

[脳情報科学が拓く AI と ICT]

② 脳情報科学と人工知能

—ネオコグニトロンから Deep Learning へ—



本武陽一 | 東京大学大学院新領域創成科学研究科

庄野 逸 | 電気通信大学大学院情報理工学研究所

田村 弘 | 大阪大学大学院生命機能研究科

岡田真人 | 東京大学大学院新領域創成科学研究科

本稿の目的と構成

ヒトはコンピュータを手にした瞬間から、ヒトの脳と同じような働きをする人工物、人工知能 (AI: Artificial Intelligence) を創ろうとしたに違いない。本稿では、ヒトがコンピュータを手にして始まった AI 研究の歴史から現状を俯瞰し、今後の脳型 AI の開発に対する、筆者らの私見を述べる。

AI 研究には大きく分けて、記号处理的 AI と、ニューラルネットワークによるパターン处理的 AI の2つのパラダイムがある。現在の AI ブームのきっかけの1つは、パターン处理的 AI であるニューラルネットワークの台頭にある。ニューラルネットワークの研究にもいくつかのブームがあり、現在は1980年代後半から90年代初頭の第2次ブームから見て、第3次ブームと考えることができる。次の章では、この第2次と第3次ブームの境期での、ニューラルネットワーク研究の周辺領域の研究を俯瞰する。続く章では、現在の Deep Learning ブームの火付け役である深層畳み込みニューラルネットワーク (DCNN: Deep Convolutional Neural Network) の原点であるネオコグニトロンの動作原理を概説する。そこで述べた歴史を振り返ることで、今後の人工知能研究において、脳科学の知見を積極的に取り入れることが必須であることを述べる。その一例として「第1次視覚野の光学計測」の章では、脳の光学計測によって判明した、第1次視覚野における入力刺激のマップ構造と DCNN のプーリング機構の関係

を述べ、最後の「視覚の脳型人工知能設計指針」の章で、今後の脳型人工知能開発の指針の1つを述べる。

ニューラルネットワークとコンピュータビジョン

本章では、第2次と第3次ブームの境期での、ニューラルネットワーク研究の周辺領域の研究を俯瞰する。図-1は、第3次 AI ブームの発端となったコンピュータビジョン (CV) 分野の発展を大まかに表した年表である。CV 分野は2000年代に入る頃からインターネットやソーシャルメディア上に蓄積されつつあった大量のデジタル画像を処理するために、画像中の物体や人の識別、顔位置の検出などの研究が盛んに行われてきた。CV 分野は機械学習分野と積極的に結びつくことによって多くの成果を上げてきている。機械学習分野は、第2次ニューラルネットワークブームから脱却し、統計的学習理論を背景にしたベイズ手法やカーネル法といった手法が盛んに研究されていた。CV 分野では、これらの機械学習の手法に、画像特有の特徴抽出手法を組み合わせることでさまざまな応用を発展させてきた。ここでは特に Histogram of Oriented Gradients^{☆1} (HOG) と Scale Invariant Feature Transform^{☆2} (SIFT) と呼ばれる特徴抽

^{☆1} Dalal, N. and Triggs, B. : Histograms of Oriented Gradients for Human Detection, Intl. Conf. Comp. Vis. & Patt. Rec., Vol.2, pp.886-893 (2005).

^{☆2} Lowe, D. G. : Distinctive Image Feature from Scale-Invariant Keypoints, Intl. J. Comp. Vis. Vol.60, No.2, pp.91-110 (2004).



