

ガボール型受容野をもつ TAM ネットワークの提案

A Proposal of TAM Network with Gabor Type Receptive Field

林 勲

Isao Hayashi

関西大学

Kansai University

ジェームズ R. ウィリアムソン

James R. Williamson

ロッキード・マーチン

Lockheed Martin Corp.

Abstract: The TAM (Topographic Attentive Mapping) Network based on a biologically-motivated neural network is an especially effective model. When the network makes an incorrect output prediction, the attentional feedback circuit modulates the learning rates, and adds a node to the category layer in order to improve the network's prediction accuracy. In this paper, Gabor type receptive field is formulated in the input layer of TAM network. By using Gabor filtering, we can detect orientation selectivity of visual images, and simulate the mechanism of early vision of brain as a computational model.

Keywords: Visual Neural Network, Gabor Function, Receptive Fields, Image Recognition

1. はじめに

一般に、人の視覚系では、網膜に入力された画像信号は受容野に対応した視細胞で処理された後、外側膝状体を介して第一次視覚野や第二次視覚野に入力される。第一次視覚野や第二次視覚野の方位選択性コラムでは、対象画像の方位、方向、色などが識別され、側頭葉や頭頂葉などのより高位レベルで対象画像の形状や動きなどが検知される [1]。視細胞の神経節細胞には、単純型細胞、複雑型細胞、超複雑型細胞が存在する。単純型細胞では、対象画像の直線成分の方位選択性を検知し、複雑型細胞ではエンドストップを検知する。また、超複雑型細胞では線分の結合や補間などの輪郭統合を行う。単純型細胞の受容野における方位選択性の表現モデルとしてガボール関数 [2-4] がある。ガボール関数は、ガウス関数と正弦・余弦関数から構成され、受容野構造として画像の任意の周波数成分を抽出するフィルタリング機能をもつ。

一方、視覚系の処理過程として、Hubel-Wiesel の階層仮説に代表される多くの視覚系ニューラルネットワーク [5] が提案されている。特に、受容野モデル [6]、V1-V2 モデル [7]、ネオコグニトロン [8]、ARTMAP [9,10]、TAM (Topographic Attentive Mapping) ネットワーク [11,12] など是有用である。TAM ネットワークは共振学習とビジランス機能を持ち、パターン認識において有効なモデルである。構造的には、入力層の特徴マップ層、基盤層、カテゴリー層、クラス層の 4 層の階層構造からなり、第一次視覚野以降の視覚前野を模擬している。

与えられた教師値と出力値に差がある場合、特徴マップ層から基盤層への興奮性学習のボトムアップ学習およびクラス層から基盤層への抑制性学習によるトップダウン学習とビジランス機能、カテゴリー層でのノード増加による特徴マップ層の特徴量抽出によって、高い学習機能を確保することができる。また、学習後には、情報量を用いたブルーニング機能により、各層の不必要なリンクとノードを削除して、ファジィルールを獲得できる [13-15]。

本論文では、特徴マップ層以前の層構造として、人の受容野構造を模擬したガボール型受容野を導入した新たな TAM ネットワークを提案する [16-18]。この層構造は網膜層、神経節細胞層、外側膝状体 (Lateral Geniculate Nucleus: LGN) 層から構成され、ガボール関数のガボールフィルタリングにより、対象画像の方位選択成分を抽出して、特徴マップ層での特徴マップデータを構成する。したがって、新たな TAM ネットワークは、網膜層、神経節細胞層、外側膝状体 (LGN) 層、特徴マップ層、基盤層、カテゴリー層、出力層の 7 層構造からなる。各層においては、ガボール型受容野により原画像の輝度情報から 16 方位選択成分を抽出し、輝度情報を正規化して特徴マップを構成する。また、視覚前野構造では、ボトムアップとトップダウンの双方向学習により、得られた特徴マップから対象画像の認識を行う。ここでは、まず、受容野構造と特徴マップ構造について議論し、輝度情報の信号処理アルゴリズムを定式化する。次に、アルファ

ベットの文字認識の例を用いて、本提案モデルの有用性と頑健性について検討する。なお、他の視覚系ニューラルネットワークと異なり、学習後のネットワーク構造から対象画像の特徴をファジィルールとして獲得できるので、獲得されたファジィルールの妥当性についても検討する。

2. ガボール関数による受容野構造

2次元正弦関数型のガボール関数 $G(x, y)$ は次のように定義される。

$$G(x, y) = Ae^{-\frac{1}{2}\left(\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right)} \times \sin(2\pi f_x x \cos \theta + 2\pi f_y y \sin \theta + \phi)$$

ただし、 A は振幅、 (μ_x, μ_y) はガボール関数の中心座標、 σ_x と σ_y は標準偏差、 f_x と f_y は周波数を表す。 $\phi = \pi/2$ の場合には、正弦関数型ガボール関数は余弦関数型ガボール関数となる。図 1 にガボール関数の例を示す。ここで、 $(\mu_x, \mu_y) = (0.0, 0.0)$ 、 $\sigma_x = 2.0$ 、 $\sigma_y = 2.0$ 、 $f_x = 1.0 \text{ Hz}$ 、 $f_y = 1.0 \text{ Hz}$ 、 $\theta = 0^\circ$ 、 $\phi = 0.0$ としている。

