

特集：オプトエレクトロニック画像処理技術

並列分散処理による能動的画像認識 ——顔の認識への応用——

大阪大学 工学部 応用物理学科*

津村 徳道, 伊東 一良, 一岡 芳樹

1. まえがき

近年、光の持つ並列性、無誘導性等の優れた特性を利用した高密度配線技術に基づく並列コンピューター^{1),2)}が研究されている。並列コンピューターにおける計算モデルとして、生物の脳の並列分散処理を模倣したニューラルネットワーク³⁾は有効である。また、記号処理に基づく認識手法が実世界においてうまく機能しないことが指摘されている現在^{4),5)}、ニューラルネットワークは、実世界の不完全な情報を扱うことのできる新しい認識手法の1つとして期待されている^{6),7)}。

本稿では、ニューラルネットワークによる画像認識技術について概説するとともに、筆者らの行った並列分散処理による能動的な顔画像の認識とその結果⁸⁾について述べる。

2. 並列分散処理モデルによる画像認識

ニューラルネットワークは、多数のユニットと呼ばれる非線形処理単位から構成されている。表1にネットワークの構造とその画像認識能力の関係についてまとめた。本節では、表1に示された各構造と画像認識能力について説明する。本稿の主題である能動型構造については、3章において説明する。

2.1 単一ユニット

図1にユニットの構造を示す。ユニットは、生物の神経細胞（ニューロン）を模倣した多入力1出力の非線形素子である。ヘービサイド関数等の非線形関数によりユニットの出力が1か0のいずれかの値をとるとした場合、ユニットの出力値が1のとき、そのユニットは活性化しているという。縦 n 画素、横 m 画素の画像の各画素の値を成分とした $n \times m$ 次元の画像ベクトルを考える。単一ユニットへの入力ベクトルを画像ベクトルとした場合、そのユニットの出力が活性化するか否かで画像ベクトルを2つのクラスのどちらかに分類することができる。しかし、この時のクラスは、ユニットの伝達関数より求まる1つの超平面により分割された領域に限られる。しかし、次に示すように複数のユニットを用いることにより、より複雑な領域を持つクラスを認識することができる。

2.2 単純階層構造

図2に単純階層型構造の一例として3層のニューラルネットワークを示す。階層型構造では、層間にはユニット間の結合が存在するが層内には結合が存在しない。第1層は入力層と呼ばれ、各ユ

表1 ネットワーク構造と画像認識能力

認識の難易 認識対象の例	易 → 難			
	印刷文字 (1文字)	印刷文字 (英数字) 局所的特徴 (エッジ等)	手書き数字 印刷文字 (漢字)	実世界の画像 (顔、文章、環境等)
ネットワーク 構造	単一ユニット	単純階層型構造	位置不変結合型構造 モジュール型構造	能動型構造

*〒565 大阪府吹田市山田丘2-1 ☎06-877-5111

特集：オプトエレクトロニクス画像処理技術

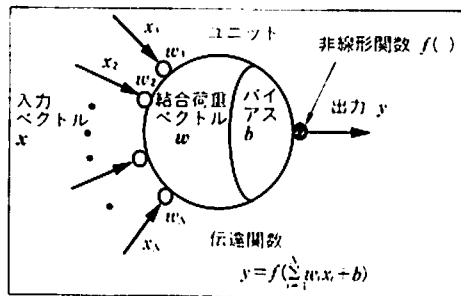


図1 ユニット

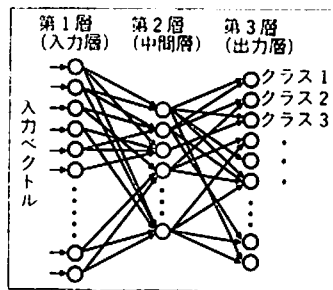


図2 単純階層型構造の一例(3層構造)

