

博士学位論文

ネオコグニトロン型神経回路モデルの
生理学的モデルとしての評価と工学的応用

平成23年6月

九州工業大学大学院生命体工学研究科

吉塚 武治

目次

第1章 序論	1
1.1 本研究の背景と目的	1
1.2 視覚に関する生理学的知見	2
1.2.1 大脳皮質の領野	2
1.2.2 視覚情報処理の経路	4
1.2.3 V1野	4
1.2.4 V2,V4,IT野	6
1.3 パターン認識器としての階層型ニューラルネットワーク	8
1.3.1 パーセプトロン	8
1.3.2 コグニトロン	11
1.3.3 ネオコグニトロン	12
1.3.4 選択的注意機構のモデル	15
1.4 本論文の構成	16
第2章 生理学的モデルとしてのネオコグニトロン	19
2.1 はじめに	19
2.2 視覚腹側経路モデルとしてのネオコグニトロン	21
2.3 ネオコグニトロンと従来のIT野のモデル	25
2.4 実験	28

2.4.1	回転, 拡大縮小, 並進移動に対する U_{C5} 細胞の反応	31
2.4.2	分割画像に対する U_{C5} 細胞の反応	33
2.5	受容野の空間結合分布の推定	36
2.6	まとめ	38
第3章	ネオコグニトロンの GPU 実装	40
3.1	はじめに	40
3.2	GPU と CUDA	42
3.3	GPU によるネオコグニトロンの実装	44
3.4	実験	46
3.5	まとめ	50
第4章	選択的注意機構による人検出	53
4.1	はじめに	53
4.2	選択的注意機構のモデル	55
4.2.1	求心性経路の細胞	56
4.2.2	遠心性経路の細胞	57
4.2.3	ゲインと閾値制御	58
4.2.4	選択敵注意機構のセグメンテーションの例	62
4.3	実験	64
4.3.1	実験環境	65
4.3.2	実験結果	66
4.4	まとめ	76
第5章	結論	81

謝辞	88
業績リスト	100

第1章 序論

1.1 本研究の背景と目的

「百聞は一見に如かず」という諺があるように、人間の視覚による情報の割合は、非常に大きく、視覚83%、聴覚11%、臭覚35%、触覚1.5%、味覚1.0%と言われている [1]. このように、外界の情報の入力器官として、視覚は重要な感覚であり、このような能力をコンピュータによって実現する事は、医療、監視システム、バーチャルリアリティ、コンピュータアニメーション、ヒューマンインターフェイス、ロボット等、広範囲な分野で応用可能である。画像処理技術の多くは、何かしらの制約条件に拘束され、環境の変化に対して不安定であった。しかし、我々人間の視覚情報処理は、環境の変化にも柔軟に対応可能である。従って、人間の脳の生理学的知見を基に、システムを構築する事は、ある意味、自然な発想である。過去にも、ニューラルネットワークを用いた画像処理技術は提案されてきたが、ネットワークの膨大さと計算コストの問題等で、応用可能な状態までには至っていない。又、安易にニューラルネットワークを用いるというコンセプトでシステムを構築する事が大半であった。そこで、我々は、人間の脳の視覚情報処理の生理学的知見を用いたニューラルネットワークであるネオコグニトロンに着目した。ネオコグニトロンは、脳の生理学的知見を用いて構築されたニューラルネットワークと

して Fukushima が提案したモデルである [2]. ネオコグニトロンは、位置ずれや、変形などの影響に頑強であるモデルとして有名であり、人間が有している柔軟な処理を実現している。本研究では、従来、工学的な手法のみでは困難であった人物認識等を実現させることを目指し、ネオコグニトロンの生理学的モデルとしての評価と工学的応用を試みる。そこで、まず、我々は、人間の脳の認識や認知を司る視覚腹側経路のモデルとして、ネオコグニトロンが妥当であるかを検討する。また、ネオコグニトロンのネットワークは、膨大であり、計算機コストに問題がある事が知られている。そこで、ネオコグニトロンの並列性に着目し、GPU による計算量のコスト削減を行う。さらに、ネオコグニトロンにフィードバック経路を追加した選択的注意機構のモデルを用いて、複雑な背景下における人の検出を行い、ネオコグニトロンの応用アプリケーションとして可能性を示す。

1.2 視覚に関する生理学的知見

本説では、本研究を理解するのに必要な前知識として、視覚に関する脳の知見と、大脳皮質の視覚腹側経路と呼ばれる経路に関して簡単に説明する [3].

1.2.1 大脳皮質の領野

大脳は、脳の中で、重要な役割を果たしており、大脳には大脳皮質と呼ばれる薄いタンパク質の層で表面が覆われている。その皮質の均質性の違いによって、大脳皮質をいくつかの領野に分類する試みが成されて

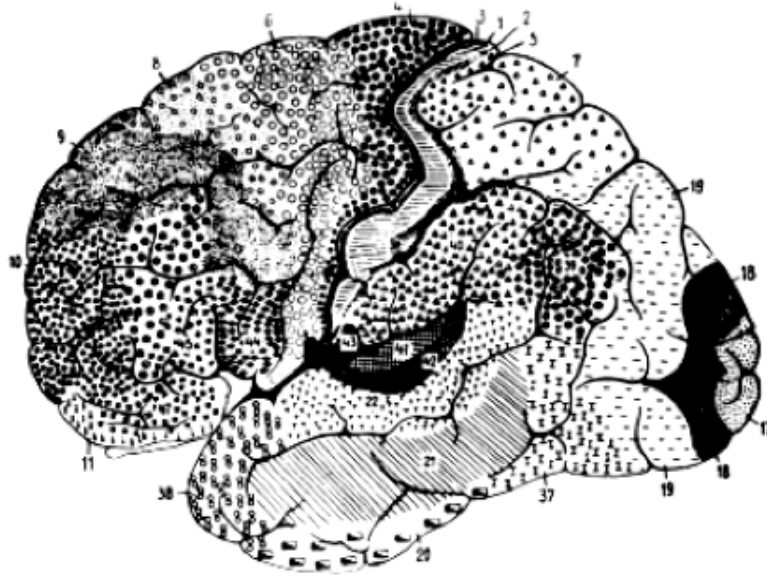


図 1.1: Brodman の脳地図 [4] .

きた. Brodman の脳地図 [4] が有名であり, 人の大脳皮質は 52 の領野に分類され, 各領野に番号がつけられている (図 1.1).

一方, 機能的にみると, 大脳皮質の各部分はそれぞれ異なった機能を分担しており, 随意運動の制御に関する運動野 (motor area), 皮質感覚や体の深部感覚の知覚に関係する体性感覚野 (somato-sensory area), 視覚情報処理を行っている視覚野 (visual area), 聴覚情報処理を行っている聴覚野 (auditory area) など, 各機能の中枢がそれぞれ異なった場所を占めている. 視覚野は, 左右の大脳半球の後部 (後頭葉) にある. Brodman の脳地図の 17 野 (area 17; striate area; striate cortex) が第 1 次視覚野 (V1=visual area I) に対応する. その前方に位置する 18 野 (area 18) および 19 野 (area 19) を視覚前野 (prestriae visual cortex) と呼ぶ. さらに視覚前野は, 細分化され, V2 (第 2 次視覚野; visual area II), V3, V3A, V4, MT (middle temporal

area), MST(medial superior temporal area)などと名づけられる領野に分けられている。視覚に関連した領野は、後頭葉だけでなく、下部側頭葉(inferotemporal cortex; 側頭連合頭の下部に位置する)の後半部のPIT野(posterior inferotemporal area), 中下部のCIT (central inferotemporal), 前半部のAIT(anterior inferotemporal area), 頭頂葉(頭頂連合野)の7a野, さらには前頭葉などにも及ぶことがわかってきている。なおPIT野, CIT野, AIT野を纏めてIT野(inferior temporal area)と呼ぶこともある。

1.2.2 視覚情報処理の経路

Van Essenらは、領野間に階層的な結合があると推定している[9]。眼球に入った信号は、網膜からまず外側膝状体(LGN)を通して脳の後頭部へと向かい、大脳皮質の一次視覚野(V1)と呼ばれる領域に達する。ここで2つの経路に分かれる。一つは腹側経路と呼ばれる経路であり、大脳半球の両側の基底部にあたる下部側頭皮質に至る、V2野→V4野→IT野(PIT, CIT, AIT)という経路である(図1.2)。もう一つの経路である背側経路は、大脳半球の頭頂部にあたる後部頭頂皮質に至る、V2野→MT野→MST野という経路である。機能的には、腹側経路が物体の”形”の認識に関する主要な経路であり、一方、背側経路は”動き”に重要な役割を果たしている。以下で腹側経路の各領域野に関して述べる。

1.2.3 V1野

HubelとWieselは、V1野における細胞を受容野の性質に基づいて単純型細胞と複雑型細胞などに分類し、これらの細胞相互間には、網膜→外

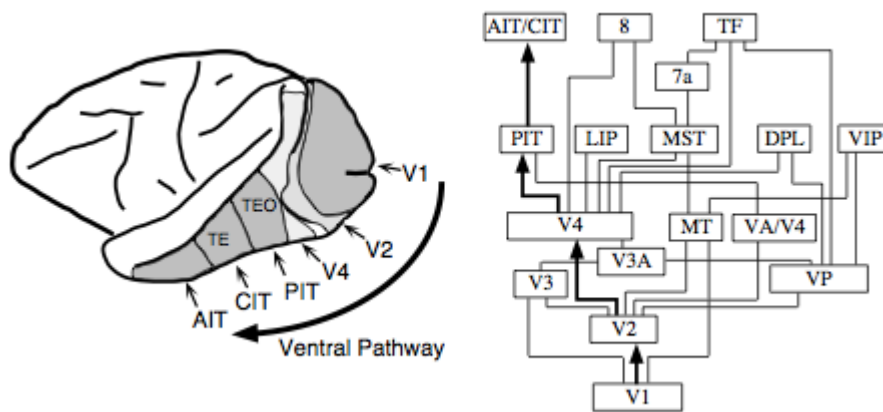


図 1.2: 領野間の階層的な結合関係と腹側経路 ([9] を一部変更).

側漆状体→単純型細胞→複雑型細胞という階層構造が存在するという階層仮説を提唱している [5, 6, 7, 8]. 単純型細胞は直線やエッジに反応するような受容野を持つことが知られている. 図 1.3 は, Hubel と Wiesel の実験によって得られた単純型細胞の受容野の例を示したものである. 図中の + 印の位置に対して刺激が入力されると発火し, - 印の部分に刺激が入力されると抑制される. これらの細胞は + と - の配置から考えて, あ

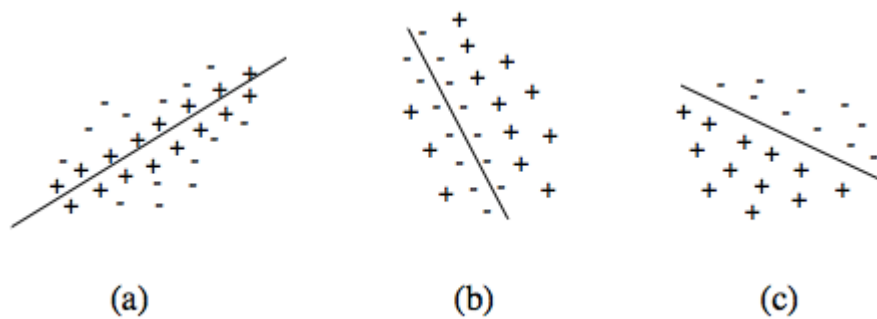


図 1.3: 単純型細胞の受容野の例 ([6] を一部変更).

る特定の傾きを持った直線やエッジを検出するのに役立っていると考えられる。ある細胞が最も強く反応する直線やエッジの方位を、その細胞の最適方位 (preferred orientation) と呼ぶ。受容野の最適方位は、細胞によって異なり、ある特定の最適方位を持った細胞が多く存在するようなことはなく、あらゆる方位のものが一様に存在していると言われている。単純型細胞では、刺激として与える直線やエッジの位置が少しでもずれると出力が抑制されてしまう。これに対して、複雑型細胞では、最適方位の直線やエッジが受容野の内部にあれば、その位置に関わらず反応するという特性を持つ。Hubel と Wiesel は、同一の最適方位を持つ複数の単純型細胞を興奮性結合によって結び付けられているモデルを複雑型細胞のモデルとして提案している [6]。このモデルにおいて、複雑型細胞は結合している単純型細胞のいずれか一つでも反応すれば複雑型細胞も発火するものと考えられている。

1.2.4 V2,V4,IT 野

V2, V4, IT 野については V1 野ほど詳しくは分かっていない。各領野の細胞の持つ受容野の大きさは、階層が上がるにつれ受容野が大きくなり、特異的に反応する特徴も複雑なものになると言われている [13, 11]。Desimone らは、AIT 野の受容野特性について調べており、線分の長さ、入り組んだ輪郭、色の組合せ、特定のテクスチャなどに特異的に反応する細胞を発見している [14]。IT 野の細胞は視覚パターンの形状認識に特に重要な領野と考えられている。IT 野を破壊されたサルは、図形弁別学習に重い障害が現れるが、視覚以外の聴覚、触覚、嗅覚による弁別学習



図 1.4: AIT 野の細胞が特異的に反応する図形パターンの例 ([21] を一部変更).

には何らかの異常は起こらないと報告されている [15]. また IT 野や, その近傍の STP 野 (superior temporal polysensory area) では, ヒトの顔やサル顔に特異的に反応する細胞が発見されている [16]. このような細胞は, 顔から目を除去したものや, 顔の画像を 16 分割して並べ替えたものには反応しない. これらの細胞はパターンの大きさが変化したりしても反応に変化はないが, 顔の向きの影響は大きく, 正面からみた顔と横からみた顔とで大きく反応が変化する [17, 18, 19]. また, Tanaka らが画像から IT 野の細胞が強く反応する特徴を探しだし, その特徴をより簡単な特徴に置き換えて最適刺激特徴を探索するという手法を用いて, AIT 野の細胞の受容野特性について調べている. それによると, 図 1.4 のようなパターンに特異的に反応する細胞が発見されている [20, 21].

1.3 パターン認識器としての階層型ニューラルネットワーク

本説では、パターン認識器としてのニューラルネットワークの紹介と、本研究で用いたネオコグニロン型神経回路モデルの概要を説明する。ニューラルネットワークを大別すると、階層型と相互結合型に分ける事ができる。本説では、階層型ニューラルネットワークに関して言及する。

1.3.1 パーセプトロン

パーセプトロンは、1950年代に、Rosenblattによって考案されたパターン認識可能なニューラルネットワークである [22, 23]。図 1.5 は、パーセプトロンの構造を示したものであり、S(Sensory)層、A(Association)層、R(Response)層と呼ばれる、3つの層からなっている。このうちS層からA層への結合は固定結合で、A層からR層の結合のみが学習により決定される。S層のユニットは網膜の視細胞のようなものを想定しており、外界からの刺激パターンは、この層に入力される。A層とR層のユニットは前の層のユニットからの重みつき入力 of 総和を計算し、これに適当な出力関数をかけたものを出力する。S層の*i*番目のユニットの出力を s_i 、A層の*j*番目のユニットの出力を a_j とし、S層の*i*番目のユニットからA層の*j*番目のユニットへの固定結合を c_{ji} とすると、 a_j は式 (1.1) で表される。

$$a_j = f \left(\sum_i c_{ji} s_i - \theta_j \right) \quad (1.1)$$

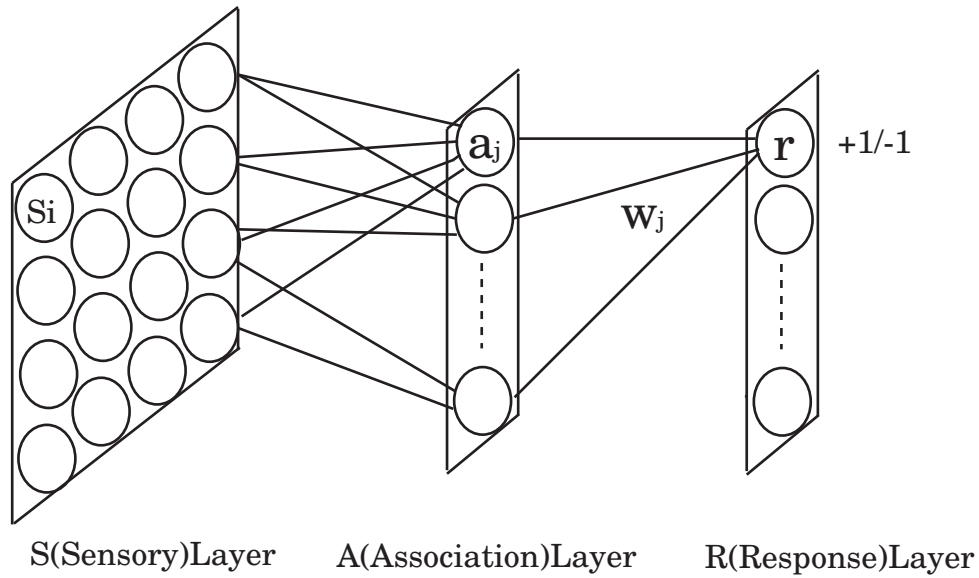


図 1.5: 3層パーセプトロン.

とかける. ただし, θ_j は A 層の j 番目のユニットの閾値であり, $f(x)$ は, 式 (1.2) で表される関数である.

$$f(x) = \begin{cases} 1 & (x > 0) \\ 0 & (x \leq 0) \end{cases} \quad (1.2)$$

R 層のユニットの出力を R , A 層の j 番目のユニットから R 層のユニットへの結合を w_j とすると R は, 式 (1.3) となる.

$$r = \text{sign} \left(\sum_j w_j a_j - \theta \right) \quad (1.3)$$

ここで, θ は R 層のユニットの閾値であり, sign は式 (1.4) で表される符号関数である.

$$\text{sign}(x) = \begin{cases} +1 & (x \geq 0) \\ -1 & (x < 0) \end{cases} \quad (1.4)$$

したがって、S層に入力されたパターンは $\{-1, +1\}$ という2種類の記号に変換されることになる。Rosenblattが提案したパーセプトロンには、誤り訂正学習と呼ばれる教師あり学習則が用いられている。この学習法則は、入力されたパターンの集合を正しく2組 $\{-1, +1\}$ に分類しようというものである。この学習則は、あるパターンが入力され、なおかつR層の出力が間違っている場合に、更新式(1.5)で重み w_j と閾値 θ を修正する。

$$\begin{aligned} w_j &= w_j + \eta r a_j \\ \theta &= \theta - \eta r \end{aligned} \tag{1.5}$$

ここで、 η は、正の定数である。Minskyらは特に学習に関連するA層とR層のみに特化した単純パーセプトロンモデルを用いて誤り訂正学習の収束定理を証明し、単純パーセプトロンの能力の限界について言及している[24]。彼らは、単純パーセプトロンをある種の論理関数を計算する機械としてとらえ、どのような学習をしても関数を計算できないことを指摘している。これは多層神経回路モデルの必要性を述べているとも言える。多層パーセプトロンの学習則はAmariによって言及され[25]、1986年にRumelhartらによって再発見された[26]。これが良く知られているバックプロパゲーション(back propagation)と呼ばれる学習則である。この学習則はユニットの出力関数を微分可能な関数に置き換え、ある結合係数を少しだけ変化させたときに出力に対してどれくらい影響があるかを計算する。この操作を全ての結合係数に対して行い、望ましい結果に向かうように結合を微小変化させる手法である。Rosenblattのパーセプトロンでは、A層の出力関数が微分不可能な関数だったために、S層からA層の結合を変化させることが不可能であるが、これを微分可能な出力関数に置き換えることによって微分のチェインルール(chain rule)が適用

でき、結合の微小変化に対する応答の変化が計算できるようになる。バックプロパゲーションを用いたパーセプトロンは、文字パターン認識等で評価され、線型分離不可能な問題に対して効果的であることが示されている。

1.3.2 コグニトロン

Fukushima の提案したコグニトロンは多層の神経回路モデルであり、各ユニットが持つ結合に空間的な局所性を持たせたモデルである (図 1.6)[27]. コグニトロンの各ユニットは前層のユニットの小領域に存在するユニット群からの重みつき入力を受け取る。この結合可能な小領域は前層の全体を覆うほど大きくはない。しかし、層を重ねるごとに入力層 u_0 のより広い範囲からの情報を受けることになり、最終的な層のユニットは入力層 u_0 全体からの情報を受けようになっている (図 1.6(a)). 図 1.6(b) はコグニトロンの基本構造を示したものであり、隣り合った 2 つの層 u_{l-1} と u_l 層のユニット間の結合のうち、 u_l 層の 1 つのユニットに至る結合のみを抜き出している。ユニット $u_l(\mathbf{n})$ は結合している小領域 A_l に含まれるユニット群 $u_l(\mathbf{n} + \boldsymbol{\nu})$ から $a_l(\boldsymbol{\nu}, \mathbf{n})$ という重みをかけた入力を受け取る (ただし $\boldsymbol{\nu} \in A_l$). ユニット $v_l(\mathbf{n})$ は、抑制性のユニットで、やはり領域 A_l から $c_l(\boldsymbol{\nu})$ という重みをつけた入力を受け取る。 $v_l(\mathbf{n})$ は抑制性の信号として $b_l(\mathbf{n})$ という重みをつけて $u_l(\mathbf{n})$ に伝達される。 $a_l(\boldsymbol{\nu}, \mathbf{n})$ と $b_l(\mathbf{n})$ が学習によって決定される結合である。コグニトロンの学習則は、Fukushima が提案した最大値検出型仮説に基づいた競合学習が用いられている [27]. あるユニットの出力がその近傍のユニットの出力に比べて大きいときに

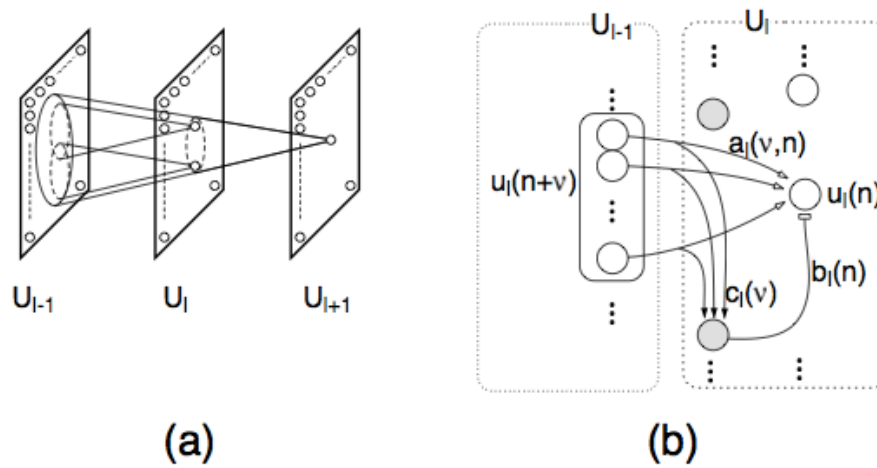


図 1.6: (a):コグニトロンにおける細胞間の結合の広がり, (b):コグニトロンの基本構造 [27].

