

deepCNN ネオコグニトロンと視覚情報処理

福島 邦彦¹

概要：高いパターン認識能力を学習によって獲得することができる手法として、深層学習 (deep learning) や CNN (convolutional neural network) が最近注目を集めている。ネオコグニトロンもそのような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。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。そのパターン認識能力を学習によって獲得していく。しかしその学習には、現在広く用いられているディープラーニングとは少し異なる学習則を用いている。ネオコグニトロンの歴史は古いですが、現在に至るまで種々の改良が加えられ発展を続けている。現在広く用いられている deep CNN との相違点に重点を置きながら、最近のネオコグニトロンを紹介する。

Deep CNN Neocognitron for Visual Information Processing

KUNIHICO FUKUSHIMA¹

1. ネオコグニトロンの概要

1.1 ネオコグニトロンの回路構造

ネオコグニトロンは、変形に強い視覚パターン認識能力を持つ人工神経回路で、パターン認識能力を学習によって獲得していく [1][2]。

ネオコグニトロンの回路構造は、脳の第一次視覚野に関する神経生理学的な知見にヒントを得て考案された。図 1 に示すように、入力層 U_0 の後ろに、特徴抽出を行なう S 細胞の層 U_S と、位置ずれを許容する働きを持つ C 細胞の層 U_C (pooling 層) とを交互に階層的に配置した多層構造を持つ。ちなみに S 細胞と C 細胞は、大脳視覚野の単純型細胞 (simple cell) と複雑型細胞 (complex cell) に似た性質を持っているので、その頭文字をとって名づけられた。

各層は、複数の細胞面 (sub-layer) から構成されている。図 1 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。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。

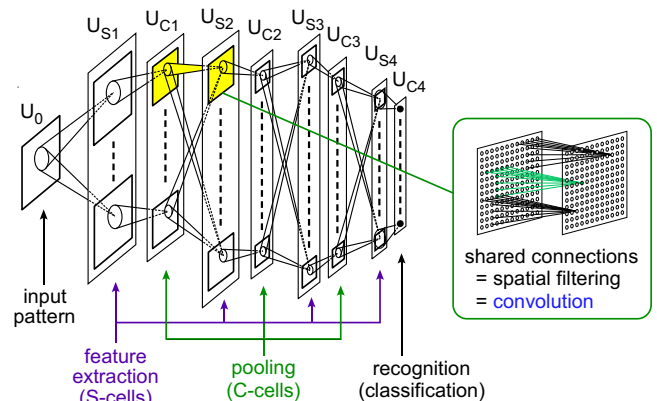


図 1 ネオコグニトロンの回路構造。

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 細胞は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、前層の出力に空間フィルターを掛けて解釈することもできる。このような構造を持つ回路を、最近では CNN (convolutional neural network) と呼んでいる。

¹ ファジィシステム研究所
Fuzzy Logic Systems Institute
fukushima@m.ieice.org
<http://personalpage.flsi.or.jp/fukushima/>

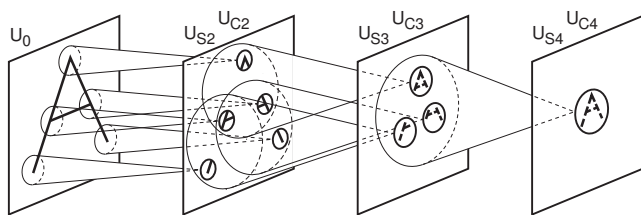


図 2 ネオコグニトロンにおけるパターン認識の原理. ([1] を変更)

1.2 特徴の抽出と統合

ネオコグニトロンでは、個々の細胞は一つ前の層から信号を受け取っているが、前層のすべての細胞から信号を受け取っているわけではない。比較的狭い範囲にある細胞だけから結合を受け取っている。しかも、信号を送り出している細胞と、信号を受け取っている細胞とは、それぞれの層の中でほぼ同じような位置を占めている。つまり retinotopy が保たれている。