ユニットが入力ベクトルの各成分を出力する。出力層と呼ばれる第3層のどのユニットが活性化するかにより入力画像の属するクラスを認識する。単純階層型構造の画像認識能力について述べる。第2層の各ユニットは、単一ユニットの場合と同じように活性化するか否かで画像ベクトルを2つの簡単なクラスに分割する。第2層の複数のクラス分けの結果をうけた第3層の各ユニットは、画像ベクトルのより複雑なクラス分割を行うことができる。一般的に、ユニットの非線形関数にシグモイド関数などの飽和型非線形関数を用いた場合、第2層のユニットの数に制限を加えない3層のニューラルネットワークでは、画像ベクトルを任意のクラスに分割する能力を持つことが知られている^{9)~10)}。しかし、実際には、以下に述べる学習に関する問題から、3層以上の階層を持ったニューラルネットワークでも印刷文字や画像中の局所的な特徴などの単純なパターンの認識しか行えないのが現状である。

学習とは、ネットワークのユニット間の結合荷重等のパラメーターを、与えられた例から決定することをいう。現在、誤差逆伝搬学習法¹¹⁾がもつ

とも強力かつ簡単な学習法としてよく用いられている。手書き文字のように、同一クラスの画像間に大きな変形を含む画像ベクトルは、画像ベクトル空間内で互いのクラスが複雑に入り組んだ分布を呈する。ネットワークにより、それらの分布を正しく学習するためには、それらの分布を十分正確に表現するだけの例の数が必要となる。しかし、そのような数の例を十分に得ることは困難である。また、仮に十分な例の数を得たとしても、大量な数の例の学習に要する時間その他の問題が起こる。また、漢字などの多くのクラスに存在する対象の認識においても同様に、学習に要する時間の問題が生じる。

しかし、次の2章3節に示すように、入力の大変形に関して出力結果が左右されない構造をあらかじめネットワークに与えることにより、少ない数の例で大きな変形のある画像に対しても認識可能なネットワークが実現できる。また、2章4節に示すように、複数のネットワークを用いた構造を用いることにより多くのクラスの存在する対象が認識可能なネットワークが実現できる。

2.3 位置不変結合階層型構造

図3に変形の大きい文字等の認識に適した位置不変結合階層型構造の一例を示す。これは、Fukushimaの提案したネオコグニトロン¹²⁾の構造である。画像を2次元のまま伝達するために、各層ではユニットが2次元面内に整列している。各層内のユニットは、いくつかのクラスターを形成している。前後する層に含まれるクラスター間は、位置不変な結合荷重を持つ接続が行われている。位置不変な結合荷重とは、2層間に図4に示すよ

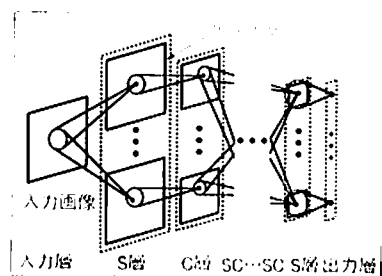


図3 位置不変結合階層型構造の一例
(ネオコグニトロン¹²⁾の構造)

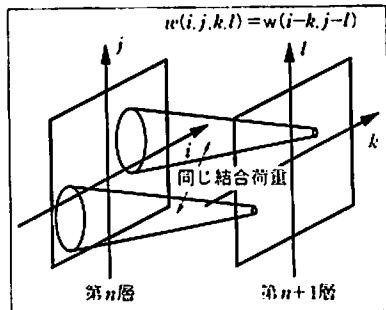


図4 位置不変結合

うな関係を持つ結合をいう。ネオコグニトロンの場合は、S層、C層と呼ばれる働きの異なる層が互い違いに配置された特殊な構造をしている。S層は、前層の局所的な特徴を検出する層であり、前層のすべてのクラスターと結合している。C層は、前層で検出された特徴の位置をぼかす働きをする層であり、前層の対応するクラスターとのみ結合している。S層の特徴検出能力と、C層の位置ずれの大きさに関する許容能力により、局所的な特徴の位置関係が大きく変化した画像に対しても変化前の画像と同じ認識結果を得る構造をしている。したがって、各クラスに対して単一の画像のみでネットワークを学習させたとしても、各画像の局所的な特徴の位置関係が大きく変化する画像に対しても同じ出力を得ることができる。