のみ結合の更新が起こるといふものである.

1.3.3 ネオコグニトロン

前節で述べたパーセプトロンやコグニトロンは, 学習させたパターンを別の場所に提示したり, パターンの拡大・縮小を行ったりすると全く別のパターンと判断されてしまう. Fukushima は, このような欠点を取り除いた神経回路モデルであるネオコグニトロンを提唱した [2, 28, 29]. ネオコグニトロンは, S細胞と名付けられたユニットの層 U_S と C細胞と名付けられたユニットの層 U_C とが交互に並べたような階層型の神経回路モデルである (図 1.7). S細胞は Hubel と Hiesel が発見した単純型細胞をモデル化したものであり, C細胞は, 複雑型細胞をモデル化したものである. S細胞又はC細胞を2次元平面状に並べたものを細胞面, 各細胞

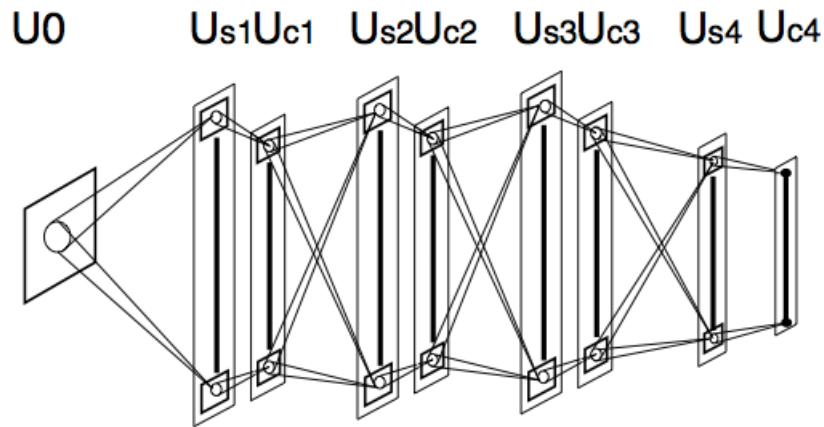


図 1.7: ネオコグニトロン²の階層構造 [2].

面を複数並べたものを細胞層，S細胞層とC細胞層をまとめたものを細胞段と呼んでいる，S細胞の役割は，特徴抽出であり，C細胞の役割は，位置ずれの許容である．また，段が上がるにつれ，細胞面の細胞密度が小さくなる(受容野が大きくなる)ように設計されており，最終層の細胞面では，1つの認識細胞のみとなる．各細胞間の結合はコグニトロンと同様に空間的な局所性を持っている．

$$U_S = \phi[\textit{Similarity} - R] \tag{1.6}$$

$$\textit{Similarity} = \frac{X \cdot Y}{|X||Y|}$$

S細胞は，学習ベクトル X と入力ベクトル Y を仮定すると，式(1.6)のように方向余弦(類似度)で表される．ここで， R は閾値であり， ϕ は半波整流型の非線形関数である．S細胞は，低次の段では，線分のような単純なパターンを抽出し，段が上がる毎により複雑なパターンを認識するように設計されている．また，S細胞の入力結合は可塑性を持ち，コグニトロンと同様に競合学習によって自己組織化される．C細胞は，S細胞の出力に

ガウス関数のような単調減少な関数を畳み込みことにより、信号をぼかしている。ここで、C細胞がない場合を考えてみよう。パターンは同じであるが、位置がずれている場合にC細胞がないと、Similarity(類似度)は0となってしまう。つまり、位置は違うが本来同じパターンであるものを異なるパターンであるとネオコグニトロンは認識してしまう。そこで、C細胞で信号をぼかすという操作を行うことによって、類似度は大きな値をとる事となり、位置ずれを許容する事になる。また、入力パターンそのものに対してぼかしの操作を行うと、パターンの微細構造を崩してしまう。従って、特徴抽出を行った後にぼかすという、一連のシーケンシャルが存在している。また、一度に位置ずれを許容しようとする、同じ特徴を有しているが全く異なったパターンも認識してしまう可能性がある。したがって、多階層で、少しずつ位置ずれを許容する事により、位置ずれから生まれる認識率の低下を包括している。多階層における、S細胞の特徴抽出とC細胞のぼかし処理と、細胞密度の減少(受容野サイズの拡大)という処理要素によって、ネオコグニトロンは、パターンの変形に比較的頑強なパターン認識を可能としている。また、最近のネオコグニトロンでは、コントラスト層と呼ばれる層が、入力層と第1段目に組み込まれ、輪郭の抽出を行っている [29]。また、近年、ネオコグニトロンと類似した多階層型のモデルが多く提案され、LecunのCNN(convolutional neural network)やPoggioらの”Standard Model”が有名であり [45, 39]、パターン認識器として良い結果を残している。

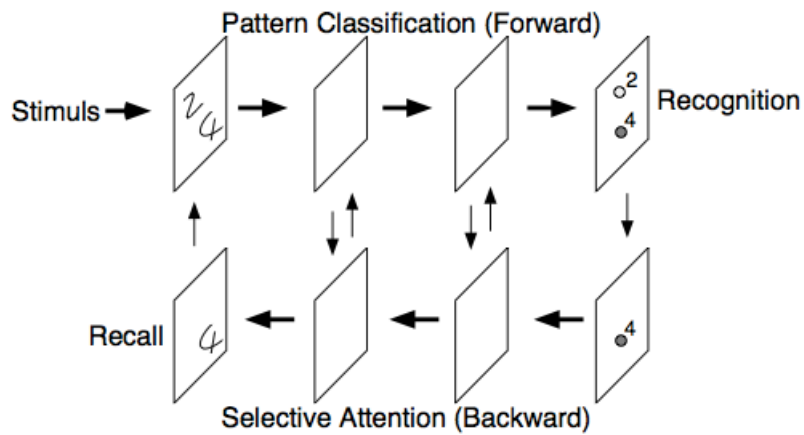


図 1.8: 選択的注意機構の概念図 [31].

1.3.4 選択的注意機構のモデル

階層性を持つ神経回路モデルとしてパーセプトロン、コグニトロン、ネオコグニトロンを例としてあげてきたが、これらの結合は一方向に信号を伝えるフィードフォワード型ネットワークであった。フィードフォワード型ネットワークは、ボトムアップ的なネットワークであるが、これとは逆にフィードバック結合を持つようなトップダウン的なネットワークも考えることが出来る。フィードバック型ネットワークの歴史は Rosenblatt の逆結合型パーセプトロンに始まる [23]。このモデルは、図 1.5 に示すようなパーセプトロンの層間の結合がフィードフォワードの結合だけでなく、R 層から A 層に信号を伝えるフィードバック結合を持つモデルである。このようなフィードバック結合があるとパーセプトロンはパターン中に含まれる特定の特徴に注意を向けるような性質を持つようになる。Fukushima はこの事実に基づき、選択的注意機構のモデルを提案している [31]。選択的注意機構のモデルは、図 1.8 のような 2 方向の信号伝達経路を持つ。

フィードフォワードの信号伝達経路は、パターン認識を行う経路として考えられる。すなわちパターンからシンボルへの変換を行っているものと言える。一方、フィードバックの経路はフィードフォワードの系とは逆にシンボルからパターンへの逆変換を行っているものと捉えることができる。すなわち認識している特定のパターンのみを抽出していることになる。工学的な応用面で考えれば複合パターンを構成する特定パターンの切り出しが可能になることを意味する。既に、工学的応用として、英字筆記体の連結文字列や [33, 34]、顔の部分パターンの切り出し [35] が行われている。

1.4 本論文の構成

本論文では、第2章でネオコグニトロンの生理学的モデルの妥当性を検討する。また、工学的応用の観点から、第3章でGPUを用いたネオコグニトロンの高速化について述べ、第4章で選択的注意機構のモデルを用いた人の検出について示す。

第2章では、ネオコグニトロンの各階層を構成する構造が生理学的にも妥当であることを示す。Logothetisらは、サルにゼムクリップ状の物体を回転させたり拡大縮小させた画像を見せて、IT野の細胞の反応を記録している [36]。彼らの実験とネオコグニトロンの最終層の反応との比較を行い、非常に良く似た反応が得られることを示す。また、BricoloとPoggioやRisenhuberとPoggioらも、このIT野のモデルを提案している [37, 38, 39]。しかし、彼らのモデルでは結合の局所性と皮質を構成する細胞の一様性については説明できない。ネオコグニトロンは細胞の持つ

結合は局所的で、構造に一様性を持つので彼らのモデルより近似されたモデルであると考えられる。

第3章では、GPUを用いたネオコグニトロン的高速化について述べる。ネオコグニトロンは、その膨大なネットワークの構造から計算コストの問題が懸念されている。近年GPUによる汎用計算の高速化技術が注目され多くの分野で広く利用されている。Poliらは、ネオコグニトロンをGPUで実装している。彼らの実装は、細胞間の結合データに関して並列実装を行い、高速化を実現しているが、細胞面サイズに関する考慮が十分ではない。そこで我々は、細胞面サイズを考慮した並列実装を行なった。その結果、密(要素の多くが0ではない)である入力データに対して、ネオコグニトロンの全ての階層にGPU実装を行うのではなく、細胞面サイズが大きい低次の階層にのみGPU実装するのが良い事を示す。

第4章では、選択的注意機構のモデルを用いた、人の検出に関して述べる。選択的注意機構のモデルは、英字筆記体の連結文字列、顔の部分パターンの切り出し等で応用可能な事がわかっている。近年、計算機能力の向上に、膨大な画像データを扱えるようになり、画像解析、画像識別、画像理解等の高度な情報処理が求められている。中でも、歩行者検出や、顔認識等、人に関連したものが、やはり実社会では要求が多い。選択的注意機構のモデルは、認識器としてのネオコグニトロンにフィードバック経路を追加する事によって、複数パターンからの1パターンの切り出し、パターンの欠落部分の修復、ノイズの除去等の認識における補助的機能を実現している。そこで、選択的注意機構のモデルのそのような機能を用いて、複雑な背景下から人の検出が可能ではないかと我々は考えている。オコグニトロンが物体の変形に比較的強いと言われている

が、あまりに複雑な形状の変化に対して許容できる訳ではない。人の形状は姿勢(各関節部)の変化によって無数に存在する。今回は、人の“立つ”という姿勢に限定し、人の形状の1つを学習パターンとして、学習パターンとは異なる人の形状を呈示し、その際のフィードバック経路の出力部に人の形状がどのように想起されるかを調べる。

第5章では、本研究の結論と将来の展望について述べる。

第2章 生理学的モデルとしての ネオコグニトロン

2.1 はじめに

サルの視覚情報処理過程には，大まかにいって二つの経路が存在することが生理学的，解剖学的に知られている．これらの経路は背側経路と腹側経路と呼ばれ [10]，物体の動きなどの認知は背側経路で処理され，物体の形状などの認知は腹側経路で処理される．サルの腹側経路は階層構造をもち，視覚情報は，網膜 → LGN → V1 野 → V2 野 → V4 野 → IT 野と処理されることが知られている [11]．

人やサルの IT 野の細胞（特に AIT 野といわれる領域の細胞）は非常に複雑な図形に対して特異的に反応することが知られている．Perret らは腹側経路の高次領域である AIT 野の細胞が顔に反応することを報告している [12]．Logothetis らはこれらの細胞の特性を調べるためにマカクザルにゼムクリップ状の 3 次元物体の画像を提示し，微小電極を用いて，そのときの IT 野の細胞の反応を測定した (図 2.1)[36]．

さらに彼らは物体を回転させた画像，拡大縮小させた画像，並進移動させた画像を提示し，これらの細胞がコードしている情報を調べた．それによると，変形に対して不変な細胞は少なく，ほとんどの細胞がある

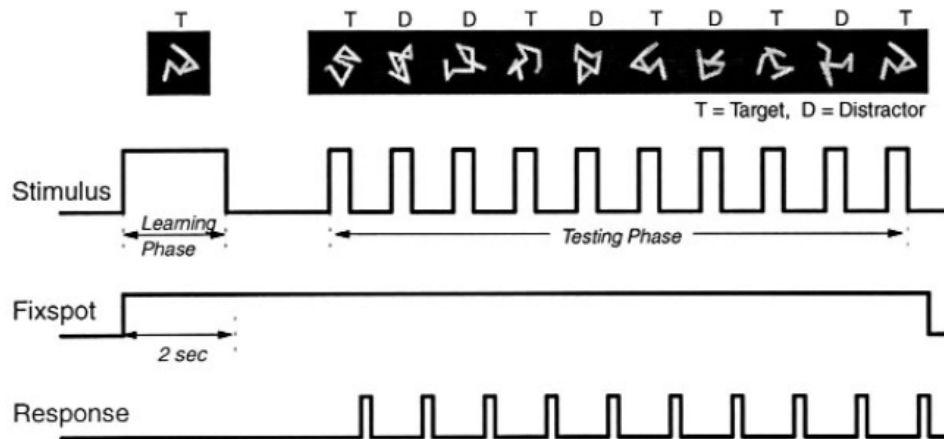


図 2.1: 弁別訓練されたマカクザルに様々な角度からターゲットオブジェクト (T) とデストラクタオブジェクト (D) をみせ IT 野での反応を測定 [36].

特定の画像に対して反応し、物体の回転などによって起こる変形の度合によって反応が徐々に小さくなるという釣鐘状の反応特性を示す事が報告されている。これを受けて、Bricolo, Riesenhuber らは、IT 野のモデルとして RBF (radial basis function : 放射基底関数) ネットワークを用いたモデルを提案し、Logothetis らの実験結果を説明した [37, 38, 39].

一方、ネオコグニトロンは、Fukushima によって提唱された階層型の神経回路モデルの一種であり、刺激パターンの位置ずれや、変形などの影響を受けにくいパターン認識のシステムを構築できることが知られている [2, 29]. そこで、ネオコグニトロンが工学的な視覚情報処理システムとして高い能力を有するだけでなく、脳の腹側経路の神経回路モデルとしても妥当かどうかを調べることにより、ネオコグニトロンの生理学的モデルとしての可能性を検討する。Bricolo[37] や Riesenhuber[38, 39]

らは RBF ネットワークを用いて、反応特性の定性的な性質を説明しており、前処理を V1 野相当、V4 野から IT 野に至る部分を RBF ネットワークとして扱っている。しかしながら、彼らのモデルでは脳の領野の各々が異なったネットワーク構造および異なった学習ルールによって構築されていることになる。これに対してネオコグニトロンは均質なネットワーク構造で同一の学習ルールによって構築される視覚システムとして捉えることができ、Logothetis らの実験を定性的にでも説明できるとすれば、Bricolo や Riesenhuber らのモデルよりも脳の構造としてはより妥当であることが期待できる。

本章では、2.2 節で腹側経路モデルとしてのネオコグニトロンを生理学的背景と我々の仮説を踏まえ説明し、2.3 節でネオコグニトロンと従来の IT 野のモデルとの比較を行う。2.4 節で Logothetis らが報告した IT 野の細胞の反応特性を定性的にネオコグニトロンで説明できることを示し、ネオコグニトロン腹側経路モデルとしての妥当性を検討する。さらに、2.5 節ではモデルから予測されることとして、V2 野に相当する高次視覚野の細胞がもつ受容野の空間結合分布を推定する。

2.2 視覚腹側経路モデルとしてのネオコグニトロン

Hubel と Wiesel の報告によれば、V1 野には主に 2 種類の細胞が存在し、それらは単純型細胞と複雑型細胞と呼ばれている [5, 6]。ネオコグニトロンは、S 細胞と C 細胞と呼ばれる 2 種類の細胞から構成される [2, 29]。単純型細胞のモデルである S 細胞は特徴を抽出し、複雑型細胞のモデルで

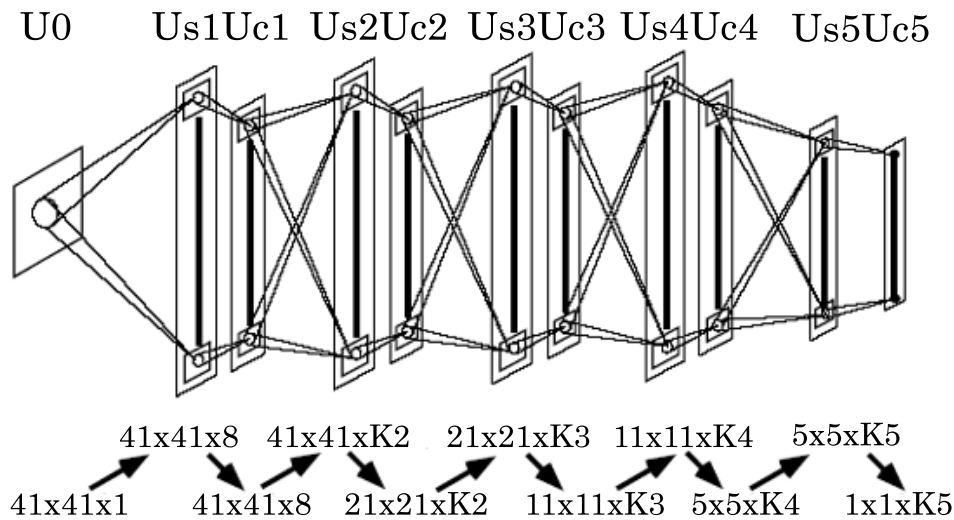


図 2.2: ネオコグニトロンの構造.

ある C 細胞は位置ずれを許容する.

図 2.2 にネオコグニトロンの構造を示す. ネオコグニトロンは, いくつかの細胞段と呼ぶ構造から構成され, 一つの細胞段は S 細胞の層 U_{Sl} と, C 細胞の層 U_{Cl} の 2 層構造からなる. 添字の l は, 階層の番号を示す. 各階層 U_{Sl} および U_{Cl} は細胞面と呼ばれる構造から構成され, 細胞面は同じ特徴の抽出を行う細胞を 2 次元平面状に並べたものである. 図 1 の下段にある数字は, 各細胞層が持つ細胞面の大きさと抽出を行う特徴数が記してある. 例えば U_{S1} 層の細胞層は, 8 個の方位の線成分を 41×41 程度の解像度で検出することを表している. 無論, 解剖学的な立場をとれば皮質内に細胞面のような構造は陽に存在するわけではない. V1 野において細胞はエッジや線分などの微小な特徴に反応することが知られているが, これらの細胞はランダムに配置されているわけではなく, 似た特徴に対して反応する細胞が柱状に配置されている. この構造はカラム構造

と呼ばれる。視覚皮質の大部分はカラム構造からなり、この構造はV1野やV2野だけでなくIT野などでも観察されると報告されている [21]。V1野において、カラムの位置は提示された刺激の位置をコードしており網膜上での像の位置関係（レチノトピー）が保存されるように配置されている。V1野のようにレチノトピーが保存されている領野において、細胞面は異なる位置にある同一な最適方位を持ったカラムを抽出し、一枚のシート状にレチノトピーを保存するように並べたものであると考えることができる [30]。

ここでS細胞、C細胞について少し詳しく述べる。第 l 段目の k 番目の細胞で、細胞面内での位置が \mathbf{n} であるS細胞の数式表現は以下の式で与えられる。

$$u_{Sl}(\mathbf{n}, k) = \frac{\theta_l}{1 - \theta_l} \cdot \varphi \left[\frac{1 + \sum_{\kappa} \sum_{\nu} a_l(\nu, \kappa, k) \cdot u_{Cl-1}(\mathbf{n} + \nu, \kappa)}{1 + \theta_l \cdot b_l(k) \cdot u_{Vl}(\mathbf{n})} - 1 \right] \quad (2.1)$$

$a_l(\nu, \kappa, k)$ は U_{Cl-1} 層の κ 番目の細胞面からの結合を表し、関数 φ は半波整流形の非線型関数で表される。 a_l と b_l は学習によって決定される結合係数である。抑制性入力である u_{Vl} は以下の式で与えられる。

$$u_{Vl}(\mathbf{n}) = \sqrt{\sum_{\kappa} \sum_{\nu} c_l(\nu) \cdot \{u_{Cl-1}(\mathbf{n} + \nu)\}^2} \quad (2.2)$$

ただし、 $c_l(\nu)$ は固定の結合定数であり、 $\|\nu\|$ に関して単調減少な結合である。一方、第 l 段目、 k 番目のC細胞面の位置 \mathbf{n} におけるC細胞の発火は以下の式で与えられる。

$$u_{Cl}(\mathbf{n}, k) = \psi \left[\sum_{\nu} d_l(\nu) \cdot u_{Sl}(\mathbf{n} + \nu, k) \right] \quad (2.3)$$

d_l は同じ l 段目の S 細胞面からの結合で, $\|\nu\|$ に関して単調で非負な結合荷重を持つ. 関数 ψ は以下で表される.

$$\psi[x] = \frac{\varphi[x]}{1 + \varphi[x]} \quad (2.4)$$

関数 φ は, 式 (2.1) と同様に半波整流形 of 非線型関数である.

ネオコグニトロン of 細胞間の結合は空間的に局所的な結合で構成されている. V1 野に対応すると考えている U_{S1} 層では局所的な特徴しか抽出しないように設計しており, 高次 (l が大きい) になるにつれ, この受容野の大きさは徐々に大きくなるようにし大域的な特徴を抽出できるようにしている. 最終層における細胞の受容野は入力面全体をカバーする大きくなる. 実際のマカクザルの V2 野と V4 野と IT 野における受容野の大きさは Tanaka と Kobatake や Rolls によって, 領野が進むにつれ大きくなることが報告されており, IT 野では, ほぼ視野全体をカバーする受容野を持つことが知られている [11, 40]. また, 図 2.2 からわかるように, S 細胞層から C 細胞層への計算を行う際に細胞面の大きさを縮小し, 細胞の密度を低くするように設定している. C 細胞は, S 細胞面の局所的な平均を計算する細胞であるので, 隣接する細胞は似たような出力を出す傾向にある. したがって, 細胞の密度を徐々に低くしても, パターンのもつ特徴の位置情報は失われにくい.