下位の層の S 細胞は、小さい受容野を持っていて、入力パターンの局所的な特徴を抽出する。例えばアルファベットの 'A' を学習したネオコグニトロンの下位の段 (U_{S2} 層) には、図 2 に示すように、文字 'A' の種々の部分パターン、すなわち局所の特徴に反応する S 細胞が並んでいる。

このような局所の特徴を抽出する S 細胞が、どのような組合せで反応しているかを見ているのが、その次の段 (U_{S3} 層) の S 細胞である。この細胞は、前の段 (U_{S2} 層) の細胞に比して、もう少し大きな受容野をもっていて、大局的な特徴を抽出している細胞であるということができる。

そのもう一段上には、この大局的な特徴を抽出する細胞の出力をたくさん集めて、もっと大きい特徴をみている細胞がある。つまり、この細胞は、大局的な特徴をみる前段の細胞の反応を介して、間接的に入力層全体の情報を捕らえていることになる。したがって、'A' というパターンが与えられると 'A' に対応した細胞が出力を出し、'B' というパターンが与えられるとまた別の細胞が出力を出す。

1.3 変形に強いパターン認識

C 細胞は、いくつかの S 細胞から固定結合を受け取っていて、そのうちのどれか一つの S 細胞が出力を出せば、C 細胞も出力を出すようになっている。これらの S 細胞はいずれも同じ形の特徴を抽出するが、特徴を抽出してくる場所 (受容野の位置) は S 細胞ごとに少しずつ異なっている。つまり、1 個の C 細胞は、わずかに場所が違ったところから、同じ特徴を抽出してくるような一群の S 細胞の出力を受け取っている。従って、入力パターンの位置が少し変わって、ある特徴の場所が入力層 U_0 (網膜) の上で少しずれると、最初とは異なる別の S 細胞がその特徴を捕らえて反応するかもしれないが、いずれにしても、これらの S 細胞の出力を全部まとめて受け取っている C 細胞は、出力を出し続けることになる。つまり C 細胞は、S 細胞が抽出し

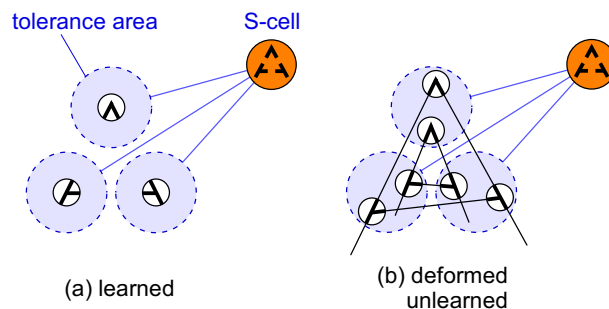


図 3 パターン 'A' の 3 個の局所的な特徴を組み合わせた大局的な特徴を抽出する中間段の S 細胞に収束する結合 [2][3].

た特徴の、位置ずれの影響を吸収する性質をもっている。

例えば図 3(a) のように、3 個の局所的な特徴を組み合わせた大局的な特徴を抽出する中間段の S 細胞に注目してみよう。この S 細胞は、このような 3 個の特徴が、図のような位置関係を保って並んでいることを学習時に覚えた細胞である。学習時には、きれいな 'A' というパターンだけしか見せられていないのである。しかし、学習が終わったあとには、C 細胞の働きによって、これらの 3 個の局所の特徴の位置が多少ずれても、図に点線の円で示した程度の位置ずれであれば許容される。したがって、図 3(b) のように、大きい 'A' というパターンを見せられても、小さいパターンを見せられても、3 個の特徴はそれぞれの点線の円内に入っているため、正しく認識する。このように、学習段階で見せられなかったパターンでも正しく認識するのである。

多層回路内には S 細胞の層と C 細胞の層とが交互に繰り返して階層的に結合されているので、入力パターンの情報は、多層回路内の各段で S 細胞による特徴の抽出と C 細胞による特徴の位置ずれの許容化を繰り返しながら上位段に送られていくことになる。この過程において、下位段で抽出された局所的な特徴は、次第に大局的な特徴に組み上げられていく。このとき、入力パターンの拡大や縮小をはじめとする種々の変形にともなう局所の特徴の相対的な位置ずれの影響は、C 細胞の働きによって少しずつ吸収されるので、最終的には入力パターンのかなりの変形に対しても影響されない出力を得ることができるのである。

