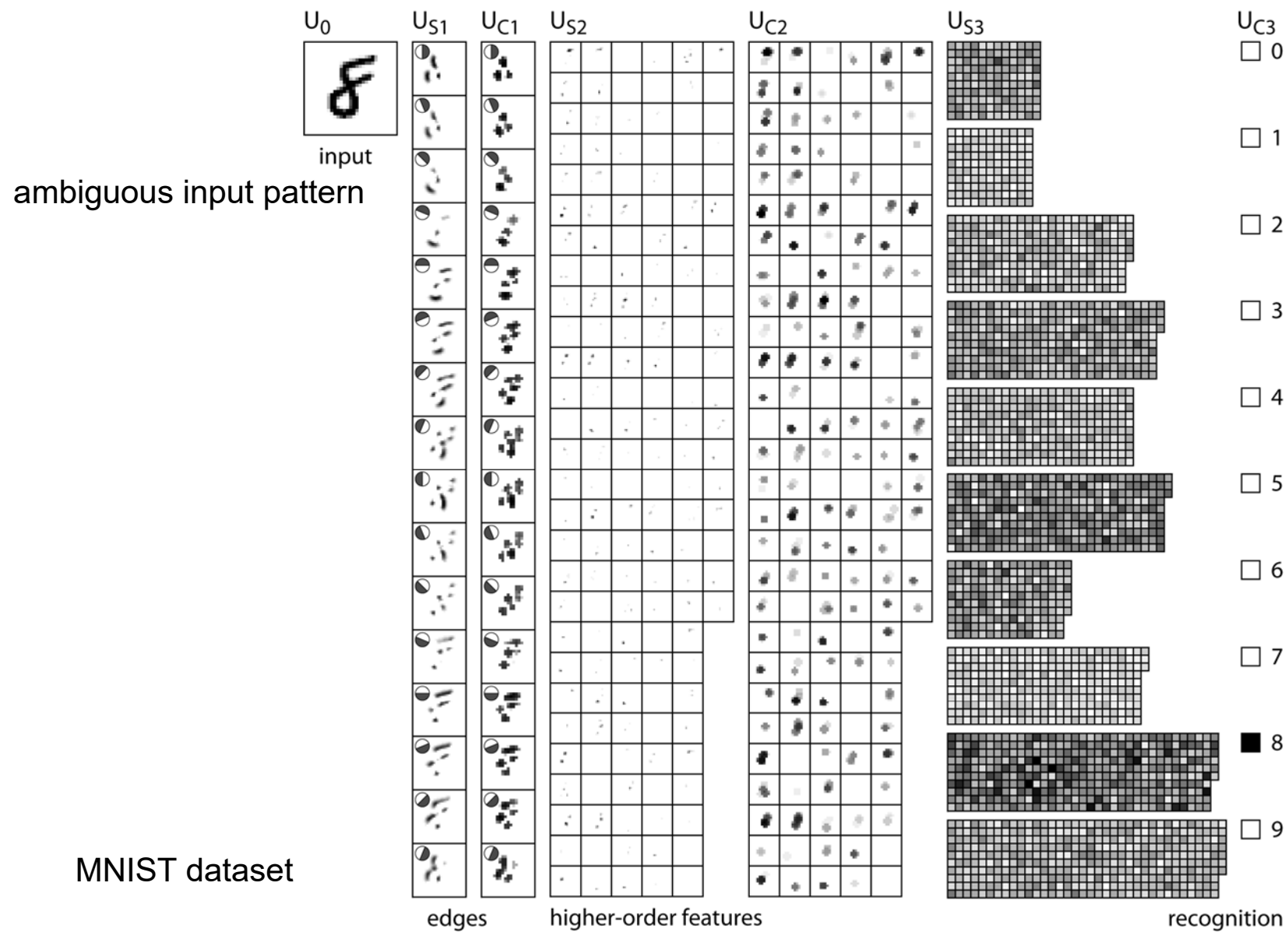
認識誤差: ほとんど不変

(Fukushima 2019)

Examples of the response of cells in neocognitron

MNIST dataset

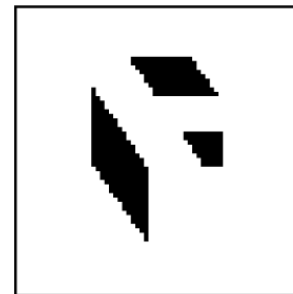
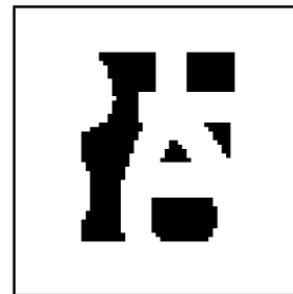
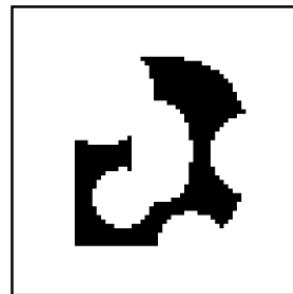
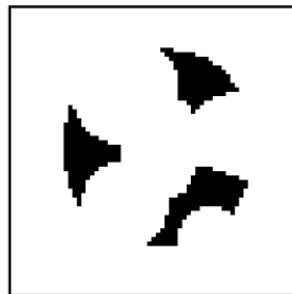
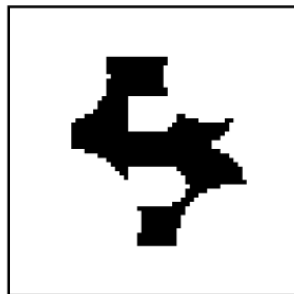
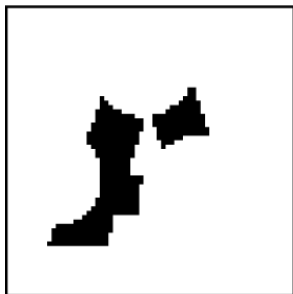


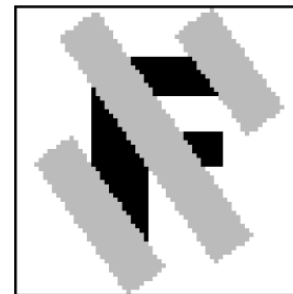
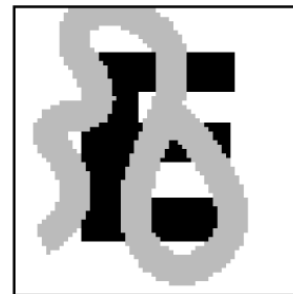
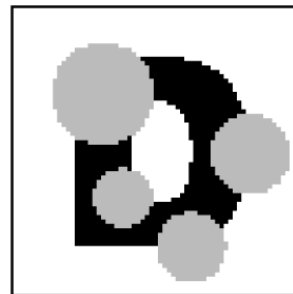
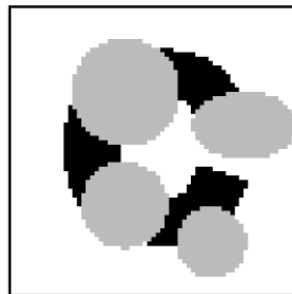
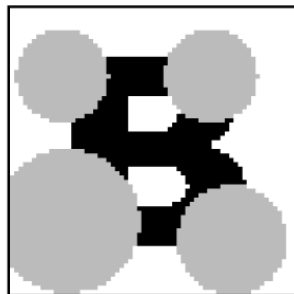
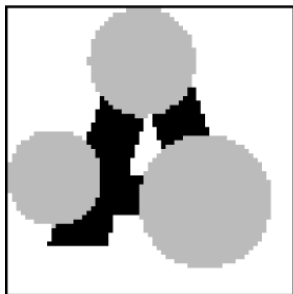


一部が隠されたパターンの認識

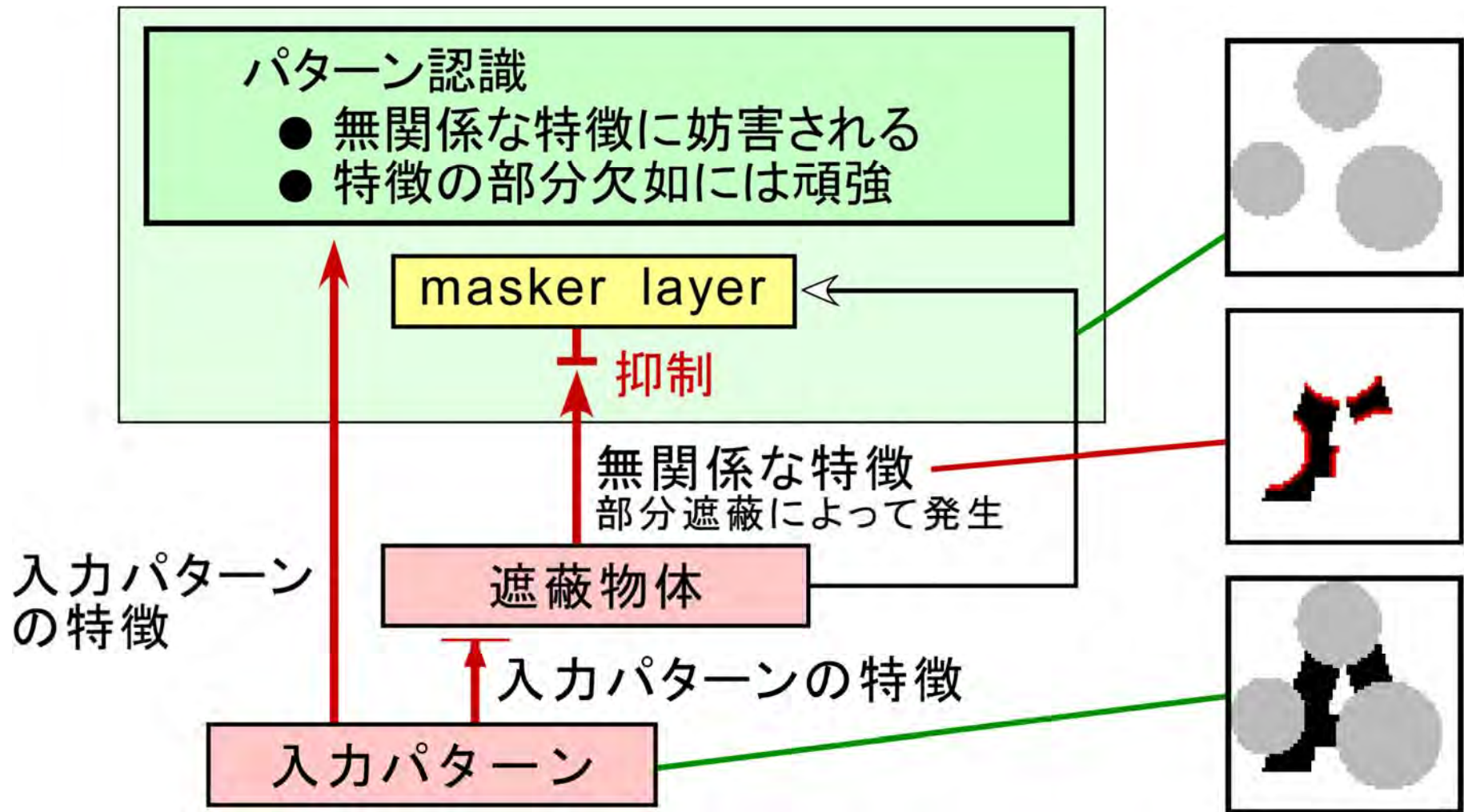
Recognition of Partly Occluded Patterns

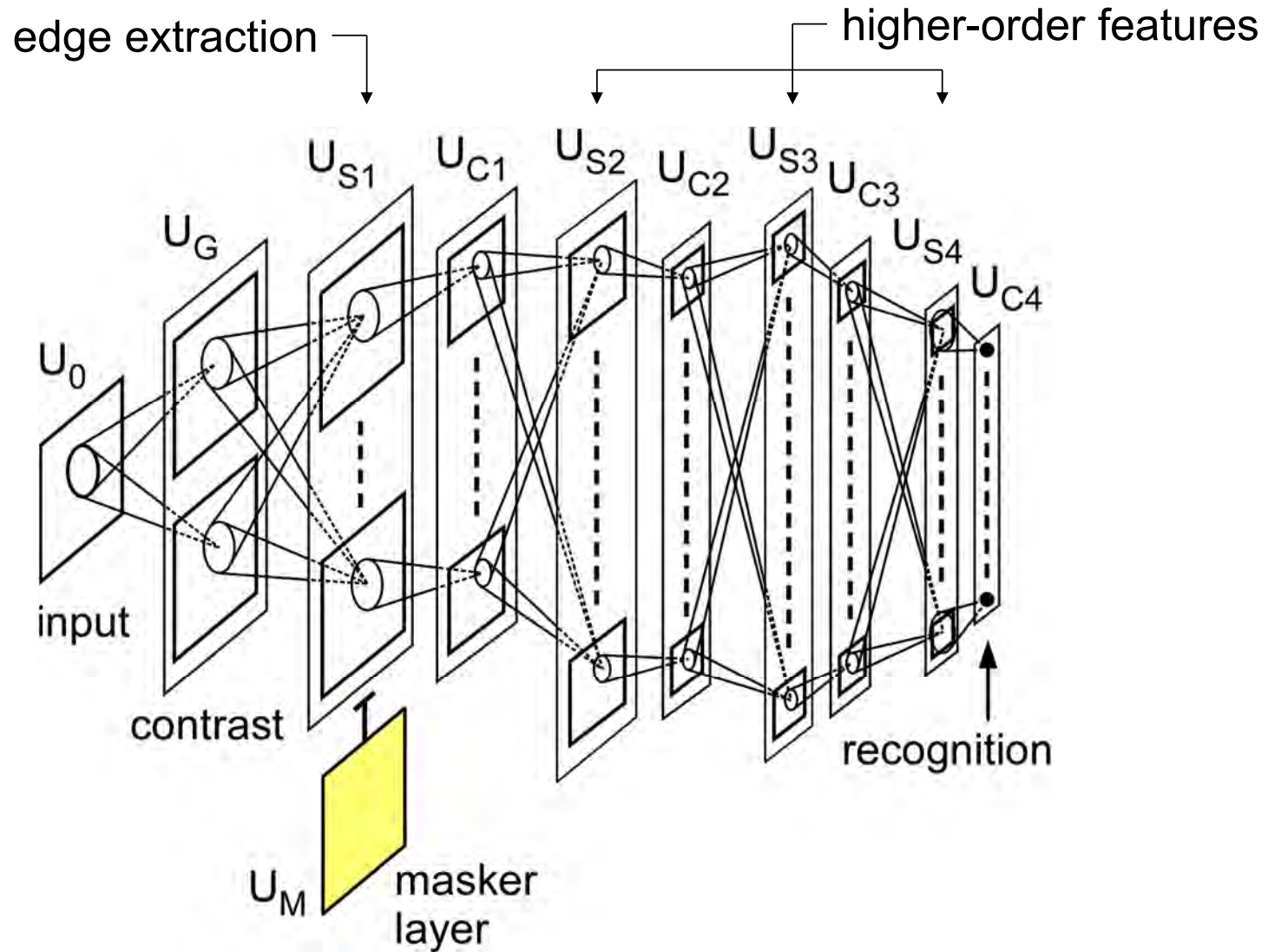
(Fukushima 2001)