さらに単純型細胞と複雑型細胞の関係については様々なモデルが提案されているが, Sakai と Tanaka は V1 野の生理学実験の結果を説明するモデルとしてカスケードモデルとノンカスケードモデルとを比較し, カスケードモデルを妥当なものとしている [41]. カスケードモデルとは Gabor フィルタのような異方性をもつフィルタの出力値を非線型関数で変調し,

それを統合することによって方位選択制を実現するモデルである。これに対してノンカスケードモデルはDOGフィルタのような等方性フィルタの出力を特定の線分方向に積をとってから統合するようなモデルである。ネオコグニトロンにおけるS細胞の役割はフィルタリングと非線型関数による変調であり、C細胞の役割はさまざまな位置のフィルタ出力の和をとることであるので、ネオコグニトロン各段、すなわち U_{Sl} 層と U_{Cl} 層の組合せはカスケードモデルの一種であると考えることができる。それゆえ、ネオコグニトロンの低次段 ($l = 1$ など) はV1野の性質をよく表しているものと考えられる。

残念ながらV2野以降の構造はV1野ほど明らかにされていないが、皮質の均質性などを考えれば、それほど急激に性質が変化しているとも考えにくい。よって、カスケードモデルのような処理をするプロセスがシーケンシャルに並んだものを腹側経路の主たる計算プロセスと捉えてもそれほど不自然ではないと考えられる。

2.3 ネオコグニトロンと従来のIT野のモデル

我々の研究に先立ってBricoloらは、IT野のモデルとしてRBF (radial basis function: 放射基底関数) ネットワークを用いたモデルを提案し、Logothetisらの実験を説明しようと試みた(図2.3)[37]。彼らのモデルでは各階層間は非常に広範囲な結合によって実現されているが、現実の腹側経路における各領野間の結合は局所的なもので構成されており、彼らのモデルにおける中間層と実際の腹側経路を構成している領野との対応関係がとれない。最終的にはIT野の細胞は全視野をカバーする大きさの

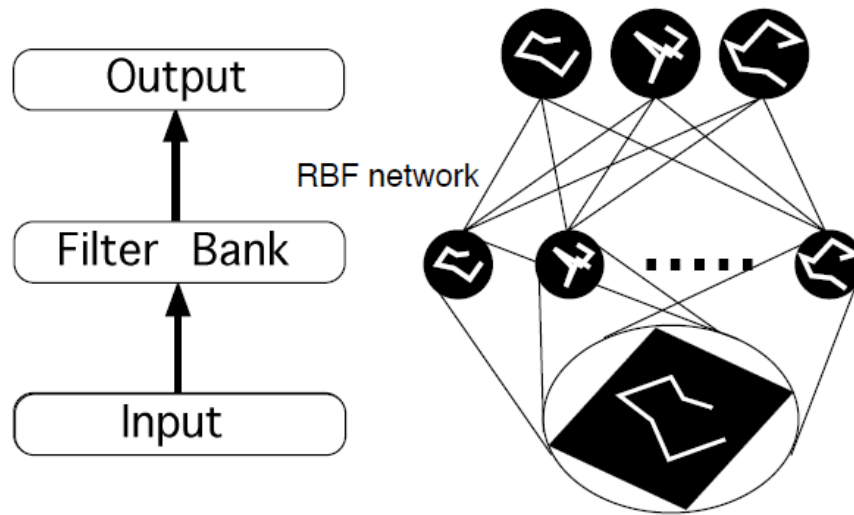


図 2.3: Bricolo らの IT 野のモデル [37].

受容野を持つようになるが，V1 野や V2 野などの受容野は局所的なものでレチノトピーが存在している [11]．このような脳の各領野の階層性を考慮すれば，形状を認識する腹側経路のモデルは，階層性をもち，各階層で局所的なイメージの情報を徐々に統合していくようなモデルであるほうが望ましい．Riesenhuber らは Bricolo らのモデルを改良し，より腹側経路のモデルとして近いものを提唱した [38]．彼らは，モデルの第 1 のプロセスであるフィルタリングのプロセスを，レチノトピーを持つ局所的なフィルタリングと抽出した部分特徴を統合するプロセスとで置き換え，それを統合した特徴を RBF ネットワークへの入力として与え，この局所的なフィルタリングのプロセスを V1 野のモデルとして位置付け，統合するプロセスを V4 野のモデルとして考えた (図 2.4)．しかし，実際の腹側経路において各領野は解剖学的には分類されているが，構成してい

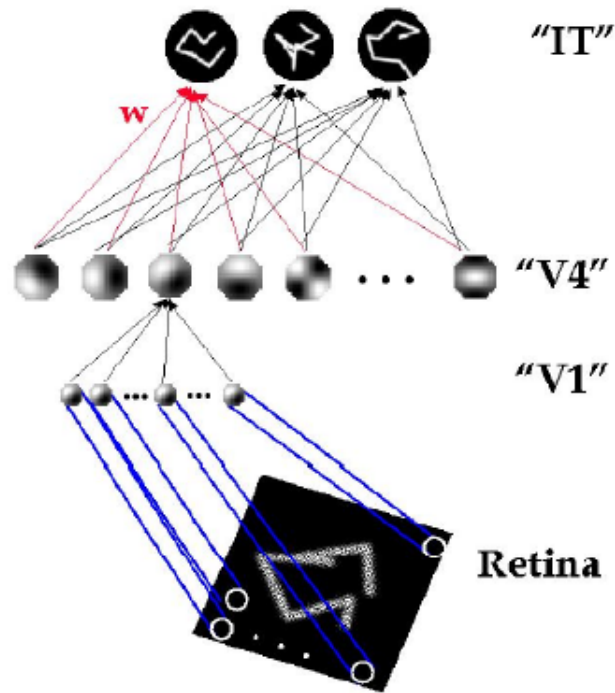


図 2.4: Riesenhuber らの IT 野のモデル (1)[38].

る皮質という意味では均質であると考えられるので、各階層をモデル化したときに、その階層毎に計算手法が著しく異なるようなモデルは望ましくない。さらに、Riesenhuber らは、単純型細胞と複雑型細胞の概念を取り入れ、特徴抽出を行なう S 細胞、それらを統合する C 細胞を交互に二つ並べた階層構造を含む IT 野のモデルを提案した (図 2.5)[39]。フィルタリングと統合というプロセスが二階層になった事によって、以前のモデルよりも均質的になったが、RBF との計算プロセスは異なったままである。これに対して我々が提案するネオコグニトロンによる IT 野のモデルでは、各階層の基本的な構造は一様であり、特徴抽出 (S 細胞) と許容 (C 細胞) という均質なプロセスである。ネオコグニトロンは、Bricolo,

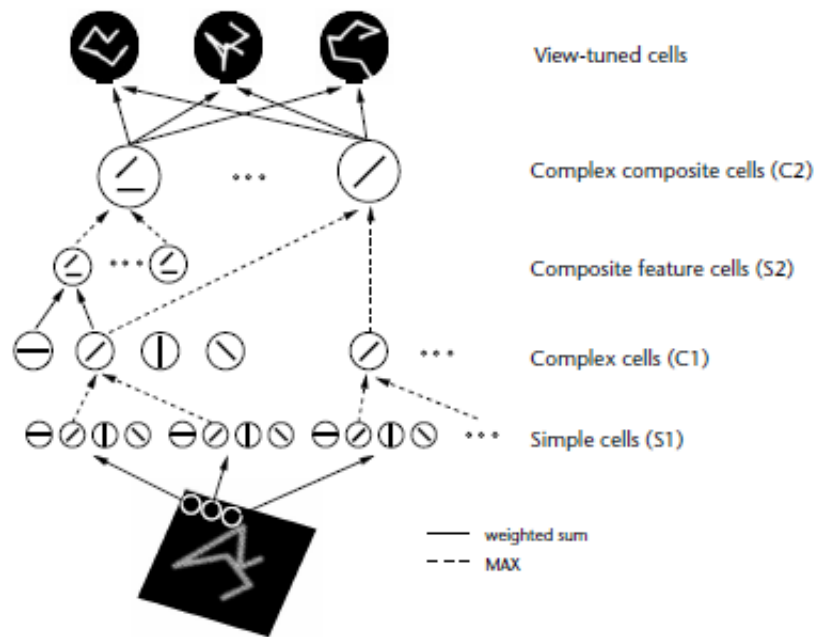


図 2.5: Riesenhuber らの IT 野のモデル (2)[39].

Riesenhuber らのモデルより，階層性，局所性，均質性の条件を十分に満たしており，生理学的な構造という意味で妥当であると我々は考える．さらに機能的な妥当性を検証するために，Logothetis らが報告した IT 野の細胞の反応特性を定性的に良く説明できることを示す．

2.4 実験

本章の実験では，Logothetis らの実験と同様にゼムクリップ状の物体をネオコグニトロンに提示し学習させ，学習した物体を回転させたもの，拡大縮小させたもの，並進移動させたもの，さらにディストラクタ（誤認識物体）を提示し，ネオコグニトロンに含まれる細胞の反応を調べた．

表 2.1: ネオコグニトロンのパラメータ.

l	0	1	2	3	4	5
S-cell plane size	-	41^2	41^2	21^2	11^2	5^2
C-cell plane size	41^2	41^2	21^2	11^2	5^2	1^2
$a_l(\boldsymbol{\nu}), c_l(\boldsymbol{\nu})$ size	-	7^2				
$d_l(\boldsymbol{\nu})$ size	-	5^2				
$c_l(\boldsymbol{\nu})$	-	$c_l(\boldsymbol{\nu}) = 1$ for all $\boldsymbol{\nu}, l$				
$d_l(\boldsymbol{\nu})$	-	$d_l(\boldsymbol{\nu}) = 1$ for all $\boldsymbol{\nu}, l$				
θ_l	-	$\theta_l = 0.53$ for all l				

さらに, Riesenhuber らが行った分割画像を用いた実験 [39] を行い, 彼らのモデルとネオコグニトロンとの比較を行った. 表 2.1 に本実験で用いたネオコグニトロンのパラメータを示す. Logothetis らの実験結果をより良く説明するように細かな調整を行うことも可能である. しかし, 本実験ではなるべく単純なパラメータを用いて, 彼らが示した IT 細胞の釣鐘状の反応特性を定性的に再現できるかを調べた.

V1 野の単純型細胞のもつ空間フィルタ特性は Gabor 関数で近似でき, 直線成分やエッジを抽出する [42]. したがって, V1 野の単純型細胞に対応する U_{S1} 層は, 入力された画像に含まれる 8 方位の直線成分を検出するように作り付けている. U_{S1} 層への入力結合である $a_1(\boldsymbol{\nu}, \kappa, k)$ は固定結合とし, $c_1(\boldsymbol{\nu})$ にはガウス関数で近似された結合分布特性を持つフィルタを用いた. これ以外の興奮性の可変結合である $a_l(\boldsymbol{\nu}, \kappa, k)$ ($l > 1$) は, ある種の競合学習によって決定される [43]. 同一層内の細胞面で同じ受容野

(中心を \mathbf{n}^* とする) をもつ細胞間で競合させ、最も反応値が高い細胞を $u_{Sl}(\mathbf{n}^*, k^*)$ とする。この場合、反応の最大値を持つものが勝者となるので以下の更新式によって結合が強化される。

$$\begin{aligned} a_l(\boldsymbol{\nu}, \kappa, k^*) \\ = a_l(\boldsymbol{\nu}, \kappa, k^*) + q \cdot u_{Cl-1}(\mathbf{n}^* + \boldsymbol{\nu}, \kappa) \end{aligned} \quad (2.5)$$

一方、抑制性の可変結合である $b_l(k)$ は以下の式によって決定した。

$$b_l(k) = \sqrt{\sum_{\kappa} \sum_{\boldsymbol{\nu}} \frac{\{a_l(\boldsymbol{\nu}, \kappa, k)\}^2}{c_l(\boldsymbol{\nu})}} \quad (2.6)$$

このように結合を決定することにより、S細胞はそれ自身が最も特異的に反応する特徴 $a_l(\boldsymbol{\nu}, \kappa, k)$ と入力特徴 u_{Cl} との距離を方向余弦で測るようになる [3]。また、可変結合 a_l , b_l の初期値はすべて 0 とし、 $q = 10.0$ と定めた。

図 2.6 は本実験に用いた入力画像のサンプルを示す。この画像を生成するために、まず 60 個のゼムクリップ状の物体を計算機上で生成し、第 1 番目の物体 (図 2.6 (a) を参照) を学習させる物体と定めた。残りの 59 個の物体はディストラクタとして用いた。図 2.6 (b) はディストラクタの一部を示す。次に、これらの物体の画像を、コンピュータグラフィックスを用いて生成した。その際、学習させる物体に関しては、物体を回転させた画像や、拡大縮小させた画像、並進移動させた画像を生成した。このうち $\pm 90^\circ$ の範囲で 30° 刻みで回転させた物体の画像と、5 種類のディストラクタを混在させたものを学習パターンとして定めた。この学習パターンを用いて、ネオコグニトロンの変結合である a_l と b_l を決定した。

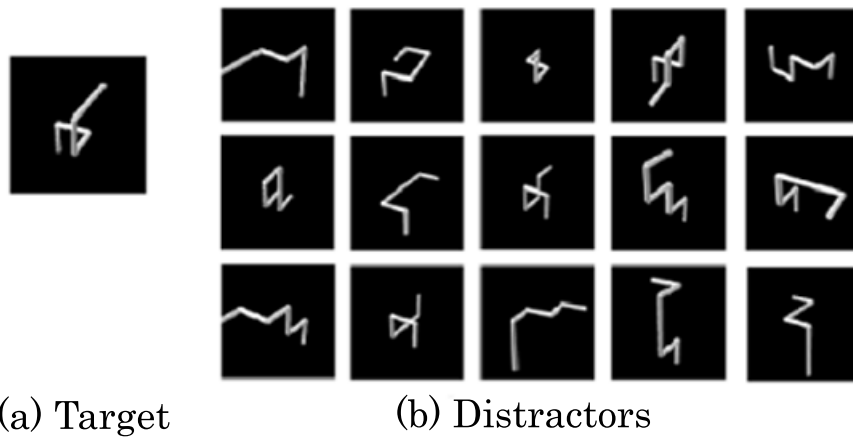


図 2.6: ターゲットオブジェクト (目標物体) とディストラクタオブジェクト (誤認識物体) の例.

2.4.1 回転, 拡大縮小, 並進移動に対する U_{C5} 細胞の反応

学習後, U_{C5} 層には複数の認識細胞が生成され, その中のある細胞の, 物体の回転, 拡大縮小, 並進移動に対する反応を調べた.

図 2.7 は, 物体の回転に対するある細胞の反応例である. この細胞は -30° 付近の画像に対し特異的に反応していることがわかる. また, 図 2.7 の棒グラフからわかる通り, ほとんどのディストラクタには反応しないか, 反応があってもその値は非常に小さい.

図 2.8 は, 物体の拡大縮小と S 細胞の閾値変化 (各段の閾値は全て同値) に対するある細胞の反応例である. この細胞では, 閾値が $\theta = 0.44$ の場合において, ディストラクタレベルは, 0.45 程度となり, おおよそ, 0.7 倍から 1.7 倍が有効範囲であり, 反応は鋭い釣鐘型の形状となった. 一方で, 閾値が $\theta = 0.94$ の場合においては, ディストラクタレベルは, 0.75 程度となり, $\theta = 0.44$ とほぼ同様な 0.7 倍から 1.8 倍が有効範囲であり,

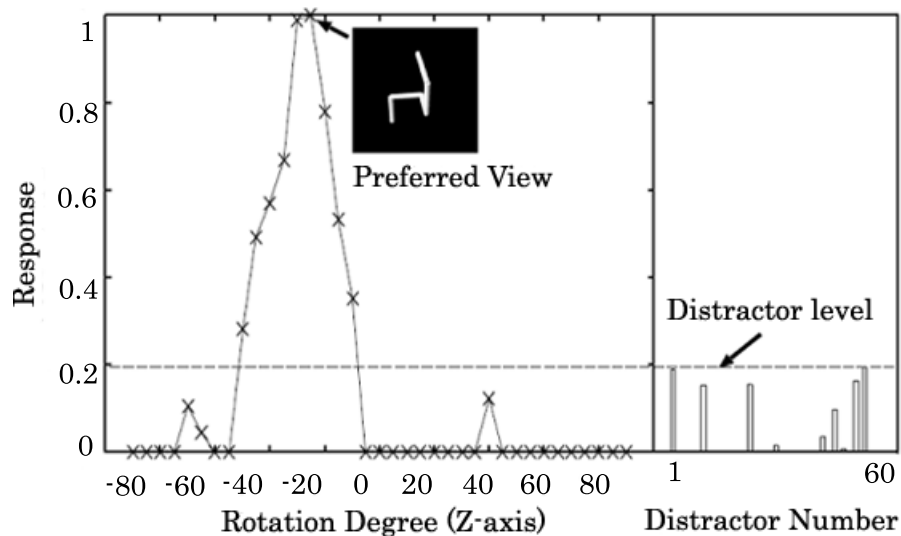


図 2.7: ターゲットオブジェクトを回転させた画像での最終層のある細胞の反応: 縦軸は, U_{C5} 層のある細胞の反応値 ([0-1] で正規化, 1 は最適反応) を示す. 横軸は, z -軸に対する回転角とディストラクタオブジェクトの番号を示している.

なだらかな釣鐘型の形状となる. つまり, 閾値を大きくすれば, 反応は鋭い形状となりディストラクタレベルは下がる. 閾値を小さく設定すれば, なだらかな形状となりディストラクタレベルは上がる. このように, Logothetis らの実験結果をより良く説明するように細かな調整を行うことも可能である.

図 2.9 は, 物体の並進移動に対するある細胞の反応例である. ネオコグニトロンは, 結合が並進対称性をもつので, C 細胞においては物体の提示位置が変化しても, 細胞の出力がそれほど変化しないが, パターンの一部が画像範囲外に出ると値は下がっているのがわかる.

この実験においてネオコグニトロンの U_{C5} 細胞の反応特性は比較的幅

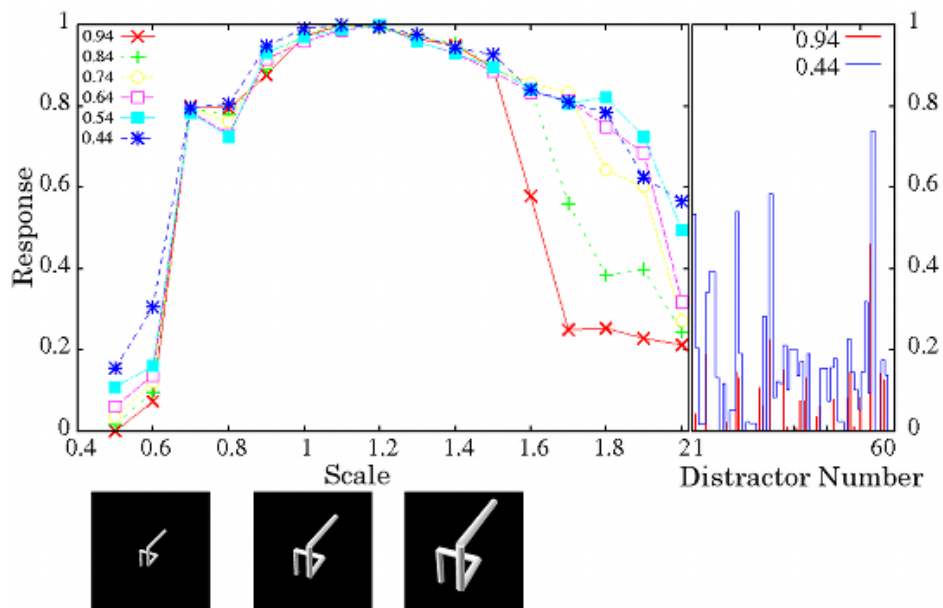


図 2.8: ターゲットオブジェクトの拡大縮小と S 細胞の閾値変化に対する細胞の反応例: 縦軸は U_{C5} 層のある細胞の反応値 ([0-1] で正規化, 1 は最適反応) を示し, 横軸は拡大率とデストラクタオブジェクトの番号を示す. なお, 各段の閾値は全て同じに設定している.

の広い釣鐘状の形を示しており, Logothetis らの実験結果を定性的によく説明しているといえる. また構造上の観点からいっても腹側経路のモデルとしての意義が大きい.

2.4.2 分割画像に対する U_{C5} 細胞の反応

図 2.10 は Riesenhuber らが行った分割画像の実験 [39] をネオコグニトロンで行った結果を示したものである. 分割画像は, 彼らの実験と同様に, 入力画像を 4, 16, 64, 256 個のタイルに分割し, ランダムにシャッ

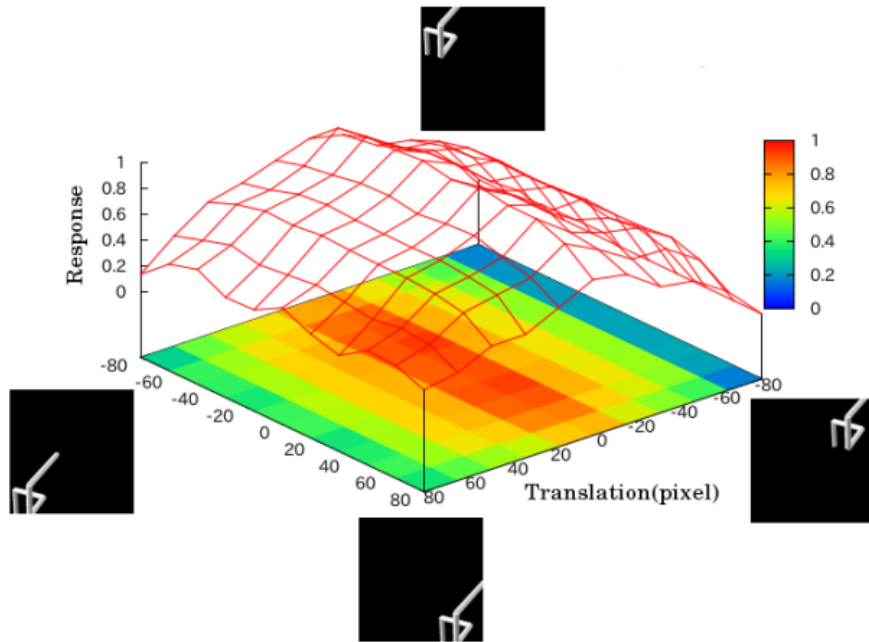


図 2.9: ターゲットオブジェクトの平行移動に対する細胞の反応例: 縦軸は, U_{C5} 層のある細胞の反応値 ([0-1] で正規化, 1 は最適反応) を表し, 並進されたピクセル数を 2 次元平面上に示し, 更に色によって反応値をプロットしている.