出手法に着目する。HOGはDalalとTriggsによって提案された特徴抽出手法であり、画像の局所的な領域に含まれるさまざまな方向のエッジ成分を、方向ごとのヒストグラムで表現する特徴抽出手法である。HOGは照明の変化などに強く、画像の大まかな特徴を記述するのに用いられる。このHOGによる特徴量とSVMとを組み合わせることで画像中の人物検出を実現している。一方のSIFTは、Loweによって提案された特徴抽出手法である。SIFTは画像中の局所的な輝度変化の大きい点をスケールごとに検出し、その周辺のエッジ成分を、回転やスケールに対して正規化を行った上で、方向ごとのヒストグラムで表現した特徴表現手法である。SIFT特徴量は、画像の回転や拡大縮小といった操作に対して不変性を持つことが知られており、撮像位置が異なる複数の画像間の位置合わせといった応用が提案されている。

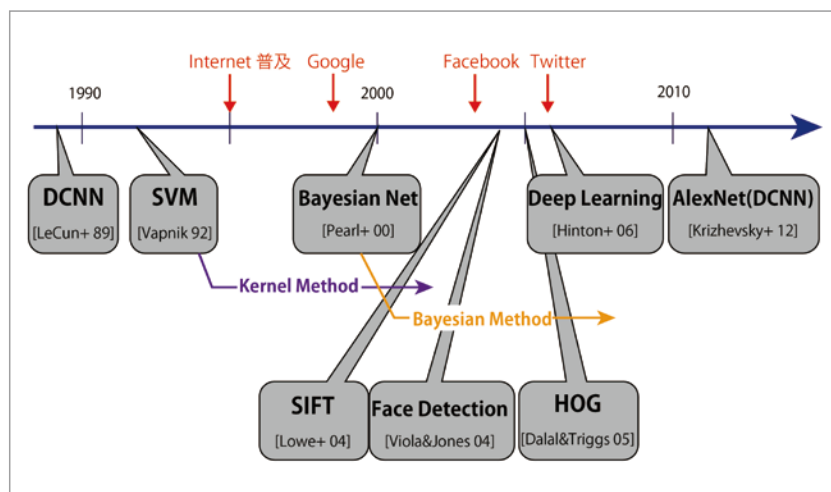
それでは、これらの特徴抽出手法と、脳科学における知見とがどのように関連するかを考えてみよう。まずHOGは、画像中の局所的なエッジ成分をヒストグラムとして表現する手法であるが、1つ1つの局所領域の表現には、すべての方位セットが含まれている。この特徴表現は第1次視覚野等の初期視覚野におけるハイパーカラム構造と呼ばれる皮質構造に類似する。さらに見方を変え、このようなハイパーカラム構造は、近年のCV分野の中核をなすDCNNや後述するネオコグニトロンといったニューラルネットワークの低階層側の構造に相当する。またSIFTは、いったん画像中のエッジ成分を計算し、その特徴を元に計算する階層的な特徴量と捉えることができる。結果としてSIFT特徴量は画像中のコーナーといった曲率が高い個所を特徴位置として抽出しやすい。脳における初期視覚野がエッジ抽出に相当するものとすれば、SIFT特

徴量は、第2次視覚野や第4次視覚野といった中期視覚野の情報表現と関係すると考えられる。これらの領域も曲率などの画像特徴に選択的に反応することが知られている。

このような視点で考えると、2000年代のCV分野における特徴抽出手法は、脳科学の分野とは、ほぼ独立に発展していながら、ネオコグニトロンやDCNNといった脳科学の知見に基づく画像認識装置が持つ階層構造を、手作業で作り込んでいったものと捉えることができる。このことは、脳科学の知見を積極的にCV分野に取り入れることができれば、AI自体の進展を加速することが可能であることを示唆している。

ネオコグニトロン

ネオコグニトロンは、Fukushimaによって提案された階層型ニューラルネットワークであり¹⁾、今日のCV分野で標準的に用いられているDCNNの原型である。ネオコグニトロンは主に2種類の性質を持つニューロンからなる。これらは、それぞれS-cellとC-cellと呼ばれるニューロンモデルで、第1次視覚野の単純型細胞と複雑型細胞とをそれぞれのモデルとしている。図-2上段に単純型細胞と複雑型細胞の特性を示す。共通点として、双方の