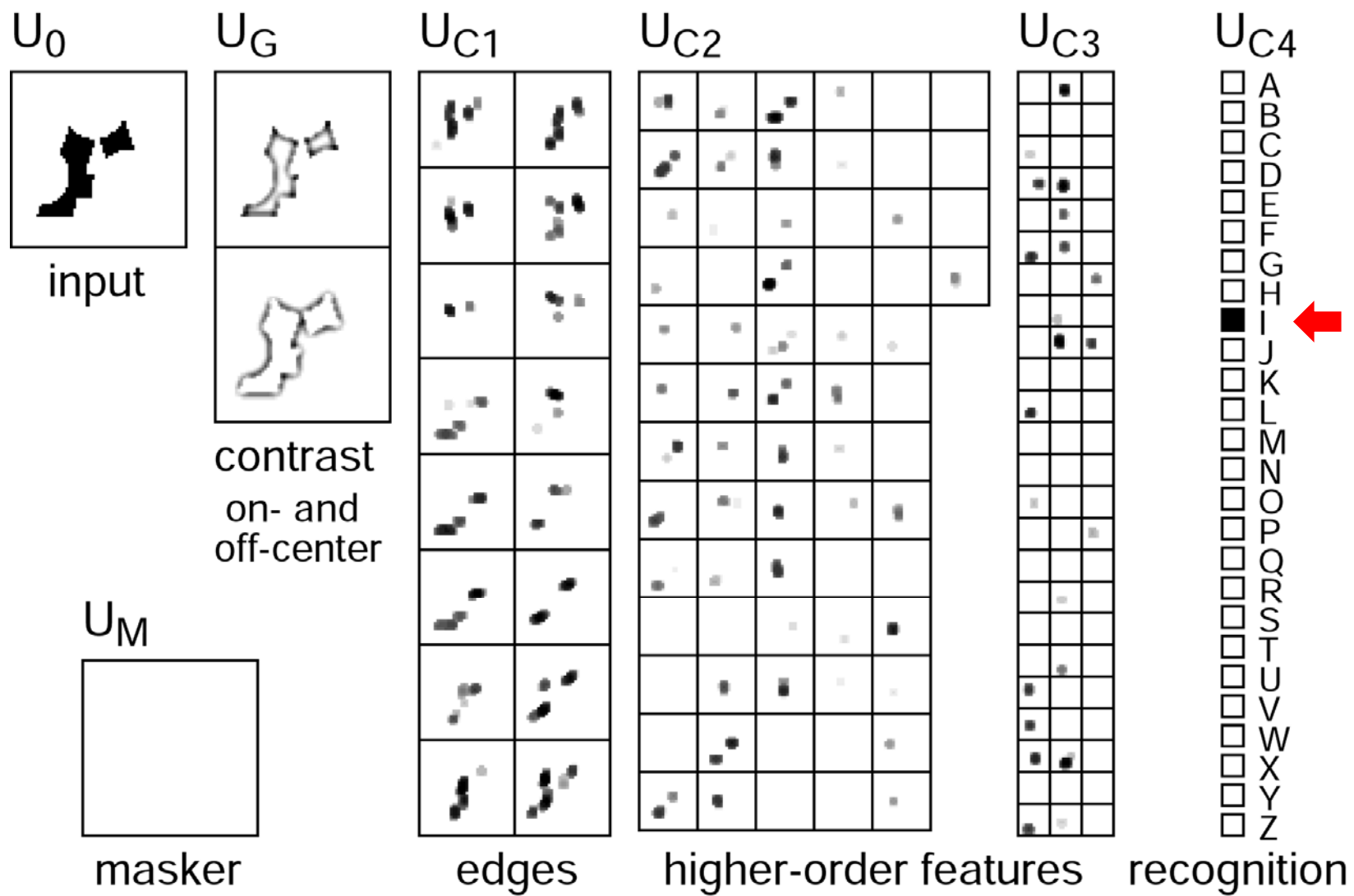


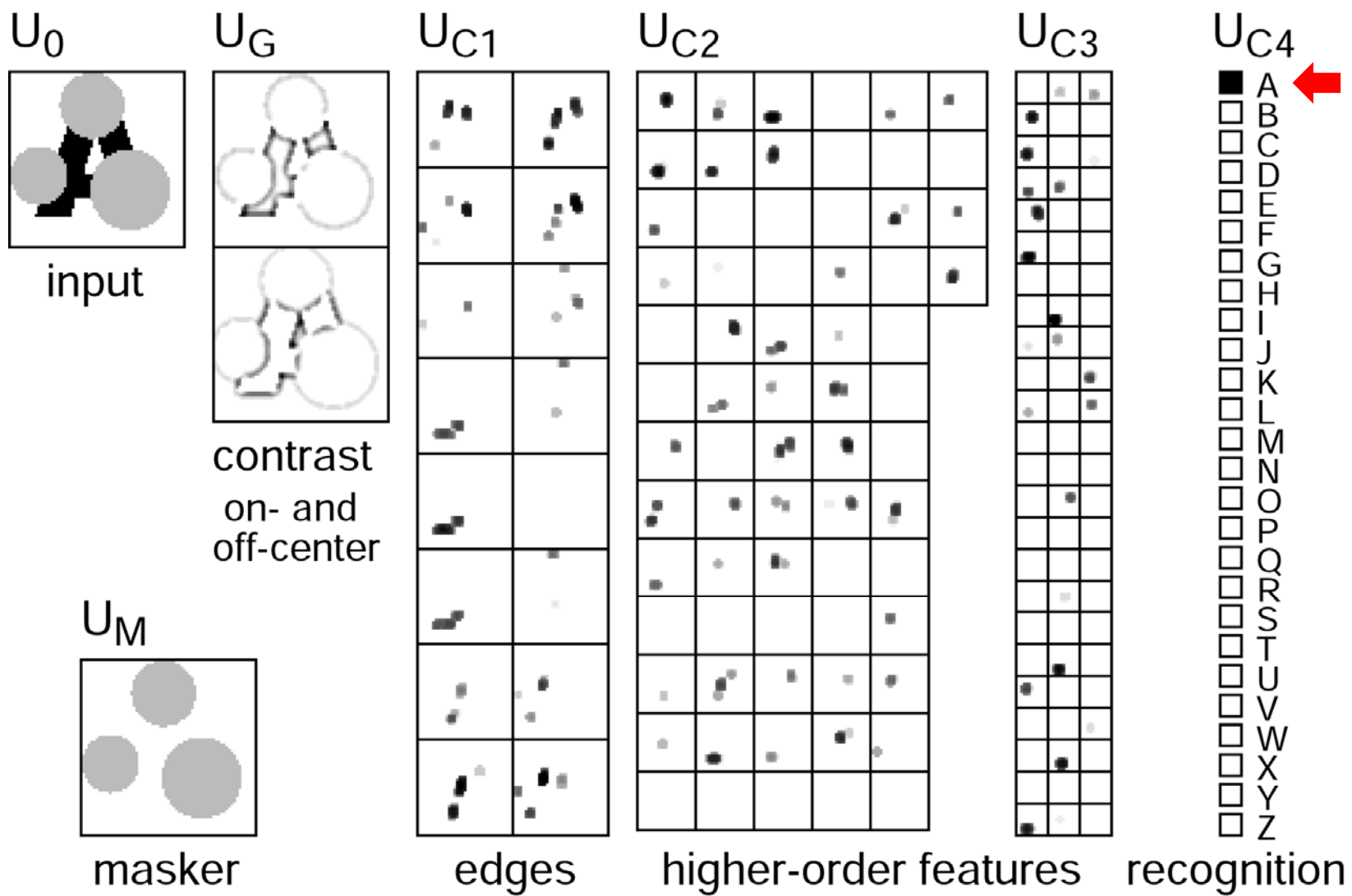
遮蔽されたパターンは 遮蔽物体が見えると認識しやすくなる

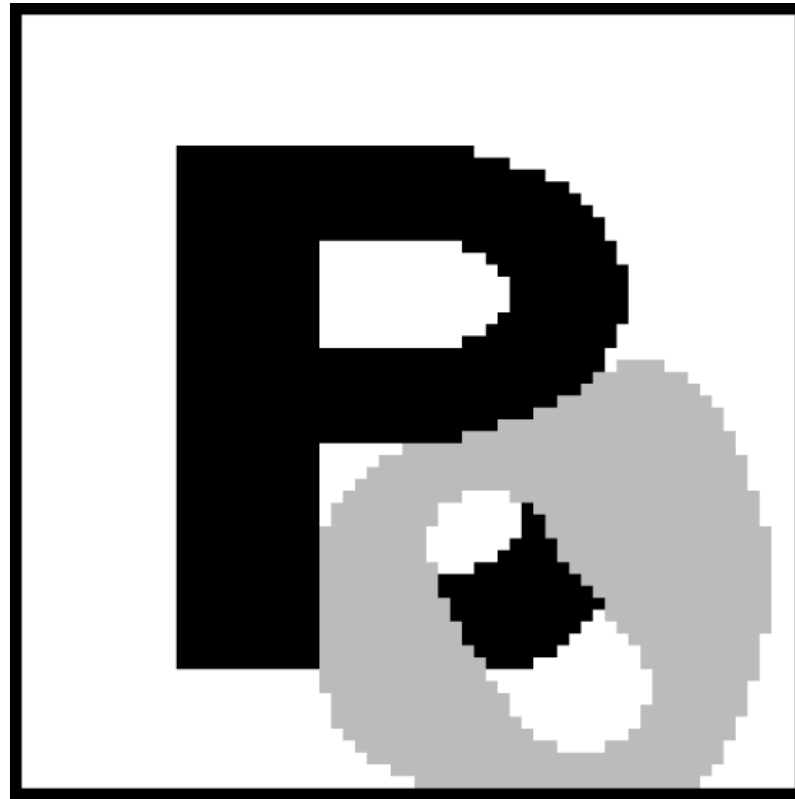




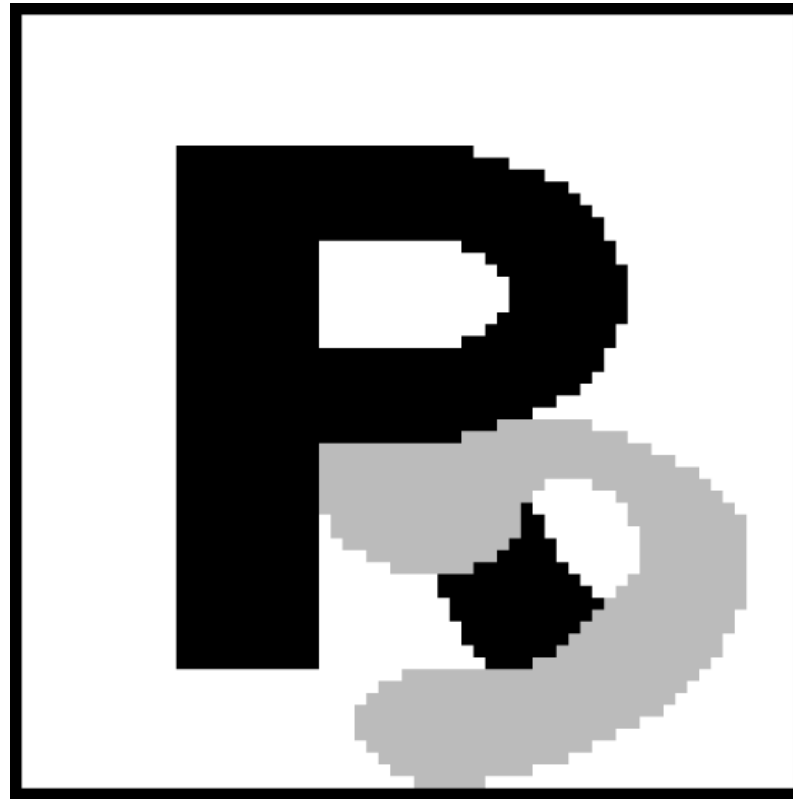
Proposed Model = Neocognitron + **Masker layer**



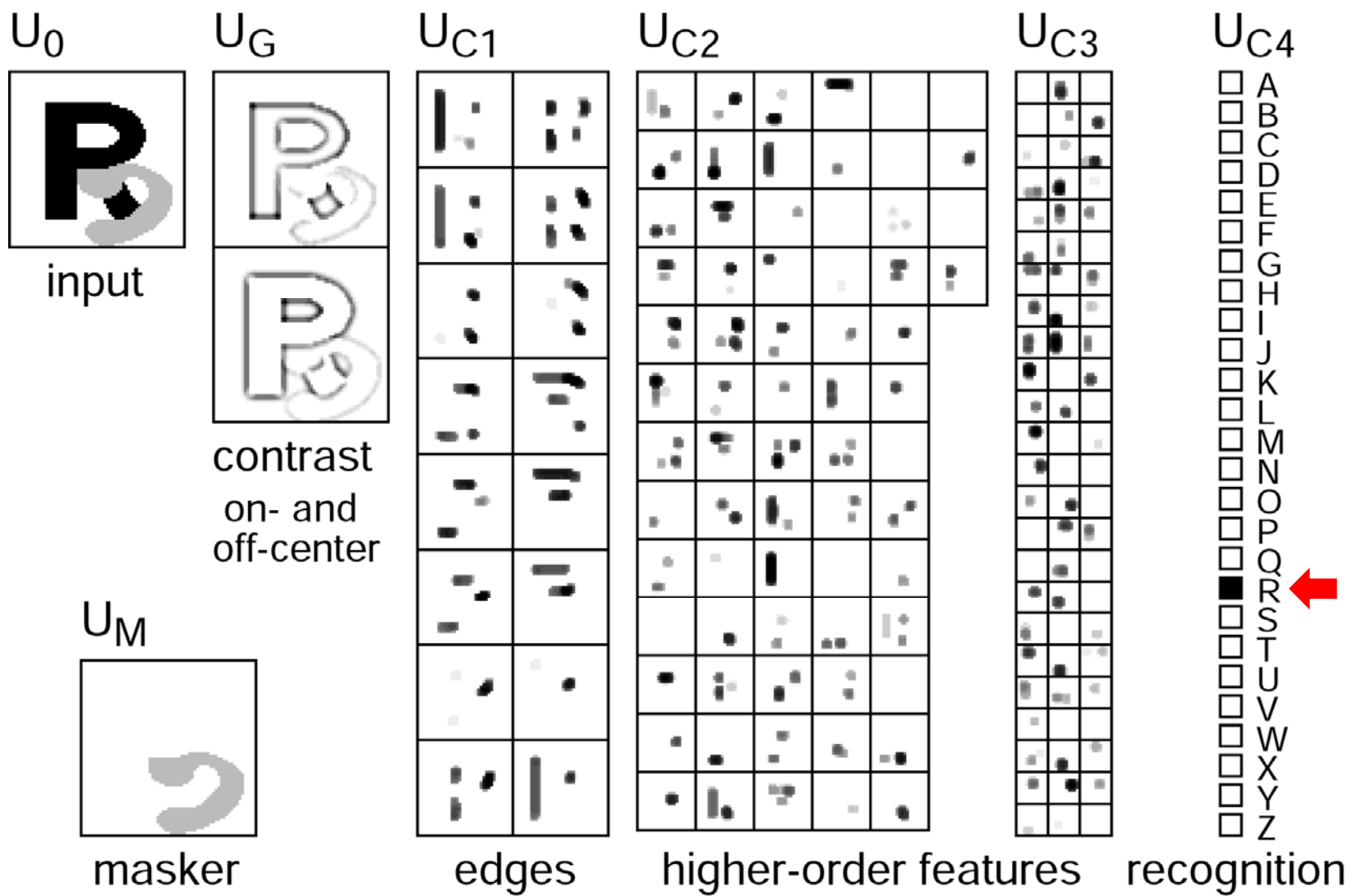


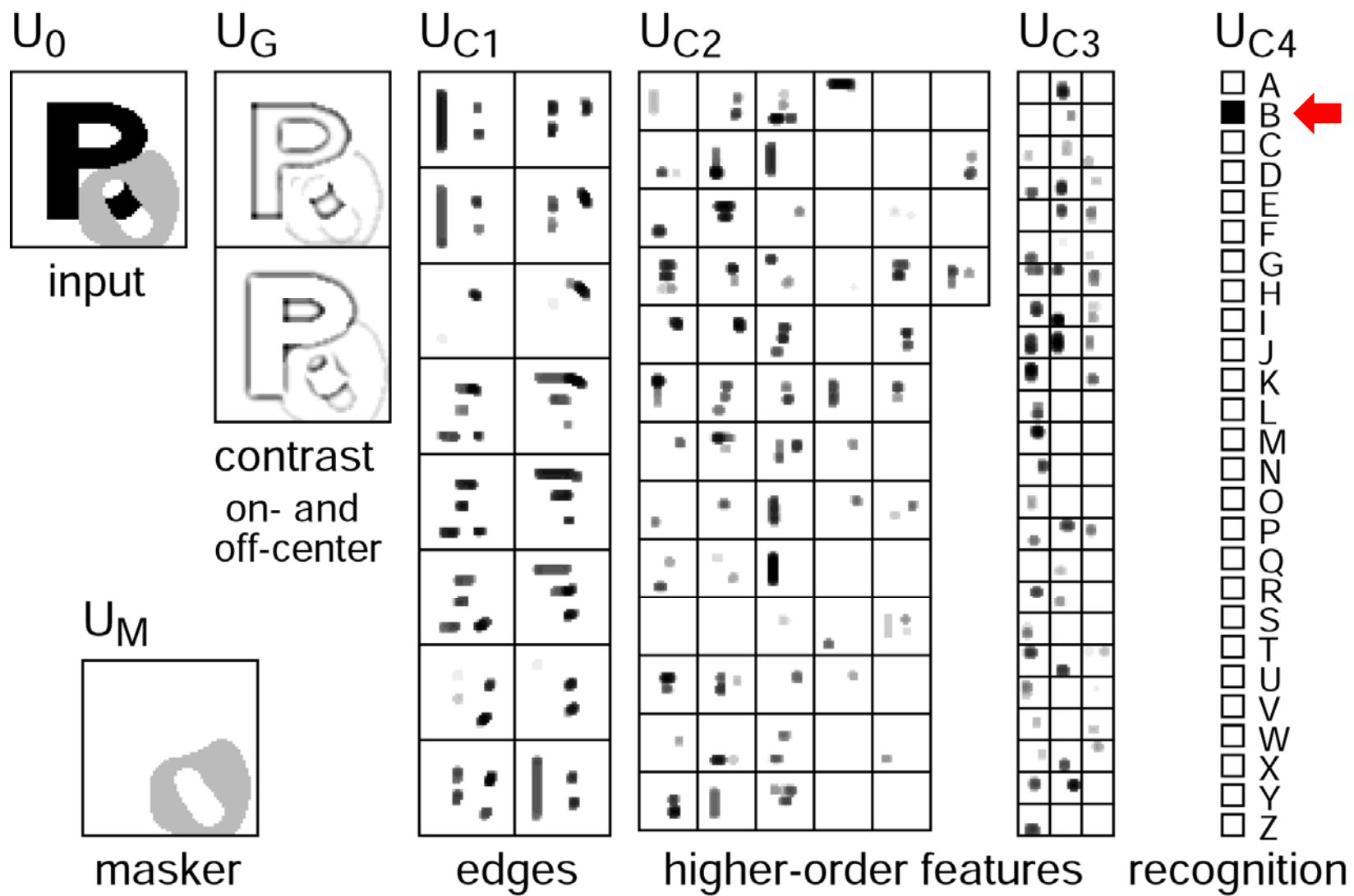


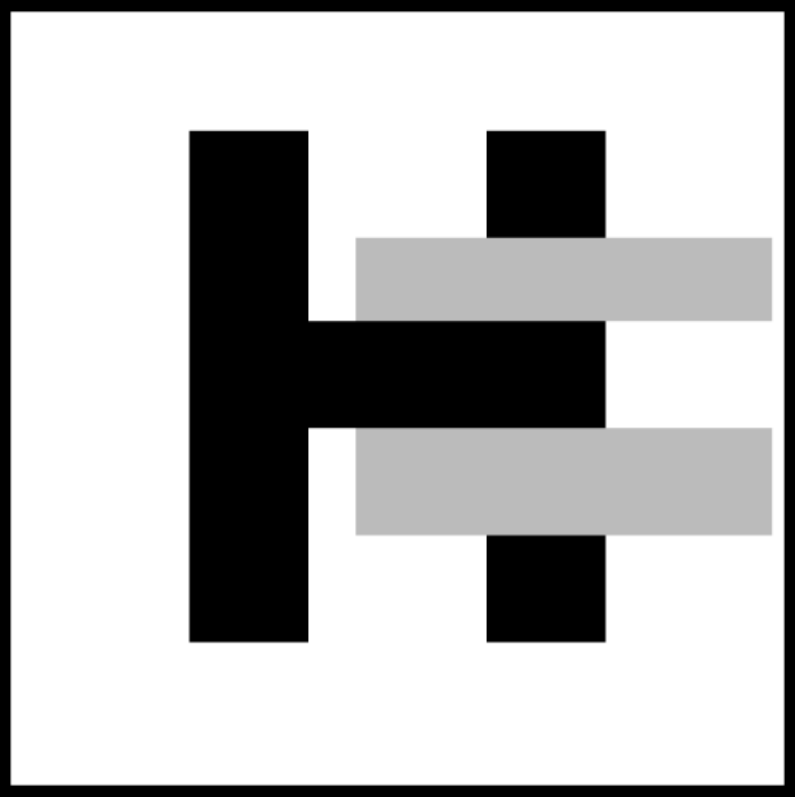
An identical pattern is perceived differently
by the placement of gray objects.

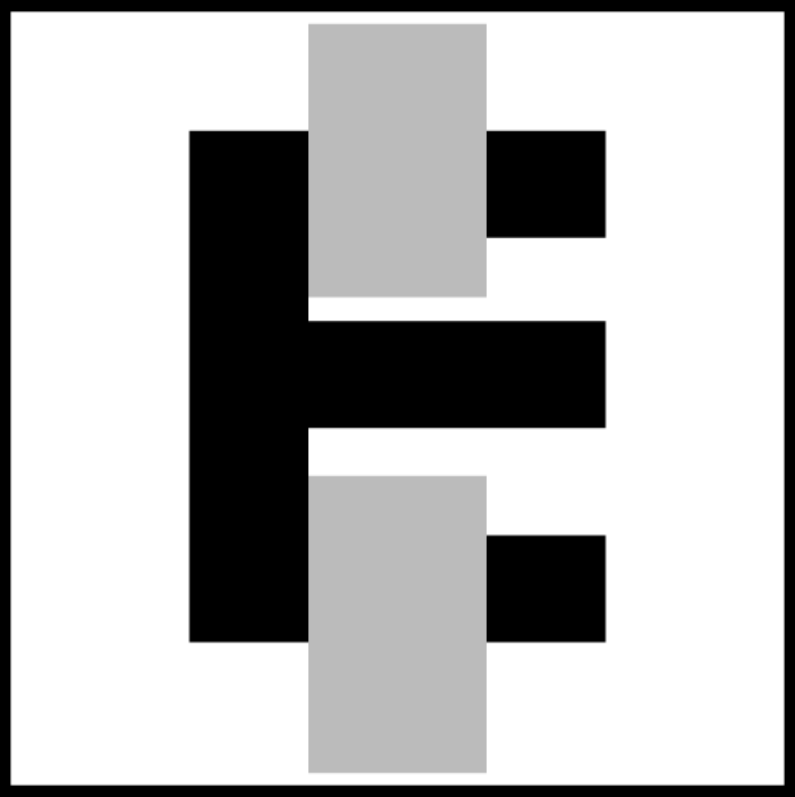


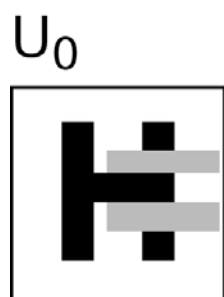
An identical pattern is perceived differently
by the placement of gray objects.



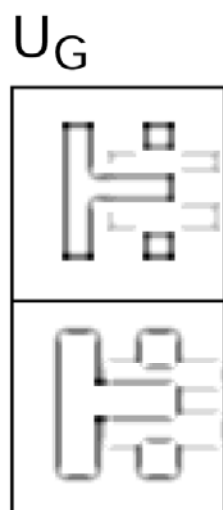




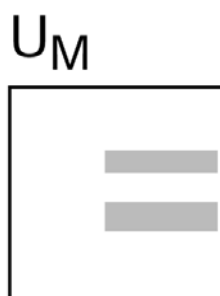




input



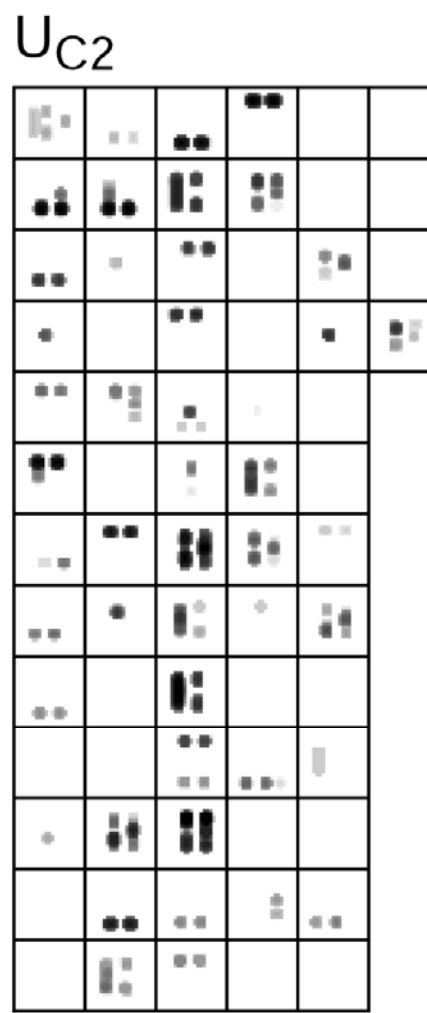
contrast
on- and
off-center



masker



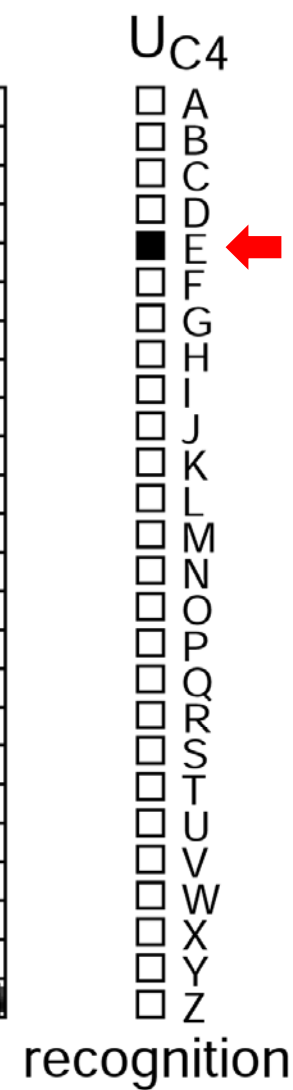
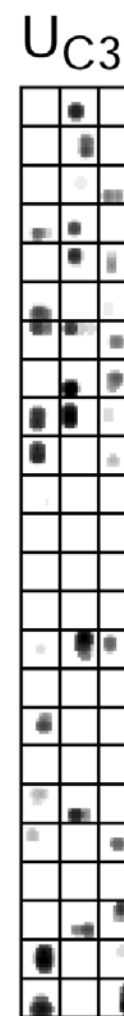
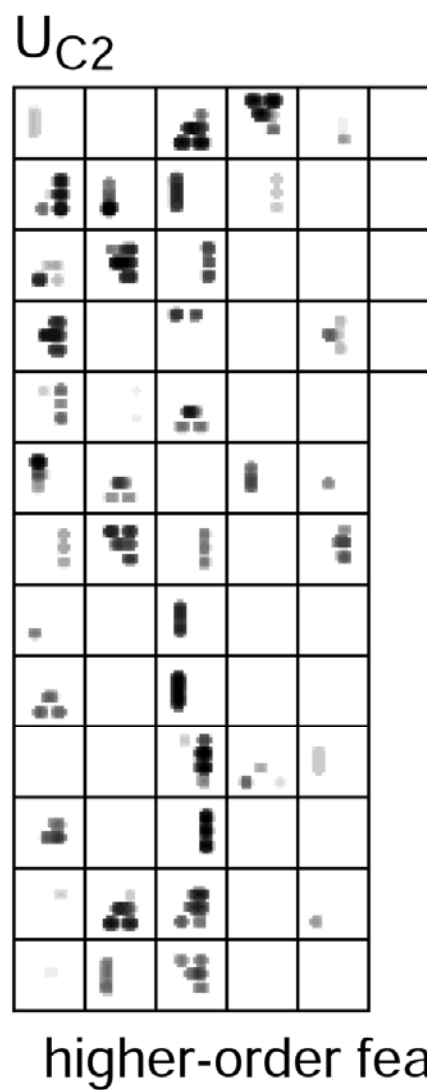
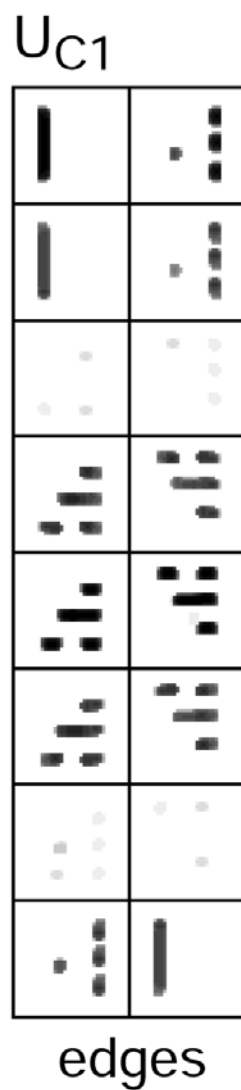
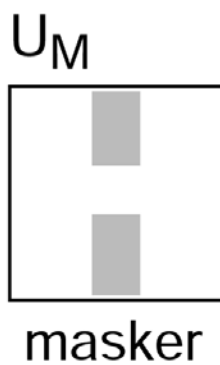
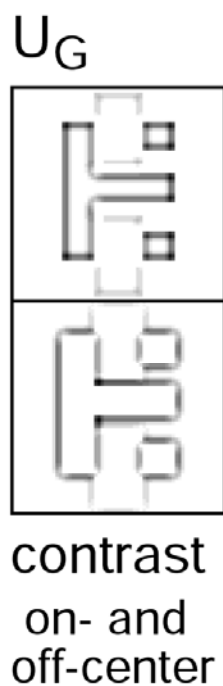
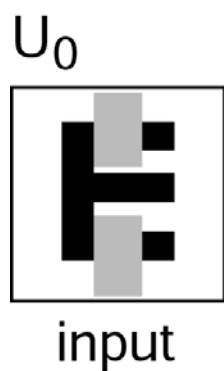
edges