ちなみに、C 細胞によるこのような操作は、CNN で pooling と呼ばれている操作とほぼ同じである。しかしネオコグニトロンでは、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。S 細胞は、二つのパターン間の類似度を、互いの重なり具合に基づいて判定している。もし C 細胞によるぼかしがないと、次段の S 細胞は、入力パターンが変形すると以前に学習したパターンとの重なり具合が減るので、別のパターンであると判定する恐れがある。しかしぼかしを行なうと、二つのパターンの間の重なりが大きくな

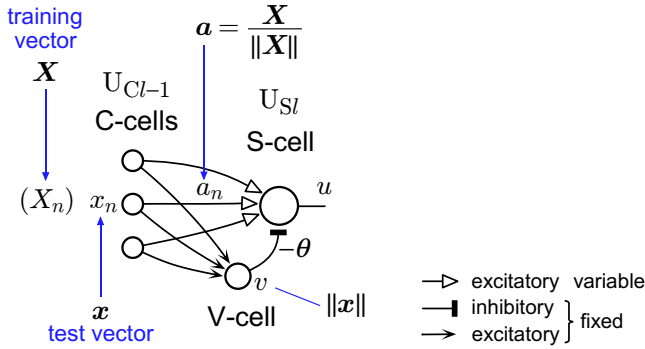


図 4 S 細胞に収束する結合 ([4] を変更).

り, S 細胞は両者を類似したパターンであると判定するようになる。

しかし, 入力層 (網膜) に与えられた入力パターンそのものに直接このようなぼかしの操作をほどこしてしまうと, 入力パターンの微細構造もぼけてしまって, 微細構造の違いを区別できなくなってしまう。そこでネオコグニトロンでは, 局所的な特徴抽出を行なった上で, ぼかしの操作をほどこすようにしている。ネオコグニトロンの中では, C 細胞層によるこのようなぼかしの操作を, 各細胞面ごとに (すなわち各特徴ごとに) 独立に行なっている。次の段の S 細胞は, このような C 細胞層のすべての細胞面の出力を同時に入力として受け取って, より大局的な特徴抽出を行なうのである。このように S 細胞は, 刺激パターンと学習パターンとの間の類似度を, C 細胞層によって作られただけパターンの重なりによって判断して反応していると解釈することもできる。

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。

1.4 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。S 細胞は, 図 4 に示すように, 複数個の C 細胞から興奮性結合を受け取っている。ここでベクトル表示を用いて, C 細胞の出力を $\mathbf{x}$, 興奮性結合の強度を $\mathbf{a}$ と表わす。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。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一群の C 細胞の出力の二乗平均値, すなわちノルム (L2-norm), $v = \|\mathbf{x}\|$ を計算している。

これらの入力を受けて S 細胞の出力は,

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

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。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 θ よりも大きければ, S 細胞は 0 でない出力を出す [2]。このように, θ 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 細胞の反応の許容領域と呼ぶ。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。

2. Add-if-Silent 則による中間層の学習

学習は, 下位の段から順次進めていく。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_{Cl-1} 層の反応が, U_{Sl} 層の S 細胞に対する学習刺激になる。この中間層の S 細胞の学習に, AiS 則 (Add-if-Silent 則) を用いるのである [2][4][5]。

もし, 前シナプス側の C 細胞が出力を出しているにもかかわらず, 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_{Sl} 層に追加する。新しく追加した S 細胞の入力結合は, 前シナプス側の細胞の出力に比例した強度とする。従って, その時点の前シナプス側細胞の反応に対して選択的に反応する S 細胞が作られる。

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, 参照ベクトルを選び出す操作である。学習パターンの集合に含まれる各パターンを 1 回ずつ提示するだけで学習は完了する。このように AiS 則は, 多くの深層学習で用いられている手法とは異なり, 繰り返し演算で結合強度を補正する必要がなく, 高速に学習を進めることができる。