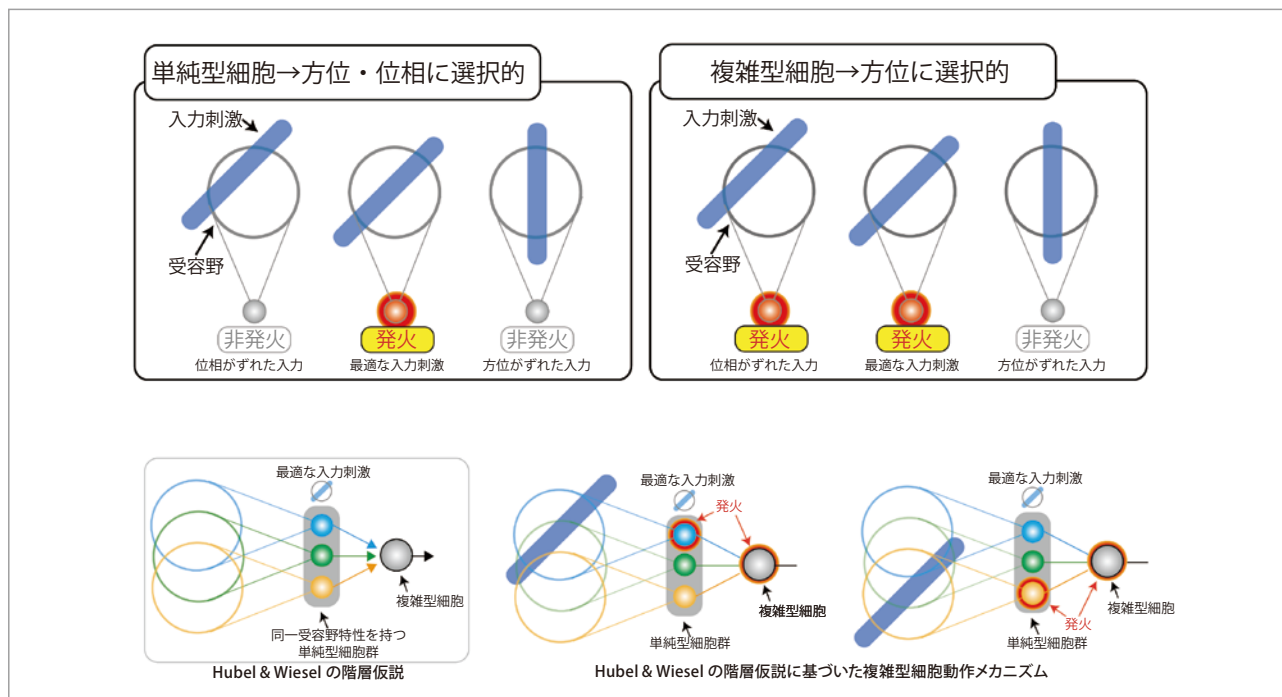
■図-1 2000年代のAIとCV分野の発展

細胞とも受容野と呼ばれる局所的な領域を視野中
 持っており、受容野内に与えられた刺激によって応
 答が変化する。また双方の細胞とも受容野内に提示
 された特定のパターンにのみ選択的に反応する。単
 純型細胞と複雑型細胞の違いはパターンの特定性の
 許容度に現れる。図-2 上段左は単純型細胞の特性
 を模式的に表したものである。図中の各円は、ある
 単純型細胞の受容野を表しており、受容野中央に
 現れた“斜め線”が選択的に反応するパターンだ
 ったとしよう。このとき単純型細胞は、位相がずれた
 パターンや方位が異なるパターンには反応しない。
 一方、複雑型細胞の反応を模式的に表したものは
 図-2 上段右図となる。複雑型細胞は、方位が異
 なるパターンには反応しないが、位相がずれたパ
 ターンに関しては反応する。HubelとWieselは、図-2
 下段に示すように、わずかにずれた位置に受容野を
 持ちながら同じ方位選択性を持つ単純型細胞群の出
 力を集めるような構造を複雑型細胞が持っていれば、
 その機能を説明できるという階層仮説を提案してい
 る²⁾。この同じ特徴選択性を持つ細胞の出力を集め