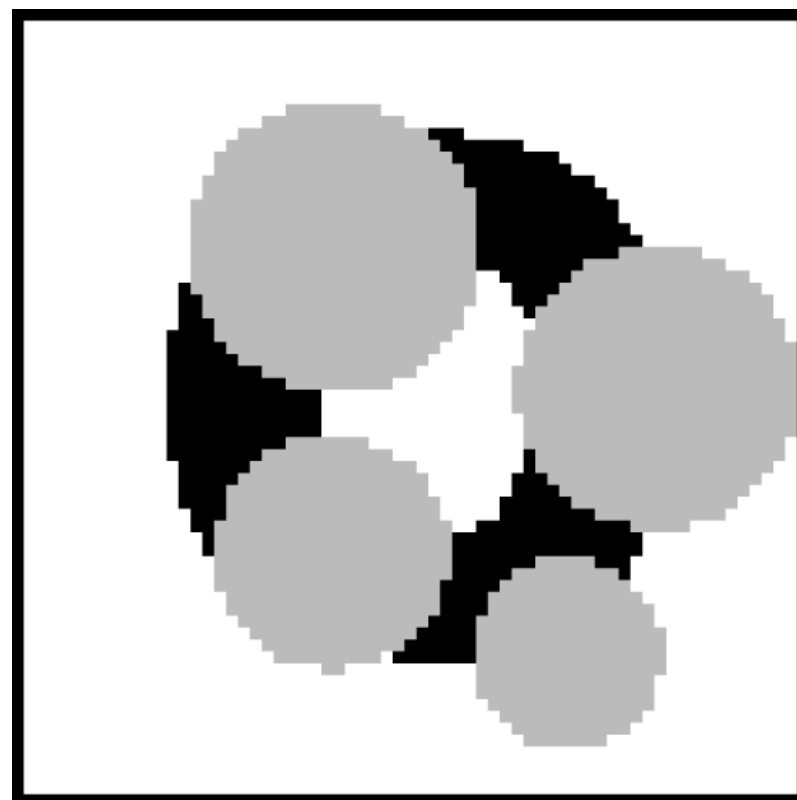


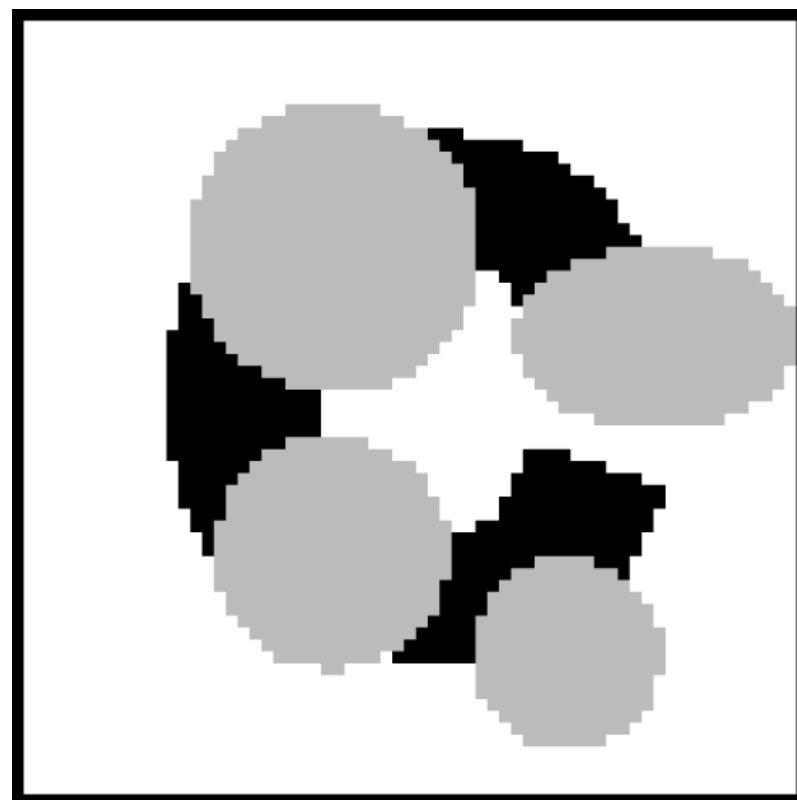
higher-order features

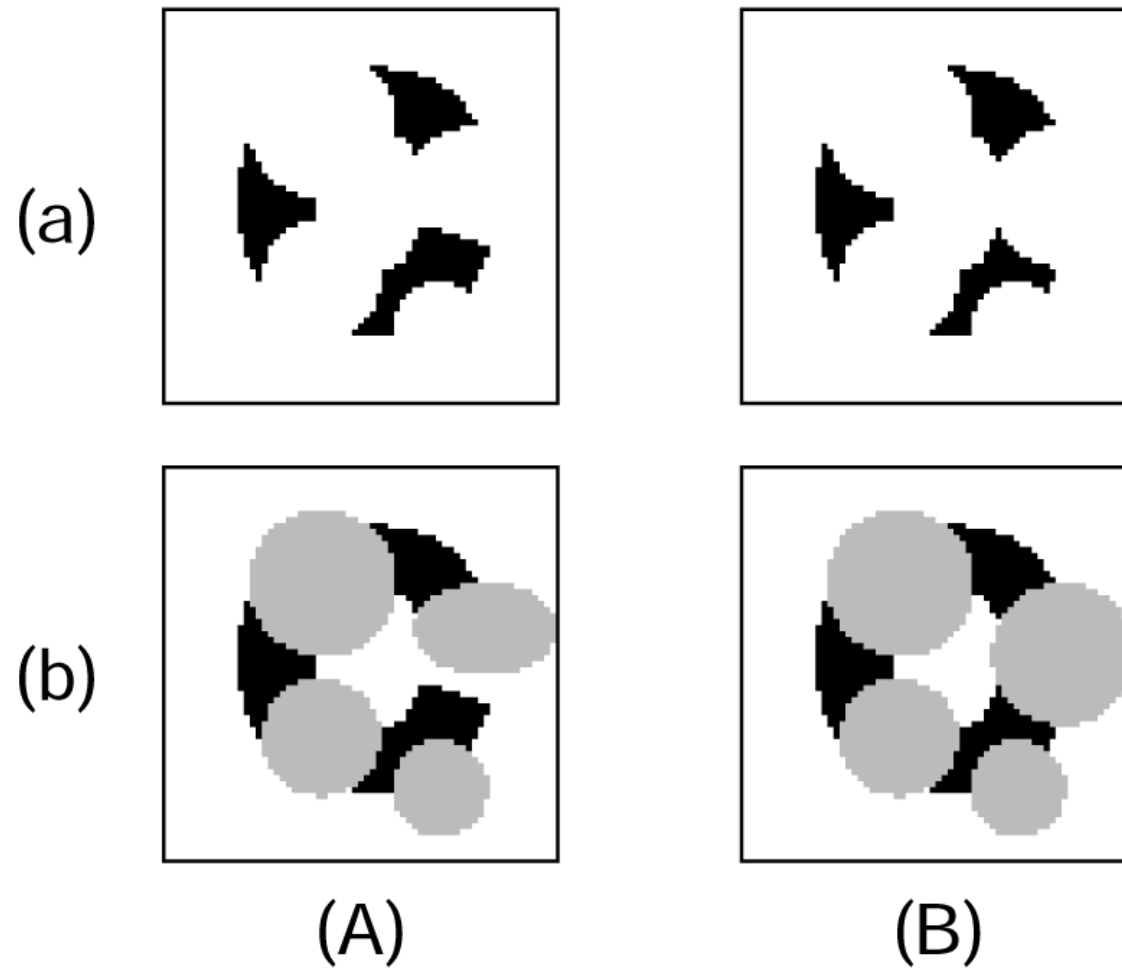


recognition



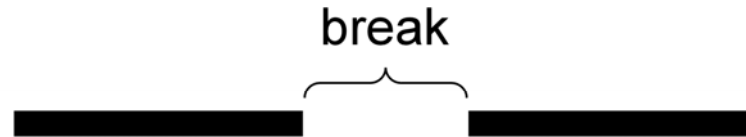






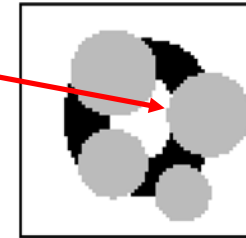
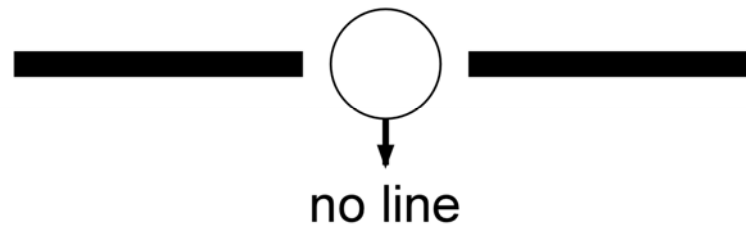
The same pattern is perceived differently through a slight difference in occluding objects.

Perception of a Break of a Line Behind an Occluding Object

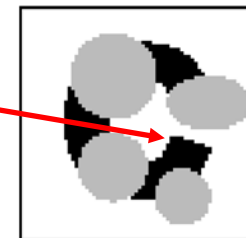
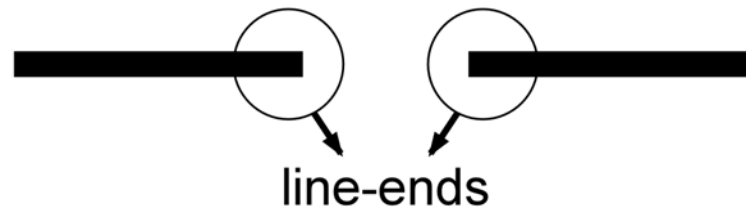


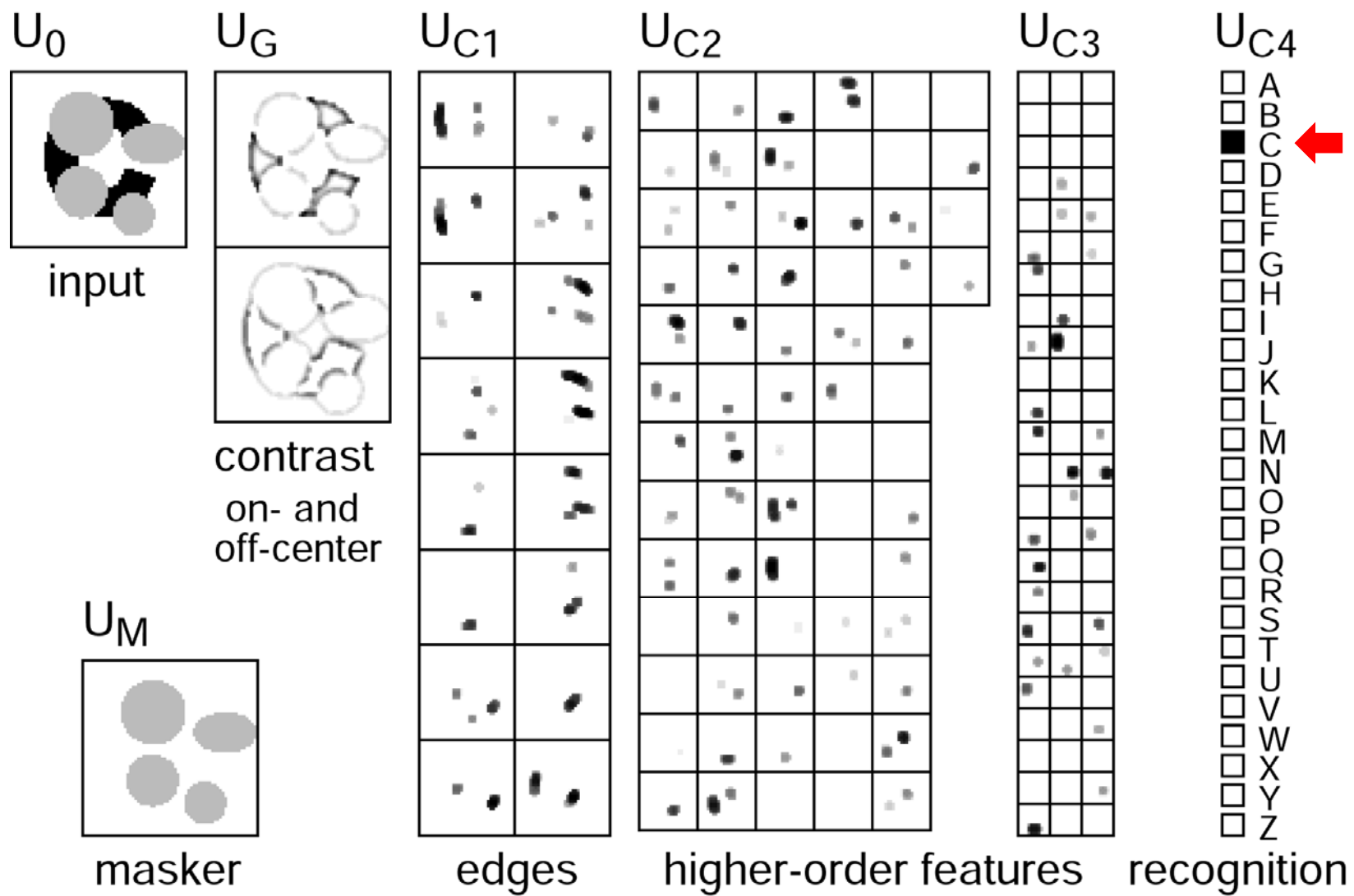
Hypothesis:

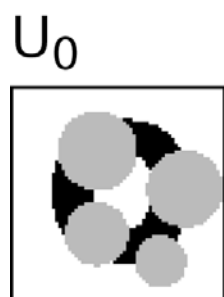
A break of a line behind an occluder is perceived,
NOT from the absence of a line component,



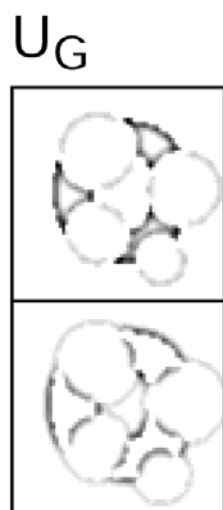
but from the existence of line-ends.



