

## Deep CNN ネオコグニトロンの学習

## Training Deep CNN Neocognitron

福島 邦彦

Kunihiko Fukushima

ファジィシステム研究所

Fuzzy Logic Systems Institute

The neocognitron is a multi-layered convolutional network that can be trained to recognize visual patterns robustly. This paper discusses the recent neocognitron focusing on differences from the conventional deep CNN (convolutional neural network) and deep learning. For training intermediate layers of the neocognitron, the learning rule called AiS (Add-if-Silent) is used. Under the AiS rule, a new cell is generated if all postsynaptic cells are silent in spite of non-silent presynaptic cells. The generated cell learns the activity of the presynaptic cells in one-shot. Once a cell is generated, its input connections do not change any more. Thus the training process is very simple and does not require time-consuming repetitive calculation. In the deepest layer, a method called Interpolating-Vector is used for classifying input patterns based on the features extracted in the intermediate layers.

## 1. はじめに

ネオコグニトロンは、階層型多層神経回路の一種で、文字認識をはじめとする視覚パターン認識に高い能力を発揮する [Fukushima 80][Fukushima 13]. つまり、入力パターンが変形したり、大きさが変わったり、あるいは位置がずれたりしても、あまり影響を受けずに正しくパターンを認識する. そのような頑強なパターン認識能力を学習によって獲得していく.

ネオコグニトロンの歴史は古いが、現在も種々の改良が加えられ発展を続けている. ここでは最近のネオコグニトロンを、通常の deep CNN (convolutional neural network) や deep learning との違いに重点を置いて紹介する.

## 2. ネオコグニトロンの概要

## 2.1 ネオコグニトロンの回路構造

ネオコグニトロンの回路構造は、第一次視覚野に関する神経生理学的な知見、すなわち Hubel-Wiesel の古典的な階層仮説にヒントを得て考案された. 図 1 に示すように、特徴抽出を行なう S 細胞の層と、位置ずれを許容する働きを持つ C 細胞の層 (pooling 層) とを交互に階層的に配置した多層神経回路である. S 細胞の層と C 細胞の層とが組になったモジュールが、何段も従属的に接続されている. ここで、第  $l$  段目の S 細胞の層を  $U_{S_l}$ , C 細胞の層を  $U_{C_l}$  と表示することにする.

各層は、複数の sub-layer (細胞面) から構成されている. 図で、太線で囲んだ各四角形が細胞面である. 細胞面内には、細胞が retinotopy を保って並んでおり、同一の入力結合を共有している (shared connection). その結果、一つの細胞面に含まれるすべての細胞は、同じ性質の受容野を持つが、受容野の位置は、細胞ごとに異なる.

S 細胞は学習によって変化する可変入力結合を持つ. 学習を終わると、S 細胞は特徴抽出細胞として働くようになり、入力層に与えられる刺激パターンの局所的な特徴を抽出する. 細胞面内のすべての S 細胞が結合を共有しているため、各細胞面は convolution の演算を行なっていることになる. あるいは、

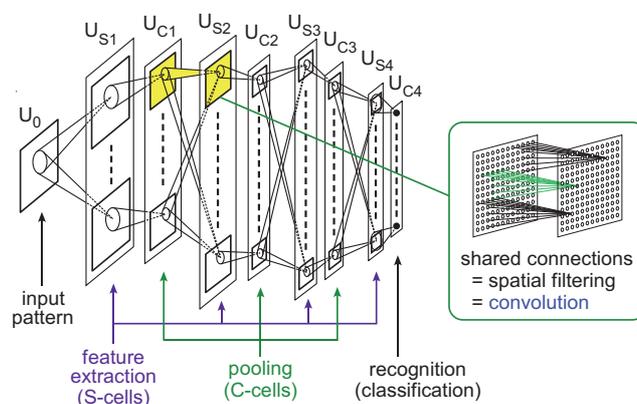


図 1: ネオコグニトロンの回路構造.

前層の出力に空間フィルター (spatial filter) を掛けてしていると解釈することもできる.

C 細胞は固定した入力結合を持ち、受容野の位置が僅かに異なる S 細胞層の反応を集めてくる (二乗平均をとる) ことによって、刺激の呈示位置ずれにあまり影響されない反応を示す. C 細胞は、最近の deep CNN とは異なり、max-pooling ではなく平均化操作によって pooling を行なっている.