る（プールする）操作を以降ではプーリングと呼ぶ。

Fukushimaは、この仮説をパターン認識に適用
 するために、各タイプの細胞を2次元平面上に配
 置した特徴マップ表現をネオコグニトロンに導入し
 た。特徴マップは、同じ特徴選択性を持つ細胞群で
 構成されるため、数理的には畳み込み（コンボリュ
 ーション）で記述できる。このように考えるとS-cell
 は、局所的な方位に関する特徴抽出をフィルタ演算
 で導出する役割を担い、C-cellは、抽出された特徴
 の位相変化を許容する位相不変的応答を局所的な空
 間平均操作等によって提供する。DCNNにおいて
 はS-cellとC-cell特徴マップによって構成される
 層は、それぞれ畳み込み層、プーリング層と呼ばれ
 る。ネオコグニトロンは、これらの層を交互に並べ
 た階層構造を持つ。畳み込み層とプーリング層とを
 接続した構造が初期視覚野の近似モデルであるとす
 れば、ネオコグニトロンは、この近似モデルで高次
 層を外挿したモデルと捉えることができる。

次にネオコグニトロンの動作原理について考えて
 みよう。図-3のようにスケールの異なる“A”と



■ 図-2 初期視覚野の単純型細胞と複雑型細胞の概要



いうパターンが入力に与えられた状況を考えてみる。ここでは“**A**”というパターンが局所的な逆V字型の形状とT字型の形状として特徴づけられるものとする。スケールの異なる“**A**”パターンの場合、これらの特徴は、ある畳み込み層の特徴マップ上では、異なった位置に検出される。プーリング層では、畳み込み層の演算結果に対して空間的な平均操作等がなされるため、これら3種類の特徴がボカされたような出力となる。この結果、上位の畳み込み層の細胞は、これら3種類の組合せ特徴を局所的な結合の範囲で捉えることができるようになり、スケールの異なる“**A**”パターンに対しても同様の反応をすることができる。これはつまり、S-cell、C-cellという構造を交互に適用することで、ネオコグニトロンが受容野内における位相普遍的応答だけでなく、受容野サイズの拡大を伴った位置不変的応答も獲得できたことを意味する。ネオコグニトロンは、このような計算原理に基づいたパターン認識を行うニューラルネットワークモデルである。