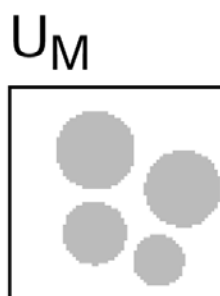




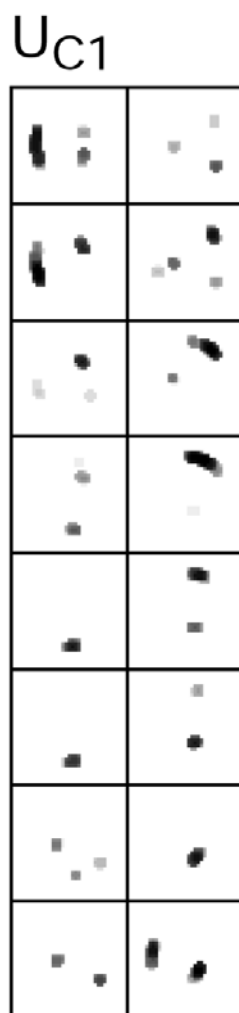
input



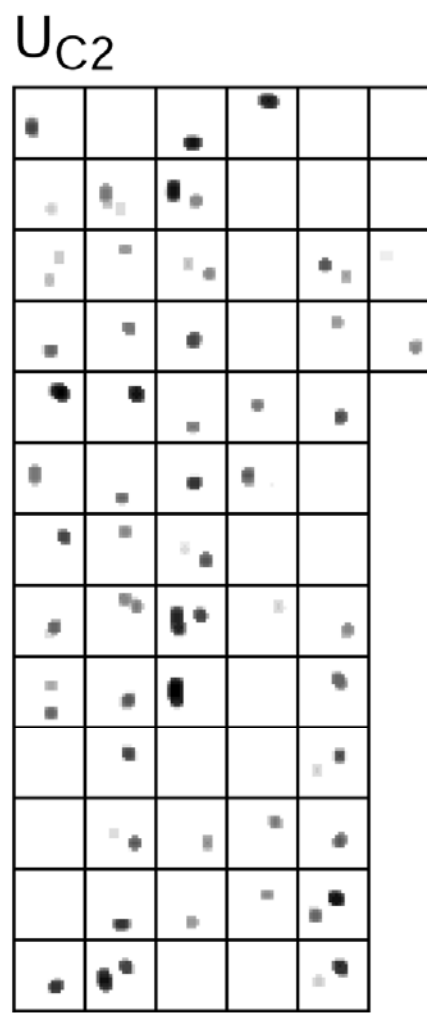
contrast
on- and
off-center



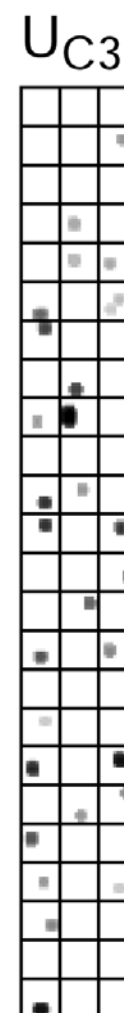
masker



edges



higher-order features



recognition



選択的注意

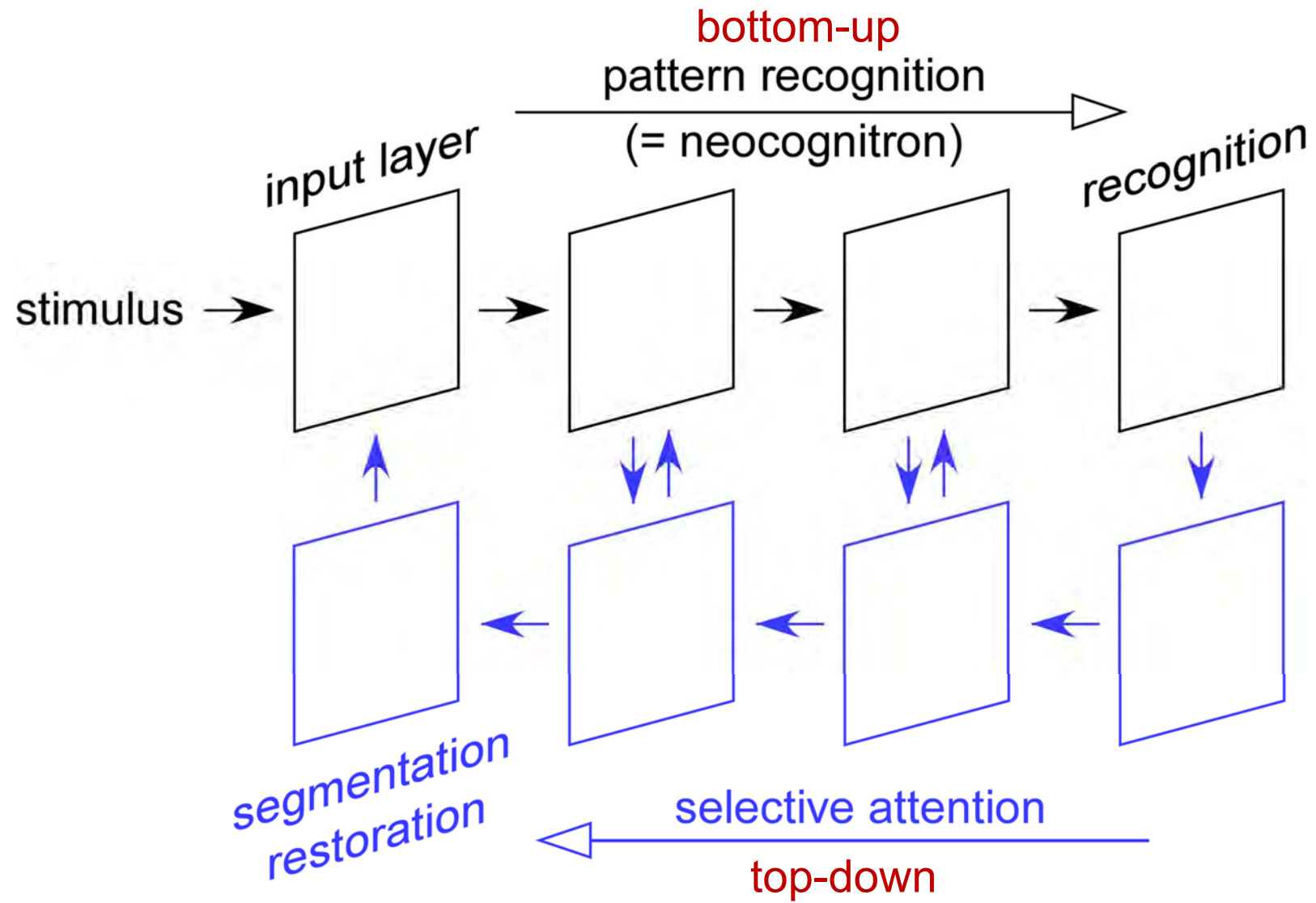
Selective Attention

— トップダウン信号の導入 —

— introducing backward signal paths —

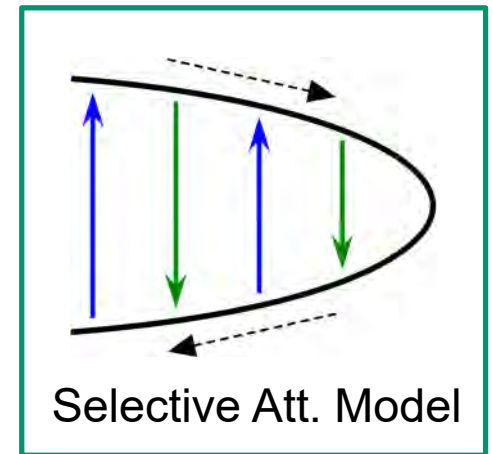
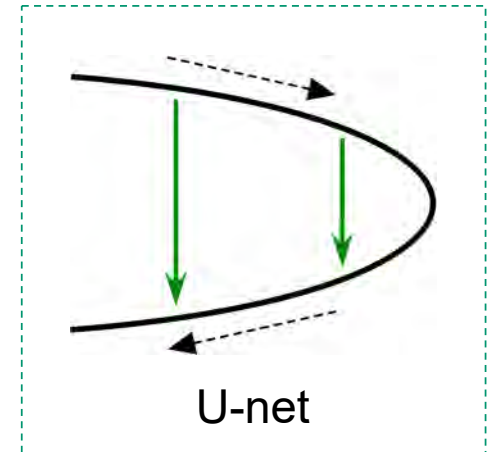
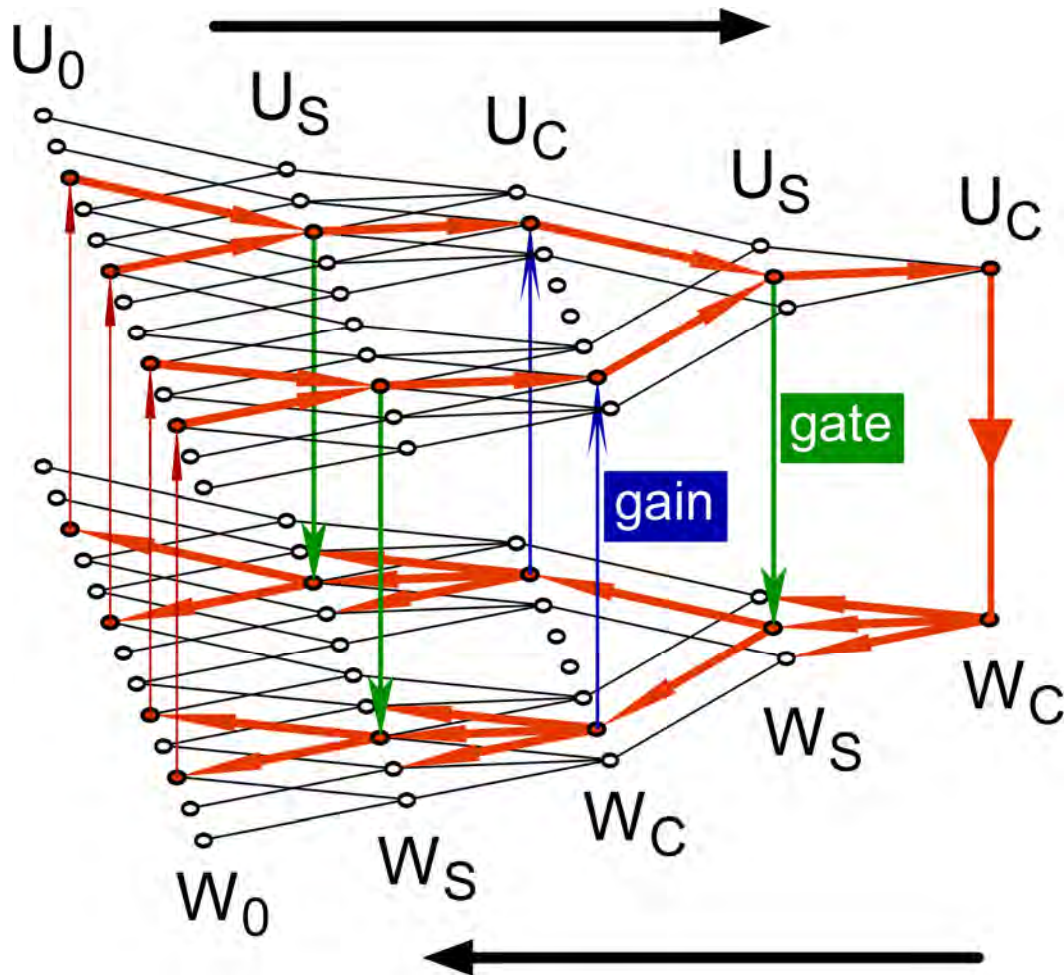
(Fukushima 1986)

Selective Attention model

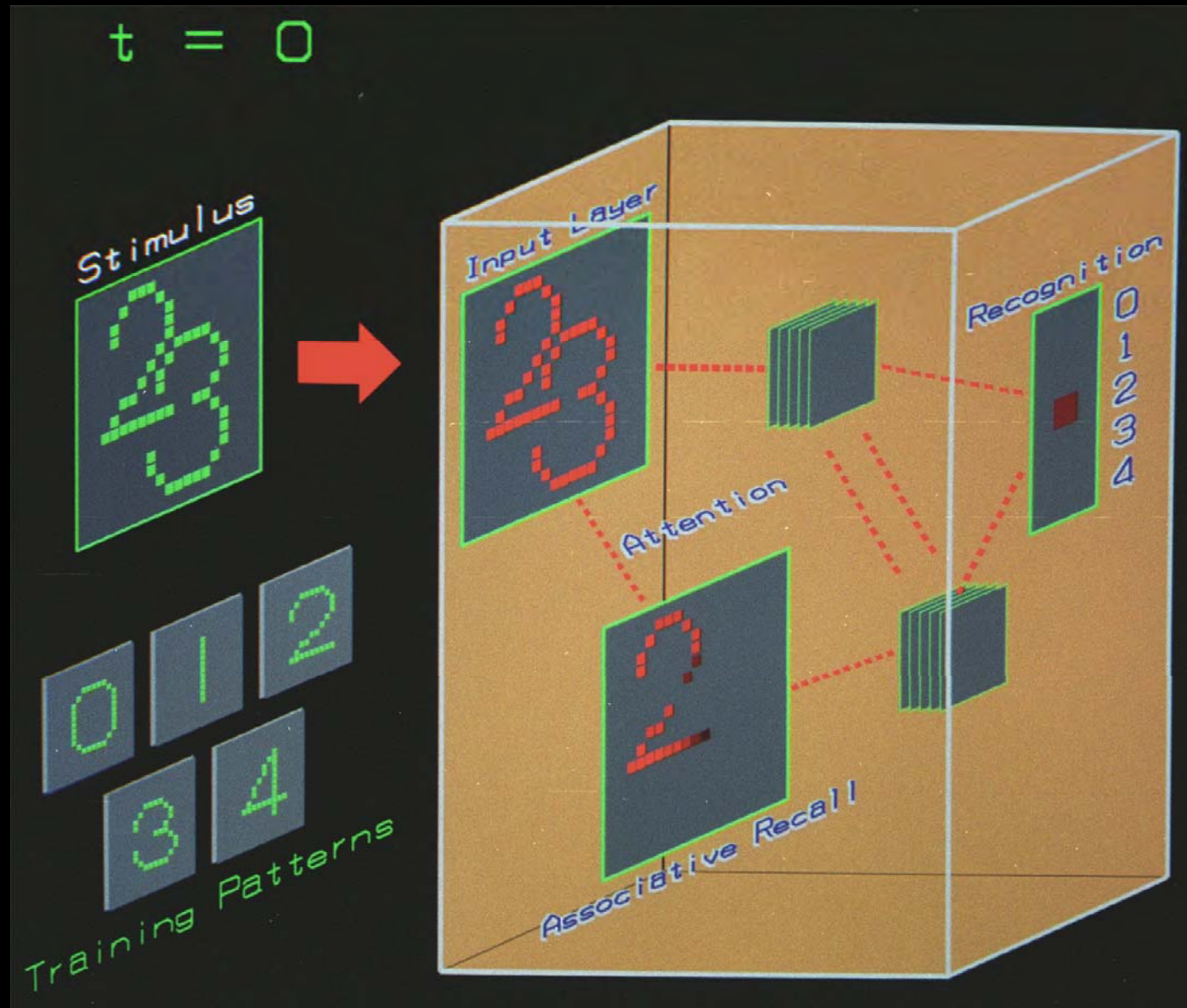


Selective Attention

--- Interaction between forward and backward signals ---

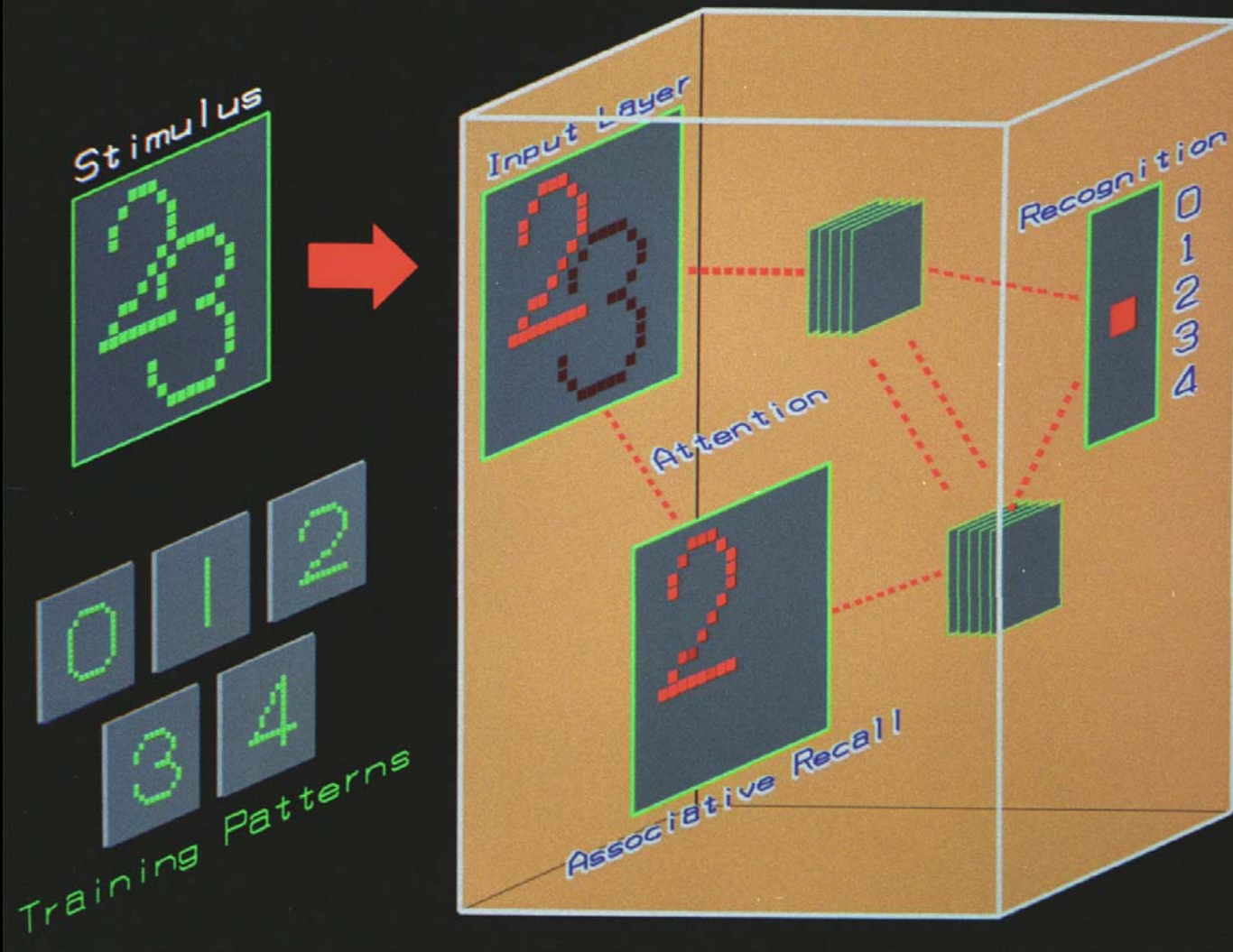


(Fukushima, 1986)

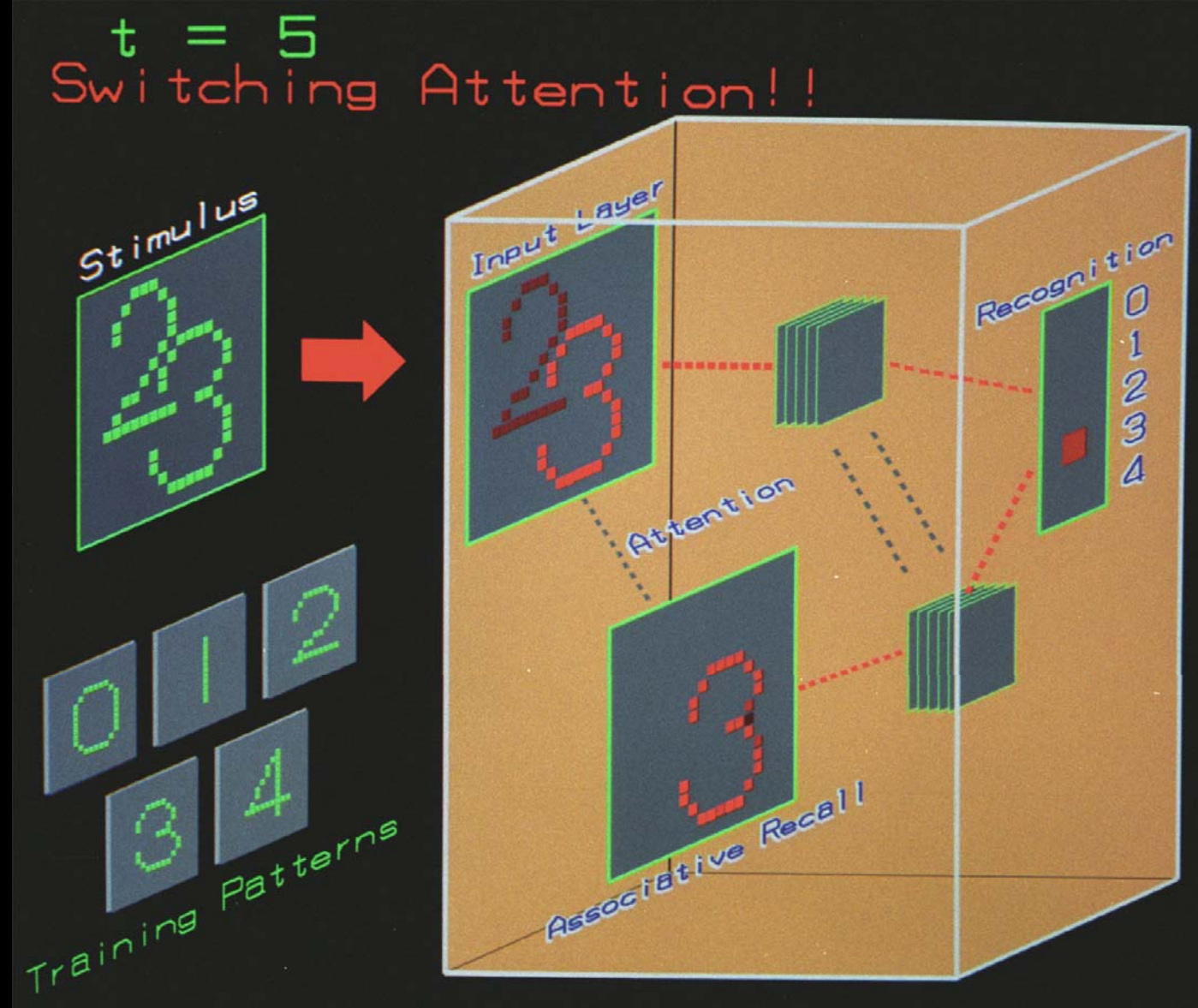


(Fukushima, 1986)

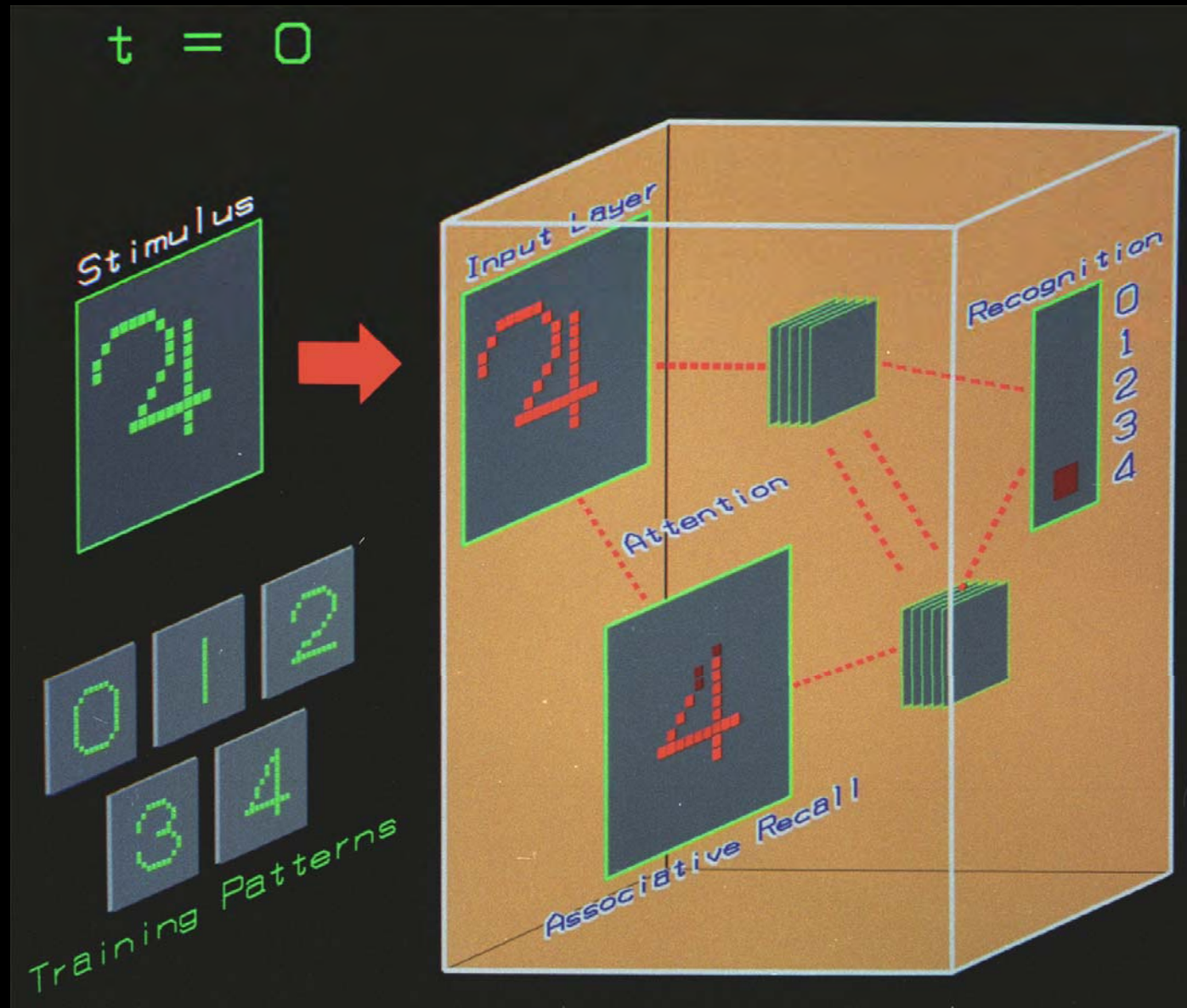
$t = 4$



(Fukushima, 1986)