C 細胞によるこの操作は、S 細胞層の反応を、それに続く C 細胞層で空間的にぼかしていると解釈することも出来る. ネオコグニトロンの各段では、演算コストを減らすために細胞の間引き (down-sampling) を行なっているが、ぼかし操作は空間周波数の高周波成分を除去する low-pass filter の役割も果たし、粗い down-sampling によって発生する aliasing noise の抑制にも貢献している. 更にまた、平均化操作によって加算的なランダムノイズを軽減する働きも生じる.

入力層に与えられた視覚パターンの情報は、多層回路の中を伝わる過程で特徴の抽出と統合が繰り返され、次第に高次の特徴が抽出されていく. 最上位層 (最も深い層) では、中間層で抽出された特徴をもとに、入力パターンを識別 (認識) する.

このようなネオコグニトロンの回路は、学習によって自己組織的に形成されていく. 学習終了後は、ネオコグニトロンは、

連絡先: email: fukushima@m.ieice.org,

URL: <http://personalpage.fsi.or.jp/fukushima/>

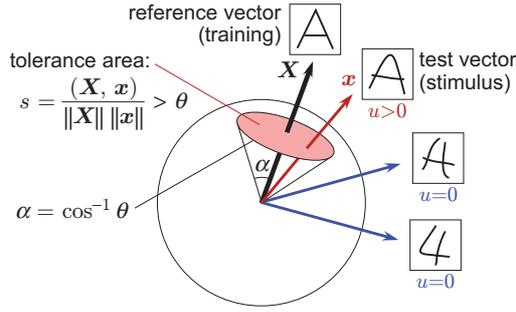


図 2: S 細胞の反応の, 特徴空間内の許容域 [Fukushima 13].

入力パターン変形, サイズの変化, 位置ずれなどの影響をあまり受けずに, 頑強にパターンを認識できるようになる.

## 2.2 S 細胞による特徴抽出

S 細胞は, 学習終了後は, 特徴抽出細胞として働く.  $U_{SI}$  層の各 S 細胞は, その前段  $U_{CI-1}$  の複数個の C 細胞から直接, 興奮性結合を受け取っている. その  $n$  番目の C 細胞の出力を  $x_n$  とし, その細胞からの結合の強度を  $a_n$  とする. ここでベクトル表示を用いて, これらを  $\mathbf{x}$ ,  $\mathbf{a}$  と表わすことにする. この S 細胞はまた, V 細胞と名付けられた細胞を介して強度  $-\theta$  ( $0 \leq \theta < 1$ ) の抑制性結合を受け取っている. V 細胞は, S 細胞が信号を受け取っているのと同じ C 細胞から, 興奮性固定結合を受け取り, これらの C 細胞の出力のノルム (L2-norm), すなわち二乗平均値を計算している. すなわち, V 細胞の出力は,  $v = \|\mathbf{x}\|$  で与えられる.

これらの入力を受けて S 細胞の出力は,

$$u = \varphi \left[ \sum_n a_n x_n - \theta v \right] / (1 - \theta) \quad (1)$$

となる. ここに,  $\varphi[\ ]$  は,  $\varphi[x] = \max(x, 0)$  で定義される rectified linear (半波整流型) 関数である.

興奮性入力結合  $\mathbf{a}$  は,  $\mathbf{a} = \mathbf{X} / \|\mathbf{X}\|$  で与えられる. ここに  $\mathbf{X}$  は, S 細胞が学習した学習ベクトル (あるいは学習ベクトルの線型和) である. 従って, 式 (1) は次のように表わすことができる.

$$u = \|\mathbf{x}\| \cdot \frac{\varphi[s - \theta]}{1 - \theta} \quad \text{where } s = \frac{(\mathbf{X}, \mathbf{x})}{\|\mathbf{X}\| \cdot \|\mathbf{x}\|} \quad (2)$$

多次元特徴空間内で  $s$  は,  $\mathbf{X}$  と  $\mathbf{x}$  との内積 (ノルムで正規化された内積) で定義される類似度を表わしている. 類似度  $s$  が  $\theta$  よりも大きければ, S 細胞は非負の出力を出す [Fukushima 13]. このように,  $\theta$  は S 細胞の閾値を決定する. 多次元空間内で  $s > \theta$  を満足する領域を, この S 細胞の許容領域と呼ぶことにする (図 2). また  $\mathbf{X}$  を, この S 細胞の参照ベクトルと呼ぶ. これは, この S 細胞に対する最適特徴 (最も強く反応する特徴) を表わしている.