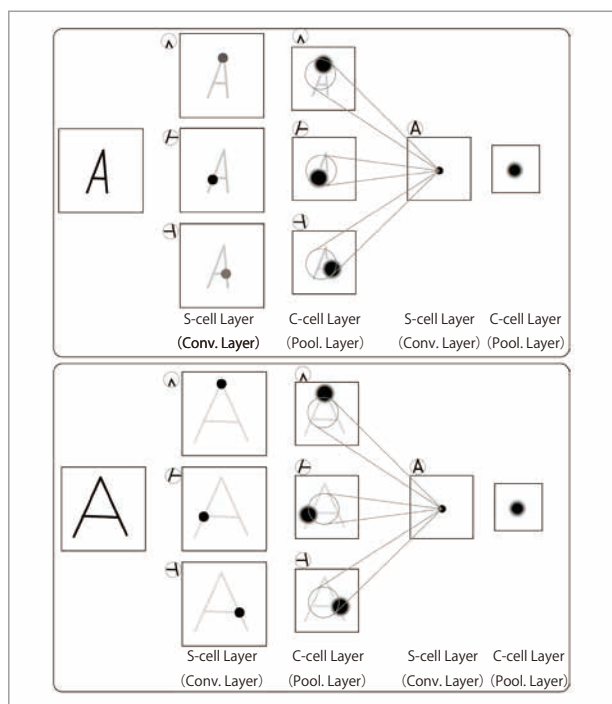
ここで注目したいのは、ネオコグニトロンの畳み

込み層は、S-cellが反応するパターンのタイプごとに、別々の2次元特徴マップを持つ3次元構造となっており、その2次元特徴マップには入力（網膜）画像上の位置関係がそのまま投射されているという点である。この構造によって、ネオコグニトロンでは畳み込み層における距離関係と、入力（網膜）画像上における距離関係が1対1に対応することになる。これを背景として、空間的な平均操作といったプーリングによって、受容野サイズの拡大を伴った位相・位置普遍的応答が獲得されたのである。ところが、脳での畳み込み層に対応する視覚皮質では、さまざまなパターン選択性を持つ単純型細胞が、2次元の同一皮質シート上にマッピングされている。このため、次章で説明するように脳の視覚皮質では、視覚皮質上における距離関係と入力画像上の距離関係の1対1対応は成立しない。したがって、ネオコグニトロンで用いられた、空間的な平均操作のみを行うような単純なプーリングは、脳の視覚系では用いられていないと考えられるのである。

第1次視覚野の光学計測

光学計測は、皮質上の各位置の神経細胞集団に最も強い反応を誘発する刺激についてのマップを、数mmの解像度で作成できる強力な手法である。神経細胞活動に基づいて消費された酸素を補うために生じる血流変化を、吸光スペクトルの変化として捉えることで、マップを得ることができる。本章では、DCNN（≡ネオコグニトロン）と異なり、脳の第1次視覚野マップが2次元空間となっていることが、脳のプーリングに与える影響について説明する。

前章で述べたように、階層的情報処理におけるプーリングの1つの役割は受容野内における位相不変的応答の獲得である。DeAngelisらにより、同じ方位選択性を持ちながらさまざまな位相選択性を持つ単純型細胞が皮質局所に混在していることが示されている³⁾。そのため、DCNNと同様にこれら近



■図-3 ネオコグニトロンの動作原理

傍細胞の活動をプールすることで位相普遍的応答は形成できる。一方、この皮質上に存在するピンホールセンター（図-4, PWC）と呼ばれる領域では話が異なってくる。PWCとは、視覚皮質がさまざまな方位選択性を持つ単純型細胞を2次元の同一皮質シート上にマッピングすることから生じる、異なる方位ドメインが皮質上の1点に集まる点（図-4）のことである。このPWCの近傍細胞の出力をプールすると異なる方位情報が混在してしまう。このような事態を避けるため、脳では方位ドメインの境界を考慮したプーリングが行われていると考えられる。

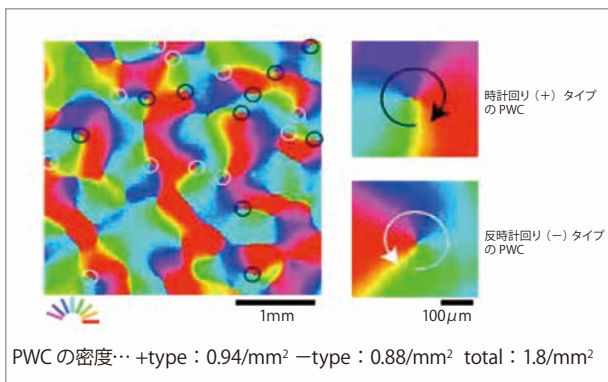
階層的情報処理におけるプーリングのもう1つの役割は、前章でも述べた受容野サイズの拡大を伴った位置不変的応答の獲得である。プーリングによって受容野サイズを拡大するためには、最適方位が同じで隣接した視野上の位置に受容野を持つ細胞からの出力をプールする必要がある。ところが、脳の第1次視覚野マップ上で、特定の方位に選択的な領域は不連続に出現するので、DCNNと違い、皮質地図上で離れた位置にある特定の神経細胞からの出力をプールする必要があると考えられるのである。実際、最適方位を共有する離れたコラムからの入力階層的神経結合において収束する様子が観察されており⁵⁾、これらコラム間では同期的スパイク発火を示すことも知られている⁶⁾。このような神経回路の性質が、選択的なプーリングに貢献していると考えられる。