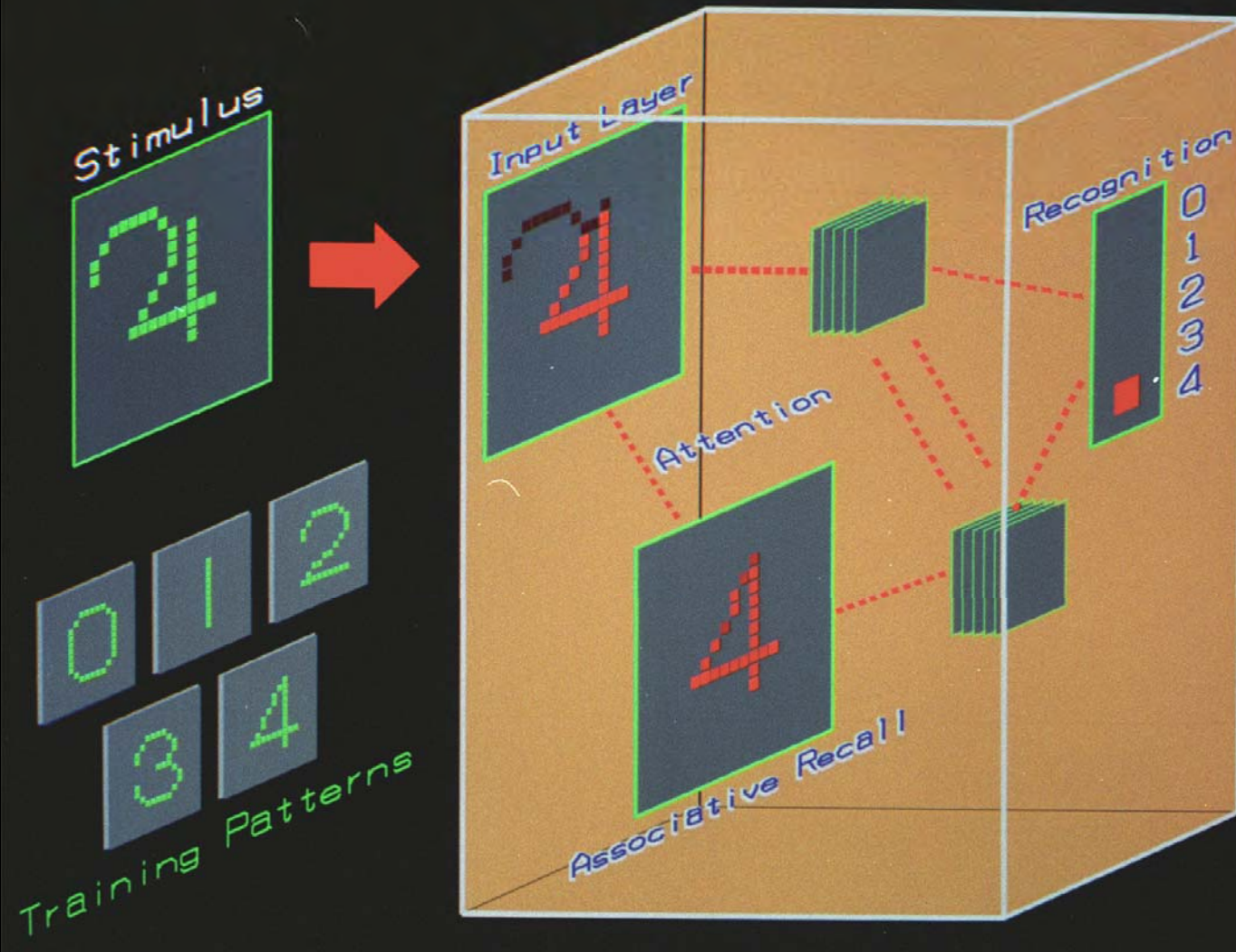


(Fukushima, 1986)



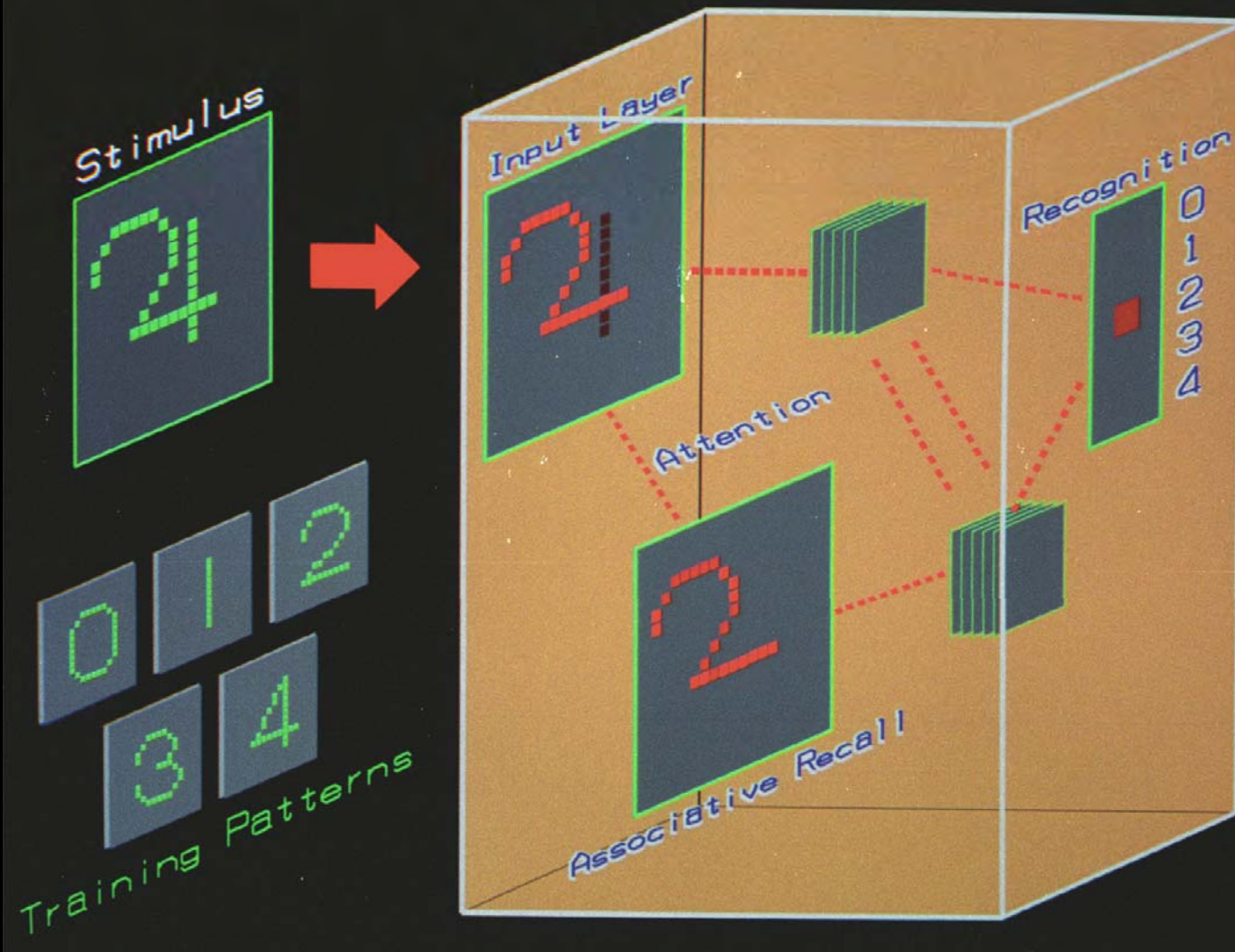
(Fukushima, 1986)

$t = 4$



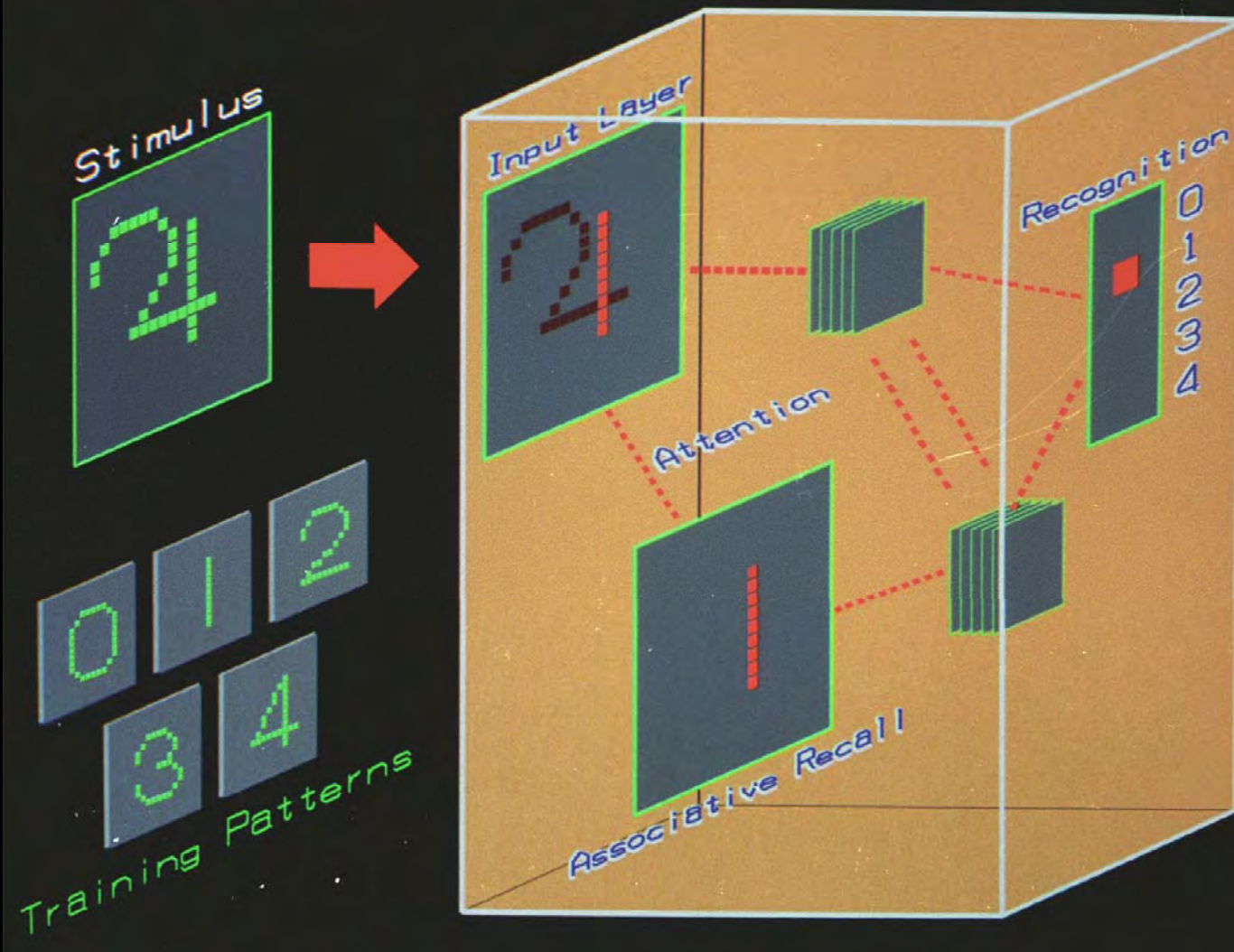
(Fukushima, 1986)

$t = 12$



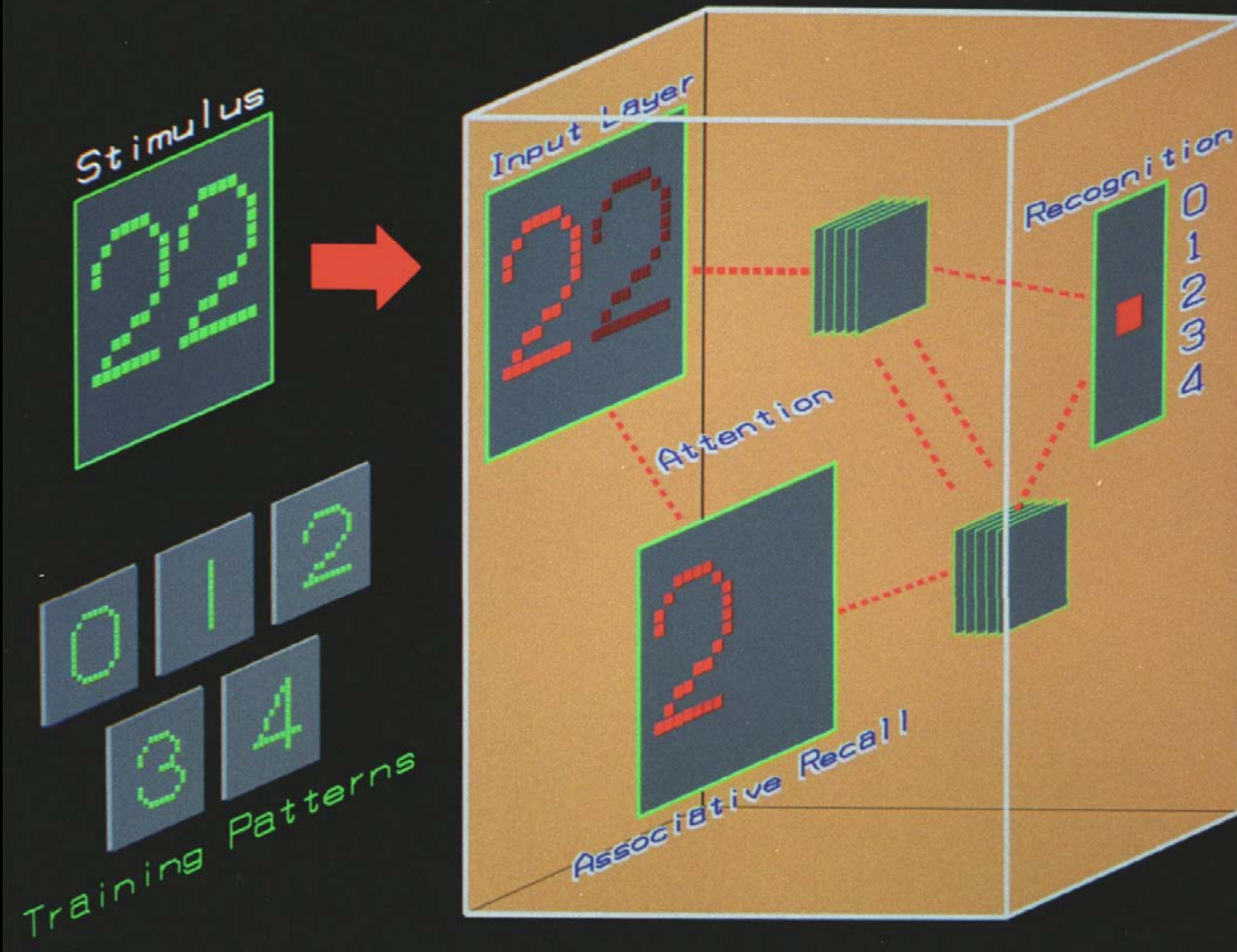
(Fukushima, 1986)

$t = 15$

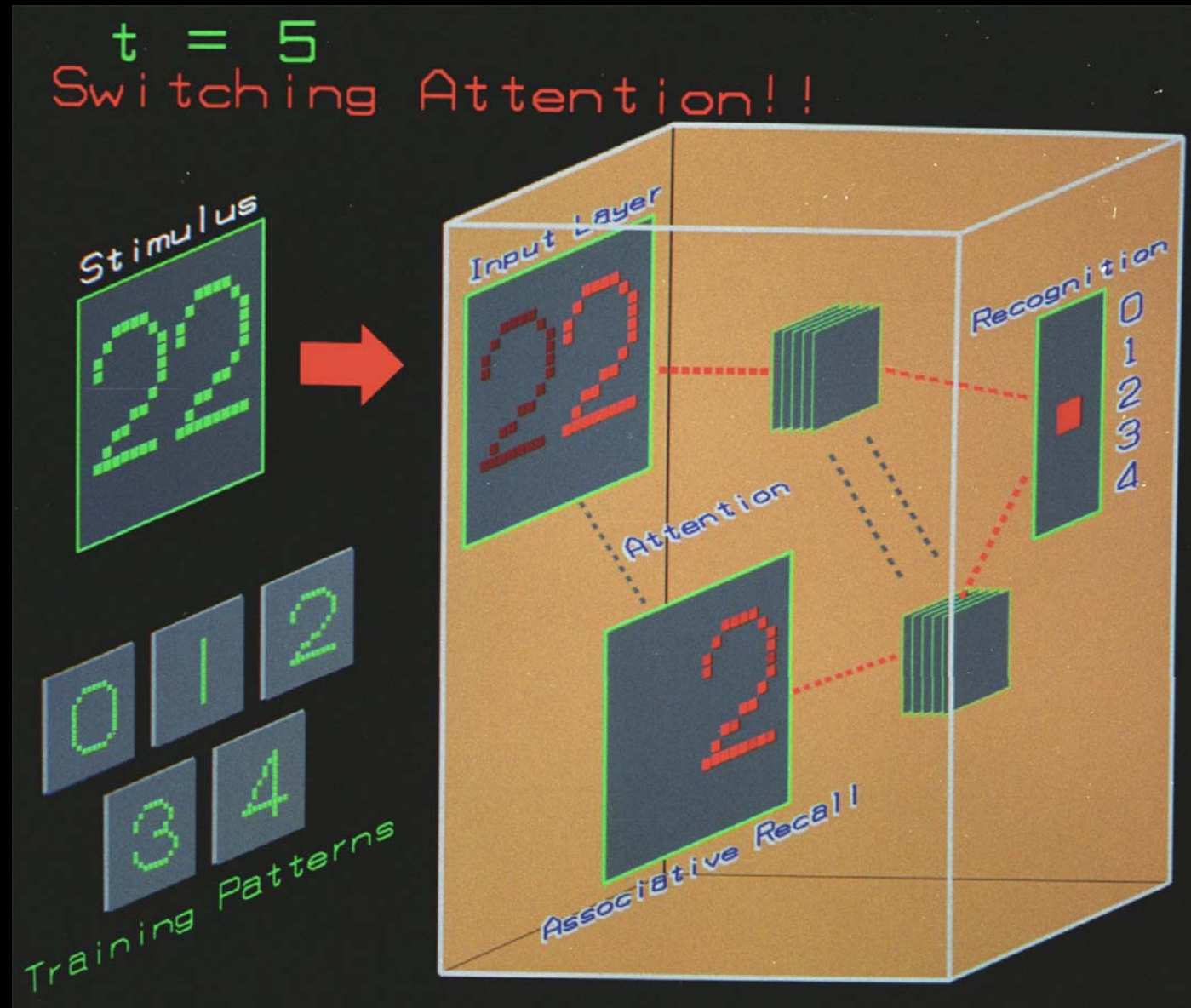


(Fukushima, 1986)