フルして生成した. さらに, 前節で用いた 60 個の画像で実験を実施し, 学習パターンに反応する細胞の反応の平均と分散を計算した. 図 2.10 の縦軸が反応を表し横軸が分割数を表す. 4 分割された画像においてネオコグニトロンは, 急激に反応が落ちている. この細胞のディストラクタのレベルはおよそ 0.2 程度なので画像の 4 分割でディストラクタと同じレベルになることがわかる. 一方, Riesenhuber らのモデルにおいてディストラクタのレベルは 0.6 程度なので, ディストラクタとの識別においては同

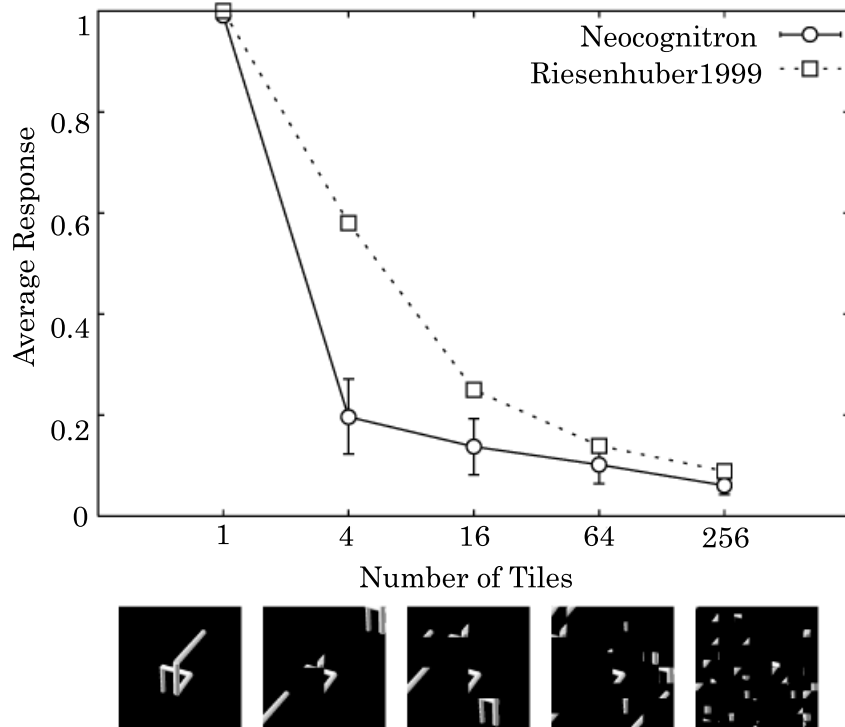


図 2.10: 物体を分割した場合の細胞の反応 (平均と分散):縦軸は図 2.7 と同様であり, 60 個の異なるターゲットオブジェクトの画像を提示した場合の平均された反応を示す. 横軸は, ターゲットオブジェクトの画像の分割数を示している. 実線は, 各分割数 (4,16,64,256) でのネオコグニトロン細胞の平均と分散を表し, 破線は, Riesenhuber らのモデルの結果を表している [39].

程度の結果を示したことになると考えられる。

2.5 受容野の空間結合分布の推定

前節までにおいて Logothetis らの実験をネオコグニトロンが定性的に上手く説明できる事を示した。本節ではこのモデルから予測できることとして V2 野以降の受容野の空間結合分布について考察してみる。

マカクザルなどの V2 野以降の受容野の空間結合分布は、形状が複雑で調査するのが困難である。ネオコグニトロンが腹側経路のモデルであると仮定する事により、V2 野以降の結合の形状をある程度推定できるかもしれない。これは、Fukushima が提案した選択的注意機構のモデル [44] を応用して、ある段の S 細胞から信号を逆行方向に流し、 U_{S1} 層に再生されたパターンを観察することによって行った。選択的注意機構のモデルの遠心性経路をそのまま用いると、求心性経路にパターンを提示していないためどの部分が発火しているのかを辿ることができず、さらに、 U_{S1} 層から U_{C1} 細胞への結合 d_i が信号を拡散させる働きを持つため、 U_{C1} 層から U_{S1} 層へ逆伝播させる際に信号が空間的にぼける。そこで、本推定方法は U_{S1} 層においてラプラシアン型の核関数と畳み込むことによって信号の尖鋭化を行い、形状を推定する工夫を施している。

図 2.11 はこの逆再生の結果を示す。1 行目は U_{S1} 層に含まれる S 細胞の空間結合分布であり、直線を抽出するような結合が再生されている。前章のはじめで説明した通り、 U_{S1} 層はあらかじめ作り付けられた直線成分を検出する層である。逆再生の結果がそれと同じであるので、この方法による推定が妥当であるといえる。2 行目は U_{S2} 層に含まれる S 細胞の空

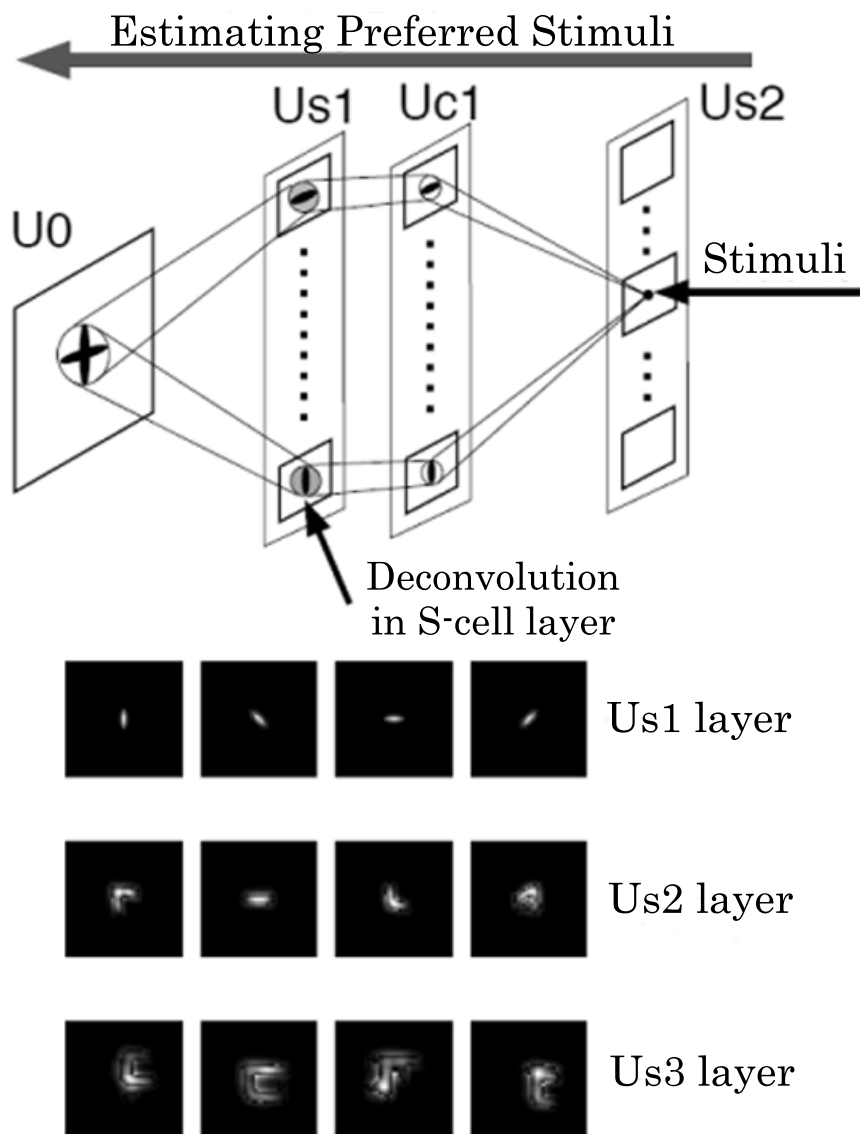


図 2.11: 受容野の空間結合分布の推定結果の例.

間結合分布を示す。これらの細胞は1行目の U_{S1} 層の細胞よりも複雑な特徴をコードしていることがわかる。3行目は U_{S3} 層のS細胞の空間結合分布を示しており、物体の角の部分などを中心に見たときの画像らしきものがコードされているのがわかる。

もし、腹側経路がネオコグニトロンでよく近似できるのであれば、V2野以降の細胞がもつ受容野の空間結合分布を上述のような方法によって推定でき、実験結果のように、V1野の形状を組み合わせたような受容野がV2野以降に存在するかもしれない。

2.6 まとめ

我々はネオコグニトロンを用いた腹側経路のモデルを提案した。このモデルは、V1野の単純型細胞と複雑型細胞のモデルを拡張したものであり、S細胞、C細胞の組み合わせは、SakaiとTanakaが示した知見[41]を支持している。我々のモデルを用いて、ゼムクリップ状の3次元物体の画像の認識シミュレーションを行なった結果、Logothetisらの実験結果[36]を定性的によく表していることが判明した。これらの結果は、ネオコグニトロンの腹側経路モデルとしての意義が大きい事を示している。また、ネオコグニトロンを用いる事で、V2野以降の受容野の複雑な空間結合分布の形状を推定できる可能性を示した。

しかしながら、ネオコグニトロンは、厳密な生理学的モデルとしては、次のような課題を有している。ネオコグニトロンは、結合に並進対称性があることを仮定しているが、この仮定を前提としない学習則の構築を行わなければならない。また、C細胞の結合は学習によって決定してい

るわけではなく、位相不変な応答を出すために作り付けられた固定結合である。したがって、C細胞の様な位相不変な結合を学習によって構成することを考えなければならない。また、陰影やオクルージョンの影響を補正すると思われるフィードバック経路の影響についても議論の余地が残されている。今後、これらの課題を解決し、生理学的モデルとしてより妥当な腹側経路モデルに近づけていく予定である。

第3章 ネオコグニトロンのGPU 実装

3.1 はじめに

近年，産学間の連携が頻繁に行われ，学術世界で得られた知見や技術の産業応用が推進されている．産業へ応用する際，処理時間，コスト，汎用性，ユーザビリティ等の要因を考慮する必要がある．我々は，生理学的知見からモデル化されたネオコグニトロンを基に画像認識アプリケーションの作成を行い，それらの要因に関して産業応用への可能性を検討していこうと考えている．まず，我々はネオコグニトロンの処理時間に注目した．

ネオコグニトロン [2][29] は，福島によって提唱された階層型の神経回路モデルの一種であり，刺激パターンの位置ずれや，変形などの影響を受けにくいパターン認識のシステムを構築できることが知られている．ネオコグニトロンの構造は，図 3.1 のように，細胞段と呼ばれる層によって構成され，細胞段は，単純型細胞のモデルである S 細胞と，複雑型細胞のモデルである C 細胞と呼ばれる 2 種類の細胞からなる 2 層構造に成っている．単純型細胞のモデルである S 細胞は特徴を抽出し，複雑型細胞のモデルである C 細胞は位置ずれを許容する．前章で説明したように，ネ

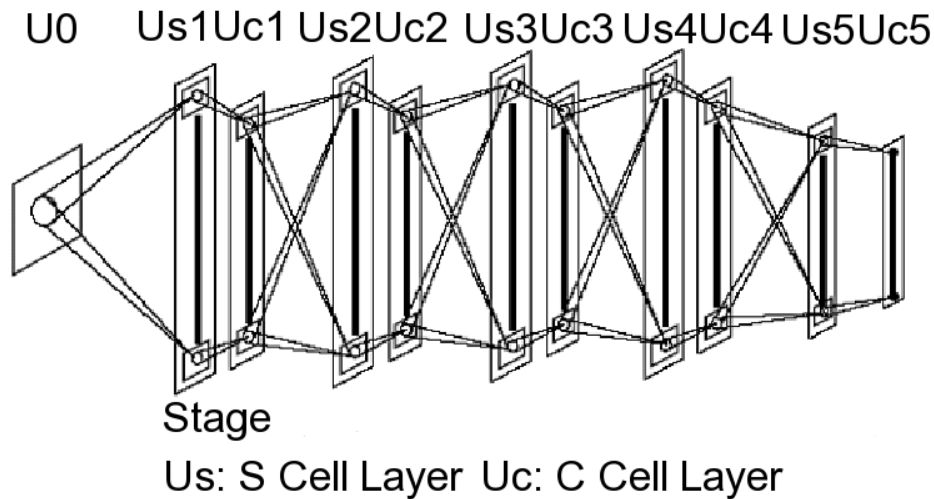


図 3.1: ネオコグニトロンの構造.

オコグニトロンは、視覚対象の認識や認知に関与している脳の視覚腹側経路のモデルとして最も良く近似されており、このモデルを利用することで、人間の持つ高度な視覚情報処理を計算機で再現可能であると期待される。しかし、ネオコグニトロンは膨大なネットワーク構造をしており、計算コストの問題が懸念されている。

GPU(Graphics Processing Unit)は、2000年代初頭に、グラフィック処理専用の演算プロセッサとして開発されたが、現在、その並列演算性能の高さとGPGPU (General-Purpose computing on GPUs) 統合開発環境CUDAの登場によるプログラマビリティの向上により、物理シミュレーションや、数値計算、画像処理・認識などの分野等、グラフィック処理以外の汎用計算にも広く使われるようになった。また、GPUは市販のPC等にも多く標準搭載されており、安価に入手できる。そこで、我々は、並列計算の性能の高さと汎用性からGPUアーキテクチャに沿ってネオコグ

ニトロンを実装する事は価値があると考えている.

ネオコグニトロン的高速化の着眼点として, 2つのデータ量を考える必要がある. 1つ目は細胞間の結合データ, 2つ目は細胞面のサイズである. Poli[52][53]らは, 顔認識にGPU/CUDAで実装したネオコグニトロンを用いている. 彼らの実装は, 細胞間の結合データ量の増加に関する高速化に主眼をおき, 細胞面サイズに関する考慮が十分でない. CUDAは, 上位からGrid, Block, Threadという3階層でスレッドを管理している(図3.2). PoliらはネオコグニトロンをCUDAのスレッドの階層と対応させて, Gridを細胞段, Blockを細胞面, Threadを細胞とみなしている. 更に, 1つの細胞間の結合分布をBlockとしているので, 細胞面サイズが大きくなるとBlock数が大きくなり, 並列化の効率が悪くなってしまう. また, 細胞間の結合データは学習段階によって獲得されるので, 彼らの実装は学習段階の高速化の観点がない. そこで, 本研究では, 細胞面のサイズに関する高速化として, 画像データを複数のBlockに分割するという画像処理分野のGPU実装で良く用いられる手法で高速化を行い, 細胞面サイズの増加に対する速度改善の結果を示す. 本章では, 3.2節でGPUとCUDAに関して簡単に述べ, 3.3節でネオコグニトロンの実装に関して説明する. 次に, 3.4節で実験と考察を行い, 3.5節で結論を述べる.

3.2 GPUとCUDA

GPU(Graphics Processing Unit)は1990年代, Microsoft社のWindowsというOSの登場をきっかけに, GUI(Graphical User Interface)のニーズの

高まりによって、グラフィック専用のプロセッサとして登場した。当初は、グラフィックアクセラレータ (Graphics Accelerator) という呼び名の拡張ボードであった。更には、DirectX[46] の API 群の登場によりグラフィックアクセラレータの性能は強化されていった。当初、グラフィックアクセラレータは、レンダリング (数値データとしての物体や図形の描画) のみしか対応しておらず、Zバッファや、テクスチャマッピング等の機能を次々に搭載していった。2000年代に入ると、座標変換やライティングがハードウェア化され、GPU という呼び名へ変わった。DirectX8の登場以降、自由に計算式を変更できるプログラマブルシェーダがサポートされ、画像処理、ニューラルネットワーク等の計算量が多い分野の汎用計算にも利用され始めた [47, 48, 49]。しかし、レンダリングやシェーディングと言ったグラフィックス特有の知識と Cg(C for Graphics), HLSL(High Level Shader Language), GLSL(OpenGL Shading Language) 等の言語が必要であり、汎用計算としては、敷居が高かった。そこで、GPGPU(General-purpose computing on graphics processing units) と呼ばれる GPU による汎用計算の動きが始まった。2007年、NVIDIA 社によって GPGPU 開発環境 CUDA[50, 51] が登場すると、多くのユーザーが興味を示し、さまざまな分野の研究で、GPU による高速化が研究されるようになった。

図 3.2 は、GPU のアーキテクチャと CUDA のスレッド構造との関係を示したものである。GPU は、SM(Streaming MultiProcessor) と呼ばれる複数の演算装置から構成され、SM は SP(Streaming Processor) と呼ばれる演算器から構成されている。CUDA のスレッド構造は、上位から Grid, Block, Thread と呼ばれる階層構造になっている。Grid は複数の Block(最大 65535) を持ち、Block は複数の Thread(最大 512) を持ってい

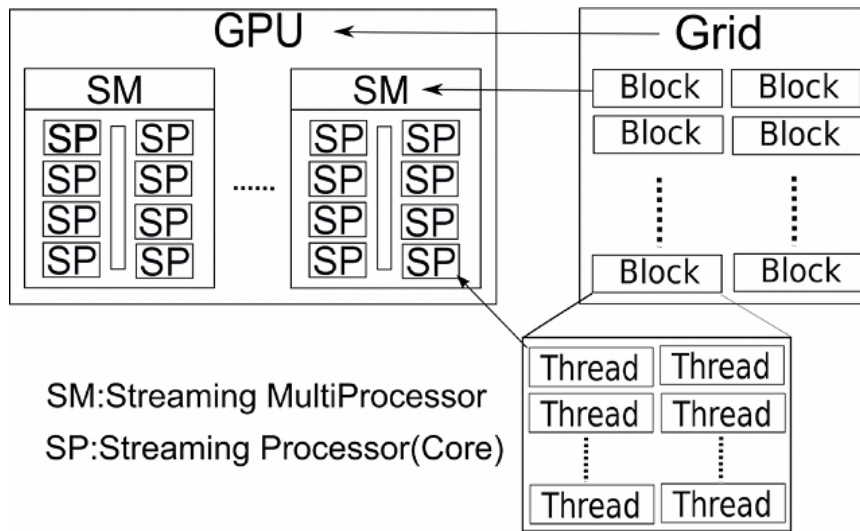


図 3.2: GPU と CUDA のスレッド構造.

る. GridはGPUで実行されるある1つのプログラム単位であり, BlockはSMに, ThreadはSPに割り当てられ実行される. 演算方式は, SIMD(Single Instruction Multiple Data)と呼ばれる処理方式が採用されている. (GPUとCUDAに関するより詳しい技術内容は[50, 51]を参照)

3.3 GPUによるネオコグニトロンの実装

GPUは, 多くのデータに対して独立した単一の処理を行うのに適し, 条件分岐等を含む複雑な処理には向かない. ネオコグニトロンはS細胞とC細胞による畳み込み演算から構成されており, 畳み込み演算はGPUで実装するのに適した演算である. ここでS細胞とC細胞について詳しく述べる. 第 l 段目の k 番目の細胞で, 細胞面内での位置が \mathbf{n} であるS細胞の数式表現は以下の式で与えられる.

$$u_{Sl}(\mathbf{n}, k) = \frac{\theta_l}{1 - \theta_l}.$$

$$\varphi \left[\frac{1 + \sum_{\kappa} \sum_{\nu} a_l(\nu, \kappa, k) \cdot u_{Cl-1}(\mathbf{n} + \nu, \kappa)}{1 + \theta_l \cdot b_l(k) \cdot u_{Vl}(\mathbf{n})} - 1 \right] \quad (3.1)$$

$a_l(\nu, \kappa, k)$ は U_{Cl-1} 層の κ 番目の細胞面からの結合を表し、関数 φ は半波整流形の非線型関数で表される。 a_l と b_l は学習によって決定される結合係数である。抑制性入力である u_{Vl} は以下の式で与えられる。

$$u_{Vl}(\mathbf{n}) = \sqrt{\sum_{\kappa} \sum_{\nu} c_l(\nu) \cdot \{u_{Cl-1}(\mathbf{n} + \nu)\}^2} \quad (3.2)$$

ただし、 $c_l(\nu)$ は固定の結合定数であり、 $\|\nu\|$ に関して単調減少な結合である。一方、第 l 段目、 k 番目の C 細胞面の位置 \mathbf{n} における C 細胞の発火は以下の式で与えられる。

$$u_{Cl}(\mathbf{n}, k) = \psi \left[\sum_{\nu} d_l(\nu) \cdot u_{Sl}(\mathbf{n} + \nu, k) \right] \quad (3.3)$$

d_l は同じ l 段目の S 細胞面からの結合で、 $\|\nu\|$ に関して単調で非負な結合荷重を持つ。関数 ψ は以下で表される。

$$\psi[x] = \frac{\varphi[x]}{1 + \varphi[x]} \quad (3.4)$$

関数 φ は、式 (3.1) と同様に半波整流形の非線型関数である。今回、GPU 側で実装した部分は S 細胞と V 細胞の畳み込み演算部分である。C 細胞に関しては、細胞密度を減少させる間引きの処理が含まれており、GPU 側で仮に処理した場合、無駄なスレッドを生成してしまうので、GPU 側での処理を行っていない。

次に細胞面のサイズと CPU と GPU の処理時間について述べる。以下に CPU と GPU との処理時間に関する関係を示す。

$$CPUTime = ISize \times KSize \times CpuCT \quad (3.5)$$

$$GPU_{Time} = \frac{ISize \times KSize \times GpuCT}{TN \times PN} + CPU \leftrightarrow GPU \quad (3.6)$$

ここでは $CpuCT$ は CPU における 1 度の積和演算時間, $GpuCT$ は GPU における 1 度の積和演算時間, $CPU \leftrightarrow GPU$ は CPU-GPU 間のデータ転送時間, $ISize$ は S 細胞の細胞面サイズ, $KSize$ は細胞間の結合分布 $a_l(\nu, \kappa, k)$ のサイズ $\|\nu\|$, TN は GPU の 1 ブロックあたりのスレッド数, PN は GPU のプロセッサ数をそれぞれ示している. $CPU_{Time} > GPU_{Time}$ となるためには, 細胞面と細胞間の結合分布のサイズが以下のような条件を満たす必要がある.

$$ISize \times KSize > \frac{CPU \leftrightarrow GPU}{CpuCT - GpuCT / (TN \times PN)} \quad (3.7)$$

従って, 式 (3.7) のような細胞面と細胞間の結合分布のサイズを満たす細胞段の S 細胞と V 細胞の畳み込み演算部分にのみ実装するのが良い. ネオコグニトロンは上位の細胞段になるにつれ細胞密度が減少し, 最終的には 1 つの認識細胞が生成する事になるので, 上位の細胞段に GPU を実装してもあまり効果を得られない可能性がある. 更に, 疎データに関して注意しなければならない. CPU 側では零の場合は計算する必要がないように条件分岐が可能である. 従って, $Z \times KSize$ (Z は零の個数) 回演算を少なくできるので GPU 側よりも速くなると考えられる.

3.4 実験

GPU で実装されたネオコグニトロンの処理時間を考察するために, 以下のような実験を行った.