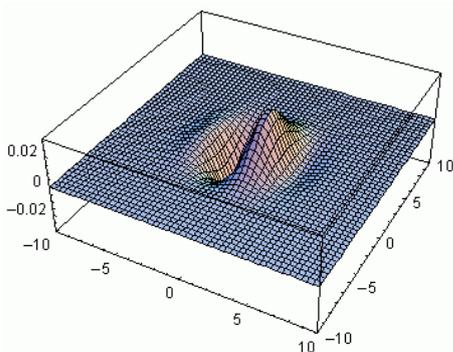


Fig. 1: Example of Gabor Function

ガボール関数を用いた方位選択成分の検出過程を図 2 に示す。まず、原画像に対してある特定座標に受容野を設定し、ガボール関数を用いたフィルタリング処理 [19] により 16 方位の方位選択成分を検出する。同じガボール型受容野を用いて、他の座標に対しても 16 方位の方位選択成分を検出する。すべての領域で方位選択成分を検出した後、抽出された輝度成分を正規化して特徴マップを構成する。

ガボールフィルタリング [19] は、脳内の第一次視覚野のモデルとして、また、特徴抽出モデルとして良く知られた一般的な手法であるが、TAM ネットワークの特徴マップの構成を説明するために必要であるので、ここでは敢えて概説することにする。いま、原画像の横幅と縦幅を

それぞれ $R_H \text{ pixels}$ 、 $R_V \text{ pixels}$ とし、座標 (p, q) における原画像の輝度を $O(p, q)$ 、 $1 \leq p \leq R_H$ 、 $1 \leq q \leq R_V$ とする。座標 (x, y) における第 i 番目の方位の方位選択成分 $C_i(x, y)$ 、 $i = 1, 2, \dots, M$ は、ガボール型受容野を用いて次のたたみ込み計算により得られる。

$$C_i(x, y) = \sum_{q=1}^{R_V} \sum_{p=1}^{R_H} G_i(x-p, y-q) \times O(p, q) \quad (1)$$

ただし、ここでは、原画像における畳み込みの積分範囲のみを表している。なお、一般に、全体の方位数 M は、 $M = 16$ とすることが勧められている。

原画像全体の第 i 番目の方位選択成分 C_i^W は、方位選択成分 $C_i(x, y)$ を $R_H \times R_V$ 個重ねることにより得られる。

$$C_i^W = \bigcup_X \bigcup_Y C_i(x, y), \quad x \in X, \quad y \in Y \quad (2)$$

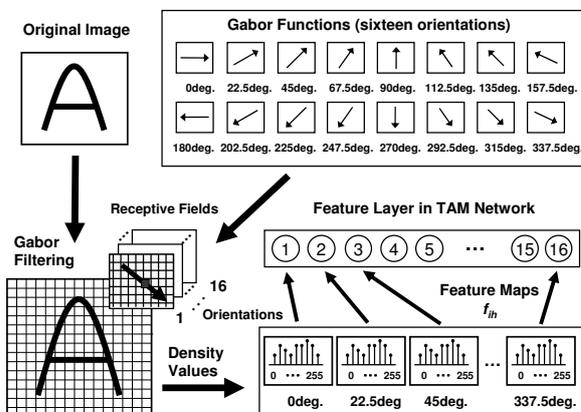


Fig. 2: Gabor Filtering Process

例として、アルファベット ‘A’ の画像の方位選択成分を抽出しよう。抽出した方位選択成分の結果を図 3 に示す。ただし、ここでは、各画素を 8 ビットの輝度で表現し、アルファベット文字を輝度 255 の白色、背景色を輝度 0 の黒色とした。パラメータは次のように設定した。

$$\begin{aligned} \sigma_x &= 2.5 & \sigma_y &= 1.92 \\ f_x &= 0.127 & f_y &= 0.127 \\ \phi &= 90.0 & A &= 1.0 \end{aligned}$$

左側のアルファベット ‘A’ の画像に対して、上段左端の画像が方位 0° の方位選択成分を表し、その右側は 22.5° 、同様に、 45.0° 、 67.5° 、 \dots 、 337.5° の方位選択成分を表す。これらの結果から、文字の方位選択成分が良好に抽出されていることがわかる。



Fig. 3: Images after Gabor Filtering

次に，標準偏差および周波数，振幅を変化させた場合の方位選択成分について議論しよう．まず，標準偏差 σ_x を変化させた場合の方位選択成分について検討した．135.0°において， σ_x を変化させた場合の方位選択成分画像とそのガボール関数を図4に示す． σ_x の値は，左から 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0 としている．ただし，他のパラメータ値を $\sigma_y = 1.92$, $f_x = f_y = 0.127$, $A = 1.0$ とした．標準偏差 σ_x の値が大きくなるほど，方位選択成分画像は不明瞭となっていることがわかる．同様に，標準偏差 σ_y を変化させた場合も方位選択成分の明瞭性は変化する．これらの結果から，標準偏差は方位選択成分の画像の明瞭性を表現するパラメータといえる．また，周波数 f_x および f_y , 振幅 A は，方位選択成分に対してそれほど感度が高くないことを確認している．