位置不変結合階層型構造において、他に、Le Cun らは、誤差逆伝搬学習法を用いて郵便番号の認識を行った¹⁹⁾。Zhang らも、同時期に、誤差逆伝搬学習法による位置不変結合階層型構造の学習を実現した¹⁹⁾。彼らは、特に、画像認識の前処理として重要な画像処理に位置不変結合階層型構造を応用している¹⁹⁾。

2.4 モジュール型構造

図5にモジュール型構造の一例を示す。これは岩田らのComb NET¹⁰⁾の構造である。誤差逆伝搬法で学習した144個の3層ネットワークモジュールと、各モジュールに対応した単一ユニットを持ち、その活性化に応じてモジュールに入力を振り分けるための一層で構成されている。モジュールに入力を振り分けるための層は、Kohonen の

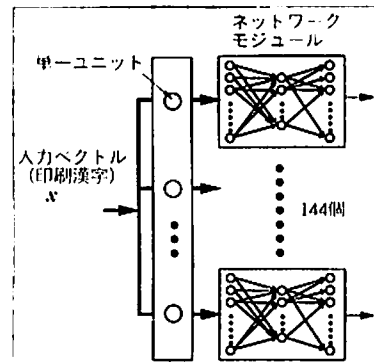


図5 モジュール型構造の一例 (Comb NET の構造)

自己組織化学習法¹⁷⁾により学習される。Iwata らは、これにより JIS 第1水準の印刷漢字2965文字の認識を行い高い認識能力を得ている。

モジュール型構造において、近年、Jacobs らにより、各モジュールに仕事を効率よく分ける学習法が提案されている^{18), 19)}。ほかに、モジュール型構造を、複数の情報の統合に用いようとする例もある²⁰⁾。

3. 並列分散処理による能動的顔画像認識

近年、ロボットビジョンの分野で、視点の移動などに基づく能動的視覚の有効性が主張されている²¹⁾。特に、屋内や野外の実世界の膨大な情報を取り扱う際、それらの情報をすべて一度に取り込むのではなく、認識に有効な情報を選択的に取り込むことが重要である。並列分散処理モデルによる画像認識技術においても、図6に示す概念図のように、外界からの情報取得部にネットワーク自身が働きかける能動型構造が提案されている^{22)~24)}。本節では、筆者らの行った視点の移動に基づく能動型構造による顔画像の認識について説明する。

3.1 能動型構造と能動的認識

図7に、津村らの提案した顔の認識を行う能動型構造を示す²⁵⁾。パターン照合、顔同定、パターン識別、カメラ制御の4つの機能モジュールと、各機能モジュールを接続し、視点の移動による能動的認識過程を推論する関係モジュールで構成されている。

以下に各モジュールの動作とともに能動的認識

特集：オプトエレクトロニック画像処理技術

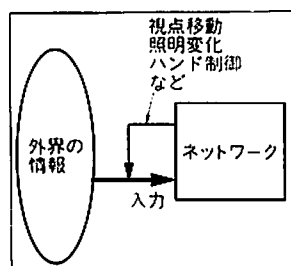


図6 能動的認識構造の概念図

処理の要点を説明する。詳しくは文献8)を参照されたい。ユニット間の結合荷重の学習はすでに完了しているものとして説明する。特徴ユニットを出力層に持つパターン照合モジュールは、モジュールのもつ複数の参照パターンと現在の視点においてカメラにより得られた局所パターンとの照合結果を関係モジュールを通して、顔同定、パターン識別モジュールに伝える。局所パターンの照合結果を受け取った顔同定モジュールは、得られた特徴ユニットの値を基に顔の同定を行う。認識すべき顔に対して、複数の候補がある場合には、候補となる顔の中からどれかの顔を仮説として取り上げる。その仮説を基に、顔同定モジュールと関