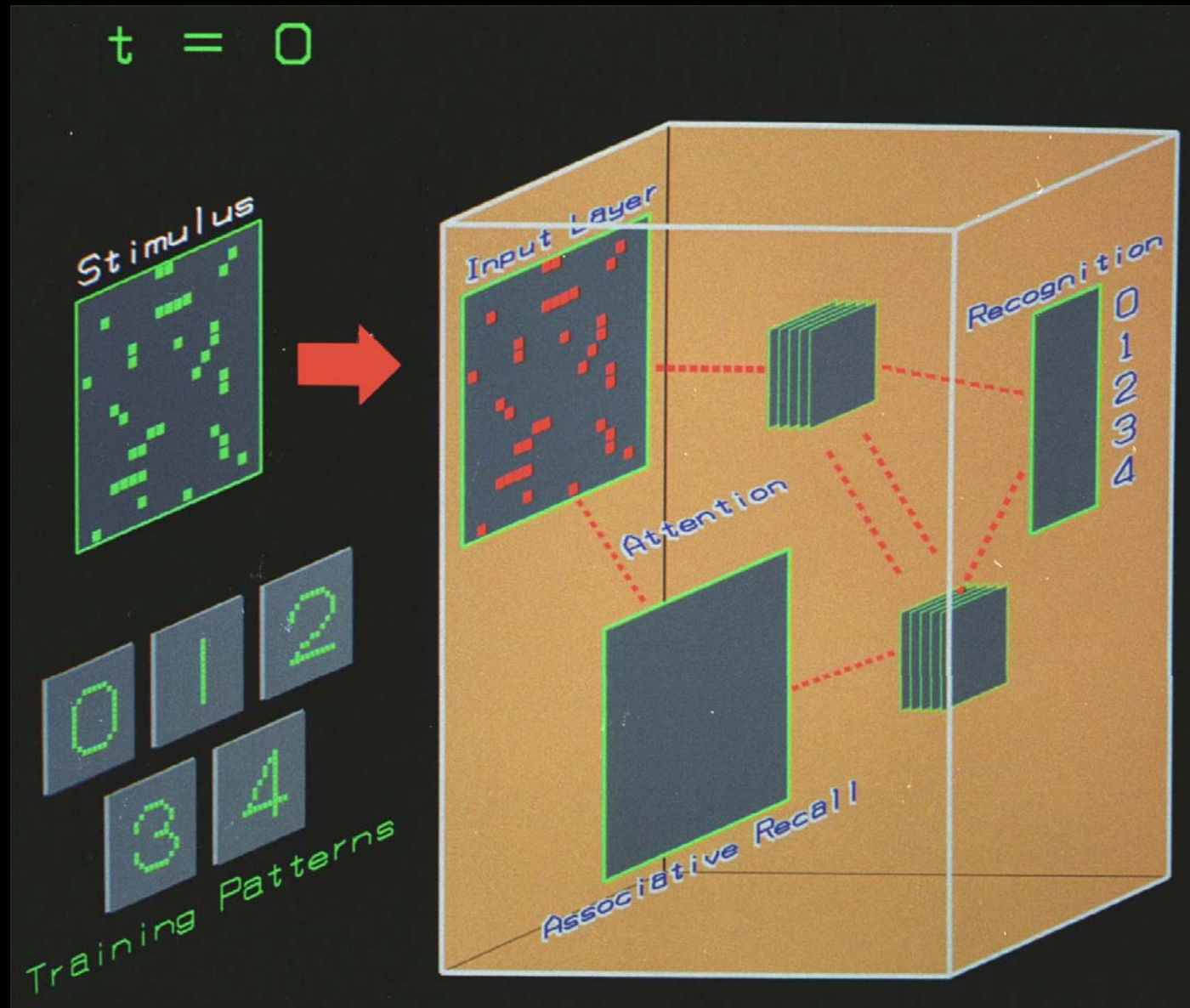
$t = 4$



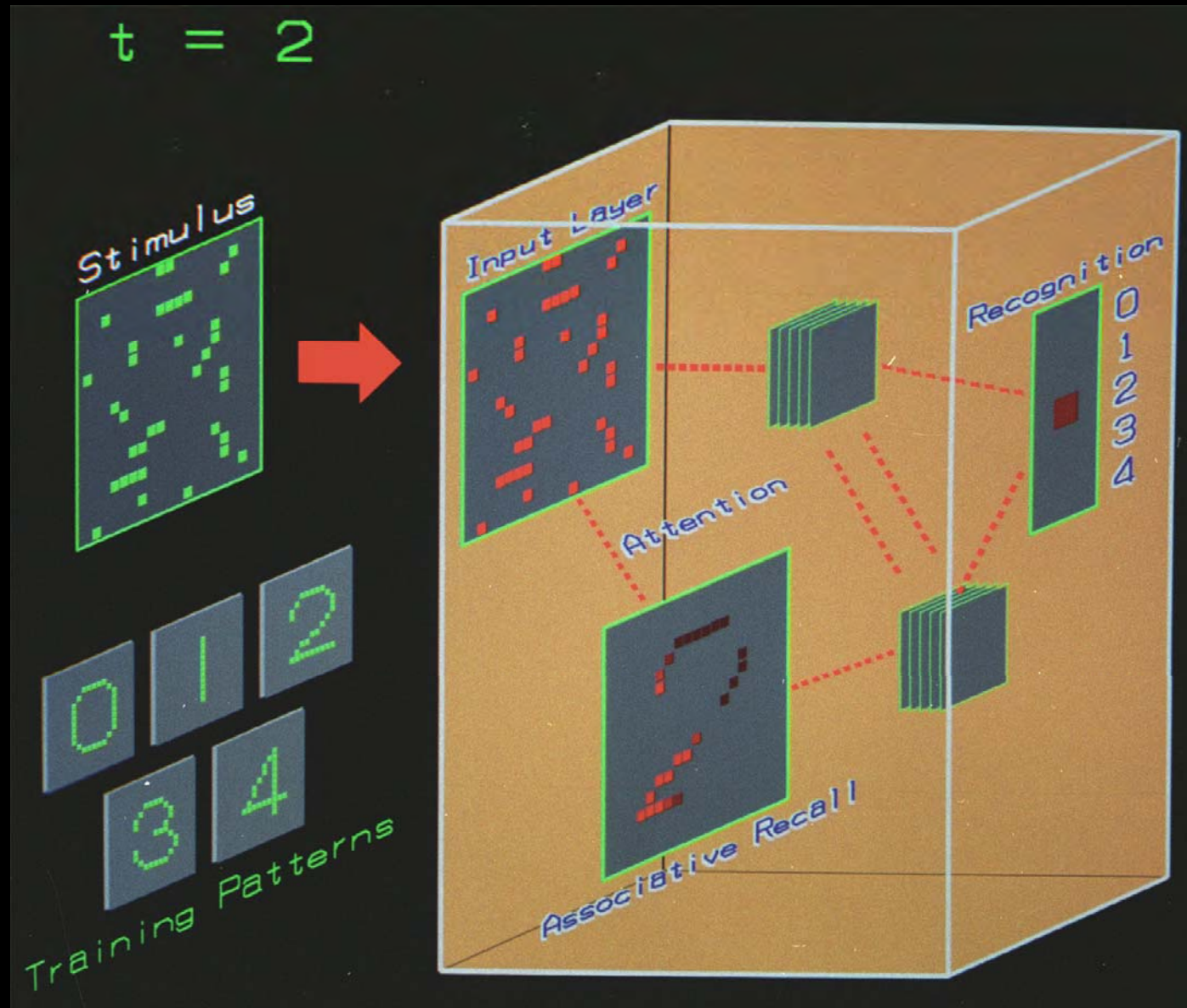
(Fukushima, 1986)



(Fukushima, 1986)

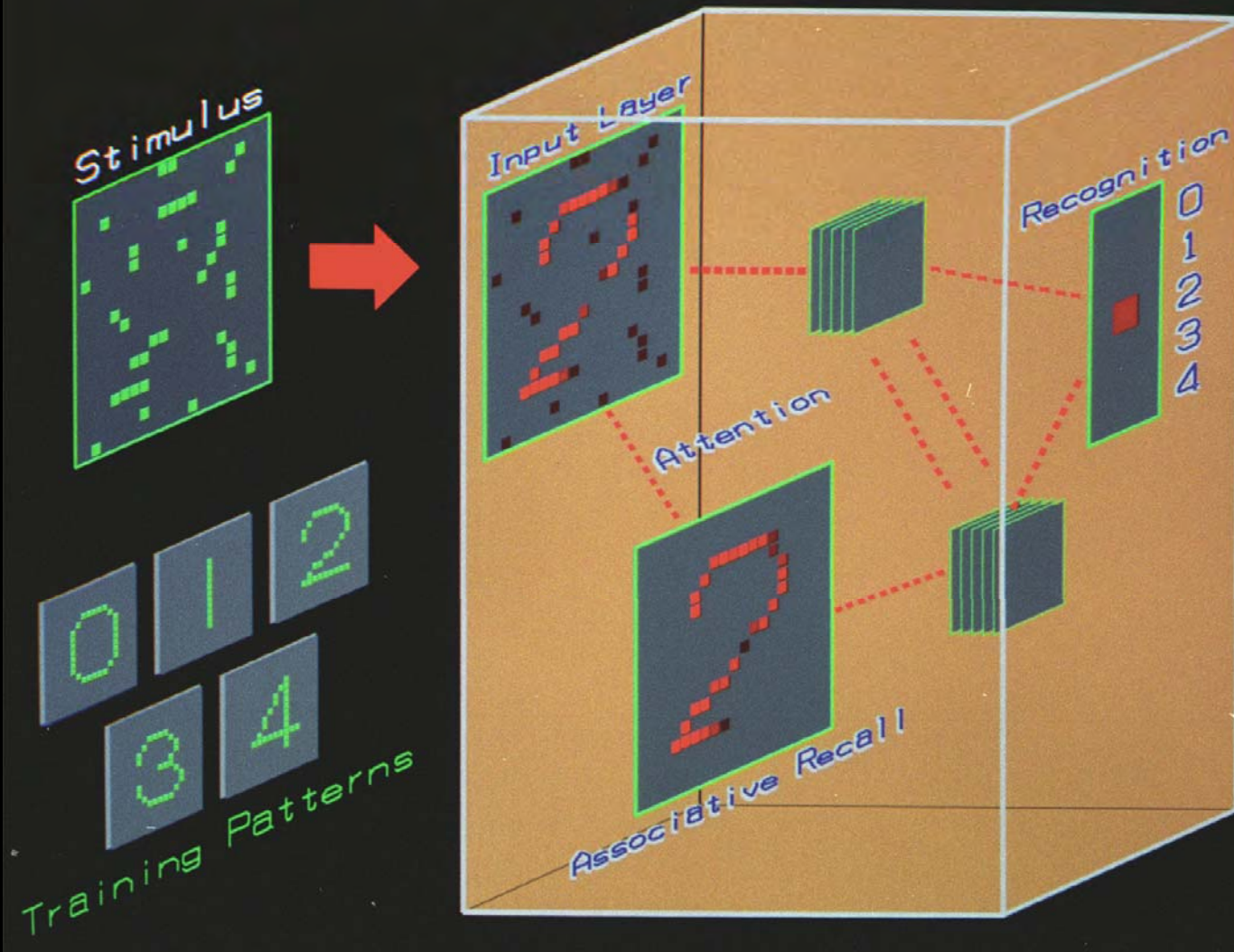


(Fukushima, 1986)



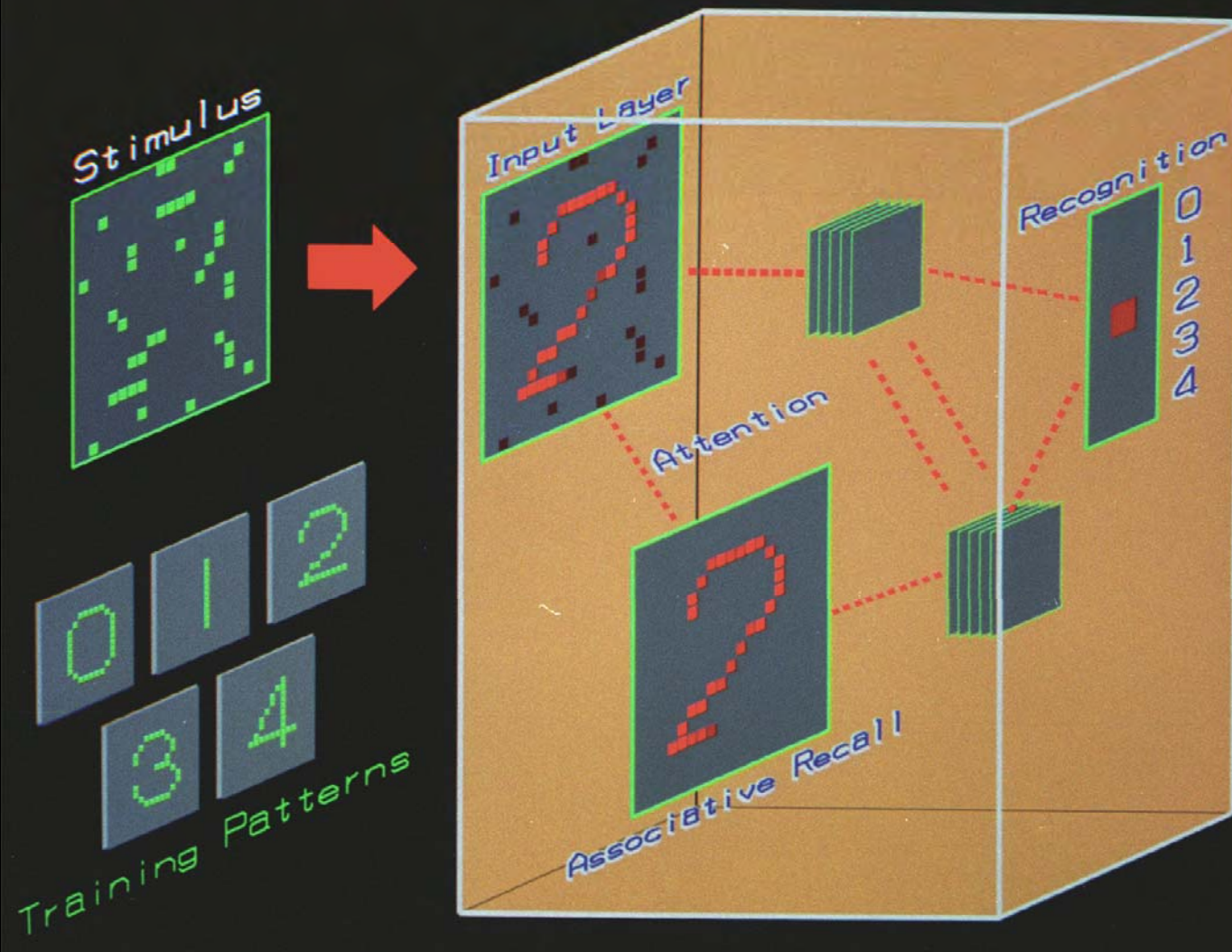
(Fukushima, 1986)

$t = 6$



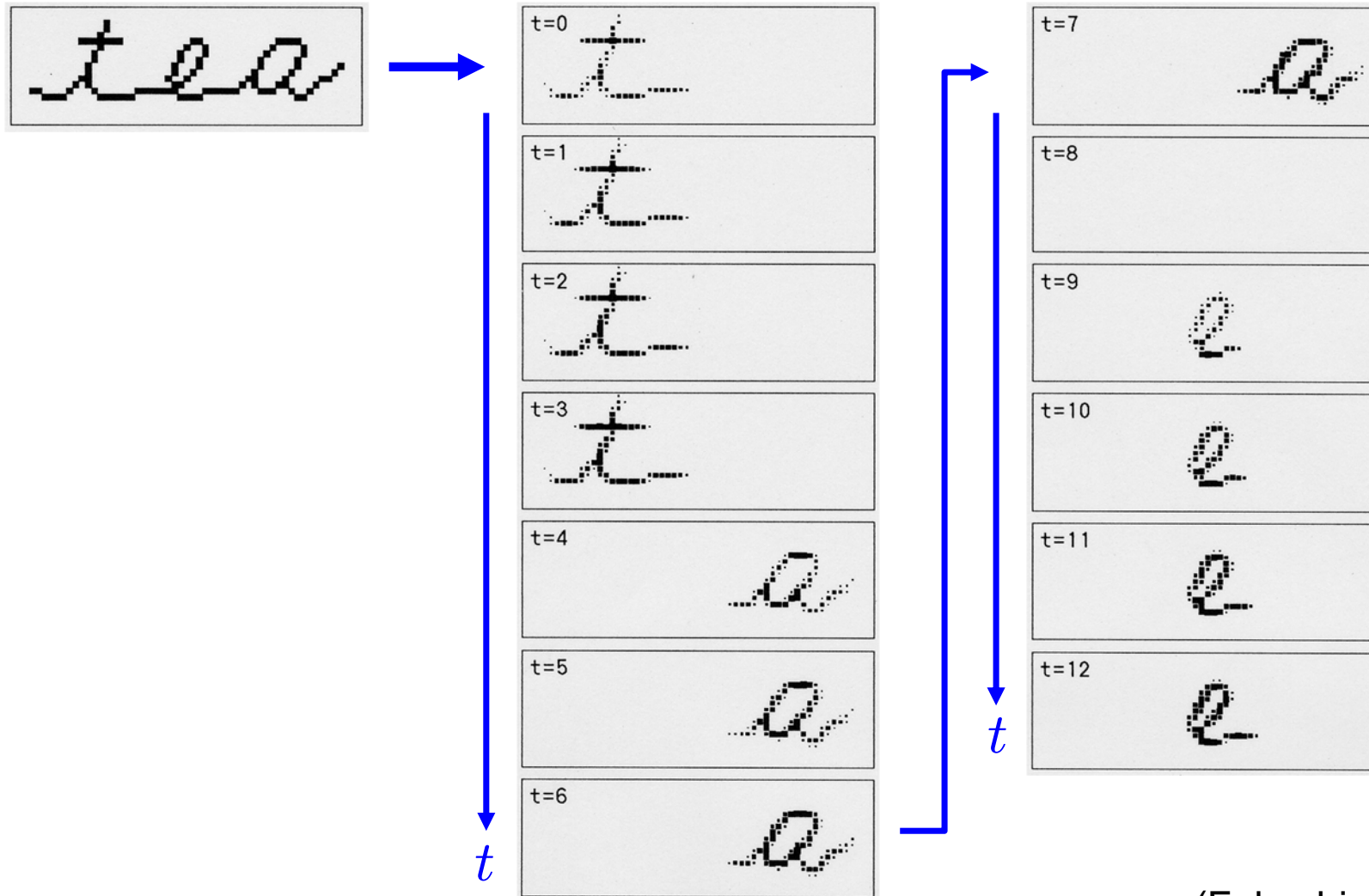
(Fukushima, 1986)

$t = 13$



(Fukushima, 1986)

Segmentation of connected characters



(Fukushima, Imagawa 1991)

xwdump

Character Recognition in Cursive Handwriting.
by T.Imagawa & K.Fukushima, Osaka University

time = 0

Success in Recognition

Uc0



Uc1



Uc2



Uc3



Uc4

☐ a
☐ e
☐ o
☒ t
☐ c



Wc0



Wc1



Wc2



Wc3

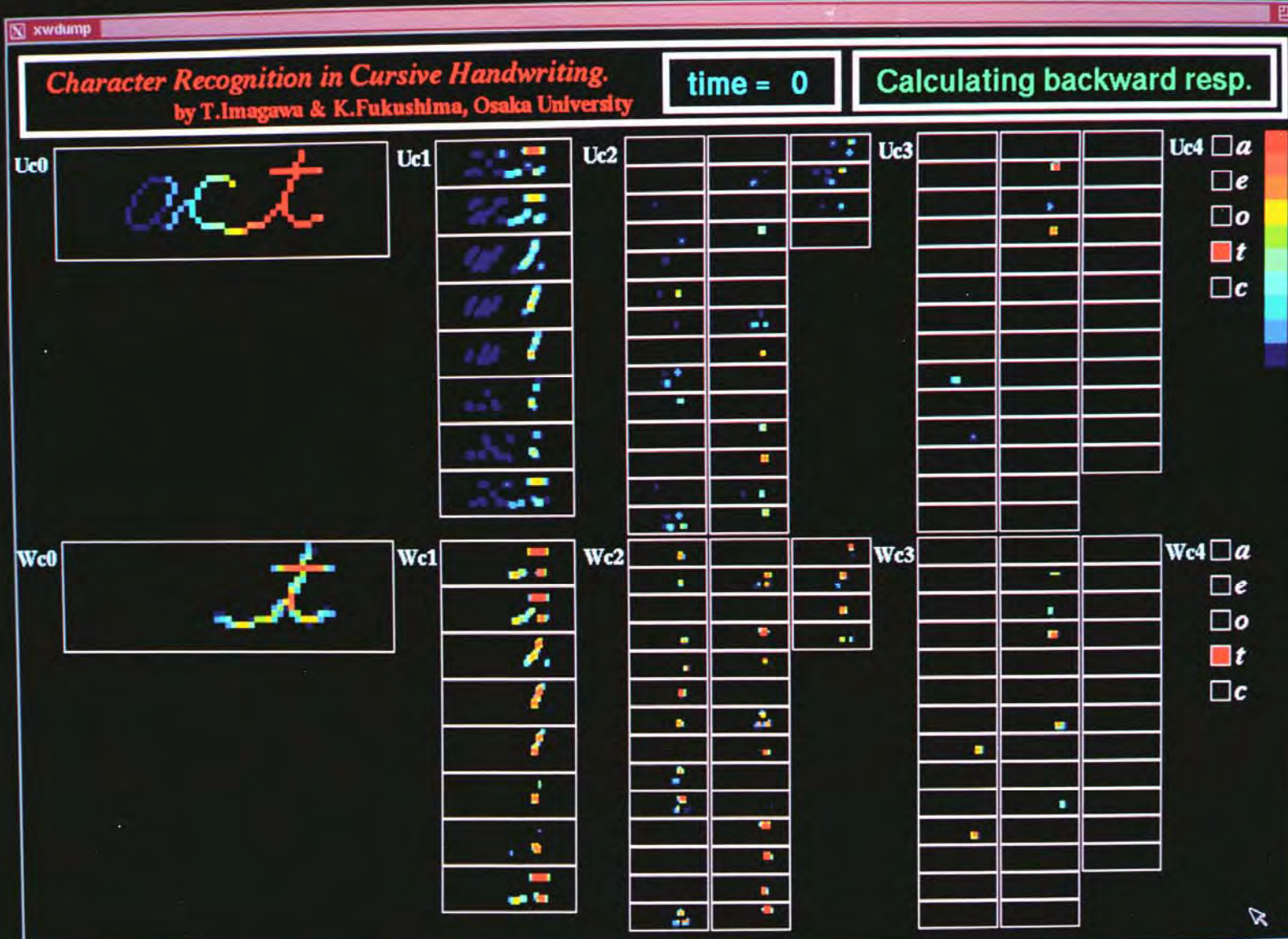


Wc4

☐ a
☐ e
☐ o
☐ t
☐ c









xwdump

Character Recognition in Cursive Handwriting.
by T.Imagawa & K.Fukushima, Osaka University

time = 2

Trying again

Uc0



Uc1



Uc2



Uc3



Uc4

☐ a
☐ e
☐ o
☐ t
☐ c



Wc0



Wc1



Wc2



Wc3



Wc4

☐ a
☐ e
☐ o
☐ t
☐ c

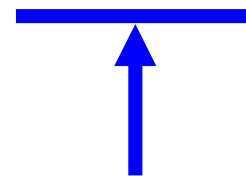
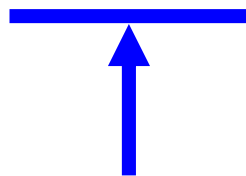