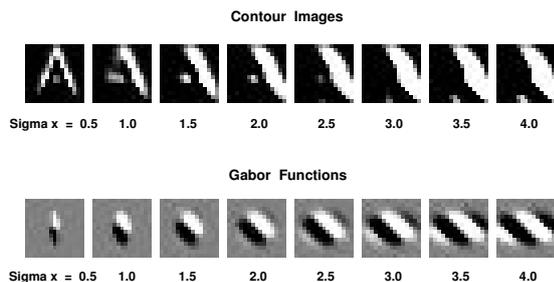


Fig. 4: Contours and Gabor Functions

3. ガボール型受容野をもつ TAM ネットワーク

ガボール型受容野構造をもつ TAM ネットワークの構成を図5に示す．本論文では，特徴マップ層以前の層構造として，新たに人の受容野構造を模擬したガボール型受容野を導入して受容野構造部を構成した．受容野構造部を構成する網膜層や神経節細胞層では，ガボール型受容野により原画像の方位選択成分を検出し，その信号は LGN 層を介して，第一視覚野において特徴マップとして構成される．この特徴マップを入力値とし，視覚前野構造部 [15] において，基盤層とカテゴリー層からクラス

層へのボトムアップ学習とクラス層から基盤層へのトップダウン学習により，原画像を認識する(図6を参照)．

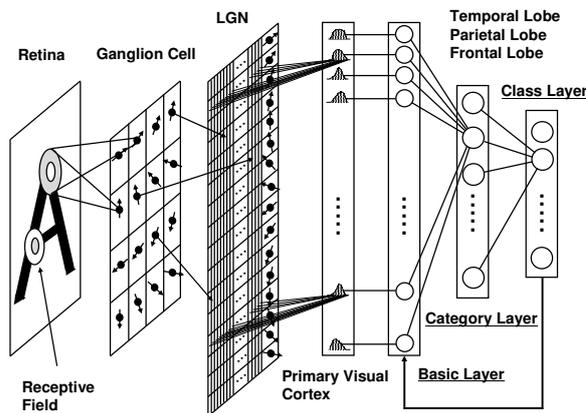


Fig. 5: TAM Network with Extensive Receptive Field

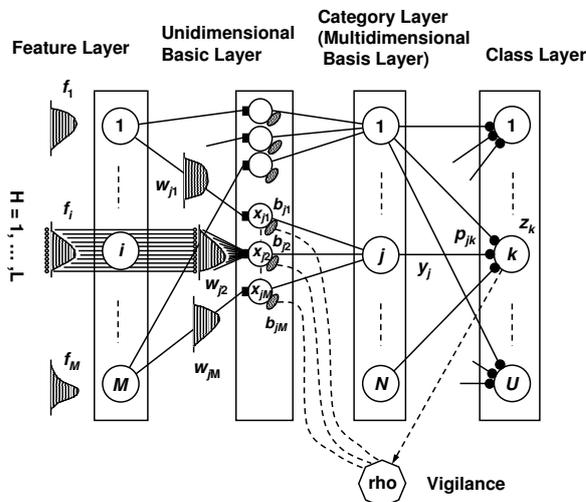


Fig. 6: Prestriate Visual Cortex in TAM Network

すでに受容野構造部の網膜層と神経節細胞層での信号処理過程は前章で説明している．すなわち，網膜層では，原画像に対して受容野のガボール関数を設定し，神経節細胞層では，(1)式と(2)式により方位選択性を計算する．一方，LGN層では，神経節細胞層からの信号を受け， M 個の方位選択成分を正規化する．LGN層において，第 i 番目の特徴マップ f_{ih} は， i 番目の方位選択成分の輝度を正規化して次のように得られる．

$$f_{ih} = \frac{\sum_{\{x,y|O_i(x,y)=h\}} C_i(x,y)}{\sum_{y=1}^{R_V} \sum_{x=1}^{R_H} C_i(x,y)}, \quad h = 1, 2, \dots, 256$$

ただし，特徴マップの離散値 h は，0～255の輝度に値

1を加えて、1~256の範囲に変換して用いている。

この f_{ih} は視覚前野構造部の特徴マップ層の特徴マップである。基盤層とカテゴリー層では、この特徴マップ f_{ih} と学習荷重 w_{jih} との活性値 x_{ji} を計算して、クラス層への入力 y_j を得る。

$$y_j = \prod_{i=1}^M x_{ji} = \prod_{i=1}^M \frac{\sum_{h=1}^L f_{ih} w_{jih}}{1 + \rho^2 b_{ji}} \quad (3)$$