このように脳の第1次視覚野におけるプーリング

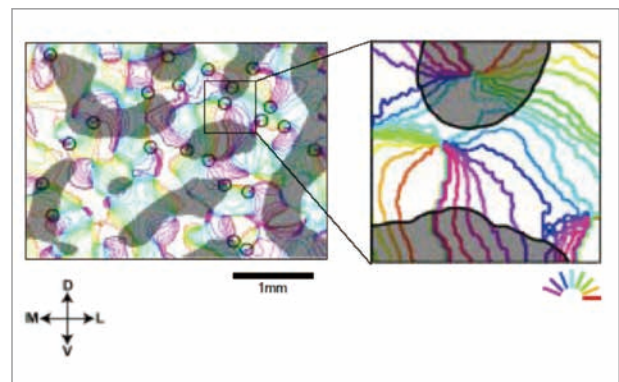
では、方位ドメインの境界を考慮したり、離れた位置にある特定の神経細胞からの出力を収束させるようなプーリング機構が必要となる。このようなプーリング機構は、“画像の隣接部位間の視覚特徴は類似する”という統計的性質よりHebb則に基づいた視覚経験依存的な神経回路調整＝学習によって、形成することができると考えられる。

ここまで方位のマッピングについて見てきたが、第1次視覚野では、方位以外にも複数の視覚情報（網膜座標、優位眼、空間周波数、色）がこの2次元皮質上にマッピングされている。たとえば、優位眼領域は連続するように皮質上でマップされているが、しばしば分断される（図-5, グレー領域と白色領域）。上で述べたように、方位マップはPWCで不連続となっている。このように複数の視覚情報を2次元皮質上にマップするには、不連続性や重なりが生じることは避けられない。

ここで、DCNNと脳の第1次視覚野との相違点をまとめる。前章で述べたように、DCNNは3次元の特徴マップ構造を持ち、そこには位相情報もマップされている。このような構造を持つDCNNでは、先に脳の第1次視覚野で見たような方位情報の混在や、視覚情報間の自由度が制約されるといった現象は生じず、そのため空間的な平均操作のみを行うような単純なプーリングが用いられてきた。このような特徴マップの構造や制約の違いは、DCNNと脳の情報処理の間に相違が存在することを示唆する。



■図-4 サル第1次視覚野最適方位マップ⁴⁾



■図-5 2つのコラムの関係⁴⁾



視覚の脳型人工知能設計指針

最後にこれまでの議論に基づき、今後の脳型人工知能開発の指針の1つを述べる。

前章で解説したように、脳の視覚系におけるプーリング機構には、DCNN で通常用いられているような単純なプーリングではなく、外界の情報に基づきその構造を Hebb 則等の学習によって形成していくような複雑なプーリング機構が採用されていると考えられることが分かった。そして、その学習する外界の構造とは、第1次視覚野においては、同じ方位は空間的に連続して現れる傾向にあるという、自然画像が持つ空間的な統計的性質であった。これまで見てきたように、脳の初期視覚野における複雑なプーリング機構の形成は、皮質の2次元性に起因するものであるとされる。このことから、脳は3次元の皮質シートを持つ DCNN よりも強力な次元圧縮の制約の元、画像特徴量の抽出を実現していることが推察される。これは、人間の脳を参考にしたプーリング機構を導入することで、現在の DCNN よりも少ない層数で同等のパフォーマンスを得ることや、錯視のように人間が環境への適応の中で獲得した視覚認知バイアスを再現するといった、より強力で人間らしい(≒効率的な)画像認識を実現できる可能性を示唆する。