## 3. 中間層の学習

ネオコグニトロンでは, 前節で述べたような回路を, 学習によって自己組織的に形成していく. 学習は, 下位の層から始めて, ある層の学習が終わった後にその 1 段上位の層の学習を始める. 可塑性を持っているのは S 細胞の入力結合で, S 細胞から C 細胞へは固定結合で結ばれている.

学習時には, 学習パターンを一つずつ入力層  $U_0$  に提示し, そのときの  $U_{CI-1}$  層の C 細胞の反応が,  $U_{SI}$  層の S 細胞に対する学習刺激になる.

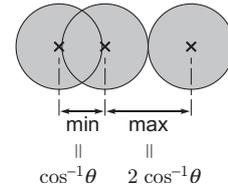


図 3: 発生した S 細胞の, 特徴空間内の分布. 灰色の円盤は, S 細胞の許容域を示す.

### 3.1 Add-if-Silent 則

最近のネオコグニトロンでは, 中間層の S 細胞の学習に, Add-if-Silent (AiS) と名付けた学習則を用いている. この学習則は, S 細胞の入力結合の強度を定めるだけでなく, 細胞そのものの発生も制御している.

もし, 前シナプス側 (信号を送り出している側) の C 細胞が出力を出しているにもかかわらず, その付近の後シナプス側のすべての S 細胞が無反応であれば, 新しい S 細胞を作成して  $U_{SI}$  層に追加する. (ちなみに Add-if-Silent という名称は, S 細胞が silent ならば, 新しい細胞を追加するという意味である). 新しく追加した S 細胞の入力結合は, 前シナプス側の C 細胞の反応に比例した強度とする. S 細胞が一旦形成されて細胞層に追加された後は, その後どのような学習パターンが与えられても, その入力結合 (参照ベクトル) は変化させない. すなわち, AiS 則による学習は, 多数の学習ベクトルの集合の中から, 有用なベクトルだけを S 細胞の参照ベクトルとして選び出す操作であると言える.

AiS 則のもとでは, 既存の S 細胞と類似した参照ベクトルをもつ S 細胞が重複して形成されることはない. 学習ベクトルに類似した参照ベクトルを持つ S 細胞が存在すれば反応出力を出すので, その学習ベクトルに対して新しい S 細胞が作られることはないからである. 従って学習時に十分多数の学習ベクトルが提示されたならば, ベクトル空間内に, S 細胞の参照ベクトルが一様に (ほぼ等間隔に) 分布するように自己組織化が進むことになる (図 3).

通常の deep learning では, 細胞間の初期結合を, 学習時に繰り返し tune-up (微調整) しながら期待値に少しずつ近づけていくので, 学習に長時間を要する. これに対して AiS 則では, 一旦出来上がった結合の微調整は不必要なので, 繰り返し演算の必要はなく, 学習速度が非常に速い. 学習時に, 各学習パターンを何回も提示する必要はなく, 各学習パターンを 1 回ずつ提示するだけで学習は完了する.

このように簡単な AiS 則で, なぜ高い認識率が得られるのであろうか. 入力パターンを最終的に識別するのは, 多層回路の中間層ではなく最上位層である. 中間層の役割は, 入力層に提示されたパターンを, 単一細胞の反応によってではなく, 多数の細胞の反応の集合, すなわち population coding によって, 正確に表現することである. population coding の場合には, 個々の細胞の反応特性が学習刺激に正確に一致していることは, 必ずしも必要ではない. その層の全細胞の集団としての反応が, 入力刺激を正確に表現してさえいれば十分なのである.

このことは, ヒトを含む哺乳動物の視覚神経系の形成に臨界期が存在し, 臨界期を過ぎると下位の視覚野 (第 1 次視覚野など) の神経系は可塑性を失うという神経生理学的な事実とも符合している. 臨界期以前に多様な視覚刺激を与えておく必要があるが, それ以降は下位の視覚野が可塑性を失っていても, 上位の視覚野の働きで, 新しい視覚パターンを学習し認識するこ

とができるのである。臨界期以降は可塑性を失っているのに、特定の視覚対象に対する神経回路の tune-up は行なわれていないはずであるが、我々人間は新しい視覚対象でも問題なく学習し認識できるようになる。ネオコグニトロンでも、中間層の自己組織化が完了した後は、最上位層の細胞だけに可塑性を持たせて、学習を進めている。