ただし、 ρ はカテゴリー層のノード数を決定するためのクラス層から基盤層へのトップダウン信号用のパラメータであり、ピジランスパラメータと呼ぶ。また、 b_{ji} は抑制荷重である。

クラス層での出力値 K は次式より計算される。

$$K = \{k | \max_k z_k\} = \{k | \max_k \sum_{j=1}^N y_j p_{jk}\} \quad (4)$$

ただし、 p_{jk} は学習荷重である。

TAM ネットワークは、カテゴリー層においてノードの増幅機能をもつ。いま、 K^* をクラス層での正しい教師値とし、 OC を教師値の認識のためのしきい値とする。 $K \neq K^*$ の場合には、ピジランスパラメータ ρ は初期値 $\rho^{(init)}$ から $z_{K^*}/z_K \geq OC$ の条件が満足されるか最大値 $\rho^{(max)}$ になるまで $\rho^{(step)}$ ごとに上昇し、最大値になった場合には、カテゴリー層のノードが1個分増加する。

If $z_{K^*}/z_K < OC$ then repeat

(a) $\rho = \rho + \rho^{(step)}$

(b) equation (3) and (4)

until either $z_{K^*}/z_K \geq OC$ or $\rho \geq \rho^{(max)}$

一方、 $z_{K^*}/z_K \geq OC$ を満足する場合には、学習モードに入り、学習荷重 w_{jih} 、 p_{jk} 、 b_{ji} を次式で更新する。

$$\Delta w_{jih} = \frac{\alpha y_j^* (1 - \lambda^{1/M}) (f_{ih} - w_{jih})}{(\alpha - 1) \lambda^{1/M} + n_j}, \quad \lambda \in (0, 1)$$

$$\Delta p_{jk} = \frac{\alpha y_j^* (z_k^* - p_{jk})}{\alpha + n_j}$$

$$\Delta b_{ji} = b_j^{(rate)} y_j^* (x_{ji} - b_{ji})$$

$$\Delta n_j = \alpha y_j^* (1 - n_j)$$

$$y_j^* = \frac{\prod_{i=1}^M x_{ji} \times \sum_{k=1}^U z_k^* p_{jk}}{\sum_{j'=1}^N \prod_{i=1}^M x_{j'i} \times \sum_{k=1}^U z_k^* p_{j'k}}$$

$$z_k^* = \begin{cases} 1 & ; \text{ if } k = K^* \\ 0 & ; \text{ otherwise} \end{cases}$$

ただし、 $b_j^{(rate)}$ は学習定数であり、 α と λ は係数である。また、 y_j^* はクラス層から基盤層へのトップダウン学習のためのフィードバック信号である。

学習は特徴マップ f_{ih} を逐次的に構成することに行い、全学習データはデータ集合を1エポックとしてエポック回数分とする。学習によって、 w_{jih} は、フィードバック信号 y_j^* を介して、 f_{ih} に近似するように学習される。また、 p_{jk} は、 z_k^* に近似するように学習され、 b_{ji} は x_{ji} に近似するように学習される。これらの学習は、Grossberg の winner-take-all 学習 [20] を用いており、ボトムアップ学習とトップダウン学習との相互学習は共振学習 [9] と呼ばれている。

一方、学習後、情報量を用いて不必要な結合リンクとノードを削除するブルーニング手法が提案されている [13-15]。このブルーニング機能により、TAM ネットワークは他の視覚系ニューラルネットワークと異なり、学習後のネットワーク構造から対象画像の特徴をファジィルールとして獲得できる。

4. 文字認識による事例

ここでは、TAM ネットワークの有用性を確認するため、原画像としてアルファベットを用いて、文字認識の事例について議論する。画像データの大きさを $15 \text{ pixels} \times 15 \text{ pixels}$ とし、電子タブレットによって作成された画像データを用いた。TAM ネットワークの有用性を次の4点から評価した。

- 全 26 文字のアルファベットの認識率を評価する。
- アルファベット A と B に対して、ガボール関数のパラメータを変化させた場合の認識率の変化を議論する。
- 全 26 文字のアルファベットに対して、ピクセルを欠如させた文字に対する認識率から、TAM ネットワークの頑健性を評価する。
- 位置をシフトしたアルファベット文字に対して、ネットワーク構造から抽出したファジィルールを用いて、他手法との相違点を議論する。

学習パラメータは次の通りである。

L	= 255	$\rho^{(init)}$	= 0.0
OC	= 0.8	$\rho^{(step)}$	= 0.1
α	= 0.0000001	$\rho^{(max)}$	= 100.0
λ	= 0.33	$b_j^{(rate)}$	= 0.01
μ_x	= 0.0	μ_y	= 0.0
σ_x	= 1.99	σ_y	= 1.92
f_x	= 0.127	f_y	= 0.127
ϕ	= 90.0	A	= 1.0