- GPU で実装した畳み込み演算部の細胞面サイズに対する CPU と GPU の処理時間の計測
- 式 (3.7) を満たすサイズの細胞段に GPU 実装した場合と全ての段に対して実装した場合の処理時間の計測

使用した CPU と GPU は以下である.

CPU Intel Core 2 Duo 2.4[GHz]

GPU1 GeForce 9400m (NotePC用GPU)

GPU2 GeForce GTX 480 (最新のGPU)

Intel Core 2 Duo2.4GHz と GeForce 9400m は MacBook5,1 に標準搭載されているものであり, GeForce GTX 480 は Nvidia 社が 2010 年春に発表した最上位モデルである. ネオコグニトロンに関しては第 2 章で説明したのと同様に入力された画像に含まれる 8 方位の直線成分を検出するように U_{S1} 層をガウス関数で作り付けている.

図 3.3 に上述の環境で測定したネオコグニトロンの S 細胞の畳み込み演算の平均処理時間を示す. 細胞間の結合分布のサイズは 7×7 で固定している. また, CPU に関しては, 密 (dense) データ (零の個数 0, 本実験では零の場合も計算) と疎 (sparse) データ (零の個数の割合が多い, 本実験では零の場合は計算しない) に対して処理時間を計測した. GPU2 が最も速く, 次に CPU(sparse) となって GPU1 を上回り, CPU(dense) が最も遅い結果となった. 図 3.4 は, 図 3.3 の CPU に対する GPU のスピードアップ率を示したものである. 疎データに対して高性能な GPU2 は速くなり, GPU1 は遅い結果となった. しかしながら, 密データの CPU に対

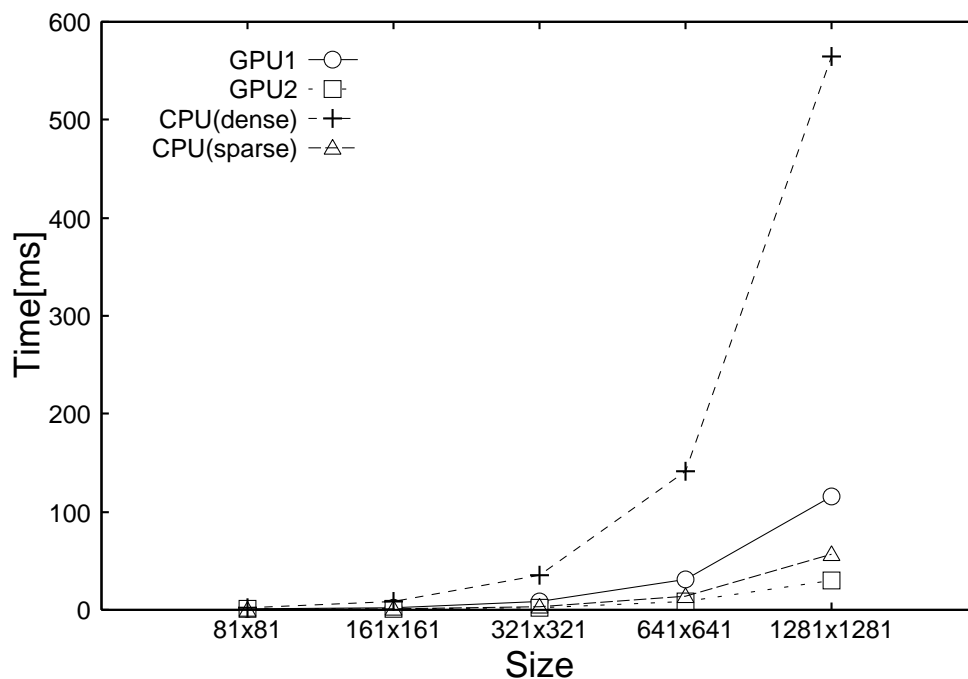


図 3.3: S 細胞の畳み込み演算における処理時間の比較.

して GPU1 と GPU2 はいずれも速くなり、 1281×1281 では GPU1 が 5 倍、GPU2 が 20 倍近く速くなっている。

次に、図 3.5 に GPU 上の総処理時間に対する CPU-GPU 間のデータ転送時間の割合を示す。CPU-GPU 間のデータ転送以外の処理は、GPU 上でのメモリ確保/解放、スレッドの演算となる。GPU1 と GPU2 共に、CPU-GPU 間データ転送に非常に大きな時間がさかれ、データサイズが大きくなるにつれその割合が大きくなり、 1281×1281 ではほぼ 1 に近い割合でデータ転送が行われているのがわかる。つまり、データ転送が GPU のボトルネックになっており、純粋な演算時間は CPU に対して非常に高速であると言える。

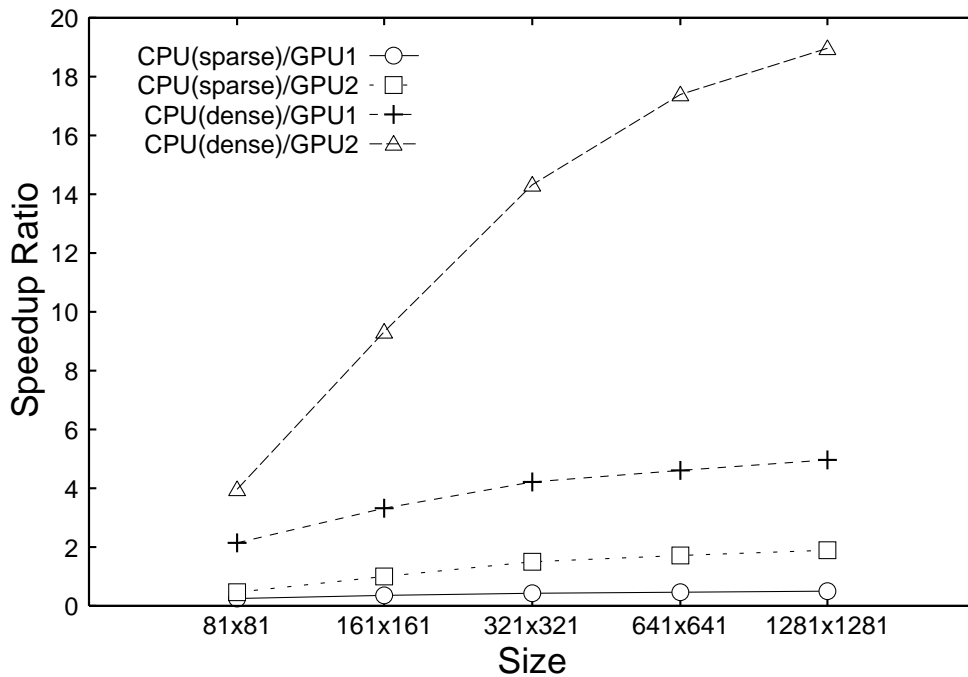


図 3.4: CPU に対する GPU のスピードアップ率.

表 3.1 に密と疎データに対する CPU 上のネオコグニトロン (細胞面のサイズ:161, 41, 31, 15, 5) の処理能力と式 (3.7) を満たすサイズの細胞段 (第 1 層) に GPU を実装した (それ以外は, 疎データに対する CPU での実装) 場合と全ての段に対して実装した場合の処理能力を示す. 密データの CPU の場合の処理能力は 8[f/s] であるが, その他の実装は約 2 倍以上に高速化された. GPU1 ではワンセグ放送のフレームレート 15[f/s] と同程度となり, 疎データの CPU の処理能力は図 3.3 の結果と同様に GPU1 の処理能力を上回った. 最新の GPU2 が最も速く, テレビやビデオ等のフレームレート 30[f/s] と同程度となった. 更に, 全ての段に対して GPU 実装するよりも, 第 1 層のみに GPU 実装した方が速くなっているのがわかる.

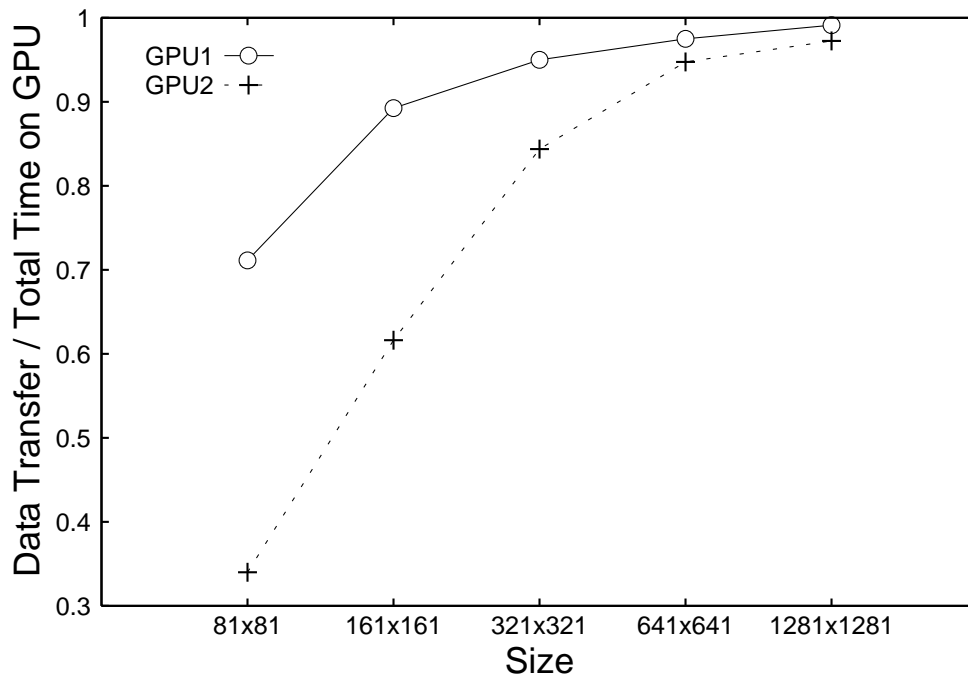


図 3.5: GPU の総処理時間に対するデータ転送時間の割合.

3.5 まとめ

GPU によるネオコグニトロンの GPU 実装を行った. Poli らの細胞間の結合データに関するネオコグニトロンの高速化に対して, 我々は細胞面サイズに関して高速化を行った. その結果, GPU(9400m) では 15[f/s] 程度, GPU(GTX480) では 30[f/s] 程度の処理能力があり, アプリケーションとして十分な応答速度を満たしている事がわかった. 密疎データに関して比較すると, 密データについては CPU の処理よりも GPU の実装が速くなる事が確認された. 疎データについては CPU 側の演算回数の工夫に対して, 低性能の GPU(9400m) の方が遅くなったが, 最新の GPU(GTX480) の方が速い結果となった. つまり, 背景差分や 2 値化等の何らかの前処理

表 3.1: ネオコグニトロンの処理能力 [frame/s]((d): 密データ, (s): 疎データ, (a): 全ての階層, (o): 第 1 階層のみ).

CPU(d)	CPU(s)	GPU1(a)	GPU1(o)	GPU2(a)	GPU2(o)
8.37	22.94	14.23	15.37	27.02	37.83

を施した画像をネオコグニトロンで使用する場合は、現在の NotePC に標準搭載されているクラスの GPU 性能では、CPU での演算の工夫に対して GPU 実装を行う意義が薄れてしまう事がわかった。しかしながら、実画像に零要素が多く含まれるとは考えにくく、実画像をネオコグニトロンの入力データとできるのであれば、十分に価値はある。また、ネオコグニトロンの全ての階層に GPU 処理を行うのではなく、細胞面と GPU の性能から算出されるサイズの細胞段にのみ GPU を実装するのが良い事を明らかにした。細胞間の結合データが多い場合は、Ploli[52][53] らが行った実装方法は有効であり、深い階層で細胞密度が低く細胞間の結合が多くなる場合は彼らの実装を用い、低次の階層で細胞面が大きく細胞間の結合が少なくなる場合及び学習段階に対しては細胞面サイズの大きさに着目した我々の実装を行うのが良いと考えられる。

我々の実装は GPU の性能を十分に引き出しているわけではない。例えば GPU のメモリ計画は実装の容易さから TextureMemory を使用したが、データ転送が高速な SharedMemory を使用すればさらなる高速化が期待される。また、ネオコグニトロンのデータ構造は CPU 側に配置しており、GPU 側に配置する事で通信回数が減り CPU-GPU 間のデータ転送を大幅に改善可能である。しかしながら、アプリケーションの用途によって

は、CPU 側から GPU のデータ構造を必要とする場合も考えられる。今後 GPU の性能を十分に引き出せる細かな実装の改善を行い、ネオコグニトロンをベースにした画像認識アプリケーションの開発を行う予定である。

第4章 選択的注意機構による人 検出

4.1 はじめに

近年，計算機能力の向上に，膨大な画像データを扱えるようになり，画像解析，画像識別，画像理解等の高度な情報処理が求められている．中でも，歩行者検出や，顔認識等，人に関連した社会的要求が多い．人を検出する能力をコンピュータで実現する事は，広範囲に応用可能な有益なアプリケーションとなりえる．従来の既存技術はある状況では動作可能だが，動的に変わる環境の変化（形状，コントラスト，複雑な背景，オクルージョン）により状況が変わると，動作不可能と，まだ課題が多い．我々は，そのような環境の変化から生じるパターンの変形という問題に対して，ネオコグニロン型神経回路モデルのパターンの変形に対する頑強性に着目している．

コンピュータビジョンの分野で，Dalal らは，HOG 特徴量を算出し，SVM(Support Vector Machine) を用いて人を検出する方法を提案した [54]．局所領域における輝度の勾配方向をヒストグラム化した特徴ベクトルであり，物体の形状を表すことが可能である．HOG は，照明や影の影響を受けにくく，幾何学的変化に頑強であるので，多くの手法が提案されて

いる [55, 56]. 一方で, ネオコグニトロン型神経回路モデルも, 環境の変化に対して比較的安定した結果を示す事が知られている. HOG とネオコグニトロンの S 細胞の特徴抽出を比較すると, 方向余弦による類似度と統計的な手法の違いはあるが, 局所性としての類似点がある. また, 近年, 統計的学習法 (Adaboost[57]) と HOG 局所特徴量を組み合わせることによって, 一つでは意味を持たない Low-level な特徴量を組み合わせることによって, 何らかの意味のある middle-level な特徴量に統合する方法が提案されている [58]. つまり, ネオコグニトロンと同様に, 局所的な特徴抽出と情報統合を行っていると考えられる. このような工学的, 数学的な工夫から生まれた技術が, 生理学的知見から構成されたネオコグニトロンの多階層による特徴統合の処理とある意味似ているのは興味深い.

ネオコグニトロンにフィードバックの経路を付加した選択的注意機構のモデルが Fukushima によって提唱されている [31]. 選択的注意機構のモデルは, 認識器としてのネオコグニトロンにフィードバック経路を追加する事によって, 複数パターンからの 1 パターンの切り出し, パターンの欠落部分の修復, ノイズの除去等の認識における補助的機能を実現している. 既に, 選択的注意機構のモデルは, 英字筆記体の連結文字列, 顔の部分パターンの切り出し等で応用可能な事がわかっている [32, 33, 34, 35]. そこで, 今回, 我々は, 選択的注意機構のモデルを用いて, 複雑な背景下からの人の検出を試みる. 人の形状は姿勢 (各関節部) の変化によって無数に存在する. オコグニトロンが物体の変形に比較的強いと言われているが, あまりに複雑な形状の変化に対して許容できる訳ではない. そこで, 我々は, ネオコグニトロンの最終層の認識細胞を姿勢単位で生成し, 各姿勢をネオコグニトロンの変形に対する頑強性でカバーしようと

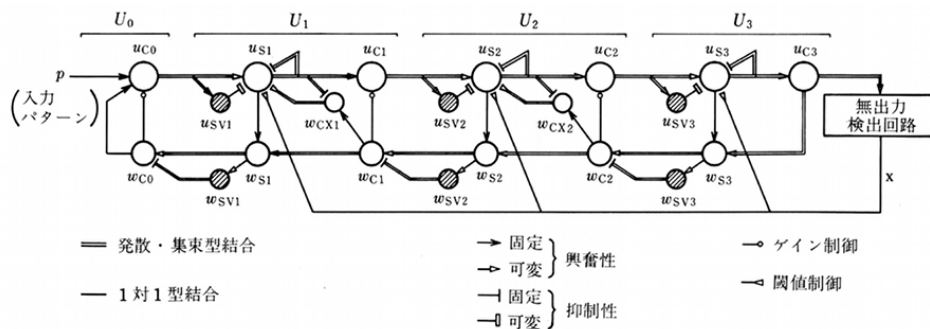


図 4.1: 選択敵注意機構の構造 [31].

考えている. そこで, まず, 今回は, 人の“立つ”という基本的な姿勢に限定し, 人の形状の1つを学習パターンとして, 学習パターンが全く異なる人の形状を呈示した際のフィードバック経路の出力部に人の形状が想起されるかを調べた.

本章では, 4.2節で選択的注意機構のモデルに関し詳しく述べ, 4.3節で実験と結果を示す.

4.2 選択的注意機構のモデル

選択的注意機構のモデルの構造は, ネオコグニトロンと同様の構造を信号の向きを逆にして, ネオコグニトロンに付加したものである(図4.1). 各細胞間の結合は, ネオコグニトロンと同様の学習アルゴリズムによって獲得され, 求心性経路のみによって行われる. 遠心性経路の細胞結合は, 求心性経路の結合をそのままコピーしたものである. 以下で, 求心性経路と遠心性経路に関して詳しく述べる.

4.2.1 求心性経路の細胞

求心性経路における u_S 細胞は、ネオコグニトロンの S 細胞と同様に、特徴抽出を行う。選択的注意機構の第 l 段目の U_S 細胞がネオコグニロンと違う点は、側抑制をかける点である。具体的には、式 (4.1) によって、一時的に \bar{u}_{Sl}^t を求めておき、この出力に対して側抑制を 1 回かけ、その結果を再び細胞の入力と操作を有限回繰り返す。

$$\bar{u}_{Sl}^t(\mathbf{n}, k) = \frac{\theta_l^t(\mathbf{n}, k)}{1 - \theta_l^t(\mathbf{n}, k)}.$$

$$\varphi \left[\frac{1 + \sum_{\kappa} \sum_{\nu} a_l(\nu, \kappa, k) \cdot u_{Cl-1}^t(\mathbf{n} + \nu, \kappa)}{1 + \theta_l^t \cdot b_l(k) \cdot u_{Vl}^t(\mathbf{n})} - 1 \right] \quad (4.1)$$

側抑制をかえる理由は、同一受容野をもつ細胞間で競合をさせることによって、ある位置に存在する特徴が単一であるようにするためのメカニズムである。 $a_l(\nu, \kappa, k)$ は U_{Cl-1} 層の κ 番目の細胞面からの結合を表し、関数 φ は半波整流形の非線型関数で表される。 a_l と b_l は学習によって決定される結合係数である。 $\theta_l^t(\mathbf{n}, k)$ は、時刻 t における閾値の制御変数である。 u_{Sl}^t 細胞の最終的な出力は、 \bar{u}_{Sl}^t に対して式 (4.2) で側抑制をかけたものである。

$$u_{Sl}^t := \varphi \left[u_{Sl}^t - \sum_{\nu} e_l(\nu) \cdot u_{Sl}^t(\mathbf{n} + \nu, k) \right]$$

$$\sum_{\kappa=1, \kappa \neq k} \sum_{\nu} e'_l(\nu) \cdot u_{Sl}^t(\mathbf{n} + \nu, \kappa) \quad (4.2)$$

$e_l(\nu)$, $e'_l(\nu)$ は、 $\|\nu\|$ は側抑制の空間結合係数の形状である。第 l 段目の細胞 u_{Vl} 細胞の出力は、 a_l の結合範囲と同じ範囲にある $l-1$ 段目の u_{Cl-1}

細胞の出力の重み付き 2 乗平均で表される.

$$u_{Vl}^t(\mathbf{n}) = \sqrt{\sum_{\kappa} \sum_{\nu} c_l(\nu) \cdot \{u_{C_{l-1}}^t(\mathbf{n} + \nu)\}^2} \quad (4.3)$$

ただし, $c_l(\nu)$ は固定の結合定数であり, $\|\nu\|$ に関して単調減少な結合である. 相対的位置ずれを吸収する役目を担う u_C 細胞の出力は, U_S 細胞の出力に対して, 前層の u_S 細胞の出力を空間的にぼかし, その結果に利得変数 $g_l(\nu, k)$ を掛けたものである.

$$u_{Cl}^t(\mathbf{n}, k) = g_l^t(\nu, k) \psi \left[\sum_{\nu} d_l(\nu) \cdot u_{Sl}(\mathbf{n} + \nu, k) \right] \quad (l \neq 0) \quad (4.4)$$

d_l は同じ l 段目の S 細胞面からの結合で, $\|\nu\|$ に関して単調で非負な結合荷重を持つ. 関数 ψ は以下で表される.

$$\psi[x] = \frac{\varphi[x]}{1 + \varphi[x]} \quad (4.5)$$

関数 φ は, 式 (2.1) と同様に半波整流形の非線型関数である. U_C 細胞層のうち入力層 U_{C0} 層は, 外部からの入力のみならず, 遠心性経路の最下位層 W_{C0} からのフィードバック信号によっても制御される. U_{C0} 細胞は,

$$u_{C0}^t(\mathbf{n}) = g_0^t(\mathbf{n}) \cdot \max[p(n), w_{C0}^{t-1}(\mathbf{n})] \quad (4.6)$$

という出力を出す. ここで, $p(n)$ は外部から入力層 U_{C0} 層へ提示された入力パターンである.