係モジュール間の並列分散処理により、次に視点を移すべき局所パターンの頭部、鼻、口などのクラスの検出が行われる。パターン識別モジュールは、顔同定モジュールと同じく関係モジュールから局所パターンの照合結果を受け取り、局所パターンがどのクラスに属するかを識別する。また、次に視点を移すべき局所パターンのクラスも識別する。現在捉えている局所パターンのクラスから、次に視点を移すべき局所パターンのクラスへの視点の移動に関する推論は、パターン識別モジュールと関係モジュール間の並列分散処理により行われる。その結果は、カメラ制御モジュールに伝達され、カメラの位置、ズームレンズの調整が行われ、次の局所パターンがネットワークに入力される。以上の認識サイクルは認識対象が同定できるまで行われる。

3.2 顔の認識実験

津村らは、認識対象として61名の平常顔の画像を用いて能動型構造の有効性を確認している。図8に、61名のうち4名の顔に関する通常の画像と、同じ人の顔から、カメラの移動とズーム操作を想定して切り出した特徴的な複数の局所パターンから

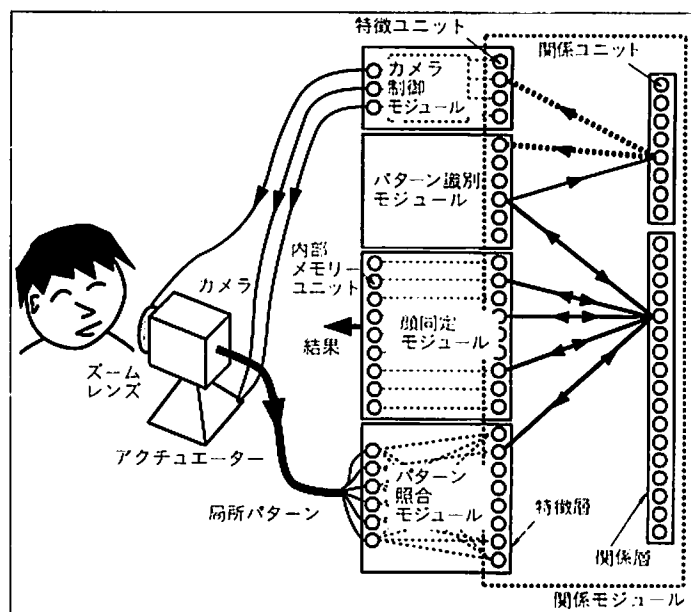


図7 顔の認識のための能動型構造

から構成されるパターン木¹³⁾を示す。パターン木は、通常、画像の特徴抽出などにより獲得される。ここでは、簡単なパターンマッチング処理により、1人につき7個の局所パターンをあらかじめ用意し、カメラ制御モジュールの出力に従ってパターン照合モジュールに提示している。

図8のパターン木内の各局所パターンの左側に、パターン照合モジュールにおいてその入力に対して活性化する特徴ユニットのシリアル番号を記している。シリアル番号の同じ局所パターンは、パターン照合モジュールにおいて区別することができない。図8には、また、認識処理が完了するまでの