### 3.2 2重しきい値

認識時に中間層で population coding の状態を作って頑強な認識をするために、S 細胞の閾値を、認識時には学習時よりも低い値に設定する [Fukushima 96]。学習時には、一つの特徴に対して、1 個（あるいはごく少数）の細胞だけしか反応しない状態を作ることによって冗長な細胞の発生を防いでいた。しかし学習終了後の認識時にも、S 細胞の閾値が学習時と同じ値に保たれていると、S 細胞は“おばあさん細胞”(grandmother cell, gnostic cell) 的な振る舞いをするようになる。つまり、一つの特徴に対して、1 個（あるいはごく少数）の細胞だけしか反応しない状態になり、回路は、変形パターンに対して汎化能力を発揮できなくなる。そこで population coding のような状態を作るために、認識時の閾値を、学習時の閾値よりも低い値に定めるのである。入力パターンの状態を、多数の細胞の集合で表現していれば、入力パターンが多少変形しても、細胞の集合としての反応状態はあり変動しないので、変形パターンに対しても頑強な認識が可能になる。

### 3.3 フィードバックを導入した Add-if-Silent

以上が中間層の学習の基本的な考え方であるが、AiS 則をネオコグニトロンに適用するためには、多少の修正が必要になる。ネオコグニトロンでは、新しい細胞の発生は、新しい細胞面の発生を意味する。前述のようにネオコグニトロンは convolutional network であり、各細胞面内の S 細胞は入力結合を共有している。この共有結合の条件は、S 細胞の結合の発生時にも、常に保たれていなければならない。学習時にも共有結合の条件を保つために、学習パターンが提示される毎に、seed cell と呼ばれる細胞を細胞面の中から選ぶ。seed cell は、結晶成長における核 (seed) のような働きをする細胞である。seed cell は、自分自身に提示された学習刺激を学習して入力結合を決定する。一つの細胞面内の細胞はすべて同一の入力結合を共有しているので、細胞面内のほかの細胞も seed cell と同一の入力結合を持つようになる。

学習時に新たな seed cell を置く位置は、前シナプス側の C 細胞が出力を出しているのに、すべての S 細胞が無反応である位置とする。このような位置が複数箇所存在した時には、前シナプス側の C 細胞の出力の総和が極大になる位置を順番に選んでいく。例えば図 4 の  $U_0$  に示すような学習パターンが提示されたときを考えてみよう。特徴  $a$  は、学習パターンの中に類似した特徴がほかに存在しないので、特徴  $a$  の中心位置に seed cell を置けばよい。しかし特徴  $b_1$  と  $b_2$  のように、類似した特徴が複数個存在した場合はどうであろうか。この 2 個の特徴は互いに類似しているのに、同一の細胞面で抽出するように学習すべきであろう。そうすると、新しく作成される細胞面 B では、seed cell を、 $b_1$  に置くべきか、それとも  $b_2$  に置くべきかが問題になる。

そこで、seed cell の位置を決めるために、以下のような方法を用いる [Fukushima 14]。前シナプス側の各 C 細胞が、後シナプス側の S 細胞の反応にどの程度関与しているかを調べるために、S 細胞から、その入力結合  $a$  に比例した強度の抑制性結合  $-a$  を介してフィードバック信号を前シナプス側の C 細胞に送る。S 細胞の反応に寄与した C 細胞の反応はこの

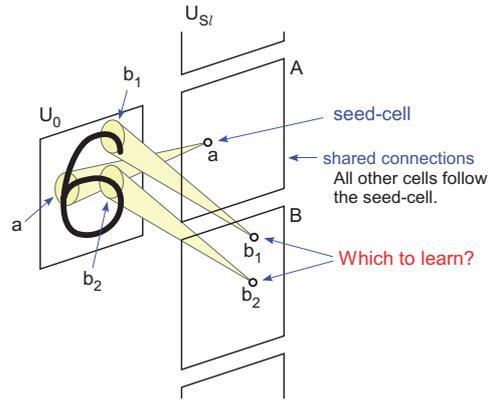


図 4: seed cell を置くべき位置 [Fukushima 15].

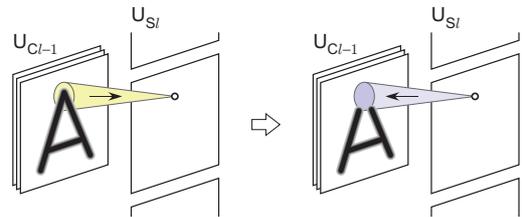


図 5: S 細胞の反応に寄与した C 細胞の反応を抑制性フィードバック信号によって抑えることによって、まだ抽出されていない特徴の存在する位置を目立たせる [Fukushima 15].