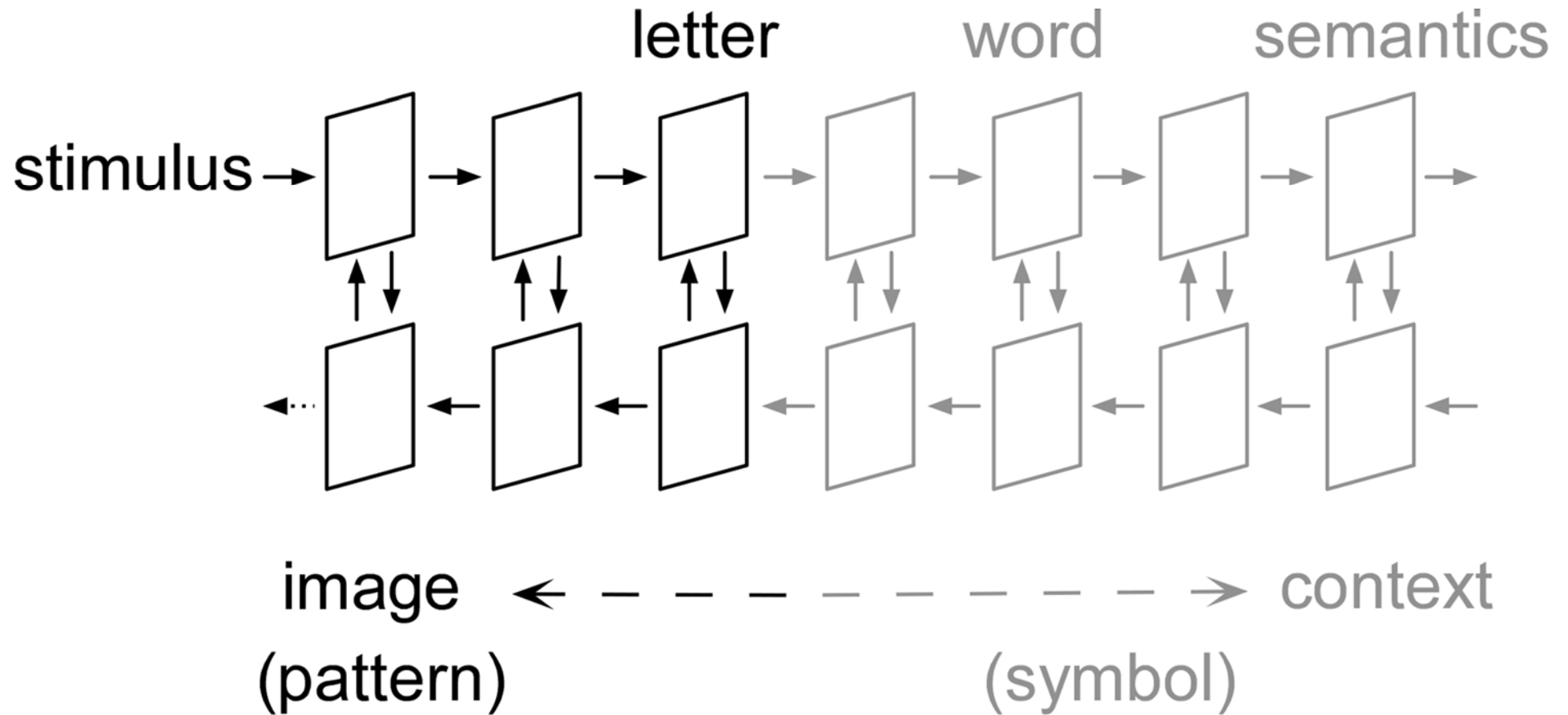
TAE CAT



identical shape

形の情報だけでは,
100%正確なパターン認識は不可能！

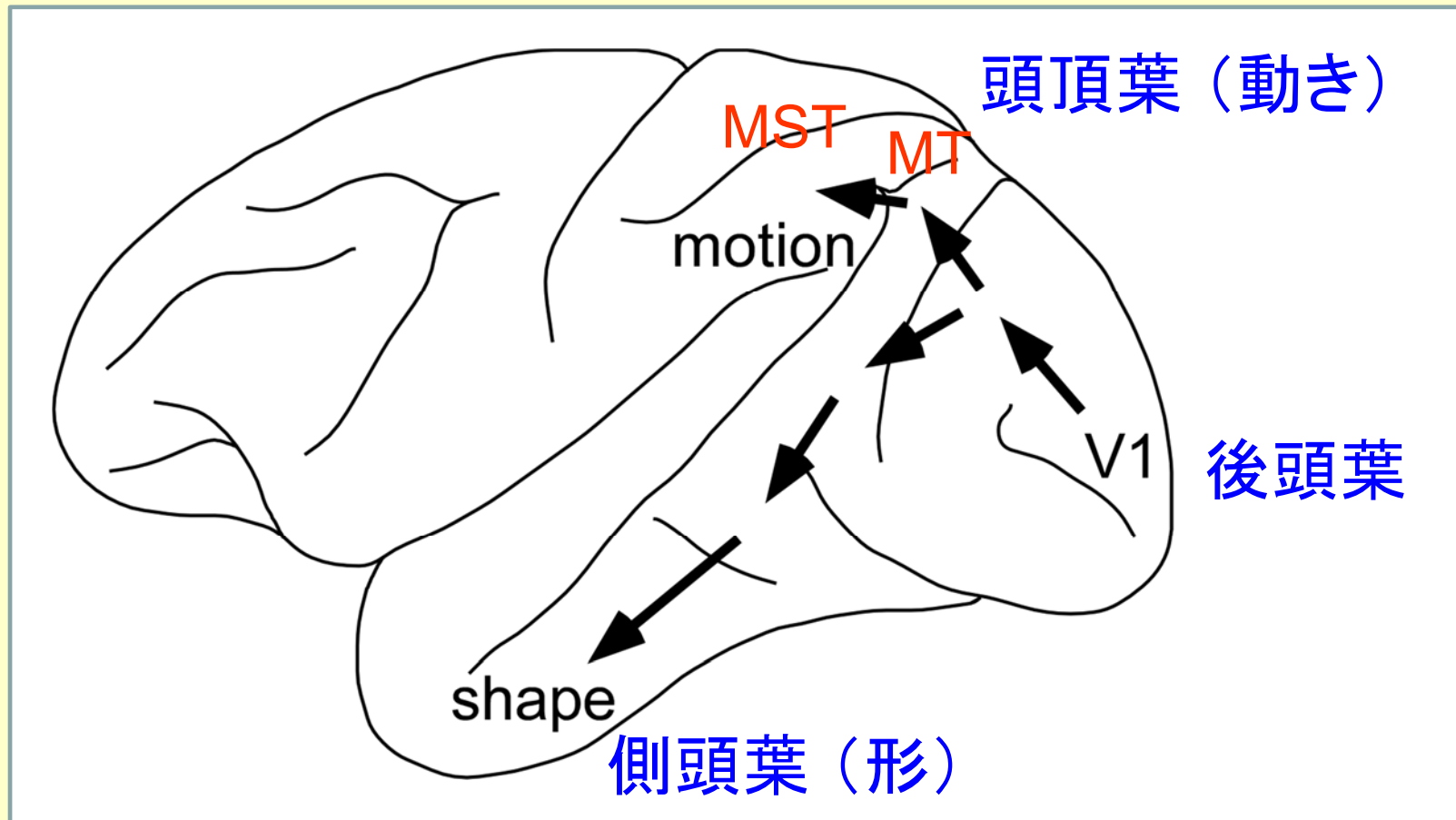
Bottom-up and top-down



Extraction of Optic Flow

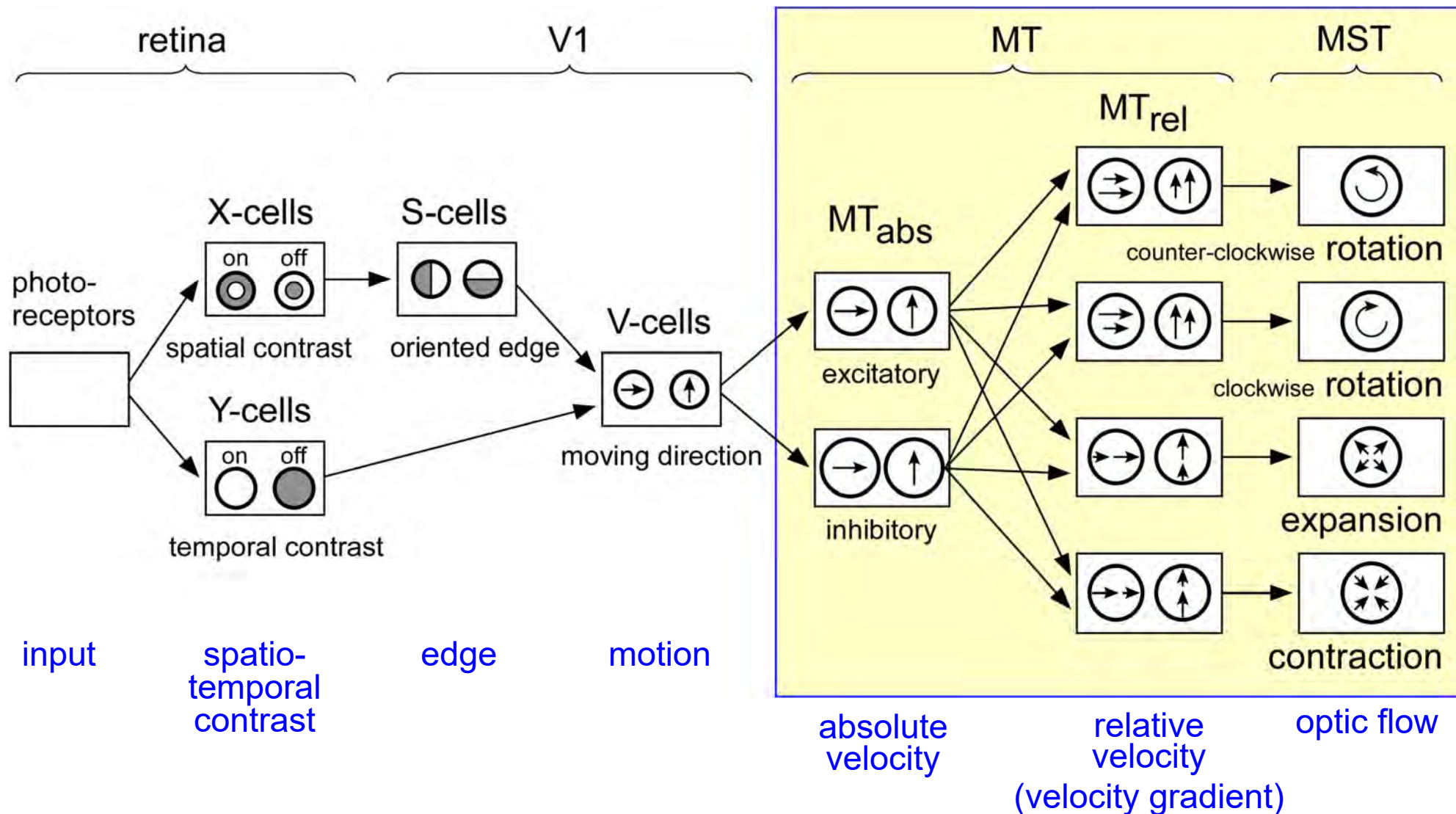
MT and MST

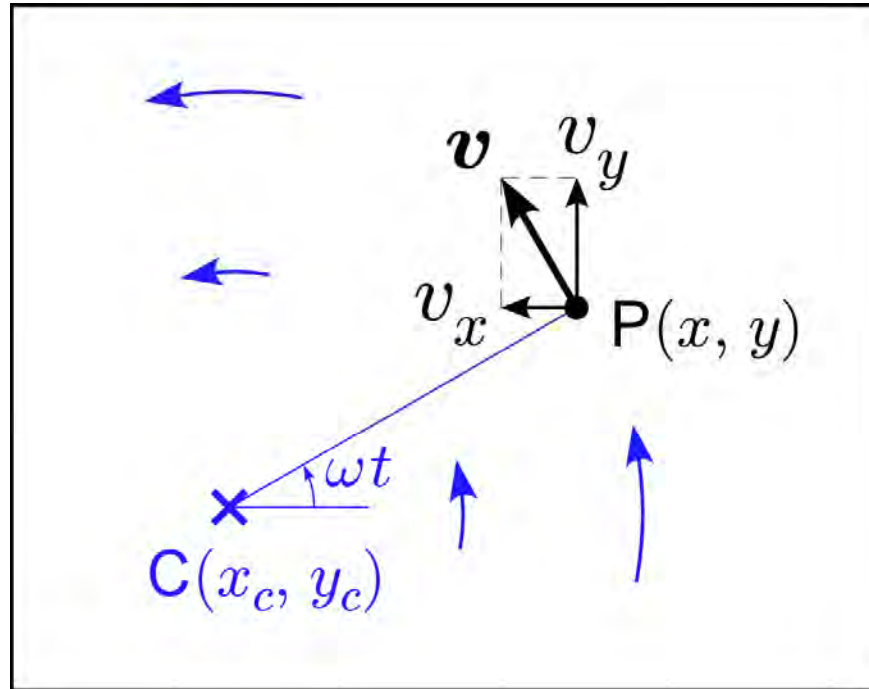
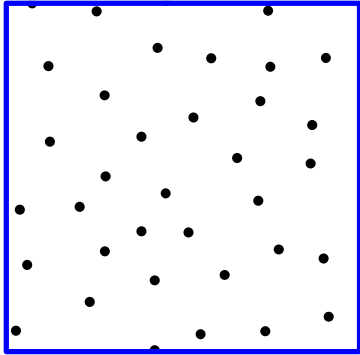
--- rotation, expansion/contraction ---



(Tohyama, Fukushima 2005)

Neural network model for extracting Optic Flow





$$\rho = \frac{\partial v_y}{\partial x} + \left(-\frac{\partial v_x}{\partial y} \right)$$

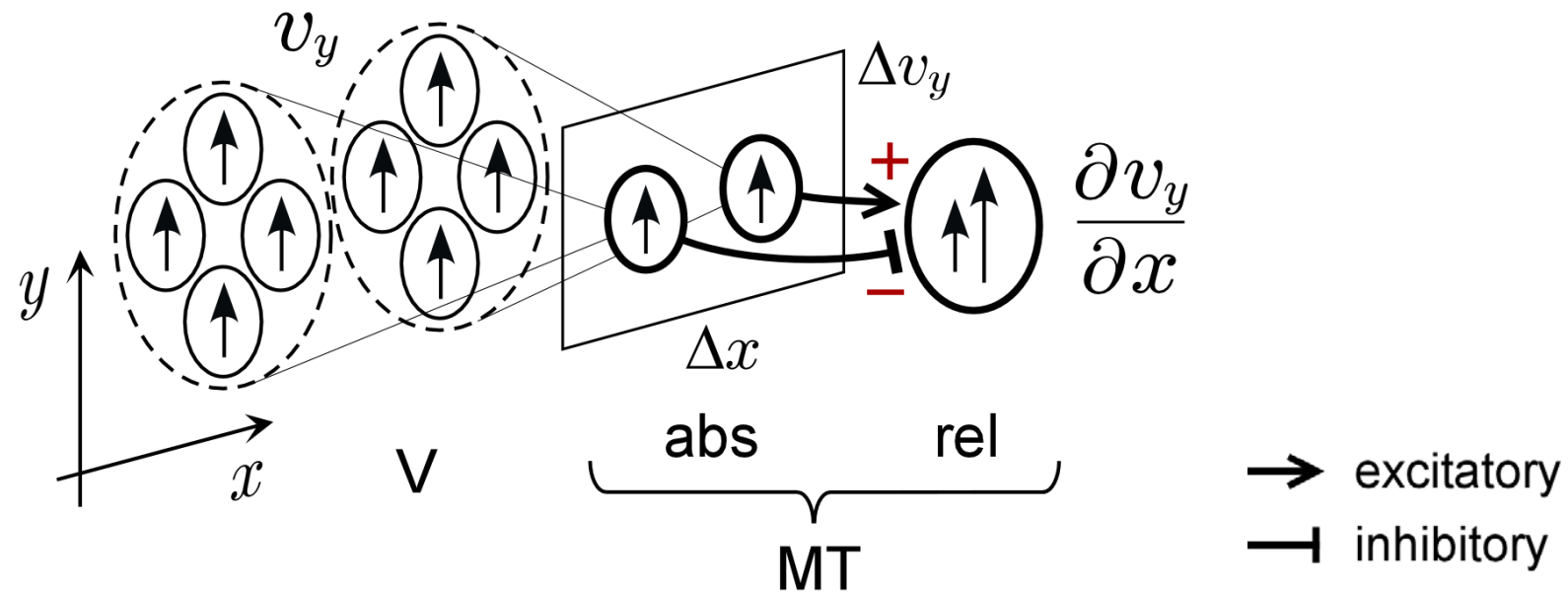
$$= 2\omega$$

$$(\text{rot } \mathbf{v} = \nabla \times \mathbf{v})$$

$$\bar{\rho} = \sum \rho$$

$$\mathbf{v} = (v_x, v_y) = (-\omega(y - y_c), \omega(x - x_c))$$

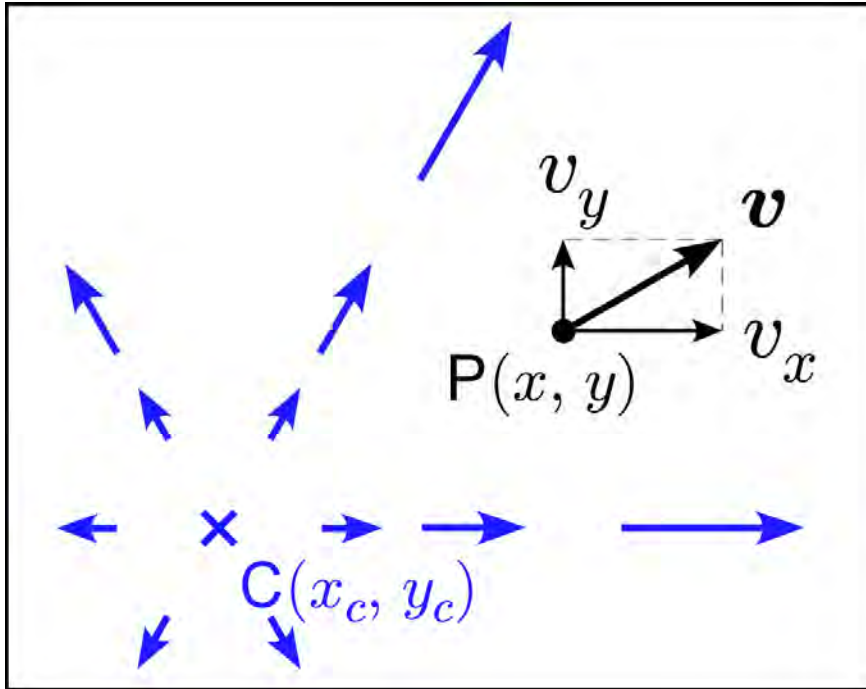
Principles of extracting rotation



rotation: $\rho = \frac{\partial v_y}{\partial x} + \left(-\frac{\partial v_x}{\partial y} \right)$

Connections converging to an MT_{rel} -cell

Extraction of relative velocity (gradient of local velocity)



$$\delta = \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y}$$

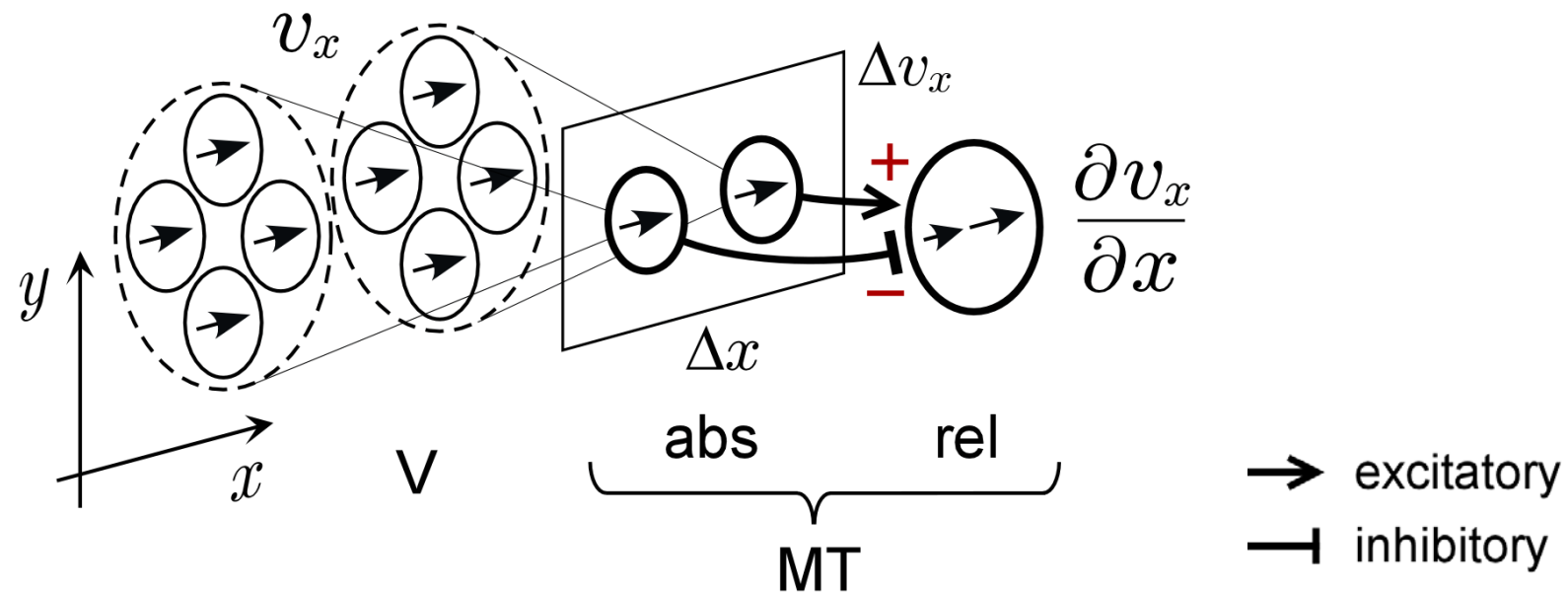
$$= 2\lambda$$

$$(\operatorname{div} \mathbf{v} = \nabla \cdot \mathbf{v})$$

$$\bar{\delta} = \sum \delta$$

$$\mathbf{v} = (v_x, v_y) = (\lambda(x - x_c), \lambda(y - y_c))$$

Principles of extracting expansion/contraction

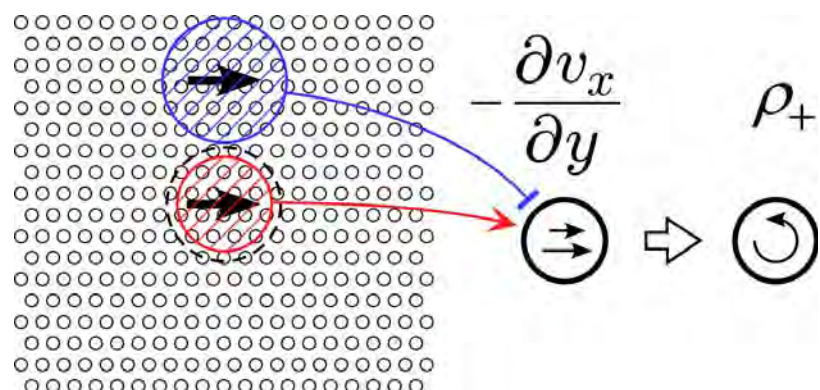


expansion: $\delta = \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y}$

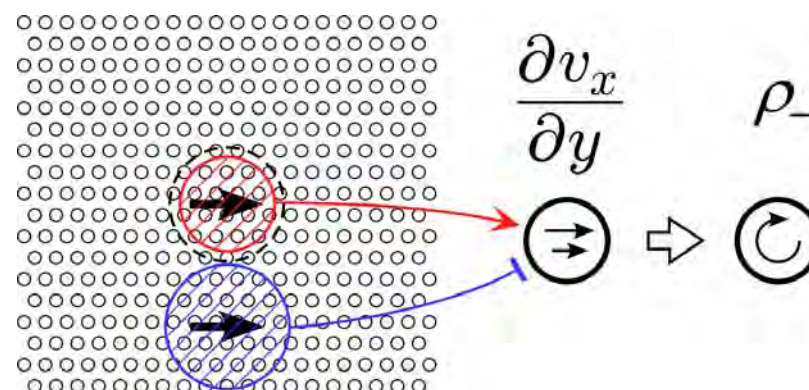
Connections converging to an MT_{rel} -cell

Extraction of relative velocity (gradient of local velocity)

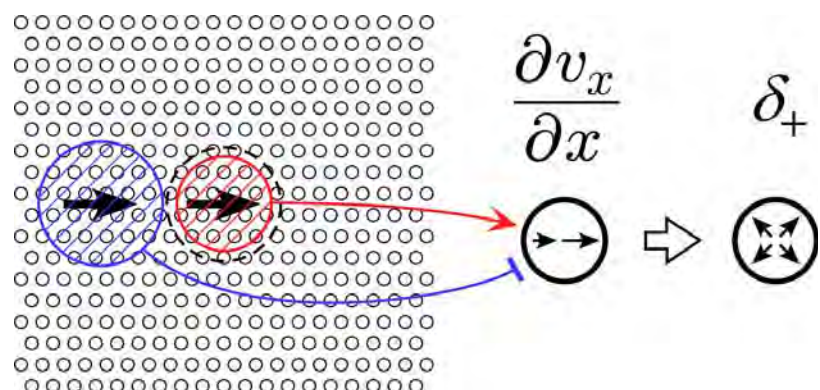
Different types of MST-cells are created simply by the difference in relative location of the inhibitory areas of MT-cells.