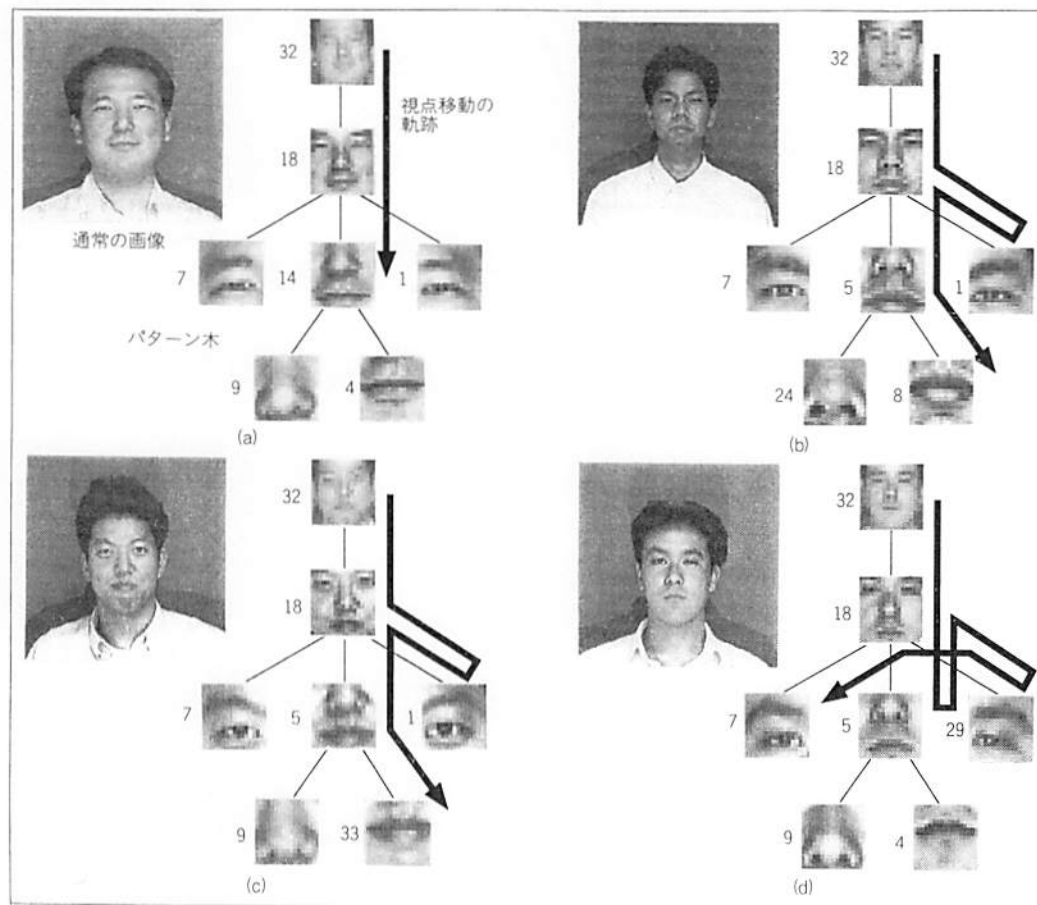


図8 選択的な視点の移動による顔の認識結果

視点移動の軌跡を各パターン木上に示している。図8(a)の顔を認識する場合には、2度目の視点移動で選択的に鼻と口の部分のパターンに視点を移動している。これは、並列分散処理による視点移動の推論により鼻と口の部分のパターンが検出されたためである。図8(b)と(c)の顔は、ユニットのシリアル番号から分かるように、鼻の部分と口の部分のパターンのみが、互いに異なる特徴ユニットを活性化させる。その結果、最終的に口のパターンまで視点の移動を行っている。一度、右目と眉の部分のパターンに視点を移動しているのは、図8(d)の顔でないことを確認するためである。図8(d)は、無駄な視点の移動をしているように見えるが、これは図8に示さなかった複数の

顔が、図8(d)と大部分の局所パターンで似ているために行われたものである。以上のように、並列分散処理による能動的な顔の認識において、選択的な視点の移動による効率的な認識処理系が実現されている。

4. むすび

ネットワークの構造と画像認識能力の関係を中心に据えて、ニューラルネットによる画像認識技術について概説した。能動的構造に関しては、筆者らの行った認識モデルにおける選択的な視点移動の例を示した。しかし、今回示した能動型構造は、基本的な選択的動作を行えているに過ぎない。今後、実際に実世界で動作する装置を考えた