AiS 則のもとでは, すでに存在する S 細胞の許容領域内に新しい細胞が形成されることはない。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, S 細胞は, その参照ベクトルがベクトル空間内に一様に分布するように作られることになる。

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によって, 正確に表現していれば十分である。従って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, 中間層の学習では, 個々の

学習パターンに応じた tuning を必ずしも必要としない。

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 [6]。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”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い。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まり変動しないので、変形パターンに対しても頑強な認識が可能になる。

3. 最上位層

上記のように、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て、次第に高次の特徴が抽出されていく。最上位層（すなわち最深層）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。

ネオコグニトロンの最上位層では、パターンを分類するために、内挿ベクトル法 (interpolating-vector) と名付けた手法を用いる [2][7]。中間層とは異なり、最上位層の特徴抽出細胞には、その細胞が学習した学習パターンのクラス名を示すラベルが付けられている。初期のネオコグニトロンでは、最大出力で反応した特徴抽出細胞のラベルを認識結果としていた (winner-take-all = WTA)。これに対して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、小さな規模の回路で、従来よりも高い認識率を得ることができる。

3.1 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) では、その前段の中間層で抽出された特徴ベクトルをもとに、入力パターンを識別する。S 細胞への入力、 U_{CL-1} 層の C 細胞の出力である。

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（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用いられた WTA (winner-take-all) 法では、テストベクトル（認識すべきベクトル）に最も近い参照ベクトルのラベルを認識結果としていた。これに対して内挿ベクトル法では、図 5 のように、同一のラ

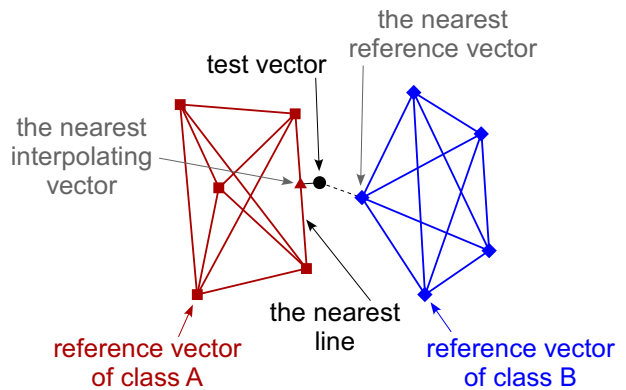


図 5 内挿ベクトル法の原理 [2]。

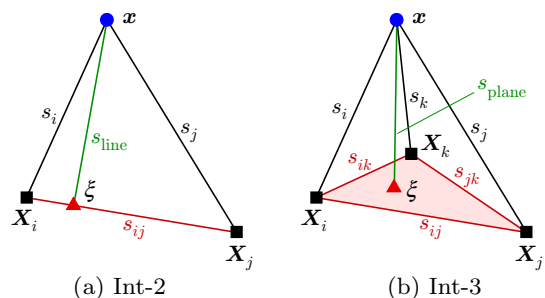


図 6 2 点からの内挿 (Int-2) と 3 点からの内挿 (Int-3) [8]。

ベルを持つ参照ベクトルの対を結ぶ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もの（すなわち類似度が最も大きいもの）を選び、その直線のラベル（クラス名）を認識結果とする [7]。

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 6(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 ξ 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）と定義するのである。

内挿ベクトル法では、ベクトルの対を結ぶ直線の代わりに、図 6(b) のように、三つ組みで形成される平面を用いることもできる [5][8]。両者を区別するために、前者を Int-2、後者を Int-3 と記すことにする。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。

3.2 最上位層の学習

参照ベクトル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、学習法をいくつか提唱しているが、ここでは *margin*ed Winner-Take-All (mWTA) と名付けた学習法を紹介しよう [9]。