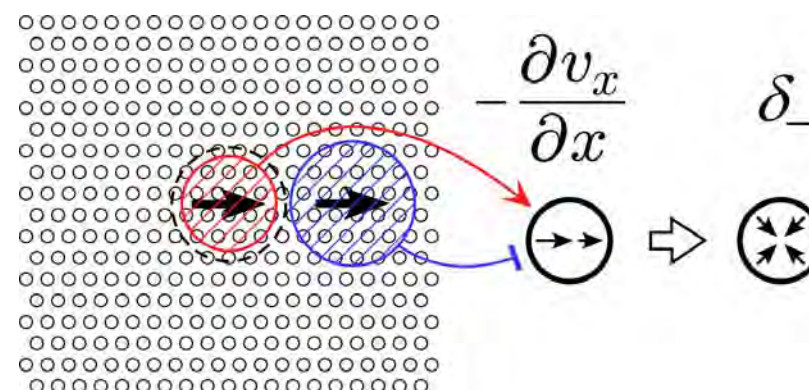
rotation counter-clockwise



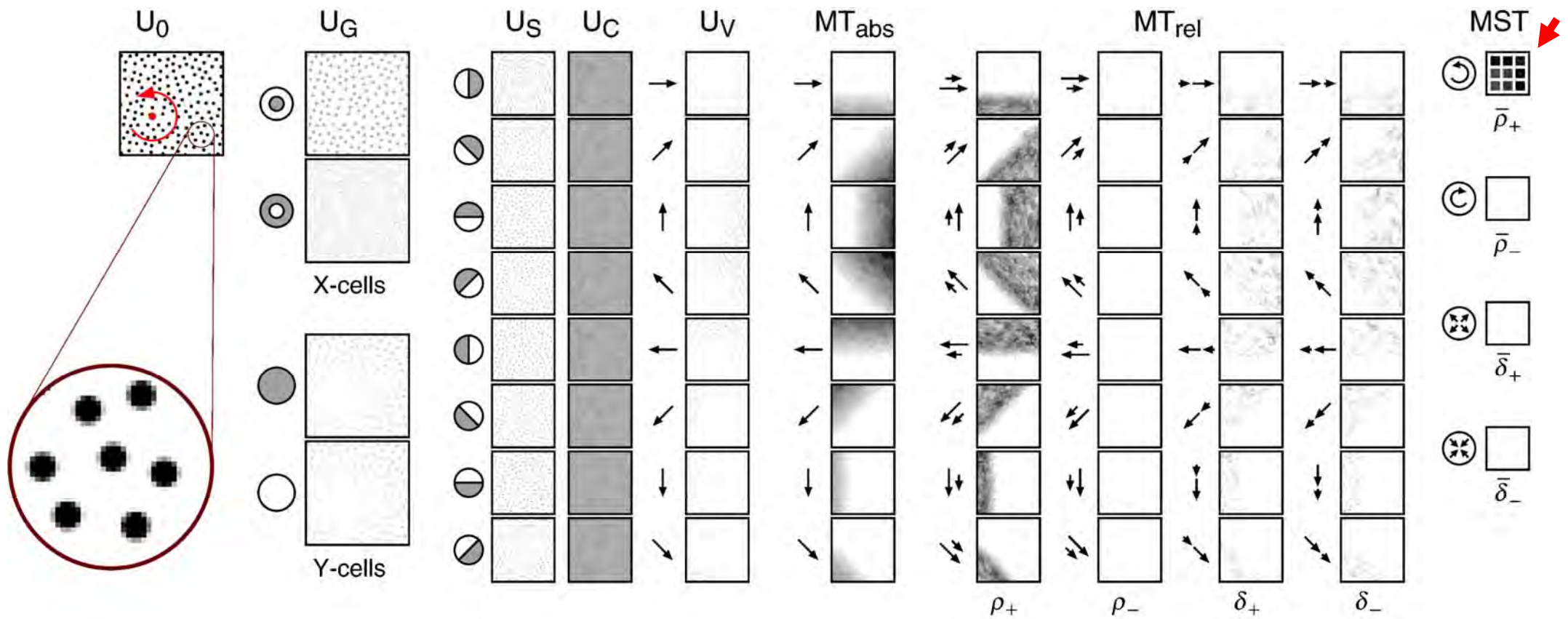
rotation clockwise



expansion

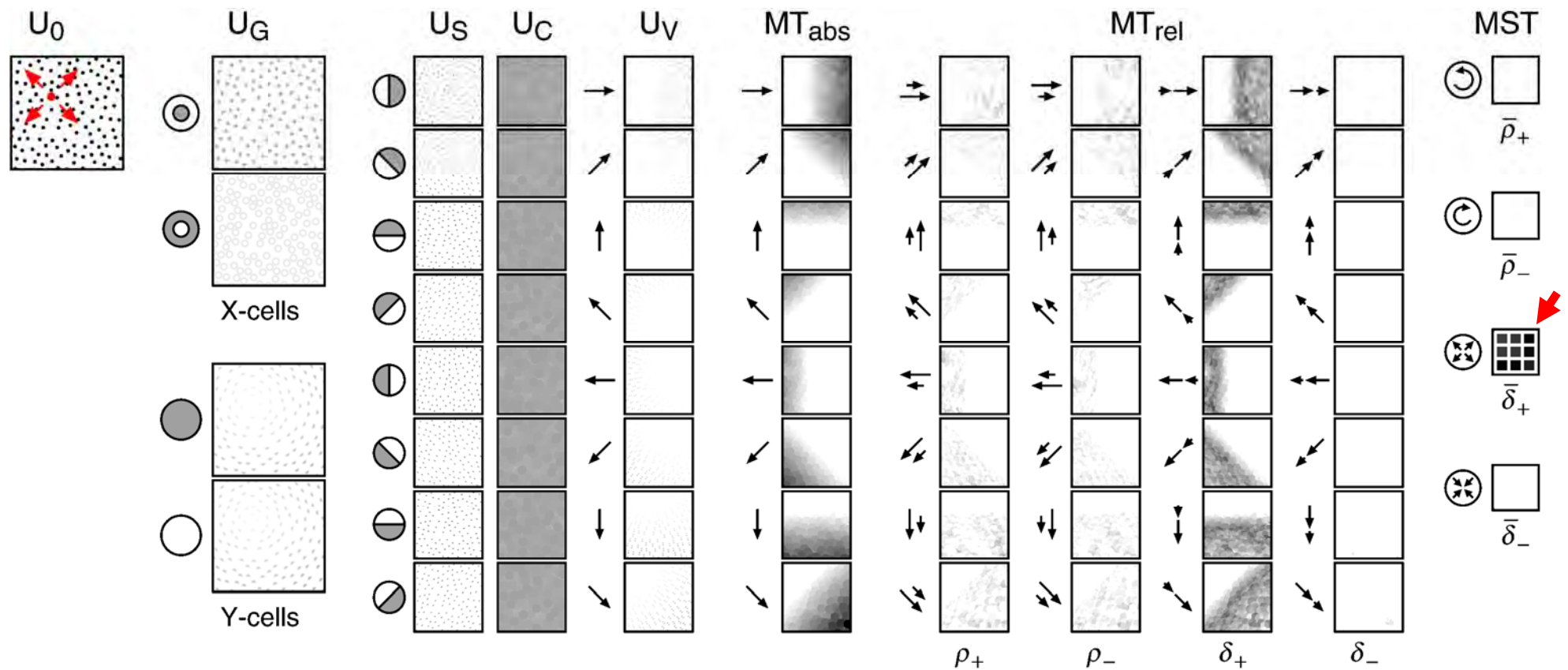


contraction



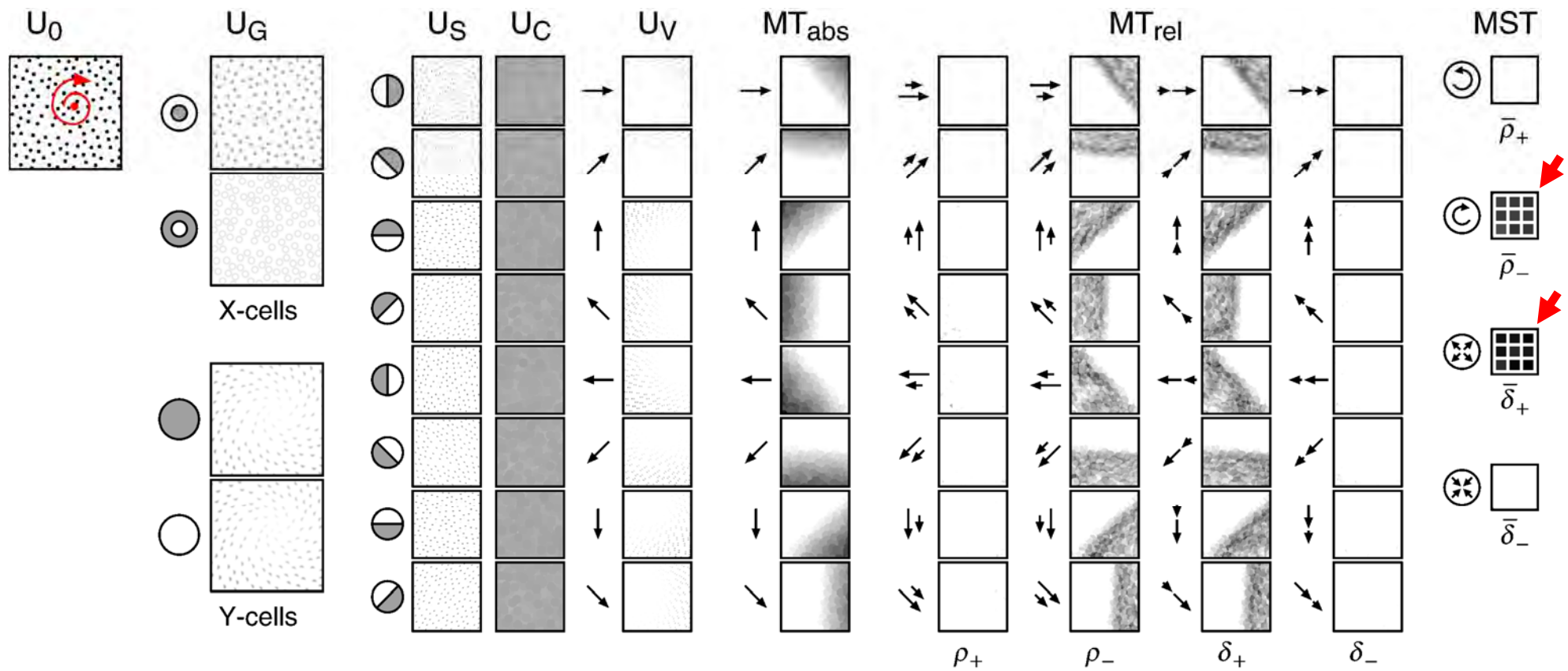
Response of the model
to a random-dot pattern rotating counter-clockwise

(Random dots are made of small black disks.)



Response of the model
to an expanding random-dot pattern

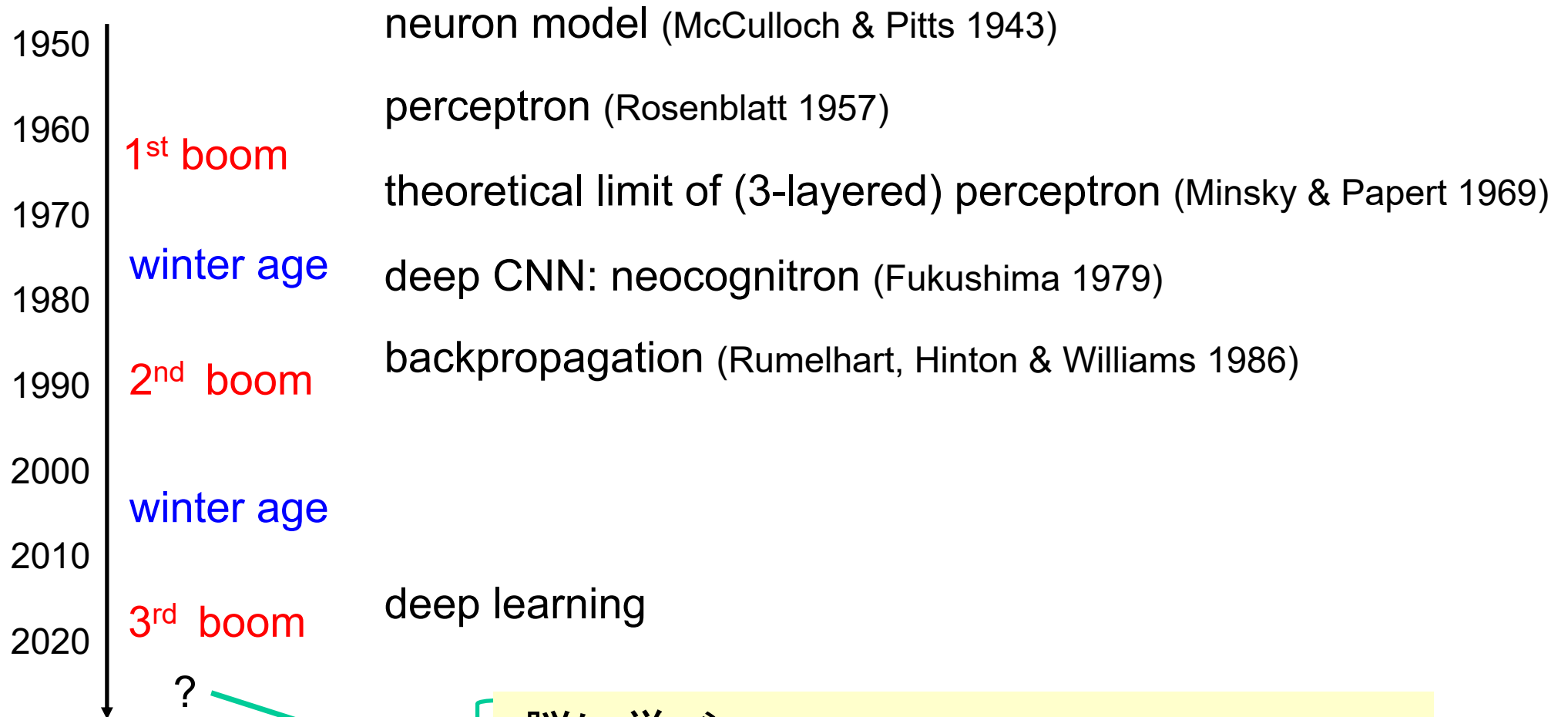
(Random dots are made of small black disks.)



Response of the model
to a spiraling random-dot pattern

(Random dots are made of small black disks.)

History of artificial neural networks



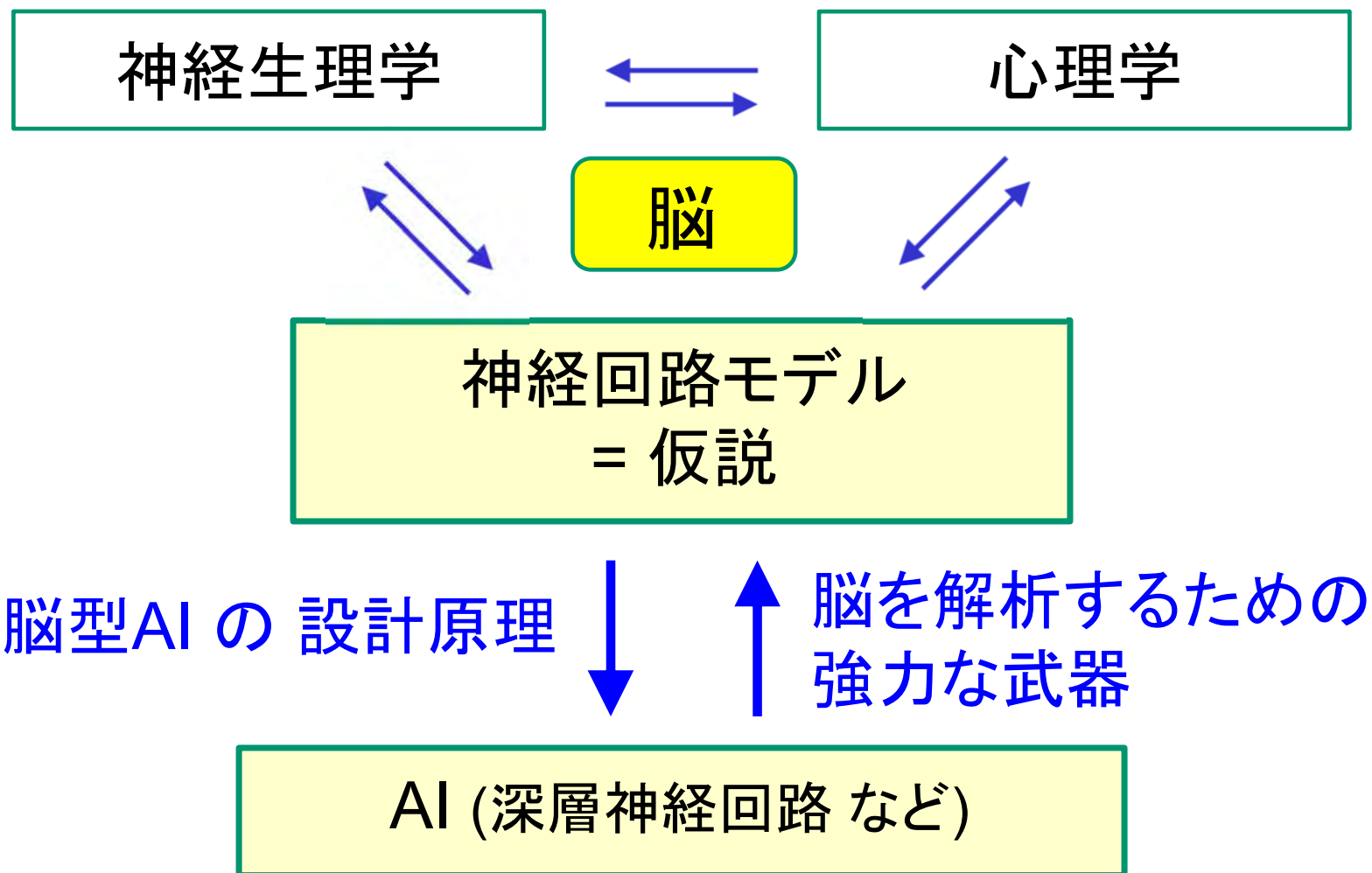
脳に学べ —— 再び冬の時代にならないために ——
脳を制御している基本原理を見つけ出そう

人間： 一を聞いて 十を知る

AI（現在）：億を聞いて 万を知る

AI（将来）：万を聞いて 億を知る

生物の脳に学ぼう



References

K. Fukushima:

“Artificial vision by deep CNN: neocognitron”,
IEEE Trans. on Systems, Man and Cybernetics,
vol. **51**, no. 1, pp. 76-90 (Jan. 2021).

doi: 10.1109/TSMC.2020.3042785

Please visit

<http://personalpage.flsi.or.jp/fukushima/>