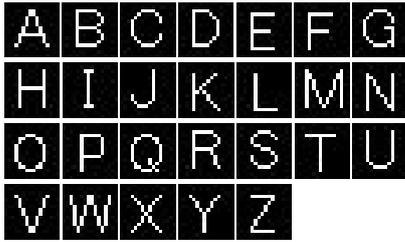


Fig. 7: Training Images of 26 Alphabets

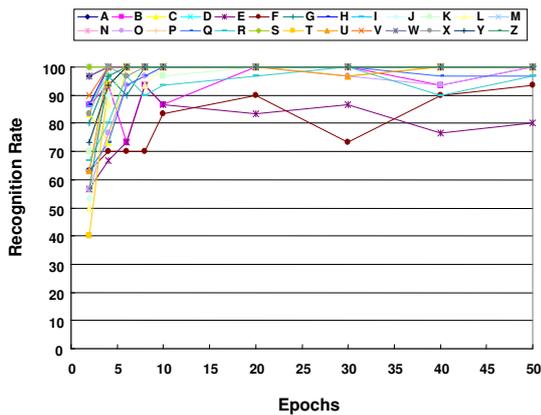


Fig. 8: Recognition Rate of 26 Alphabets

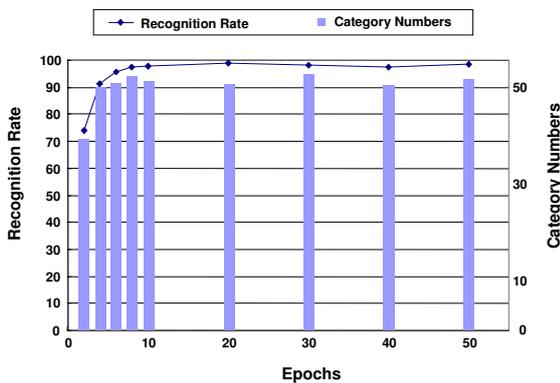


Fig. 9: Recognition Rate and Category Numbers of Alphabets



Fig. 10: Training Images of 'A' and 'B'

まず、全 26 文字のアルファベットの認識率について議論する．学習用画像を図 7 に示し、認識率の結果を図 8 と図 9 に示す．図 8 はエポック回数に対する各アルファベットの認識率を表し、図 9 は全アルファベットの認識率とカテゴリーのノード数を表す．なお、認識率は乱数によりデータ順序を変えた 30 回の試行の平均値として求めた．エポック回数は学習回数を表しているので、4~8 回程度の少ない学習回数で、すでに認識率が 90% を越えていることがわかる．また、多くのアルファベットがほぼ 80% 以上の高い認識率を示している．特に、'A'、'D'、'H'、'M'、'N'、'S'、'V'、'W'、'X' の認識率は 98% を越えており、これらのアルファベットの構造から、文字が直線的に構成されていて、その方位選択性が認識しやすい場合には、特に高い認識率を示すことがわかる．



Fig. 11: Checking Images of 'A'

次に、アルファベットを 'A' と 'B' に限定して、ガボール関数のパラメータ変化に対する認識率を議論した．学習用画像を図 10 に示し、'A' の評価用画像を図 11 に示す．評価用画像は、学習用画像に対して位置をシフトした文字、回転した文字、ノイズを加えた文字、ピクセルを欠如した文字の合計 4 種類の画像を用意した．'B' についても、同様な評価用画像を作成した．

ガボール関数のパラメータの変化範囲を次に示す．

$$\begin{aligned} \sigma_x &= 0.5, 0.75, \dots, 4.0 & \sigma_y &= 0.5, 0.75, \dots, 4.0 \\ f_x &= 0.01, 0.015, \dots, 0.3 & f_y &= 0.01, 0.015, \dots, 0.3 \\ A &= 0.1, 0.2, \dots, 1.0 \end{aligned}$$

認識率の結果を図 12 に示す．(a) は標準偏差 σ_x の変化に対する学習用画像 (TRD) と評価用画像 (CHD) の認識率を表す．他のパラメータの値は、 $\sigma_y = 1.92$ 、 $f_x = f_y = 0.127$ 、 $A = 1.0$ とした．(b) は標準偏差 σ_y の変化に対する認識率を表し、他のパラメータの値

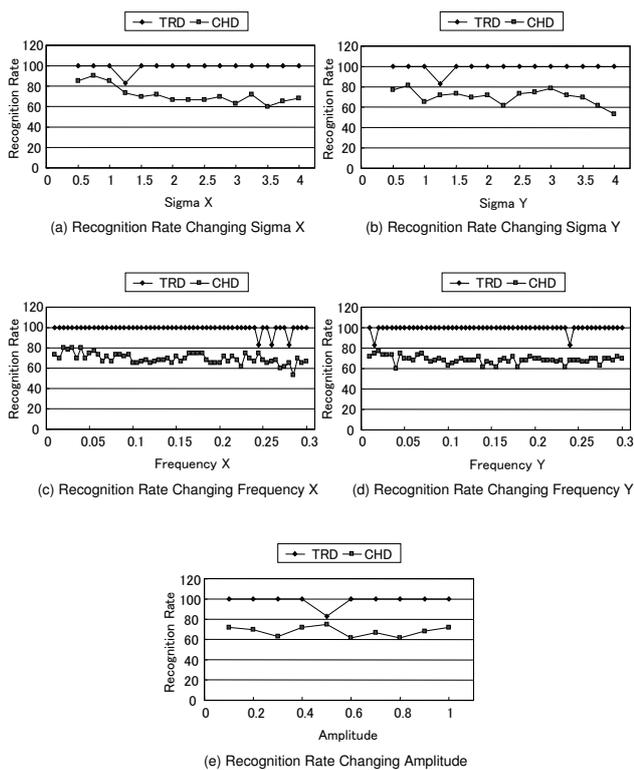


Fig. 12: Recognition Rate of 'A' and 'B'