4.2.2 遠心性経路の細胞

ある w_s 細胞からその下段に位置する w_C 細胞への結合強度は, 対応する位置の u_s 細胞に向かってくる求心性の結合強度に比例した値になるよ

うに，学習の際に連動して調節されるようにした．従って w_s 細胞から下向する遠心性信号は下段の u_C 細胞から u_s 細胞への信号の流れと同じ経路を逆方向に流れることになる．同様に抑制性細胞の w_{SV} 細胞から下向する信号も対応する位置に存在する細胞への求心性経路を逆にたどることになる． w_C 細胞の出力は式 (4.7) で表される．

$$w_{Cl}^t(\mathbf{n}, k) = \psi \left[\alpha_l \cdot \left\{ \sum_{\kappa=1}^{K_{l+1}} \sum_{\nu \in A_{l+1}} a_{l+1}(\nu, k, \kappa) \cdot w_{Sl+1}^t(\mathbf{n} - \nu, \kappa) - \sum_{\nu \in A_{l+1}} C_l(\nu) \cdot w_{Vl+1}^t(\mathbf{n} - \nu) \right\} \right] \quad (4.7)$$

α_l は飽和のレベルを決める定数である．抑制性細胞 w_V は，求心性経路の u_V 細胞とは異なり，線形の入出力特性を持ち，式 (4.8) で表される．

$$w_{Vl+1}^t(\mathbf{n}) = \theta_{l+1}^t \cdot \sum_{\kappa=1}^{K_{l+1}} b_{l+1}(\kappa) \cdot w_{Sl+1}^t(\mathbf{n} + \kappa) \quad (4.8)$$

ただし θ_l^0 は閾値の制御変数 $\theta_l^t(\mathbf{n}, k)$ の初期値である．最後に， w_s 細胞の出力は式 (4.9) で表される．

$$w_{Sl+1}^t(\mathbf{n}, k) = \min \left[u_{Sl}^t(\mathbf{n}, k), \alpha'_l \cdot \sum_{\nu \in D_l} d_l(\nu) w_{Sl+1}^t(\mathbf{n} - \nu, k) \right] \quad (4.9)$$

ここで α'_l は正の定数である．細胞の出力は，ちょうど求心性の信号と遠心性の信号の論理積のようなものになっており，求心性信号がないところは発火しないようなゲートの作用を持っている．

4.2.3 ゲインと閾値制御

遠心性信号は求心性信号の影響を受けながら伝わっていくが，遠心性信号も単に求心性信号の影響を受けるだけではなく，求心性信号に対し

て影響を与えている。求心性信号の u_C 細胞の応答特性を示す式 (4.4) には、ゲイン（利得）変数 $g_l^t(\mathbf{n}, k)$ が設けられている。これは、細胞の応答の強度を制御するための変数で、遠心性経路によって制御される。遠心性経路による利得は、現在認識しているパターンの持っているべき特徴の位置に高い利得がかけられるので、結果としてパターンに含まれる雑音成分の除去を行なうことになる。一方、認識が成功した場合でも学習パターンが持っていた全ての特徴が検出されているわけではない。これはパターンの変形や雑音の影響などが原因として考えられる。この場合、検出されなかった特徴を検出してやるためには、その特徴があるべき位置の u_S 細胞の閾値を下げてやる必要がある。

まず、ゲイン制御に関して述べる。 u_C 細胞に対する利得のうち遠心性経路によって制御される利得 $g_{Bl}^t(\mathbf{n}, k)$ は、促進作用を及ぼす成分 $g_{B1l}(\mathbf{n}, k)$ 、疲労作用を及ぼす成分 $g_{B2l}(\mathbf{n}, k)$ から成っており、式 (4.10) で表される。

$$g_{Bl}^t(\mathbf{n}, k) = 1.0 + \alpha_{B1l} \cdot g_{B1l}(\mathbf{n}, k) - g_{B2l}(\mathbf{n}, k) \quad (4.10)$$

α_{B1l} は $g_{Bl}^t(\mathbf{n}, k)$ の強度を定める 1 より大きい定数である。また、 g_{B1l} と g_{B2l} は、以下のように変化する。

- $w_{Cl}^t(\mathbf{n}, k) > 0$ のとき

$$g_{B1l}^t(\mathbf{n}, k) = \gamma_l \cdot {}^t_{B1l}(\mathbf{n}, k) + (1.0 - \gamma_l) \cdot w_{Cl}^{t-1}(\mathbf{n}, k) \quad (4.11)$$

$$g_{B2l}^t(\mathbf{n}, k) = \gamma_l \cdot {}^t_{B2l}(\mathbf{n}, k) + (1.0 - \gamma_l) \cdot w_{Cl}^{t-1}(\mathbf{n}, k) \quad (4.12)$$

- $w_{Cl}^t(\mathbf{n}, k) = 0$ のとき

$$g_{B1l}^t(\mathbf{n}, k) = \gamma_{1l} \cdot {}^t_{B1l}(\mathbf{n}, k) \quad (4.13)$$

$$g_{B2l}^t(\mathbf{n}, k) = \gamma_{2l} \cdot g_{B2l}^t(\mathbf{n}, k) \quad (4.14)$$

ここで、 γ_l , γ_{1l} , γ_{2l} は 0 以上, 1 以下の変化定数である。 g_{B1l} と g_{B2l} の初期値は 0.0 と仮定して、遠心性信号が少しでも存在すれば同一の変化定数 $(1.0 - \gamma_l)$ で遠心性信号の値に近づいていく。遠心性信号が切れると、 g_{B1l} と g_{B2l} は、それぞれ異なる変化定数で減少する、 γ_{1l} は 0.0 に γ_{2l} は 1.0 に近い値に定めている。従って、注意の転換によって遠心性経路の信号が遮断された時には、促通作用をもたらす g_{B1l} は急激に減少し、疲労の作用をもたらす g_{B2l} の効果だけが残る。また、同じカテゴリーに属するパターンが続けて認識されにくくなることを避けるために、上位の層に進むに従って γ_{1l} の値は小さく、 γ_{2l} の値は大きくしている。このようにして、認識されたパターンは γ_{1l} の効果によって順次抑制されていくので、呈示された各パターンに対して順番に注意を向けていくことが可能となる。また疲労で細胞の出力が下がっていても、式 (4.10), (4.11), (4.12) の様にゲインを得る式が和の形で示されているので、疲労効果でゲインが下がっていてもこのゲイン制御信号がくれば u_C 細胞を再び発火させ易くすることが出来る。

次に閾値制御に関して述べる。認識が成功した場合でも、必ずしもパターンに含まれる部分特徴がすべて検出されているとは限らない。検出されない原因としては、部分特徴の変形や欠損が考えられる。このような場合、 W_{C0} 層に再構成されるパターンは、その部分特徴を欠いたものとなる。そこで、存在すべき特徴が検出されていない場合には、検出が行われなかった付近に存在する特徴検出細胞の閾値を下げ、特徴を検出しやすくする。 w_{CX} 細胞は存在すべき特徴を抽出できなかったことを検

出する細胞である。ところで、遠心性経路に信号が流れているにもかかわらず、対応する位置の求心性経路に信号が流れていないときには、存在すべき特徴が検出されていないと考えられる。そこで、 w_{CX} 細胞は遠心性経路の信号と求心性経路の信号を比較し、その差に応じて出力を出す。 w_{CX} 細胞の出力は式 (4.15) で表される。

$$w_{CXl}^t(\mathbf{n}, k) = \varphi \left[w_{Cl}^t(\mathbf{n}, k) - \sum_{\boldsymbol{\nu} \in D_l'} (\boldsymbol{\nu}) \cdot u_{Sl}^t(\mathbf{n} + \boldsymbol{\nu}, k) \right] \quad (4.15)$$

w_C 細胞の応答から同じ段の u_S 細胞の応答の和を差し引いたものになっている。 w_C 細胞と u_C 細胞の差を用いていないのは、 u_C 細胞の出力が利得制御信号の影響で変化するためである。また、このようにして w_{CX} 細胞が応答すると、閾値制御信号が変化ようになっており、その変化は式 (4.16) で表される。

$$x_{Sl}^t(\mathbf{n}, k) = \beta_l \cdot x_{Sl}^{t-1}(\mathbf{n}, k) + \beta_l' \cdot \sum_{\boldsymbol{\nu} \in D_l} d_l(\boldsymbol{\nu}) \cdot w_{CXl}^{t-1}(\mathbf{n} - \boldsymbol{\nu}, k) \quad (4.16)$$

この信号によって、特徴抽出細胞の閾値 $\theta_l^t(\mathbf{n}, k)$ は次のように制御される。

$$\theta_l^t(\mathbf{n}, k) = \frac{\theta_l^0(\mathbf{n}, k)}{(1 - \theta_l^0(\mathbf{n}, k))(1 - x_{Sl}^t(\mathbf{n}, k) + x_{Xl}^t) + \theta_l^0(\mathbf{n}, k)} \quad (4.17)$$

ここで、 x_{Xl}^t は後述するもう一つの閾値制御信号で、最上位層で出力が出なかった場合に出力される。結果的に、この w_{CX} 細胞が応答した場合には、その付近に位置する細胞の閾値 $\theta_l^t(\mathbf{n}, k)$ が下がるようになっている。

入力層にパターンを呈示した場合、最上位層において、どの u_C 細胞も出力を出さない場合がある。このように、最上位層で出力が出ないことが検出されると式 (4.18) にしたがって、閾値制御信号 x_{Xl}^t が全段に対し



図 4.2: 学習パターン.

て出力される. その結果, 式 (4.17) に従い, すべての段の特徴抽出細胞 u_S の閾値を下げ, 応答が出やすい状態にする.

$$x_{Xl}^t = \begin{cases} x_{Xl}^{t-1} + \beta_{Xl} & \text{if } u_{Cl}^{t-1}(\kappa) = 0 \text{ for all } \kappa \\ \beta'_{Xl} \cdot x_{Xl}^{t-1} & \text{else} \end{cases} \quad (4.18)$$

ここで, は最上位段の u_C 細胞の出力を表わす.

4.2.4 選択敵注意機構のセグメンテーションの例

すでに, 1, 2, 4(図 4.2) というパターンを選択的注意機構のモデルは学習していると仮定する. 図 4.3 に示されるように, 2, 4 のような複数パターンが入力層に呈示されると, 最終層では, 2, または, 4 のパターンが認識される. 遠心性信号の流れは, 認識されたパターン 4 の逆経路を辿っていく事となる. しかしながら, 単に, 信号を逆に辿っていけば良いわけではない. なぜなら, 求心性経路では, ある C 細胞に対して結合している全ての S 細胞の中で 1 つでも反応すれば, その C 細胞は反応する. したがって, 遠心性経路では, C 細胞が反応すると, その C 細胞と結合があるすべての S 細胞が反応し, 信号が拡散する事となる. そのため, 求心性経路における S 細胞の反応と遠心性経路の信号の論理積を取る事によって, 信号の拡散を防いでいる (求心性信号から遠心性信

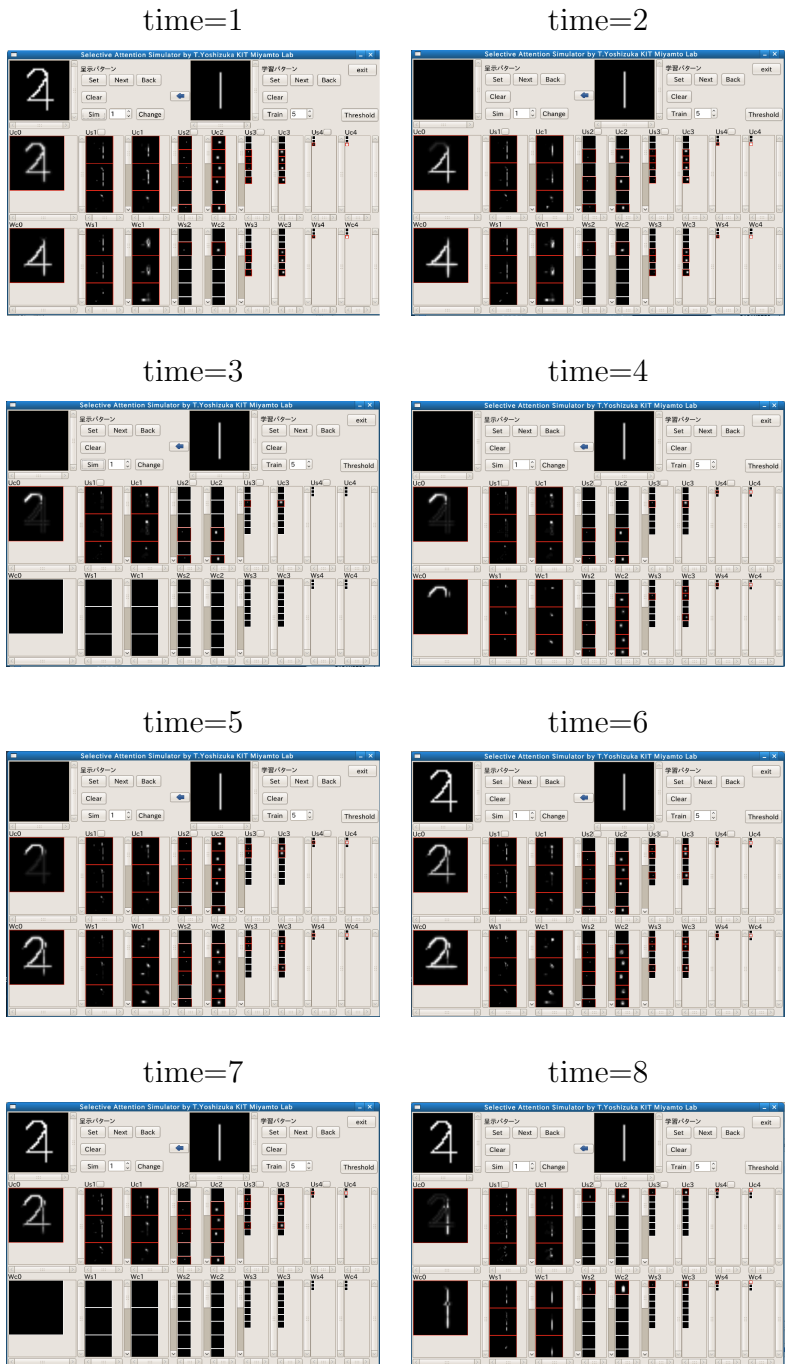


図 4.3: 1, 2, 4 の複合文字のセグメンテーション.

号の作用)。7遠心性経路から4というパターンが想起されると、次に求心性経路へ信号を流す際に、パターン4に関する特徴に関して、促通作用が働き、パターン4の信号が強くなる。逆に、その他のパターン(2)は、抑制される。これは、遠心性経路のC細胞に対応する求心性経路のC細胞の反応に対して、ゲイン(利得)を掛ける事によって、信号の増幅と減少をおこなっている(遠心性信号から求心性信号への作用)。従って、パターン4に対して注意を向ける事が可能となる。遠心性信号から求心性信号への作用は、その他に閾値制御がある。求心性経路において、パターンが認識されず、処理が停止されてしまうのを防ぐために、無出力検出回路が組み込まれており、最終層が無出力の場合、各S細胞の閾値を変化させて、類似度の許容度を変換させ、認識されやすくしている。また、パターンの一部欠落等によって、本来存在する特徴を抽出できない場合、言い換えると、遠心性経路で反応した細胞に対応する求心性経路の細胞が無反応である場合には、その無反応の細胞の閾値を下げる事によって特徴を抽出しやすくし、パターンの修復を行う作用がある。[図4.3: time=1~2:4が認識され注意が向けられた。time=3:注意の切り替え(遠心性経路を遮断する)。time=4:2が認識され注意が向けられた(一部欠落)。time=5~6:欠落部分が修復した。time=7:注意の切り替え(遠心性経路を遮断する)。time=8:1が認識された。]

4.3 実験

前節までに説明した選択的注意機構のモデルを用いて人の検出を実現する為に、今回、我々は、人の“立つ”という基本的な姿勢に限定し、人

の形状の1つを学習パターンとして、学習パターンが全く異なる人の形状を呈示した際のフィードバック経路の出力部に人の形状が想起されるかを調べた。

実験は以下の3つ入力画像の条件で実施し、 w_{C0} 層にどのような結果が想起されるかを調べた。

- 背景なしの68パターンの異なる人の形状のシルエット画像を入力とした場合
- 複雑な背景のスケッチ画に上述のシルエット画像を足した画像を入力とした場合
- 複雑な背景下に人物が存在する実画像を入力とした場合

以下で、選択的注意機構のモデルの構成と入力パターン等実験環境述べ、次に、シミュレーション結果を示す。

4.3.1 実験環境

システム環境は、Mac Book5,1(OS:Mac OS X 10.6.6, CPU:Intel Core2 Duo2.4GHz, Memory:4GM)上で、プログラミング言語はGNU/Cを、描画ライブラリにGDK/GTKを使用した。

入力パターンに関しては、FLOP DEDIGN(デザイン事務所)が作成した68パターンの形状が異なる人のシルエット画像を用いた(図4.4)[59]。また、人の姿勢は、”立っている”という状態のみに限定した。まず、これらの画像からエッジの抽出を行なった。エッジの抽出方法は、CannyFilter, SobelFilter等、さまざまなアルゴリズムが存在するが、今回は、SobelFilter

を使用した。尚、最近のネオコグニトロンの研究でも Mexican hat 型の受容野特性を持つコントラスト層と呼ばれる層により輪郭の抽出を行っている。これらのエッジ抽出された画像を選択的注意機構のモデルの入力パターンとした。また、学習パターンには服装や関節部位に特徴のある姿勢と直立不動な姿勢の2つのシルエット画像を選んだ (図 4.5)。

次に選択的注意機能のモデルのパラメータに関して述べる (表 4.1, 表 4.2)。背景なしの 68 パターンの人のシルエット画像を入力とした場合においては、選択的注意機構のモデルの各階層の細胞面のサイズ及び細胞間の結合分布のサイズは、2通りのサイズで行った。それ以外の入力条件では、 81^2 時のパラメータを使用している。また、閾値には認識と学習時に使用する2つのものが存在する。これは、初期のネオコグニトロンでは、認識時も学習時も同じ値を使用していたが、認識時に閾値が低いのはいいが結合強化時にも閾値が低いと十分な数の細胞を生成できないという問題があった。そこで、Fukushima と Tanigawa らは、ネオコグニトロンに競合学習を用いて学習させる時に、閾値を学習時において高く設定し、認識時においては低く設定することを提唱した [61]。本章の選択的注意機構のモデルでは、その閾値設定の方法を利用している。

4.3.2 実験結果

背景なしの 68 パターンの人のシルエット画像を入力とした場合

まず、最初に任意の学習パターンから、その他の 68 パターンの人のシルエット画像から得られたエッジ画像を入力パターンとして、 w_0 に人の形状がどのように想起されたかの結果を示す。細胞面サイズは 161^2 と 81^2

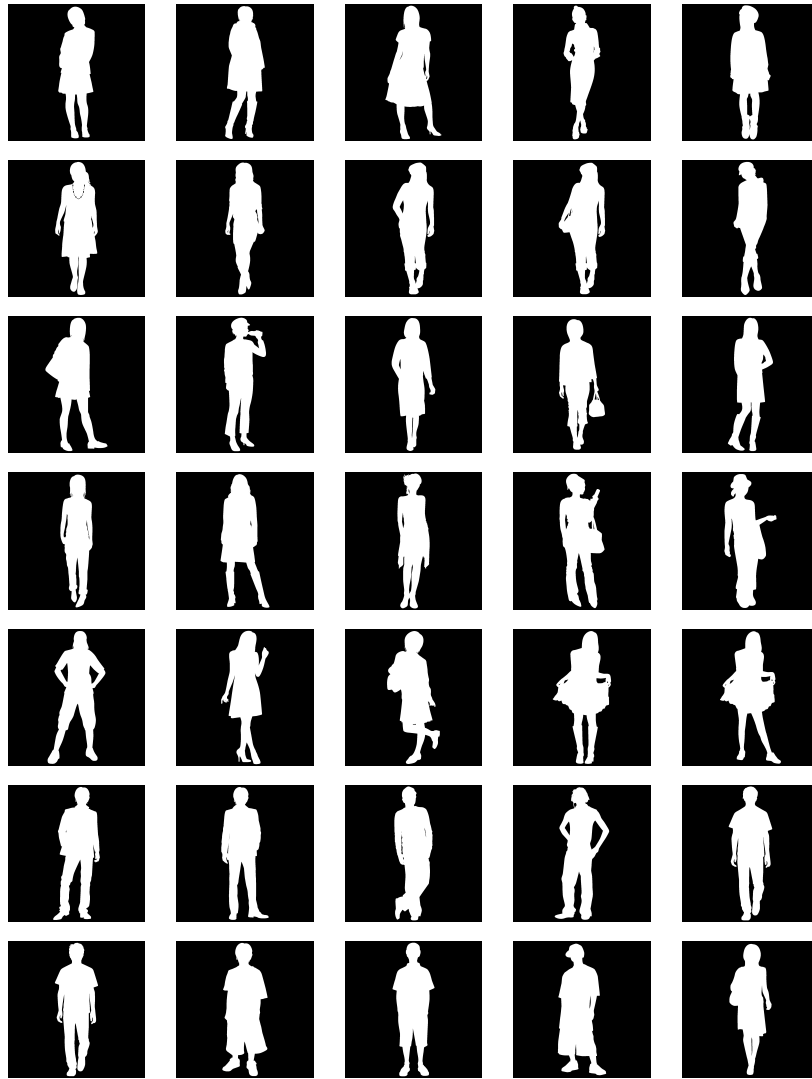


図 4.4: パターンに用いた人のシルエット画像の例 [59].



図 4.5: 学習パターン.

の2つの条件で行った. 又, 実験結果を分類すると4つのパターンが想起される結果となった. 以下に, それぞれのパターンの実験結果を示す.

- 上半身(頭と肩周辺)と下半身(足)
- 上半身(頭と肩周辺)のみ
- 下半身(足と腰周辺)のみ
- 一部のエッジの想起

図 4.6 は, 上半身(頭と肩周辺)と下半身(足)が想起した結果である. $time = 1$ においては, 胴体部分と思われる一部のパターンが w_{C0} に想起した. その後, 学習パターンに従って, パターンは復元していき, $time = 5$ では, 上半身(頭と肩周辺)と下半身(足)が想起した. $time = 6 \sim time = 10$ で, はっきりと上半身(頭と肩周辺)と下半身(足)のパターンが想起されたのがわかる. $time = 11$ 以降では, それ以上のパターンの想起は行われなかった. 学習パターンと入力パターンは異なる形状をしているのがはっきりとわかり, ネオコグニトロン型神経回路の変形に対して比較的安定した認識が可能という特徴が顕著に示されていると考える.



図 4.6: 上半身(頭と肩周辺)と下半身(足)の想起。



図 4.7: 上半身 (頭と肩周辺) の想起.

表 4.1: 選択的注意機構のパラメータ (入力パターンサイズ:161x161).

l	0	1	2	3	4	5
Us-Ws cell plane size	-	161^2	81^2	41^2	21^2	9^2
Uc-Wc cell plane size	161^2	81^2	41^2	21^2	9^2	1^2
$a_l(\boldsymbol{\nu}), c_l(\boldsymbol{\nu})$ size	-	5^2	7^2	7^2	9^2	9^2
$d_l(\boldsymbol{\nu})$ size	-	5^2	7^2	7^2	9^2	9^2
$c_l(\boldsymbol{\nu})$	-	$c_l(\boldsymbol{\nu}) = 1$ for all $\boldsymbol{\nu}, l$				
$d_l(\boldsymbol{\nu})$	-	$d_l(\boldsymbol{\nu}) = 1$ for all $\boldsymbol{\nu}, l$				
$\theta_l(\text{recognition})$	-	$\theta_l = 0.64$ for all l				
$\theta_l(\text{train})$	-	$\theta_l = 0.74$ for all l				

図 4.7 は, 上半身 (頭と肩周辺) のみが想起した結果である. $time = 1$ においては, 胴体部分と思われる一部のパターンが w_{C0} に想起した. その後, 学習パターンに従って, パターンは復元していき, $time = 5$ では, 上半身 (頭と肩周辺) と下半身 (足) が想起した. $time = 6 \sim time = 10$ で, はっきりと上半身 (頭と肩周辺) と下半身 (足) のパターンが想起されたのがわかる. $time = 11$ 以降では, それ以上のパターンの想起は行われなかった. 図 4.7 の入力パターンは, 学習パターンの上半身 (頭と肩周辺) の形状の類似が顕著であり, 下半身部分のパターンは想起されなかった. 学習パターンの衣服部分の形状がないのが一つの理由として考えられる.