フィードバック信号によって抑制されるはずである (図 5)。このような抑制性フィードバック信号を、すべての S 細胞から、前シナプス側の C 細胞に向けて送る。抑制されずに残っている C 細胞があれば、その位置に存在する特徴を抽出する S 細胞はまだ作られていないことになる。そこでその位置に、seed cell を置くことにする。しかし、その seed cell が学習するのは、フィードバックで抑制された C 細胞の反応ではなく、抑制される以前の C 細胞の反応パターンである。この考え方は、前シナプス側の C 細胞層の反応を再現できるように後シナプス側の S 細胞を学習させるという意味で、autoencoder の考え方と類似したところがある。

## 4. 最上位層

### 4.1 内挿ベクトル法によるパターン識別

最上位層  $U_{SL}$  ( $L$  段目の S 細胞層) の S 細胞への入力、その前の層  $U_{CL-1}$  の C 細胞の出力である。  $U_{SL}$  層の S 細胞の閾値は  $\theta = 0$  に設定されているので、その出力は、式 (2) より、  $u = \|\mathbf{x}\| \cdot s$  で与えられる。

1 文字ずつ切り出された文字を認識するような場合には、計算量の節約のために、ほかの層とは異なり  $U_{SL}$  層では、入力信号 (前層の C 細胞の反応) の分析は、V 細胞の出力が最大になる位置 (retinotopic location) でのみ行なうようにしている。

$U_{SL}$  層の S 細胞の学習法は後述するが、教師あり学習によって学習ベクトルのセットを学習し、各 S 細胞 (すなわち、その参照ベクトル) には、その細胞が学習したパターン (学習ベクトル) のクラス名を表わすラベルが付けられる。

初期のネオコグニトロンでは、最大出力で反応した S 細胞のラベルを認識結果としていた (WTA = Winner-Take-All)。すなわち、テストベクトル (認識すべきパターンに対する前層の C 細胞の反応) に最も近い参照ベクトルをもつ S 細胞のラ

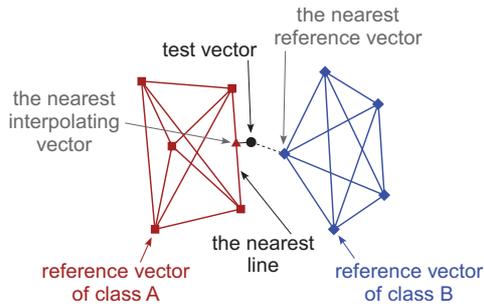


図 6: 内挿ベクトル法の原理 [Fukushima 13][Fukushima 07].

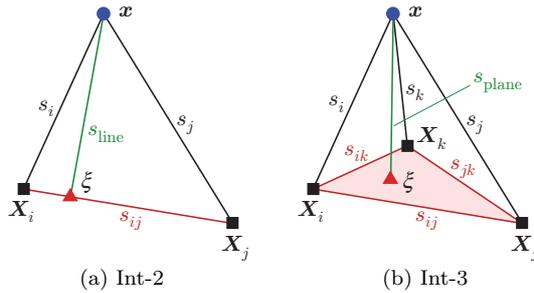


図 7: 2 点からの内挿 (Int-2) と 3 点からの内挿 (Int-3) [Fukushima 15].

ベクトルを認識結果としていた。

これに対して、最近のネオコグニトロン最上位層では、パターンを識別するために、内挿ベクトル法 (Interpolating-Vector) を用いている [Fukushima 13][Fukushima 07]. 内挿ベクトル法では、複数個の細胞の反応を組み合わせることによって認識結果を決定する。その結果、従来よりも少ない数の参照ベクトルで、従来よりも高い認識率を得ることができる。

内挿ベクトル法では、図 6 に示すように、同一のラベルを持つ参照ベクトルの対 (ついで) を結ぶ直線を考え、すべての直線の中で、テストベクトルとの距離が最も近い直線 (すなわち類似度が最も大きい直線) を選び、その直線のラベル (クラス名) を認識結果とする。

テストベクトルと直線との距離は、以下のように定義する。2 個の参照ベクトルを結ぶ直線上に内挿ベクトルと名付けた仮想的なベクトルが密に存在する状況を想定する。内挿ベクトルは、この 2 個の参照ベクトルの線型結合で与えられるベクトルである。図 7(a) のように、直線に沿って並んでいる内挿ベクトルの中で、テストベクトルとの間の類似度が最大になるものを選び出す。その内挿ベクトル  $\xi$  とテストベクトル  $x$  との類似度  $s_{line}$  を、テストベクトルと直線との類似度 (距離) と定義する。