は $\sigma_x = 1.99$, $f_x = f_y = 0.127$, $A = 1.0$ とした。同様に, (c) と (d) は, それぞれ周波数 f_x と周波数 f_y の変化に対する認識率を表し, $\sigma_x = 1.99$, $\sigma_y = 1.92$, $f_x = f_y = 0.127$, $A = 1.0$ とした。(e) は振幅 A の変化に対する認識率を表し, $\sigma_x = 1.99$, $\sigma_y = 1.92$, $f_x = f_y = 0.127$ とした。全ての認識率は乱数によりデータ順序を変えた 30 回の試行の平均値である。

(a) では, 学習用画像の認識率は全般的に高いが, 評価用画像の認識率は, $\sigma_x = 0.75$ のとき 90.0% で最大となり, 標準偏差の増加に伴い徐々に低下している。図 4 から, 標準偏差値の変化は方位選択成分の明瞭性に影響を与え, また, 標準偏差は値が小さいほど明瞭度が高いので, 標準偏差を比較的小さな範囲で調整した場合, 高い認識率を得る本結果は納得できる。同様に, 標準偏差 σ_y に関して, $\sigma_y = 0.75$ のとき認識率が 81.7% で最大となり, 標準偏差が比較的小さな範囲で高い認識率を得た。一方, 周波数 f_x と f_y に関しては, 周波数の変化に対する学習用画像の認識率は非常に高く, 評価用画像に対しても, ほぼ 70% 前後の認識率を示している。振幅 A に対しては, 学習用画像と評価用画像に対する認識率はあまり変化がなかった。これらの結果から, TAM ネットワークでは, 認識率に最も影響を与えるパラメータは標準偏差であり, 認識率の向上には, 標準偏差 σ_x

と σ_y の調整が重要であることがわかる。

次に, TAM ネットワークの頑健性を議論した。全 26 文字のアルファベットに対して, 通常の文字から 10%, 20%, 30% のピクセルを欠如したピクセル欠如文字の認識率を検証した。学習用画像として図 7 を用い, 評価用画像として, それぞれ異なる 10 種類のピクセル欠如文字を用いた。

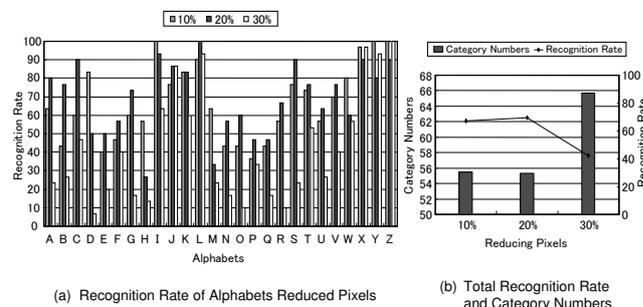


Fig. 13: Result of Reducing Pixels

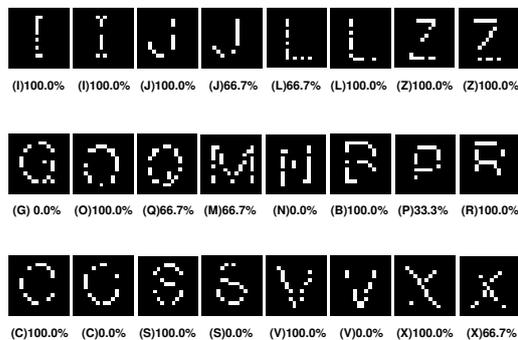


Fig. 14: Result of 30% Reducing Pixels

結果の認識率を図 13 に示す。(a) は 10%, 20%, 30% のピクセルを欠如したアルファベットのそれぞれの認識率を示し, (b) は全 26 文字のアルファベットの認識率とカテゴリーのノード数を示す。なお, 認識率は乱数によりデータ順序を変えた 30 回の試行の平均値として求めた。10% と 20% のピクセル欠如文字に対して, 認識率は, それぞれ 67.05%, 69.36% と比較的良好な結果を示したが, 30% のピクセル欠如文字の認識率は 42.18% と低い値となった。しかし, 30% のピクセル欠如文字において, 'I', 'J', 'K', 'L', 'X', 'Y', 'Z' の認識率は 60% を越えており, アルファベット文字が直線的に構成される場合, 方位選択性の抽出により比較的良好に認識できる。また, カテゴリーのノード数の平均値は, 30% の場合に 65.73 個と多く, 10% と 20% よりも学習が困難であったことを示している。それを示す例として, 30% の