特集：オプトエレクトロニック画像処理技術

場合、位置不変階層型構造やモジュール型構造をうまく取り入れた統合的な構造の構築が望まれる。

参考文献

- 1) D. G. ファイテルソン (光演算研究会訳): 光コンピューティング, 森北出版, (1991)
- 2) H. F. Jorden and M. J. Murdocca: "Scanning the Special Issue on Optical Computing Systems", Proc. IEEE 82, pp. 1591~1594 (1994)
- 3) D. E. Rumelhart, J. L. McClelland, and PDP Research Group: Parallel Distributed Processing, MIT Press (1986)
- 4) R. A. Brooks: "Intelligence without representation", Artificial Intelligence 47, pp. 139~160 (1991)
- 5) 北野宏明: 「AI における科学革命」, 人工知能学会誌 8, pp. 744~751 (1993)
- 6) P. Smolensky: "Connectionist AI, Symbolic AI, and the Brain", Artificial Intelligence Review 1, pp. 95~109 (1987)
- 7) G. E. Hinton, ed.: Connectionist Symbol Processing, MIT Press (1991)
- 8) 津村徳道, 伊東一良, 一岡芳樹: 「並列分散処理による能動的画像認識モデル: 顔の認識」, 光学 23, pp. 750~757 (1994)
- 9) G. Cybenko: "Approximations by superpositions of a sigmoidal function", Math. Control, Signals, and Systems 2, pp. 303~314 (1989)
- 10) B. Irie and S. Miyake: "Capabilities of Three-layered Perceptrons", Proc. IEEE International Conference on Neural Networks 1, pp. 641~648 (1988)
- 11) D. E. Rumelhart, G. E. Hinton, and R. J. Hinton: "Learning internal representations by back propagating errors", Nature 323, pp. 533~536 (1986)
- 12) K. Fukushima: "Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position", Biological Cybernetics 36, pp. 193~202 (1980)
- 13) Y. LeCun, B. Boser, J. S. Denker, D. Hendeorn, R. E. Howard, W. Hubbard, and L. D. Jackel: "Back-propagation Applied to Handwritten Zip Code Recognition", Neural Computation 1, pp. 541~551 (1989)
- 14) W. Zhang, K. Itoh, J. Tanida, and Y. Ichioka: "Parallel distributed processing model with local space-invariant interconnections and its optical architecture", Applied Optics 29, pp. 4790~4797 (1990)
- 15) W. Zhang, A. Hasegawa, K. Itoh, and Y. Ichioka: "Image Processing of Human Corneal Endothelium Based on a Learning Neural Network", Applied Optics 30, pp. 4211~4217 (1991)
- 16) 岩田 彰, 當麻孝志, 松尾啓志, 鈴木宜夫: 「大規模4層ニューラルネットワーク Comb NET」, 電子情報通信学会論文誌 J73-D-11, pp. 1261~1267 (1990)
- 17) T. Kohonen: "Self-Organized Formation of Topologically Correct Feature Maps", Biological Cybernetics 43, pp. 59~69 (1982)
- 18) R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton: "Adaptive Mixtures of Local Experts", Neural Computation 3, pp. 79~87 (1991)
- 19) M. I. Jordan and R. A. Jacobs: "Hierarchical Mixtures of Experts and the EM Algorithm", Neural Computation 6, pp. 181~214 (1994)
- 20) 「複数のニューラル・ネット・モデルを統合した学習システム」, ニューロコンピュータ最前線, 日経 BP 社, pp.102~113 (1988)
- 21) 浅田 聡: 「アクティブ・ロボットビジョン」, 光技術コンタクト, 30, pp. 20~27 (1992)
- 22) K. Fukushima: "Neural network model for selective attention in visual pattern recognition and associative recall", Appl. Opt. 26, pp. 4985~4992 (1987)
- 23) K. Nakano, M. Niizuma, and T. Omori: "Model of Neural Visual System with Self-Organizing Cells", Biological Cybernetics 60, pp. 195~202 (1989)
- 24) 西美保子, 大関一徳, 桜井紀之, 大森隆司: 「記号とパターン統合による画像理解のための一方式——パターン的な知識と記号的な知識の相互変換——」, 計測自動制御学会論文集 30, pp. 1368~1377 (1994)
- 25) P. J. Burt: "Smart Sensing within a Pyramid Vision Machine", Proceedings of the IEEE 76, pp.1006~1015 (1988)