

place cells grid cells .

face cells .
.
.

.

invariant repr. complex motion

.

.

?

.

.

.

'Gabor filters'

Biol. Cybernetics 36, 193-202 (1980)

Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position

Kunihiko Fukushima

NHK Broadcasting Science Research Laboratories, Kinuta, Setagaya, Tokyo, Japan Uo ', ~' Usl -----> Ucl t~-~i Us2~ Uc2 ~ Us3----* Uc3 T Science Research Laboratories, Kinuta, Setagaya, Tokyo, Japan

Neocognitron: rationale

Fig. 5. An example of the interconnections between ceils and the response of the cells after completion of self-organization

pulse or instantantan (or instantaneous mean frequency) (or in the New York) and the New York) security and the firing of the actual biological neurons. Neocognitron: activation rule Here, *al(k z_ 1, v, kl)* and *bz(kl)* represent the efficieniciration indic

forced by the same amounts as those for their representative. These relations can be a can be quantity of the can be quantity of the \sim N-laggonitron: Lagro S-cells constitute a column in an IV and the collumn in an IV and the collumn in an IV and the collumn in an IV and T other S-cells in the S-plane, from which the S-plane, from which the repre-Neocognitron: learning rule forced by the same amounts as those for their repre-

 $\int_{0}^{\infty} f(x) dx$ and $\int_{0}^{\infty} f(x) dx$ and be a *non-negative* Let can $u_{\text{SI}}(v_l, n)$ be selected as a representative. \mathcal{L} Let cell $u_{\text{SI}}(k_l, \hat{\mathbf{n}})$ be selected as a representative.

afferent to the S-cells of the S-cells of the $S-$

 $S_{\rm eff}$ is defined here closely resembles that of σ columns defined here closely resembles that of σ

 $S_{\rm 3-1}$ Since $S_{\rm 3-1}$, and $S_{\rm 3-1}$ and $S_{\rm 3-1}$ and $S_{\rm 3-1}$ and $S_{\rm 3-1}$ and $S_{\rm 3-1}$

Aal(kz_ l, v,[q)=ql.cz_ l(v).Ucl_ l(k~_ l,fi + v), (7)

Abt([q) = (qz/2). *Vcl_* l(fi), (8)

forced by the amount shown below:

pressed as follows.

$$
\Delta a_l(k_{l-1},\mathbf{v},\hat{k}_l) = q_l \cdot c_{l-1}(\mathbf{v}) \cdot u_{cl-1}(k_{l-1},\hat{\mathbf{n}}+\mathbf{v}), \quad \overbrace{\mathbf{4 - H} \mathbf{B}} \mathbf{b} \mathbf{b} \mathbf{a} \mathbf{n} \mathbf{learning}
$$

where $\mathbf{r} = \mathbf{r} \cdot \mathbf{r}$ and $\mathbf{r} = \mathbf{r} \cdot \mathbf{r}$ and $\mathbf{r} = \mathbf{r} \cdot \mathbf{r}$ and $\mathbf{r} = \mathbf{r} \cdot \mathbf{r}$ reinforcement. T_{max} to presence, the s cells there is justicely in \mathcal{L}_{max} sentatives. Hence, there is a possibility that a number of candidates appear in a single S-plane. If two or more candidates appear in a single S-plane, only the one \vert which is yielding the largest output among them is \Box such that the such that the S-pairs. In From each S-column, every time when a stimulus pattern is presented, the S-cell which is yielding the largest output is chosen as a candidate for the represelected as the representative from that S-plane. In $\frac{1}{2}$ as $\frac{1}{2}$ comput is chosen as a candidate for the repre- $\frac{1}{2}$ sentatives. Hence, there is a possibility that a number sinatures appear in $\frac{1}{\sqrt{2}}$ is vielding the largest output among them is appearance of the *algorithm* **and the set of the set of**

such that the S-cells show very weak orientation of the S-cells show very weak orientation of the S-cells show

Abt([q) = (qz/2). *Vcl_* l(fi), (8)

\blacksquare S-plane to S-plan

Neocognitron: performance

Fig. 6. Some examples of distorted stimulus patterns which the neocognitron has correctly recognized, and the response of the final layer of the network

in Fig. 6), repeatedly to the input layer U 0. The Fig. 7. A display of an example of the response of all the individual cells in the neocomitron cells in the neocognitron

'AlexNet' (Krizhevsky, Sutskever & Hinton 2012)

This isn't a good model of perception

The invariant representations produced by deep convnets are…

Images are not bags of features (BagNet - Brendel & Bethge 2019)

original

texturised images

Relative spatial relationships are important

Pareidolia

Reference frame effects in perception

Diamond or square?

How to form invariant object representations?

Reference frames require *structured representations*

The meaning of the triangular symbol in fig. 1 is quite complex. It stands for two rules:

1. Multiply the activity level in the retinabased unit by the activity level in the mapping unit and send the product to the object-based unit.

2. Multiply the activity level in the retinabased unit by the activity level in the objectbased unit and send the product to the mapping unit.