中間層の学習と同様に、S 細胞は学習時に発生させる。学習ベクトルに対して最大出力を出した S 細胞が、学習ベクトルと異なるラベルを持っていれば、識別は間違いであると判定し、新しい S 細胞を発生させる。但し最大出力細胞（すなわち学習ベクトルに最も近い参照ベクトル）の探索において、学習ベクトルとは異なるラベルをもつ参照ベクトル（S 細胞）にハンディキャップを持たせて、実際の距離よりも一定割合だけ近くにあると仮定して探索を行なうのである。このようなハンディキャップのもとでも、最も近い参照ベクトルが学習ベクトルと同じラベルを持つ場合にのみ、mWTA による識別が正しいと定義する。

学習時には、学習ベクトル $\mathbf{x}$ が提示されるごとに mWTA で識別してみて、もし識別結果に誤りがあれば、新しい S 細胞（参照ベクトル）を発生させ、 $\mathbf{x}$ を新しい S 細胞の参照ベクトルとする。

中間層とは異なり最上位層では、参照ベクトルの tuning を行なうことによって、認識誤差を大幅に減少させることができる。tuning 操作では、内挿ベクトル法 (Int-3) で選ばれた平面（最も距離の近い平面）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ことによって参照ベクトル（S 細胞の入力結合）を修正するのである [8][9]。

しかし学習を開始したばかりで、まだ十分な数の参照ベクトルが発生していない初期段階では、たとえある学習ベクトルが正しく認識されたとしても、そのときに使用さ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。この段階で tuning を行なおうとして、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弊害が生じる。そこで、学習過程を 2 段階に分けて、第 1 段階では tuning は行なわず、mWTA によって S 細胞（参照ベクトル）の発生だけを行ない、学習がある程度進んだ第 2 段階になって始めて、tuning を開始するようにしている。

4. むすび

ネオコグニトロンによるパターン認識の基本的な考え方と、最近の研究状況を紹介した。ここでは、視覚パターン認識を中心に解説したが、ネオコグニトロンの考え方を更に発展させた種々のシステムも開発されている。例えば、部分的に遮蔽されたパターンの認識 [10]、アモダル補完を行なう神経回路 [11]、選択的注意機構のモデル [12]、などである。それらの概要は、例えば文献 [2] などにも記載さ

れているので、興味のある方は参照していただきたい。

参考文献

- [1] K. Fukushima: Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybernetics*, **36**[4], 193–202 (1980).
- [2] K. Fukushima: Artificial vision by multi-layered neural networks: neocognitron and its advances. *Neural Networks*, **37**, 103–119 (2013).
- [3] K. Fukushima: A neural network for visual pattern recognition. *IEEE Computer*, **21**[3], 65–75 (March 1988).
- [4] K. Fukushima: Training multi-layered neural network neocognitron. *Neural Networks*, **40**, 18–31 (2013).
- [5] K. Fukushima: One-shot learning with feedback for multi-layered convolutional network. *ICANN 2014, LNCS 8681*, eds: S. Wermter, et al., 291–298, Springer (2014).
- [6] K. Fukushima, M. Tanigawa: Use of different thresholds in learning and recognition. *Neurocomputing*, **11**[1], 1–17 (1996).
- [7] K. Fukushima: Interpolating vectors for robust pattern recognition. *Neural Networks*, **20**[8], 904–916 (2007).
- [8] K. Fukushima, H. Shouno: Deep convolutional network neocognitron: improved interpolating-vector. *IJCNN 2015*, Killarney, Ireland, pp. 1603–1610 (July 2015).
- [9] K. Fukushima: Margined winner-take-all: new learning rule for pattern recognition. *IJCNN 2016*, Vancouver, Canada, pp. 977–984 (N-16157), (2016).
- [10] K. Fukushima: “Recognition of partly occluded patterns: a neural network model”, *Biological Cybernetics*, **84**[4], pp. 251–259 (2001).
- [11] K. Fukushima: “Neural network model for completing occluded contours”, *Neural Networks*, **23**[4], pp. 465–582 (May 2010).
- [12] K. Fukushima: “Neural network model for selective attention in visual pattern recognition and associative recall”, *Applied Optics*, **26**[23], pp. 4985–4992 (Dec. 1987).