図 4.8 は, 下半身 (足と腰周辺) のみが想起した結果である. $time = 1$ においては, 足と腕の一部と思われるパターンが w_{C0} に想起した. その後, 学習パターンに従って, パターンは腰の部分がのびていくが, $time = 4$

表 4.2: 選択的注意機構のパラメータ (入力パターンサイズ:81x81).

l	0	1	2	3	4	5
Us-Ws cell plane size	-	81^2	41^2	21^2	11^2	5^2
Uc-Wc cell plane size	81^2	41^2	21^2	11^2	5^2	1^2
$a_l(\boldsymbol{\nu}), c_l(\boldsymbol{\nu})$ size	-	3^2	5^2	5^2	5^2	5^2
$d_l(\boldsymbol{\nu})$ size	-	3^2	5^2	5^2	5^2	5^2
$c_l(\boldsymbol{\nu})$	-	$c_l(\boldsymbol{\nu}) = 1$ for all $\boldsymbol{\nu}, l$				
$d_l(\boldsymbol{\nu})$	-	$d_l(\boldsymbol{\nu}) = 1$ for all $\boldsymbol{\nu}, l$				
$\theta_l(simulation)$	-	$\theta_l = 0.64$ for all l				
$\theta_l(train)$	-	$\theta_l = 0.74$ for all l				

では、腕と思われる線がなくなった。これは、腕の部位が学習パターンに存在しないために、なくなったと思われる。その後、腕がない側面から線が伸びていくが、最終的に、上半身の想起にはいたらなかった。

最後に、図 4.9 は一部のエッジのみが想起され、人の形状が想起されなかった結果である。どのパターンも、学習パターンの形状といちじるしく異なっており、入力パターンの変形に対するネオコグニロン型神経回路モデルの許容度を超過しているため、想起が行われなかったと考える。

表 4.3 は、68 パターンの形状が異なる入力パターンに対して、 W_{C0} 層の 4 つ想起パターン(”上半身(頭と肩周辺)と下半身(足)”, ”上半身(頭と肩周辺)のみ”, ”下半身(足と腰周辺)のみ”, ”一部のエッジ”)の想起率を調べたものであり、異なる 2 つの学習パターンと 2 つの細胞面のサイズ ($161^2, 81^2$) で実験を行った。表中の想起率は、”上半身(頭と肩周

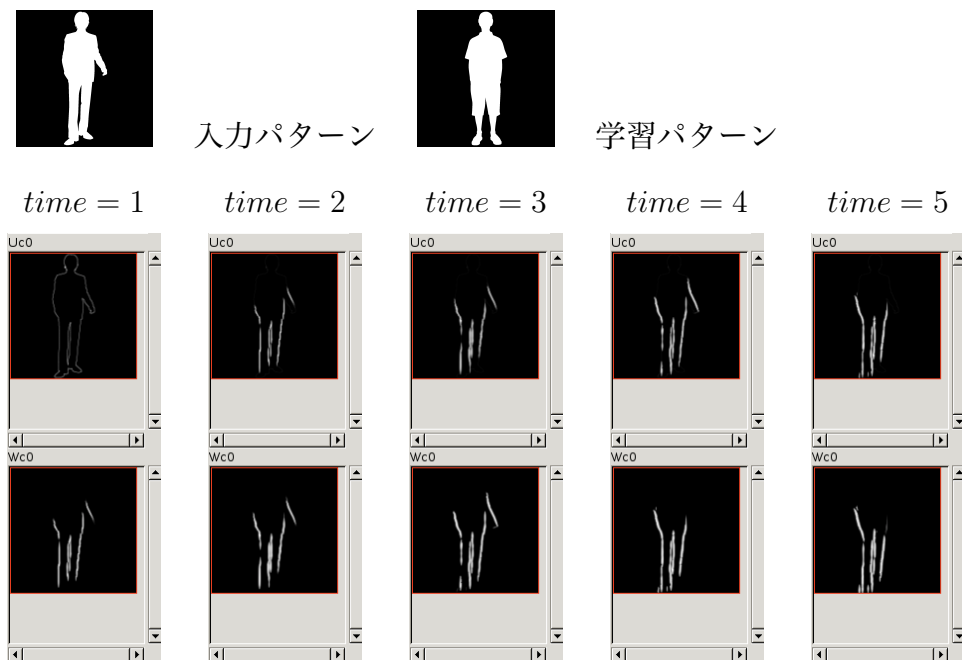


図 4.8: 下半身 (足と腰周辺) の想起.

辺) と下半身 (足) ”, ” 上半身 (頭と肩周辺) のみ ”, ” 下半身 (足と腰周辺) のみ ” が想起した場合の 68 パターン全体に対する割合であり, 我々はその場合において, 人の形状を想起したと仮定した. ” 上半身 (頭と肩周辺) のみ ”, ” 下半身 (足と腰周辺) のみ ” の想起パターンが人の形状を想起したと言えるのかという議論が出てくると思われるが, 今回は, ひとつの目安として, データを提示している. まず, 選択的注意機構の入力層 w_{C0} のサイズによる違いを考察すると, 2つの学習パターンともに 161^2 のサイズの検出率が悪い結果となった. サイズを大きくした事により, 形状の微細構造を認識できるようになったが, 学習パターンの類似度の許容度が低くなったためと考えられる. さらに, 81^2 サイズでは, 3%の検出率の違いであったが, 161^2 サイズでは, 学習パターン 1 が学習パターン

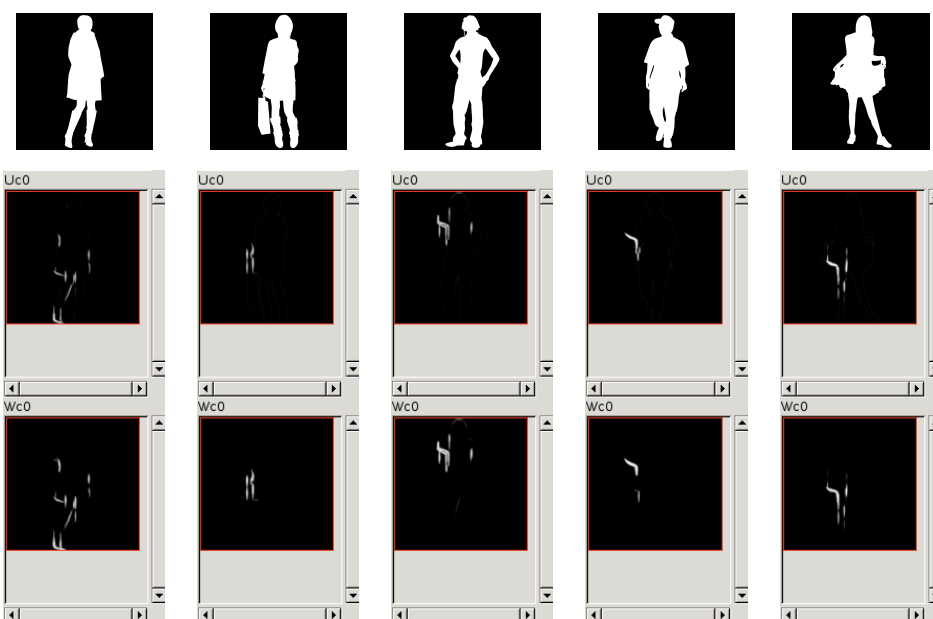


図 4.9: 一部のエッジの想起.

2より、検出率がおおよそ2倍近く悪くなっている。これは、学習パターン1の形状は、パターン2よりも複雑な形状をしているためと考えられる。また、上半身と下半身の想起率に着目すると、学習パターン1では、サイズの変化によって、上半身と下半身の想起率が逆転している。81²のサイズでは、下半身の想起率が高いが、161²サイズでは、上半身の想起率が高い。おそらく、サイズを大きくする事によって、下半身の衣服の微細構造を学習によって獲得し、下半身の類似度の許容度が著しく下がったためと考えられる。つまり、上半身の形状より下半身の形状の方が複雑な（特徴がある）ために、上半身の方の想起率が高くなったと考えられる。学習パターン2では、上半身の形状の方が複雑な（特徴がある）ために、上半身の想起率が下がったと思われる。”立つ”という基本的な姿勢の形状では、頭部の形状と足下の形状に特徴があり変形も比較的少な

表 4.3: 各部位のパターンの想起率：1:学習パターン 1, 2:学習パターン 2 を表している。また、サイズは、選択的注意機構のモデルの入力サイズである。



	サイズ	上半身	下半身	上半身+下半身	一部	想起率
1	161 ²	20.6	1.5	26.5	51.5	48.5
	81 ²	1.5	27.9	58.8	11.8	88.2
2	161 ²	2.9	38.2	45.6	13.2	86.8
	81 ²	13.2	22.1	55.9	8.8	91.2

いので、上半身と下半身が想起されるのはある意味、自然と考えられる。

複雑な背景下におけるスケッチ画と実画像を入力とした場合

表 4.3 中でもっとも想起率が高かった状態 (81² サイズで直立不動な姿勢の学習パターン 2) で、複雑背景下における実験を行った。まずは、シルエット画像をスケッチ画 (エッジが正確に取得できている) に無作為に合成したものを、入力パターンとした結果を示す。図 4.10 はその結果を示したもので、複雑な背景を抑制し、人の形状を想起している事が確認できた。

図 4.11 は実画像データを用いたものである。被験者が一人の場合と、被験者二人が重なった場合は、鮮明ではないが、人の形状を想起可能であっ

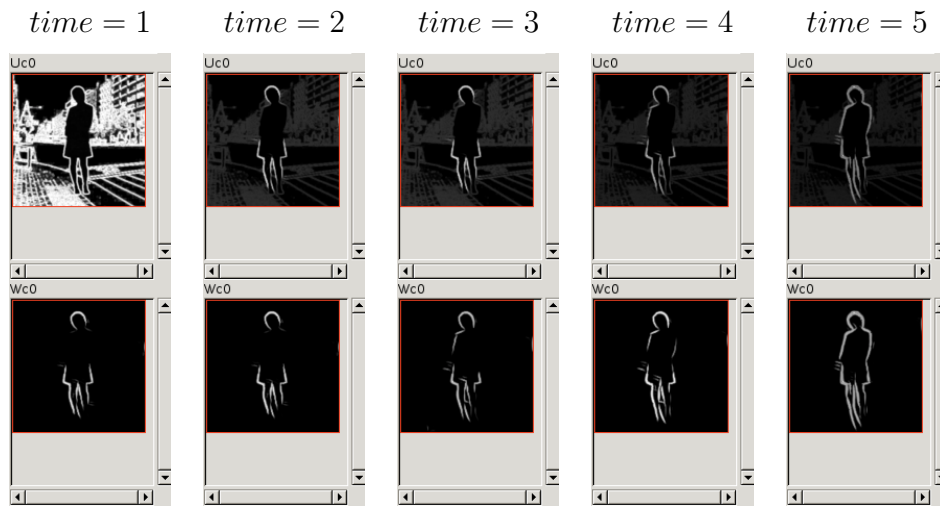


図 4.10: 複雑背景下における人の検出 (学習パターンを縮小).

た. しかしながら, 被験者二人が隣あった場合は人の形状を想起できなかった. 恐らく, 被験者二人によって一つの形状と見なされた複雑な背景のエッジが影響したものと考えられる.

図 4.12 は MIT CBCL Pedestrian Database[60] のデータを入力としたものである. エッジが鮮明な場合は, 人の形状を想起できたが, 不鮮明な場合は想起できなかった. しかし, 不鮮明な場合の入力画像を見ると, 背景に複数の影が存在したりと非常に複雑な背景となり, エッジ抽出された画像は, 我々人間の視覚によって判断しても人の形状とは認識し難たい.

4.4 まとめ

本章では, 選択的注意機構のモデルを用いて, 無数のパターンが存在する人の検出を実現するために, 我々は, ”立つ”, ”座る”等, 各姿勢の



図 4.11: 実画像での複雑背景下における人の検出。

典型的なパターンを学習パターンとして学習しネオコグニトロンの最終層の認識細胞を姿勢単位で生成する事によって、各姿勢をネオコグニトロンの変形に対する頑強性でカバーしようと考えている。まず、今回は、人の“立つ”という基本的な姿勢に限定し、人の形状の1つを学習パターンとして、学習パターンが全く異なる人の形状を呈示した際のフィードバック経路の出力部に人の形状が想起されるかを調べた。選択的注意機構のモデルの入力は形状が異なる68パターンの人のシルエット画像から抽出されたエッジ画像を用いた。更に、68のパターンから任意に1つを選び、学習パターンとした。その結果、大別すると、”上半身(頭と肩周辺)と下半身(足)”、”上半身(頭と肩周辺)”、”下半身(足と腰周辺)”、”

一部のエッジ”という，4種類の結果が w_{c0} 層に想起された。”上半身(頭と肩周辺)と下半身(足)”が想起された場合に人の形状を想起できたと仮定すると，最も良いもので，58%の想起率を示し，上半身(頭と肩周辺)のみと下半身(足と腰周辺)のみが w_c 層が想起された場合を加えると91%の想起率を示した。また，ネオコグニトロンの細胞面のサイズ(解像度)を 81^2 と 161^2 変化させると，サイズが低い方が，人の形状を想起しやすく，サイズが高い方が想起しにくかった。サイズが高い方では，服や姿勢のパターンの微細構造を抽出できるようになった代わりに変形の許容度が下がり，逆に人のパターンを想起し難い結果となった。更に，学習パターンには服装や関節部位に特徴のある姿勢と直立不動な姿勢の2つのシルエット画像を用いたが，直立不動な姿勢のパターンを学習パターンとした方が人の形状を想起しやすかった。立つという姿勢では，頭部と足部の形状に特徴があり，また変形も比較的少ない。一方で，胴周りは，服装や携帯物や，腕の姿勢によって無数に姿勢が存在する。したがって，ある意味，特徴が少ない直立不動な姿勢の方が学習パターンとして良いという結果になった。更に，スケッチ画による鮮明なエッジの複雑な背景下において人の形状が想起可能かを調べた結果，背景を抑制し人の形状を想起する事が可能であった。また，実画像実験では，複雑背景下のもとで複数の人が並んでいる状態等では，パターンの想起は進まず，背景のエッジに引きつられる事が多かった。また，MIT CBCL Pedestrian Databaseのデータでは，エッジが鮮明な場合は，人の形状を想起できたが，複雑な背景下でエッジが不鮮明な場合はほぼ不可能であった。しかしながら，そもそもエッジが不鮮明な場合は，我々人間も人の形状を認識するのが不可能である。

今回は、”立つ”という姿勢に関して実験を行ったが、今後、その他の姿勢に関する検討する必要がある。また、ネオコグニトロン of 細胞面サイズに関しては、上述したように、サイズによって想起率が変わってくるので、ネオコグニトロンに多重スケール性を持たせる等の改良が必要であると思われる。実画像に関しては、そもそもエッジが不鮮明な場合は、我々人間も人の形状を認識するのが不可能な場合が存在する。したがって、今回エッジという輝度情報のみを使用したか、影や両眼視差等のその他の特徴量と加える必要があるのかもしれない。または、ステレオ視等の前処理によって、複雑な背景におけるエッジの不鮮明さをある程度除去する等の方法が考えられる。今後、これらの課題を解決し、選択的注意機構のモデルを用いて人の検出を実現したいと考えている。

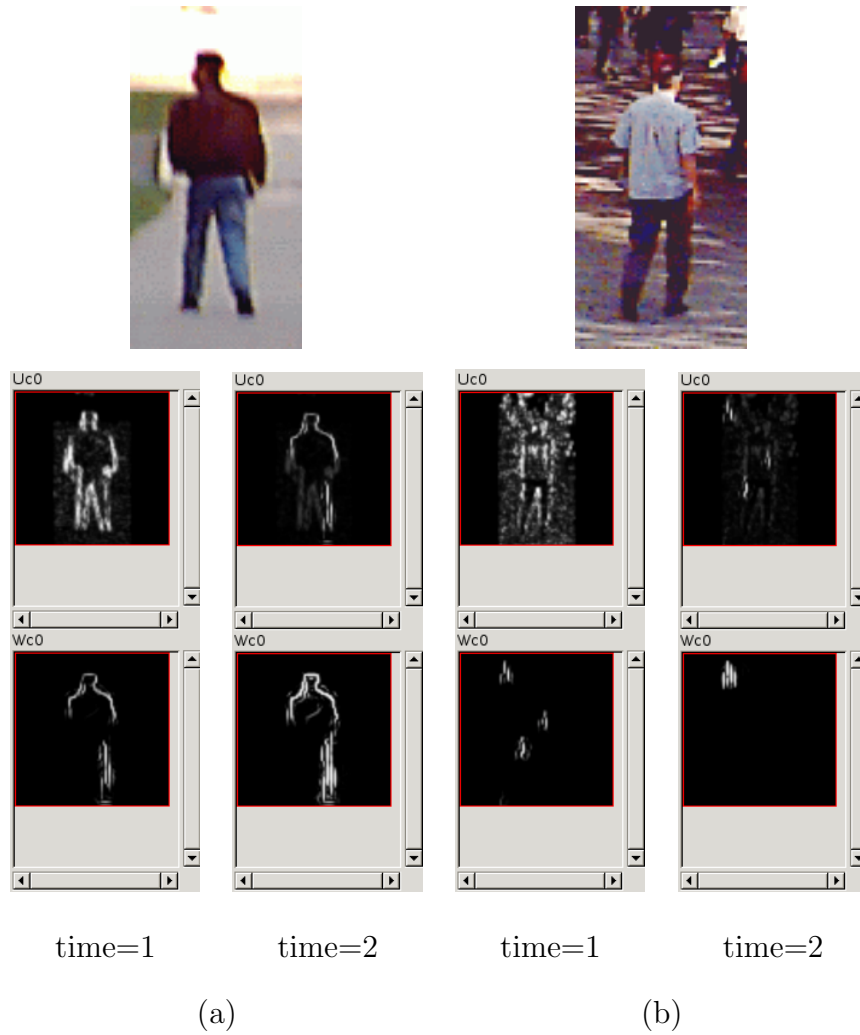


図 4.12: 実画像での複雑背景下における人の検出:MIT CBCL Pedestrian Database[60]. (a) エッジが鮮明な場合 (b) エッジが不鮮明な場合.

第5章 結論

本論文では、生理学的知見とネオコグニトロンの比較から工学的応用までを広く議論した。

第2章では、我々はネオコグニロンを用いた腹側経路のモデルを提案した。このモデルは、V1野の単純型細胞と複雑型細胞のモデルを拡張したものであり、S細胞、C細胞の組み合わせは、Sakai と Tanaka が示した知見 [41] を支持している。我々のモデルを用いて、ゼムクリップ状の3次元物体の画像の認識シミュレーションを行なった結果、Logothetis らの実験結果 [36] を定性的によく表していることが判明した。これらの結果は、ネオコグニトロンの腹側経路モデルとしての意義が大きい事を示している。また、ネオコグニロンを用いる事で、V2野以降の受容野の複雑な空間結合分布の形状を推定できる可能性を示した。しかしながら、ネオコグニロンは、厳密な生理学的モデルとしては、次のような課題を有している。ネオコグニロンは、結合に並進対称性があることを仮定しているが、この仮定を前提としない学習則の構築を行わなければならない。また、C細胞の結合は学習によって決定しているわけではなく、位相不変な応答を出すために作り付けられた固定結合である。したがって、C細胞の様な位相不変な結合を学習によって構成することを考えなければならない。また、陰影やオクルージョンの影響を補正すると思われる

フィードバック経路の影響についても議論の余地が残されている。今後、これらの課題を解決し、生理学的モデルとしてより妥当な腹側経路モデルに近づけていく予定である。

第3章では、GPUによるネオコグニトロンのGPU実装を行った。Poliらの細胞間の結合データに関するネオコグニトロンの高速化に対して、我々は細胞面サイズに関して高速化を行った。その結果、GPU(9400m)では15[f/s]程度、GPU(GTX480)では30[f/s]程度の処理能力があり、アプリケーションとして十分な応答速度を満たしている事がわかった。密疎データに関して比較すると、密データについてはCPUの処理よりもGPUの実装が速くなる事が確認された。疎データについてはCPU側の演算回数工夫に対して、低性能のGPU(9400m)の方が遅くなったが、最新のGPU(GTX480)の方が速い結果となった。つまり、背景差分や2値化等の何らかの前処理を施した画像をネオコグニトロンで使用する場合は、現在のNotePCに標準搭載されているクラスのGPU性能では、CPUでの演算の工夫に対してGPU実装を行う意義が薄れてしまう事がわかった。しかしながら、実画像に零要素が多く含まれるとは考えにくく、実画像をネオコグニトロンの入力データとできるのであれば、十分に価値はある。また、ネオコグニトロンの全ての階層にGPU処理を行うのではなく、細胞面とGPUの性能から算出されるサイズの細胞段にのみGPUを実装するのが良い事を明らかにした。細胞間の結合データが多い場合は、Poli[52][53]らが行った実装方法は有効であり、深い階層で細胞密度が低く細胞間の結合が多くなる場合は彼らの実装を用い、低次の階層で細胞面が大きく細胞間の結合が少なくなる場合及び学習段階に対しては細胞面サイズの大きさに着目した我々の実装を行うのが良いと考えられる。我々の

実装は GPU の性能を十分に引き出しているわけではない。例えば GPU のメモリ計画は実装の容易さから TextureMemory を使用したが、データ転送が高速な SharedMemory を使用すればさらなる高速化が期待される。また、ネオコグニトロンのデータ構造は CPU 側に配置しており、GPU 側に配置する事で通信回数が減り CPU-GPU 間のデータ転送を大幅に改善可能である。しかしながら、アプリケーションの用途によっては、CPU 側から GPU のデータ構造を必要とする場合も考えられる。今後 GPU の性能を十分に引き出せる細かな実装の改善を行い、ネオコグニトロンをベースにした画像認識アプリケーションの開発を行う予定である。