ピクセル欠如文字の一例を図 14 に示す．なお，各文字の下方に正しい文字記号と認識率を示した．これらの文字の解読は困難であり，TAM ネットワークの認識率が高くない結果も理解できる．

最後に，ネットワーク構造から獲得したファジィルールの精度について議論し，他の視覚系ニューラルネットワークとの相違点を明らかにする．TAM ネットワークでは，情報量を用いたブルーニング手法により，不必要なリンクとノードとを削除して，ネットワークの構造をファジィルールとして獲得できる．いま，カテゴリー層の w_{ji} をファジィルールの前件部のファジィ集合と解釈し， p_{jk} を後件部の実数値と解釈すれば，ブルーニング後の TAM ネットワークは下記のファジィルールとして表現される．

$$\left. \begin{array}{l} r_1: \text{ if } f_1 \text{ is } w_{11} \text{ and } \dots \text{ and } f_M \text{ is } w_{1M} \\ \quad \text{then } C_1 = p_{11}, \dots, C_U = p_{1U} \\ \vdots \\ r_j: \text{ if } f_1 \text{ is } w_{j1} \text{ and } \dots \text{ and } f_M \text{ is } w_{jM} \\ \quad \text{then } C_1 = p_{j1}, \dots, C_U = p_{jU} \\ \vdots \\ r_N: \text{ if } f_1 \text{ is } w_{N1} \text{ and } \dots \text{ and } f_M \text{ is } w_{NM} \\ \quad \text{then } C_1 = p_{N1}, \dots, C_U = p_{NU} \end{array} \right\} (5)$$

ただし， C_k ， $k = 1, 2, \dots, U$ はクラス層のクラスであり， r_j ， $j = 1, 2, \dots, N$ はファジィルールの番号である．

受容野構造では，ガボール関数により原画像の位置に依存することなく，原画像の方位選択成分を抽出できる．この特徴を用いると，原画像から獲得したファジィルールは，原画像を視覚サイズ内で他の位置にシフトさせた別文字から獲得したファジィルールと一致するはずである．このファジィルールの一致性を確認するため，次のような検討を行った．いま，図 10 の‘A’と‘B’を原画像として，図 11 の最初から 9 文字分の‘A’を‘シフト文字 A’として（同様に‘B’の‘シフト文字 B’も構成して），(5) 式のメンバーシップ関数 w_{ji} により相関係数を計算し，ファジィルールの一致度を計算した．なお，比較対象として，図 11 の残りの 21 文字分のファジィルールを用いた．

結果を表 1 に示す．8 *Shifts* では，‘A’の原画像に対して，‘シフト文字 A’から図 11 の最上左の‘A’を除いた残り 8 文字を‘シフト文字 A’として相関係数を求めた．11 *Shifts* では，‘A’の原画像に対して，上記の 8 文字の‘シフト文字 A’に図 10 の原画像の‘A’の 3 文字分を加えた合計 11 文字を‘シフト文字 A’として相関係数を求めた．また，14 *Shifts* では，‘A’と‘B’の原画像に対して，上記の 11 *Shifts* の 11 文字に図 10 の原画像の‘B’の 3 文字分を加えた合計 14 文字を‘シフト文字

A とシフト文字 B’として相関係数を求めた．結果から，8 *Shifts* では，原画像とシフト文字とのファジィルールの相関係数が高く，これらのファジィルールはほぼ一致していることがわかる．また，11 *Shifts* や 14 *Shifts* の場合には，無相関となっているが，原画像とシフト文字とのファジィルールの相関係数は若干高く，総合的に判断して，シフト文字のファジィルールは，比較対象のファジィルールよりも原画像のファジィルールに類似しているといえる．

表 1: Comparison of Fuzzy Rules

Correlation Coefficient	8 Shifts	11 Shifts	14 Shifts
Orig. Image (Fig.10) and Sift Characters	0.71	-0.0022	0.0187
Orig. Image (Fig.10) and Other Characters	-0.008	0.0188	-0.0179

4. おわりに

本研究では，受容野モデルとしてガボール関数を定義し，ガボール型受容野を構成する新たな TAM ネットワークを提案した．また，TAM ネットワークの有用性を文字認識の例を用いて検討した．なお，本研究は，平成 17 年度関西大学学術研究助成基金（共同研究）において，研究課題「道路維持管理分野における建設プロダクトモデルの構築と適用性に関する研究」として研究費を受けた成果の一部として公表するものである．また，本研究の一部は，文部科学省科学研究助成金（基盤研究（C））課題番号 18500181 の一環として行われた．

参考文献

- [1] 日本視覚学会（編）：視覚情報処理ハンドブック，朝倉書店（2000）
- [2] S.Marčelja: Mathematical description of the responses of simple cortical cells, *Optical Society of America*, Vol.70, No.11, pp.1297-1300 (1980)
- [3] J.G.Daugman: Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, *Optical Society of America*, Vol.2, No.7, pp.1160-1169 (1985)