最後に、初期視覚野だけでなく、高次視覚野のプーリング機構と DCNN の関係に言及する。高次視覚野で学習する外界の構造は空間的な統計性だけでないことが示唆されている。たとえば、高次視覚野では、刺激提示の時間的相関が、ニューロンの発火パターンに変換されているという Miyashita らによる知見^{☆3☆4}があり、高次視覚野では単なる画像の空間的な統計性ではなく、実験環境の背後にあるより複雑な統計性を脳内に表現している可能性が考えられている。このような時間的相関などの情報に対

する普遍的応答の必要性は、DCNN のプーリング機構に、より大きな発展を促す可能性を秘めている。

このように、人間の脳のプーリング機構は、DCNN などの画像認識アルゴリズムが持つプーリング機構にはない複雑な構造を持つことが分かった。この人間の脳における複雑なプーリング機構の存在は、人間の知覚表象の形成に大きな影響を与えていると考えられるものであり、人間と同様に外界を認識することを目標とする脳型人工知能の実現において、このプーリング機構の構造を解明し、DCNN に組み込むことは必須のことであると考えられる。

参考文献

- 1) Fukushima, K. : Neocognitron : A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position, *Biol. Cyber.* Vol.36, No.4, pp.193-202 (1980).
- 2) Hubel, D. H. and Wiesel, T. N. : Receptive Fields and Functional Architecture of Monkey Striate Cortex, *J. Physiol.*, Vol.195, No.1, pp.215-243, (1968).
- 3) DeAngelis, G. C., Ghose, G. M., Ohzawa, I. and Freeman, R. D. : Functional Micro-organization of Primary Visual Cortex: Receptive Field Analysis of Nearby Neurons, *J. Neurosci.*, Vol.19, pp.4046-4064 (1999).
- 4) Okamoto, T., Ikezoe, K., Tamura, H., Watanabe, M., Aihara, K. and Fujita, I. : Predicted Contextual Modulation Varies with Distance from Pinwheel Centers in the Orientation Preference Map, *Sci. Rep.* 1:114 (2011).
- 5) Gilbert, C. D. and Wiesel, T. N. : Columnar Specificity of Intrinsic Horizontal and Corticocortical Connections in Cat Visual Cortex, *J. Neurosci.*, Vol.9, pp.2432-2442 (1989).
- 6) Ts'o, D. Y., Gilbert, C. D. and Wiesel, T. N. : Relationships between Horizontal Interactions and Functional Architecture in Cat Striate Cortex as Revealed by Cross-correlation Analysis, *J. Neurosci.* Vol.6, pp.1160-1170 (1986).

(2017年10月4日受付)

本武 陽一 mototake@mns.k.u-tokyo.ac.jp

東京大学大学院新領域創成科学研究科特任研究員。東京大学大学院総合文化研究科博士(学術)。日本人工知能学会、日本物理学会、各会員。

庄野 逸(正会員) shouno@uec.ac.jp

電気通信大学大学院情報理工学研究所教授。本会数理モデル化と問題解決研究会主査。

田村 弘 tamura@fbs.osaka-u.ac.jp

大阪大学大学院生命機能研究科准教授。大阪大学大学院医学研究科博士(医学)。視覚認識の神経基盤、皮質局所回路、多ニューロン同時計測手法に関する研究に従事。

岡田 真人(正会員) okada@edu.k.u-tokyo.ac.jp

東京大学大学院新領域創成科学研究科教授。博士(理学)。文部科学省科研費・新学術領域研究「スパースモデリング」領域代表。

☆3 Miyashita, Y. and Chang, H. S. : Neuronal Correlate of Pictorial Short-term Memory in the Primate Temporal, *Nature*, 331, pp.68-70 (1988).

☆4 Miyashita, Y. : Neuronal Correlate of Visual Associative Long-term Memory in the Primate Temporal Cortex, *Nature*, 335, pp.817-820 (1988).