第 4 章では、選択的注意機構のモデルを用いた人の検出に関して述べた。人の形状は姿勢 (各関節部) の変化によって無数に存在する。オコグニトロンが物体の変形に比較的強いと言われているが、あまりに複雑な形状の変化に対して許容できる訳ではない。そこで、我々は、“立つ”、“座る”等、各姿勢の典型的なパターンを学習パターンとして学習し、ネオコグニトロンの最終層の認識細胞を姿勢単位で生成する事によって、各姿勢をネオコグニトロンの変形に対する頑強性でカバーしようと考えている。今回は、人の“立つ”という基本的な姿勢に限定し、人の形状の 1 つを学習パターンとして、学習パターンが全く異なる人の形状を呈示した際のフィードバック経路の出力部に人の形状が想起されるかを調べた。選択的注意機構のモデルの入力は形状が異なる 68 パターンの人のシルエット画像から抽出されたエッジ画像を用いた。更に、68 のパターンから任意に 1 つを選び、学習パターンとした。その結果、大別すると、“上半身 (頭と肩周辺) と下半身 (足)”、“上半身 (頭と肩周辺)”、“下半身 (足と腰周辺)”、“一部のエッジ”という、4 種類の結果が w_{c0} 層に想起された。”

上半身(頭と肩周辺)と下半身(足)”が想起された場合に人の形状を想起できたと仮定すると、最も良いもので、58%の想起率を示し、上半身(頭と肩周辺)のみと下半身(足と腰周辺)のみが w_c 層が想起された場合を加えると91%の想起率を示した。更に、スケッチ画による鮮明なエッジの複雑な背景下において人の形状が想起可能かを調べた結果、背景を抑制し人の形状を想起する事が可能であった。また、実画像実験では、エッジが鮮明な場合は、人の形状をある程度想起できたが、複雑な背景下でエッジが不鮮明な場合はほぼ不可能であった。しかしながら、そもそもエッジが不鮮明な場合は、我々人間も人の形状を認識するのが不可能である。今回は、”立つ”という姿勢に関して実験を行ったが、今後、その他の姿勢に関しても調べ、人の姿勢の識別や人と非人対象における識別に関して研究を行う必要がある。実画像に関しては、そもそもエッジが不鮮明な場合は、我々人間も人の形状を認識するのが不可能な場合が存在する。したがって、今回エッジという輝度情報のみを使用したか、影や両眼視差等のその他の特徴量と加える必要があるのかもしれない。または、ステレオ視等の前処理によって、複雑な背景におけるエッジの不鮮明さをある程度、除去する等の方法が考えられる。今後、これらの課題を解決し、選択的注意機構のモデルを用いて人の検出を実現したいと考えている。

今回、我々が用いたモデルは福島が考案した初期のモデルである。最近のネオコグニトロンには、外側膝状体に対応するLG層と呼ばれる層を第1細胞段の前に配置し、認識率が向上したことが報告されている。外側膝状体には、同心円状のオン中心型やオフ中心型の受容野が存在する事が報告されており、LG層の受容野も同様に、メキシカンハット型の固

定結合で、オン型とオフ型の受容野を形成している [29]. LG 層によって、ある特定の空間周波数成分を遮断することができ、輝度値が一定である領域の出力は 0 となる. 更に、C 細胞では、周辺抑制を行うように改良されている. ネオコグニトロンの C 細胞には、パターンの位置ずれによる認識の許容度を上げるためにぼかしの操作を行っているが、同じ特徴が 2 個隣接して存在する場合、単一の特徴がぼかされたものなのか、それとも隣接した 2 個の特徴がぼかされたものなのかの判別が困難になることがある. そこで、興奮性結合の周辺に抑制性の領域を配置し、もとの特徴の個数を容易に検出することが出来るように改良されている [29]. また、学習に関しても ”Winner-Kill-Loser” という新しい学習則が提案されている [62]. 従来のネオコグニトロンの S 細胞の学習は、”Winner-Take-All” 型の競合学習を用いている. この学習則では、学習パターンに複数の細胞が反応した際、勝者となった 1 つの細胞の結合を強化するのみであったが、反応したその他の細胞は、類似した冗長な存在として残る事になる. ”Winner-Kill-Loser” では、勝者となった 1 つの細胞以外の反応した細胞を消去する事で、細胞の参照ベクトルが均一に配置される事となる. 今後の課題として、これら改良されたネオコグニトロンを用いる必要がある.

実世界の 3 次元物体は、脳の視覚情報処理において、2 次元 (Image 2D Model) 又は 3 次元 (Structural Description Model) で処理されるのか? という 2 つの対立する知見が存在している. 2 次元での情報処理は、視点依存であり、物体をさまざまな角度から捉えた、いくつかの 2 次元画像の組み合わせによって、3 次元物体を認識しているという説である. Palmer らは、最も良くその物体を表している景観 (典型的景観) が存在する事を

示し、そのような景観の認識力が最も良いと報告している [63]. つまり、典型的景観の存在が、視点依存性を示唆していると考えられる. また、Ullman らは、一つの物体に対して複数の景観からの画像を組み合わせれば他の景観も認識できる事を計算理論で証明した [64, 65]. 生理学的には、第 2 章で説明した、Logothetis らの実験結果 [36] が視点依存性を支持している. 一方で、3 次元での情報処理は、視点非依存であり、3 次元物体は、物体の基本要素となるような部品によって表現されるという説である. Marr は、一般化円錐と呼ばれる円錐と円柱を拡張したような 3 次元部品をいくつか組合わせる事によって 3 次元物体のモデルを作り、それをもとにしてパターン認識を行うと考えた [66]. また、Biederman は、3 次元物体を構成する基本部品となるジオンを仮定し、その構造的な組み合わせで対象が記述できると考えた [67]. 生理学的には、Uka らは、サルの生理学実験によって、IT 野に両眼視差に選択的なニューロンの存在を報告し、IT 野に至る視覚腹側経路が 2 次元から 3 次元へ復元する過程に関与をしているという仮説を支持している [68, 69]. また、Yamane らは、進化的アルゴリズムにより生成された視覚刺激に対するサルの IT 野のニューロンの応答を調べ、3 次元形状に選択的なニューロンが IT 野に存在すると報告している [70]. ネオコグニトロンは、2 次元での情報処理を支持しており、Logothetis らの生理学的知見を上手く説明でき工学的にも高い認識能力を示している. しかし、複雑な背景や環境な変化に対して高度な物体認識を行うためには、3 次元に関する生理学的知見の存在が示す通り、3 次元情報処理の機構を組み込む等の改良を行う必要があると我々は考えている.

最後に、現在、脳の生理学的研究は盛んに行われており、一步一步着

実に脳の情報処理の仕組みが解明されてきているが、まだ、明らかにされていない事も多く、視覚のシステムひとつとっても満足なものはない。そのように、明らかにされている部分と明らかにされていない部分に対する仮説を基に、ひとつの情報処理のシーケンシャルを実現する事には、生理系、数理系、工学系という別学問の相互関係による議論を深める事に置いて意義がある。ネオコグニトロンは生理学的、工学的なモデルという2つの側面を持ったモデルであり、視覚系の学問の分野を横断できる価値のあるモデルと考える。また、ネオコグニトロンは1980年に発表された古いモデルであるが、その柔軟な処理能力は実に魅力的である。過去において、計算機能力の問題で、その実利性に疑問がもたれ、工学的に広く利用される事はなかったが、昨今の計算機能力の向上により、見直されて良いと考える。脳の視覚情報処理にヒントを得たネオコグニトロンというモデルを工学的に応用し、更には産業界の応用アプリケーションとして実用化する事は、学術世界の知見を社会に還元していく試みとして非常に意義がある研究と考えている。

謝辞

本研究を進めるにあたり、終始御指導下さいました宮本弘之准教授に心から御礼申し上げます。また、「魅力ある大学院教育」イニシアティブ(大学院 GP) 出稽古修行プログラムにて、ネオコグニトン型神経回路モデルに関してご教授いただき、また、共著者としてさまざまな面でサポートしていただきました事、東京大学 岡田真人 教授、電気通信大学 庄野逸 准教授に厚く御礼申し上げます。ファジィシステム研究所 客員研究員(兼務: 関西大学 研究員) 福島邦彦先生におきまして、共著者として、研究のご指導をしていただき、心から御礼申し上げます。本論文の副査をして頂きました九州工業大学大学院 生命体工学研究科 脳情報専攻 松岡清利教授ならびに花沢明俊准教授に感謝致します。最後に、共に研究を励んできた宮本研究室の多くの仲間へ深く感謝の意を表すると共に、様々な点で研究をご援助下さいました九州工業大学大学院 生命体工学研究科 脳情報専攻の皆様へ感謝の意を表します。

参考文献

- [1] 教育機器編集委員会. (1972): 産業教育機器システム便覧, 日科技連出版社.
- [2] Fukushima, K. (1980): Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, *Biological Cybernetics*, Vol.36, No.4, pp.193-202.
- [3] 福島邦彦. (1989): 神経回路と情報処理, 朝倉書房.
- [4] Broadman, K. (1909): Vergleichende Lokalizationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues. J.A.Barth, Leipzig.
- [5] Hubel, D.H. and Wiesel, T.N. (1959): Receptive fields of single neurones in the cat's striate cortex, *J.Physiol.(London)*, Vol.148, pp.574-591.
- [6] Hubel, D.H. and Wiesel, T.N. (1962): Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, *J.Physiol.(London)*, Vol.160, No.1, pp.106-154.

- [7] Hubel, D.H. and Wiesel, T.N. (1965): Receptive fields and functional architecture in two nonstriate visual areas(18 and 19) of the cat., *J.Physiol.(London)*, Vol.28, No.2 pp.229–289.
- [8] Hubel, D.H. and Wiesel, T.N. (1968): Receptive fields and functional architecture of monkey striate cortex, *J.Physiol.(London)*, Vol.195, No.1, pp.215–243.
- [9] Felleman, D.J. and Van Essen, D.C. (1991): Distributed hierarchical processing in primate cerebral cortex. *Cereb.Cortex*, Vol.1, pp.1-47.
- [10] Felleman, D.J. and Van Essen, D.C. (1991): Distributed hierarchical processing in the primate cerebral cortex, *Cerebral Cortex*, Vol.1, pp.1–47.
- [11] Kobatake, E. and Tanaka, K. (1994): Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex, *Journal of Neurophysiology*, Vol.71, No.3, pp.856–867.
- [12] Perret, D.I., Mistlin, A.J. and Chitty, A.J. (1987): Visual neurons responsive to faces, *Trends in Neuroscience*, Vol.10, pp.358–364.
- [13] Gross,C.G. and Rocha-Miranda,C.E. and Bender,D.B.K. (1972): Visual Properties of Neurons in Inferotemporal Corte

of the Macaque, Journal of Neurophysiology, Vol.71, No.3, pp.96–111.

- [14] Desimone, R. Albright, T.D. Gross, C.G. and Bruce, C. (1995): Stimulus-selective Properties of Inferior Temporal Neurons in the Macaque. J.Neuroscience, Vol.4, pp.2051–2062.
- [15] Iwai, E. Mishkin, M. (1969): Further evidence on the locus of the visual area in the temporal lobe of the monkey, Exp.Neurol., Vol.25, No.4, pp.585–594.
- [16] Bruce, C. Desimone, R. and Gross, C.G. (1981): Visual Properties of Neurons in a Polysensory Area in Superior Temporal Sulcus of the Macaque. J.Neurophysiol., Vol.46, No.2, pp.369–384.
- [17] Perret, D.I. Rolls, E.T. Caan, W. (1982): Visual neurones responsive to faces in the monkey temporal cortex, Exp. Brain Res., Vol.47, No.3, pp.329–342.
- [18] Perret, D.I. Smith, P.A.J. Potter, D.D., Mistlin, A.J., Head, A.S., Milner, A.D., Jeeves, M.A. (1984): Neurones responsive to faces in the temporal cortex: studied of functional organization, sensitivity to identity and relation to perception, Human Neurobiol., Vol.3, pp.197–208.

- [19] Perret, D.I. Smith, P.A.J. Potter, D.D., Mistlin, A.J., Head, A.S., Milner, A.D., Jeeves, M.A. (1985): Visual cells in the temporal cortex sensitive to face view and gaze direction, Proc. Roy. Soc. Lond., B, Vol.223, No.1232, pp.293–317.
- [20] Tanaka, K. (1993): Neuronal Mechanisms of Object Recognition, Science, Vol.262, pp.685–688.
- [21] Tanaka, K. (1996): Inferotemporal cortex and object vision, Annual Review of Neuroscience, Vol.19, pp.109–139.
- [22] Rosenblatt, F. (1958): The perceptron: A probabilistic model for information storage and organization in the brain. Psychol.Rev., Vol.65, No.6, pp.386–408.
- [23] Rosenblatt, F. (1962): Principles of Neurodynamics. Spartan Books, Washington D. C.
- [24] Minsky, M. and Papert, S. (1969): Perceptrons. MIT Press, Cambridge, MA.
- [25] Amari, S. (1967): A Theory of Adaptive Pattern Classifiers. IEE Trans. EC-16, pp.279–307.
- [26] Rumelhart, D.E., McClelland, J.L., and PDP Research Group. (1986): Parallel Distributed Processing: Explorations in Microstructure of Cognition. MIT Press, Cambridge, MA London.

- [27] Fukushima, K. (1975): Cognitron: A Self-Organizing MultiLayerd Neural Network Model, Biological Cybernetics, Vol.20, No3, pp.121–136.
- [28] Fukushima, K. Miyake, S. (1982): Neocognitron: a new algorithm for pattern recognition tolerant of deformations and shifts in position, Pattern Recogn., Vol.15, No.6, pp.455–469.
- [29] Fukushima, K. (2003): Neocognitron for handwritten digit recognition, Neurocomputing, Vol.51, pp. 161–180
- [30] 福島邦彦. (1994): 神経回路理論, ブレインサイエンス, Vol.5, No.3, pp.285–292.
- [31] Fukushima, K. (1989): Neural Network Model for Selective Attention in Visual Pattern Recognition and Associative Recall, Applied Optics, Vol.26. No.23, pp.4985–4992.
- [32] Fukushima, K. Imagawa, T. (1993): Recognition and segmentation of connected characters with selective attention, Neural Networks, Vol.6, No.1, pp. 33–41.
- [33] 庄野逸, 福島邦彦. (1994): 折れ点処理回路を用いた選択的注意機構による英字筆記体連結文字列認識, 電子情報通信学会論文誌 D-II, J77-D-II[5], pp. 940–950.
- [34] Fukushima, K. Kimura, E., Shouno, H. (1998): Neocognitron with improved bend-extractors: recognition of handwritten

digits in the real world, *Neural Computing & Applications*,
7, pp. 260–272.

- [35] 橋本英樹, 福島邦彦. (1997): 選択的注意機構による顔の部分パターンの認識と切出し, *電子情報通信学会論文誌 D-II*, J80-D-II[8], pp. 2194–2202.
- [36] Logothetis, N.K. and Sheinberg, D.L. (1996): Visual object recognition, *Annual Review of Neuroscience*, Vol.19, pp.577–621.
- [37] Bricolo, E., Poggio, T. and Logothetis, N.K. (1997): 3D object recognition: a model of view-tuned neurons. In M. Mozer, M.I. Jordan and T. Petsche (Ed.), *Advances in Neural Information Processing Systems*, Vol.9. (pp.41–47) Cambridge, MA: The MIT Press.
- [38] Riesenhuber, M. and Poggio, T. (1998): Just one view: invariances in inferotemporal cell tuning. In M.I. Jordan, M.J. Kearns and S.A. Solla (Ed.), *Advances in Neural Information Processing Systems*, Vol.10. (pp.215–221) Cambridge, MA: The MIT Press.
- [39] Riesenhuber, M. and Poggio, T. (1999): Hierarchical models of object recognition in cortex, *Nature Neuroscience*, Vol.2, pp.1019–1025.

- [40] Rolls, E.T. (1994): Brain mechanisms for invariant visual recognition and learning, *Behavioural Processes*, Vol.33, pp.113–138.
- [41] Sakai, K. and Tanaka, S. (1997): Computational analysis of the nonlinear spatial structure in V1 complex cells, *日本神経回路学会第8回全国大会講演論文集*, pp.79–80.
- [42] Jones, J.P. and Palmer, L.A. (1987): An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex, *Journal of Neurophysiology*, Vol.58, No.6, pp.1233–1258.
- [43] Fukushima, K. and Wake, N. (1992): An improved learning algorithm for the neocognitron. In I. Aleksander, J. Taylor (Ed.), *Artificial Neural Networks*, Vol.1 of 2. (pp.497–505) Amsterdam: North-Holland.
- [44] Fukushima, K. (1987): Neural network model for selective attention in visual pattern recognition and associative recall, *Applied Optics*, Vol.26, No.23, pp.4985–4992.
- [45] LeCun, Y. Boser, B. Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., and Jackel, L.D.(1989): Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, Vol.1, No.4, pp.541–551.

- [46] Microsoft. DirectX デベロッパースセンター:
<http://msdn.microsoft.com/ja-jp/directx/default.aspx>.
- [47] Kyoung, SO, Jung,KC. (2004): GPU implementation of neural networks, *Pattern Recognition* 37, 6, pp.1311-1314.
- [48] Rolfes, T. (2004): Artificial Neural Networks on Programmable Graphics Hardware, in *Game Programming Gems 4*, pp.373-378.
- [49] Bohn, C.A. (1998): Kohonen Feature Mapping Through Graphics Hardware. In *Proceedings of 3rd Int. Conference on Computational Intelligence and Neurosciences 1998*.
- [50] “CUDA Programming Guide” :
http://www.nvidia.com/object/cuda_develop.html
- [51] “CUDA ZONE” :
http://www.nvidia.com/object/cuda_home.html
- [52] Poli,G. Saito, J.H. , Mari, J.F. , Zorzan, M.R. (2008): Processing Neocognitron of Face Recognition on High Performance Environment Based on GPU with CUDA Architecture, *SBAC-PAD, 20th International Symposium on Computer Architecture and High Performance Computing*, pp.81-88.
- [53] Poli,G and Saito,JH. (2010): Parallel Face Recognition Processing using Neocognitron Neural Net-

work and GPU with CUDA High Performance Architecture, Face Recognition, Milos Oravec (Ed.), ISBN:978-953-307-060-5, INTECH. Available from: <http://sciyo.com/articles/show/title/parallel-face-recognition-processing-using-neocognitron-neural-network-and-gpu-with-cuda-high-perfor>.

- [54] Dalal, N.,Triggs,B. (2005): Histograms of oriented gradients for human detection”,Proc.of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.886–893.
- [55] Suard,F. and Broggi,A. (2006): Pedestrian Detection using Infrared images and Histograms of Oriented Gradients, IEEE Symposium on Intelligent Vehicule, pp. 206–212.
- [56] Zhu, Q. Avidan,S., Yeh, M. and Cheng,K. (2006): Fast Human De- tection Using a Cascade of Histograms of Oriented Gradients, IEEE Computer Vision and Pattern Recognition, Vol. 2, pp. 1491-1498.
- [57] Freund,Y. and Schapire, R.E. (1997): A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55(1), pp.119–139.

- [58] 三井相和, 山内悠嗣, 藤吉弘亘. (2008): Joint HOG 特徴を用いた 2 段階 AdaBoost による人検出. 第 14 回画像センシングシンポジウム SSII08, IN1-06.
- [59] FLOP DEDIGN: <http://www.flopdesign.com/>
- [60] MIT CBCL Pedestrian Database: <http://cbcl.mit.edu/cbcl/software-datasets/PedestrianData.html>
- [61] Fukushima, K. Tanigawa, M. (1996): Use of Different Threshold in Learning and Recognition, *Neurocomputing*, Vol.11, No.1, pp.1-17.
- [62] Fukushima, K. (2010): Neocognitron trained with winner-kill-loser rule, *Neural Networks*, 23[7], pp. 926-938.
- [63] Palmer, S.E., Rosch, E., and Chase, P. (1981): Canonical perspective and the perception of objects. In Long, J. and Baddeley, A.(ed.), *Attention and Performance*, Vol.9, Hillsdale, NJ: Erlbaum, pp.135-151.
- [64] Ullman, S. and Basri, R. (1991): Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13, pp.992-1005.
- [65] Ullman, S. (1998): Three-dimensional object recognition based on the combination of views. *Cognition*, 67(1), pp.21-44.

- [66] 乾敏郎, 安藤広志 (訳). D. Marr 著.(1987): ビジョン - 視覚の計算理論と脳内表現-. 産業図書.
- [67] Biederman, I. (1987): Recognition-by-components: a theory of human image understanding. *Psychological Review*, 94, pp.115-147.
- [68] Uka, T., Tanaka, H., Kato, M., Fujita, I. (1999): Behavioral evidence for visual perception of 3-dimensional surface structures in monkeys. *Vision Res.*, 39, pp.2399-2410.
- [69] Uka, T., Tanaka, H., Yoshiyama, K., Kato, M and Fujita, I. (2000): Disparity Selectivity of Neurons in Inferior Temporal Cortex. *J. Neurophysiol.*, 84, pp.120-132.
- [70] Yamane, Y., Carlson, E.T., Bowman, K.C., Zhihong, W. and Connor, C.E. (2008): A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nature Neuroscience*, 11, pp.1352-1360.

業績リスト

I 原著論文

1. 吉塚 武治, 庄野 逸, 宮本 弘之, 岡田 真人, 福島 邦彦: 「ネオコグニトロンによる視覚腹側経路のモデル化」日本神経回路学会誌, Vol. 14, No. 4, pp.266-272, (2007年12月)

II 国際会議発表論文

1. Yoshizuka, T. Shimizu, M. Miyamoto, H.: 「Toward Human Arm Attention and Recognition Using a Computer-Vision- and Neocognitron-Based Approach」Proc. ICONIP2007 Lecture Notes in Computer Science, Vol. 4984, pp.1071-1080, (2007年11月)

III 国内学会口頭発表

1. 吉塚武治, 宮本弘之: 「ネオコグニトロンのGPU実装」信学技報, Vol. 110, No. 246, pp.47-51, (2010年10月)

IV その他

1. Tsumura, T. Yoshizuka, T. Nojirino, T. Noma, T.: T4: a motion-capture-based goal-directed real-time responsive lo-

comotion engine Proc. Computer Animation 2001 IEEE Computer Society, pp.52-60, (2001 年 11 月)

2. Shimizu, M. Yoshizuka, T. Miyamoto, H. : 「A Mobile Robot Controlled by Gesture with Stereo Vision」 Proc. BrainIT2006, p.49, (2006 年 9 月)

3. Shimizu, M. Yoshizuka, T. Miyamoto, H. : 「Markerless Gesture Recognition under Cluttered Background」 Proc. SCIS&ISIS2008, pp.1557-1561, (2008 年 9 月)