- [4] D.C.Lee: Adaptive processing for feature extraction: Application of two-dimensional Gabor function, *Remote Sensing*, Vol.17, No.4, pp.319-334 (2001)
- [5] S.Grossberg: How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex, *Spatial Vision*, Vol.12, No.2, pp.163-185 (1999)
- [6] 岡島, 今岡: 情報量最大化と生体視覚細胞の受容野, *電子情報通信学会論文誌 A*, No.J83-A, Vol.6, pp.620-628 (2000)
- [7] H.Neumann and W.Sepp: Recurrent V1-V2 interaction in early visual boundary processing, *Biological Cybernetics*, Vol.81, pp.425-444 (1999)
- [8] K.Fukushima: Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, *Biological Cybernetics*, Vol.36, pp.193-202 (1980)
- [9] G.A.Carpenter, S.Grossberg and J.Reynolds: ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network, *Neural Networks*, Vol.4, pp.565-588 (1991)
- [10] G.A.Carpenter, S.Grossberg, N.Markuzon, J.Reynolds and D.B.Rosen: Fuzzy ARTMAP: A neural network architecture for incremental learning of analog multidimensional maps, *IEEE Transactions on Neural Networks*, Vol.3, pp.698-713 (1992)
- [11] J.R.Williamson: Self-organization of topographic mixture networks using attentional feedback, *Technical Report CAS/CNS*, Vol.TR-99-0273 (1999)
- [12] J.R.Williamson: Self-organization of topographic mixture networks using attentional feedback, *Neural Computation*, Vol.13, pp.563-593 (2001)
- [13] 林, 前田: TAM Network のブルーニング手法の一提案, 第 16 回ファジィシステムシンポジウム講演論文集, pp.127-130 (2000)
- [14] I.Hayashi and J.R.Williamson: Acquisition of fuzzy knowledge from topographic mixture networks with attentional feedback, *Proceeding of the International Joint Conference on Neural Networks (IJCNN'01)*, pp.1386-1391, Washington DC, U.S.A., July 15-19 (2001)
- [15] 林, Williamson: TAM Network のブルーニング手法の提案, *システム制御情報学会論文誌*, Vol.17, No.2, pp.81-88 (2004)
- [16] I.Hayashi and H.Maeda: A formulation of fuzzy TAM network with Gabor type receptive fields, *Proceeding of the 4th International Symposium on Advanced Intelligent Systems (ISIS2003)*, pp.620-623, Jeju, Korea, September 25-28 (2003)
- [17] 林, 前田: ガボール関数を用いた TAM ネットワークの受容野入力構造, 第 20 回ファジィシステムシンポジウム講演論文集, pp.410-413 (2004)
- [18] I.Hayashi, H.Maeda and J.R.Williamson: A formulation of receptive field type input layer for TAM network using Gabor function, *Proceeding of 2004 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE2004)*, No.1335, Budapest, Hungary, July 25-29 (2004)
- [19] J.G.Daugman: Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression, *IEEE Transaction on Acoustics, Speech, and Signal Processing*, Vol.36, No.7, pp.1169-1179 (1988)
- [20] S.Grossberg: Contour enhancement, short term memory, and constancies in reverberating neural networks, *Studies in Applied Mathematics*, No.L11, pp.213-257 (1973)

ガボール型受容野をもつ TAM ネットワークの提案

林 勲

ジェームズ R. ウィリアムソン

和文要旨

人の視覚系では、網膜に入力された画像信号は受容野に対応した視細胞で処理され、外側膝状体を介して第一次視覚野に入力される。視覚系処理過程を表現する Hubel-Wiesel の階層仮説モデルの一例として TAM ネットワークがある。TAM ネットワークは 4 層の階層構造からなり、第一次視覚野以降の視覚前野を模擬している。与えられた教師値と出力値に差がある場合、共振学習、ビジランス機能、中間層へのノード増加によって、高い学習機能を確保することができる。一方、受容野における方位選択性モデルとしてガボール関数があり、画像の任意の周波数成分を抽出するガボールフィルタリングを構成できる。

本論文では、入力層以前にガボール型受容野層を導入した新たな TAM ネットワークを提案する。受容野層は網膜層、神経節細胞層、外側膝状体 (LGN) 層から構成され、ガボールフィルタリングを用いて対象画像の方位選択成分を抽出し、輝度情報を正規化して特徴マップを構成する。ここでは、受容野構造と特徴マップ構造について議論し、輝度情報の信号処理アルゴリズムを定式化する。また、アルファベットの世界認識の例を用いて、本モデルの有用性と頑健性について検討する。なお、他の Hubel-Wiesel モデルと異なり、ネットワーク構造から画像特徴をファジィルールとして獲得できるので、獲得されたファジィルールの妥当性についても検討する。

キーワード：視覚系ニューラルネットワーク、ガボール関数、受容野、画像認識