Dynamic routing (Olshausen, Anderson, Van Essen 1993)

Dynamic routing circuit

Pattern matching via dynamic routing

Factorization of shape, color and position (Paxon Frady)

 \mathbf{u}^{x_i} = horizontal position x_i

 v^{y_i} = vertical position y_j

 \mathbf{w}_c = color channel c

$$
\mathbf{s} = \sum_{i,j,c} I(x_i, y_j, c) \mathbf{u}^{x_i} \mathbf{v}^{y_j} \mathbf{w}_c
$$

Given **s**, find **x**, **y**, **c** and **p** via resonator:

 $\hat{\mathbf{x}}_{t+1} = g(\mathbf{X} \mathbf{X^\top}^\top \left(\mathbf{s} \otimes \hat{\mathbf{y}}^{-1}_t \otimes \hat{\mathbf{c}}^{-1}_t \otimes \hat{\mathbf{p}}^{-1}_t))$ $\hat{\mathbf{y}}_{t+1} = g(\mathbf{Y}\mathbf{Y}_\top^\top \left(\mathbf{s} \otimes \hat{\mathbf{x}}_t^{-1} \otimes \hat{\mathbf{c}}_t^{-1} \otimes \hat{\mathbf{p}}_t^{-1}) \right)$ $\hat{\mathbf{c}}_{t+1} = g(\mathbf{C} \mathbf{C}^\top \ (\mathbf{s} \otimes \hat{\mathbf{x}}_t^{-1} \otimes \hat{\mathbf{y}}_t^{-1} \otimes \hat{\mathbf{p}}_t^{-1}))$ $\hat{\mathbf{p}}_{t+1} = g(\mathbf{P}\mathbf{P}^{\top}\left(\mathbf{s} \otimes \hat{\mathbf{x}}^{-1}_t \otimes \hat{\mathbf{y}}^{-1}_t \otimes \hat{\mathbf{c}}^{-1}_t)\right)$

horizontal position vertical position color pattern

Visual scene analysis via factorization of HD vectors (Paxon Frady)

Extension to translation, rotation and scaling transformed by arbitrary rigid transforms, composed of translation, rotation, scale, extension to transiation, rotation and so Extension to translation rotation and scaling existence of alternative interpretations.

$$
\mathbf{s} = \sum_i \mathbf{c}_{c_i} \odot \mathbf{h}^{x_i} \odot \mathbf{v}^{y_i} \odot \mathbf{\Lambda}^{-1} (\mathbf{r}^{r_i} \odot \mathbf{m}^{m_i} \odot \mathbf{d}_{p_i})
$$

$$
\mathbf{\hat{l}}(t+1) = \mathbf{\Lambda}^{-1}(\mathbf{\hat{r}}(t) \odot \mathbf{\hat{m}}(t) \odot \mathbf{\hat{d}}(t)),
$$

$$
\mathbf{\hat{p}}(t+1) = \mathbf{\Lambda}(\mathbf{s} \odot \mathbf{\hat{c}}^*(t) \odot \mathbf{\hat{h}}^*(t) \odot \mathbf{\hat{v}}^*(t))
$$

Renner, et al. (2024). Neuromorphic visual scene understanding with resonator networks. *Nature Machine Intelligence*. the network is bisected into the network is bisected into the using Cartesian and one log-polarized into the lo

Factorization of Visual Scenes with Convolutional Sparse Coding and Resonator Networks.

(Kymn, Mazelet, Kleyko & Olshausen. NICE 2024 Proceedings)

Performance

Learning to separate shape and transformations via Lie group operators and sparse coding (Ho Yin Chau, Yubei Chen, Frank Qiu) iniy to beparate briape and transionination parameters, we used a batch size of 100 images, = 0*.*1, = 30, and learning rate ⌘↵ = 0*.*001 for alpha and ⌘ (*T*) = ^p *T* for . Figure 6 shows the 10 learned with our training algorithm, and Figure 7 shows the

> $= e^{i \mathbf{r} \cdot \mathbf{r} \cdot \mathbf{r}} \mathbf{\Phi} \cdot \mathbf{r} + \epsilon$ $\mathbf{I} = \mathbf{T}(s) \, \boldsymbol{\Phi} \, \alpha + \epsilon$ $=e^{\mathbf{A}s}\,\mathbf{\Phi}\,\alpha+\epsilon$

$$
\mathbf{T}(s) = e^{\mathbf{A}s} \n= \mathbf{W} e^{\mathbf{\Sigma}s} \mathbf{W}^T = \mathbf{W} \mathbf{R}(s) \mathbf{W}^T
$$

$$
\mathbf{\Sigma} = \begin{bmatrix} 0 & -\omega_1 \\ \omega_1 & 0 & & \\ & \ddots & & \\ & & \omega_{D/2} & 0 \end{bmatrix} \qquad \mathbf{R}(s) = \begin{bmatrix} \cos(\omega_1 s) & -\sin(\omega_1 s) & \\ \sin(\omega_1 s) & \cos(\omega_1 s) & \\ & \ddots & \\ & & \cos(\omega_{D/2} s) & -\sin(\omega_{D/2} s) \\ & & \sin(\omega_{D/2} s) & \cos(\omega_{D/2} s) \end{bmatrix}
$$

Let *^I* ² ^R*^D* be the input image. We model *^I* as

$$
I = WR(s) WT \Phi \alpha + \epsilon
$$

\n
$$
\Theta \Theta
$$

\n
$$
\Theta \Theta
$$

\n
$$
\Theta \Theta
$$

\n
$$
\Theta \Theta
$$

\n
$$
\Theta \Theta
$$

\n
$$
\Theta \Theta
$$

Results