内挿ベクトル法では、ベクトルの対 (ついで) を結ぶ直線の代わりに、三つ組みで形成される面を用いることもできる (図 7(b)) [Fukushima 14][Fukushima 15]. 両者を区別するために、前者を Int-2、後者を Int-3 と記す。Int-3 のほうが Int-2 に比して計算量は増加するが認識率を向上させることができるので、最近のネオコグニトロンでは通常 Int-3 用いている。ちなみに、Int-2 でさえも、WTA や SVM (support vector machine) よりも、はるかに高い認識率を達成することができる。

## 4.2 最上位層の学習

最上位層の学習は、中間層の学習終了後に開始する。最上位層の参照ベクトル (すなわち S 細胞) は学習によって作成

するのであるが、内挿ベクトル法での認識に要する計算量は、参照ベクトル数が増えると急激に増加する。従って、できるだけ少ない数の参照ベクトルで高い認識率を得られるような学習法が求められる。筆者は種々の学習法をいくつか提唱しているが、ここでは、それらに共通する考え方だけを説明しよう。

学習は 2 段階に分けて進める。各段階で、学習セットのすべてのベクトル (学習パターン) を、1 回ずつ提示する。学習ベクトルが提示されるごとに、認識してみる。もし認識結果に誤りであれば、最上位層内に新しい S 細胞 (参照ベクトル) を発生させ、その時の学習ベクトルを、新しい S 細胞の参照ベクトルとする。これは中間層に対する AiS に似た手法であるが、中間層の場合とは異なって学習ベクトルのクラス名を示すラベルを考慮している。

認識結果に誤りがあった場合に新しい細胞を発生させる点は、学習の第 1 段階でも第 2 段階でも同じである。しかし提示された学習ベクトルに対する認識結果が正しかった場合には、第 1 段階では既存の参照ベクトルに対しては何も行なわないが、学習の第 2 段階では結合の tune-up を行なう。

中間層では一旦作られた結合の tune-up は必要なかったが、最上位層では結合の tune-up は不可欠である。tune-up を行なわないと認識率を向上させることは難しい。tune-up にはいろいろな手法が考えられるが、例えば Int-3 を用いる場合には、Int-3 で選ばれた面を形成する 3 個の参照ベクトルを、いずれも学習ベクトルに近づける方向に移動させる [Fukushima 15].

このように、2 段階学習では、十分な数の参照ベクトルが発生するまでは tune-up を開始しないのである。作られた参照ベクトルがまだ十分な数に達していない第 1 段階では、たとえある学習ベクトルがたまたま正しく認識されたとしても、そのときに選ばれた参照ベクトルと学習ベクトルの類似度が高いとは、必ずしも言えない。従ってこの段階で参照ベクトルに学習ベクトルを加算してしまうと、類似度の低いベクトルが加算される恐れがあり、あまり役に立たない参照ベクトルが作られる可能性が高い。従って、参照ベクトルの tune-up は、十分な数の参照ベクトルが形成された後に開始するのである。

**謝辞** 本研究の一部は、科学研究費補助金 (基盤研究 (C) 25330300) の助成を受けて行なわれた。

## 参考文献

- [Fukushima 80] Fukushima, K.: Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, *Biol Cybernetics*, Vol 36, No. 4, pp. 193–202 (1980).
- [Fukushima 13] Fukushima, K.: Artificial vision by multi-layered neural networks: neocognitron and its advances. *Neural Networks*, Vol. 37, pp. 103–119 (2013).
- [Fukushima 15] Fukushima, K., and Shouno, H.: Deep convolutional network neocognitron: improved interpolating-vector, *IJCNN 2015*, pp. 1603–1610 (2015).
- [Fukushima 96] Fukushima, K., and Tanigawa, M.: Use of different thresholds in learning and recognition, *Neurocomputing*, Vol. 11, No 1, pp. 1–17 (1996).
- [Fukushima 14] Fukushima, K.: One-shot learning with feedback for multi-layered convolutional network, *ICANN 2014, LNCS 8681*, eds: S. Wermter, et al., pp. 291–298, Springer (2014).
- [Fukushima 07] Fukushima, K.: Interpolating vectors for robust pattern recognition, *Neural Networks*, Vol. 20, No. 8, pp. 904–916 (2007).