広範囲受容野をもつTAMネットワークの提案

A Proposal of TAM Network with Extensive 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　TAM (Topographic Attentive Mapping) network is a biologically-motivated neural network with Gabor function type receptive fields. However, the sturucture of receptive fields is a mono-layer, and there is a lack of performance for rotation characteristic of images. In this paper, we formulate a new TAM network structure with extensive and multilayer receptive fields. We also show the usefulness of TAM nework using some examples of character recognitions.

1. はじめに

人間の視覚情報は，受容野に対応した網膜の視細胞で処理され，外側膝状体を介し，第一視覚野で対象画像の方位選択性を抽出して，より高位レベルで対象画像の回転性や対称性などが検知される [1]．この視覚処理過程は，Hubel-Wieselの階層仮説に代表され，多くの有用な視覚系モデルが提案されている [2–4]．例えば，福島は，最近では自己学習型モデルであるネオコグニトロン [5] を改良し，ぼかしの効果を導入したend-stopped cellを実現して，対象画像の対象軸性を抽出するモデルを提案している [6]．一方，人間の視細胞には，単純型細胞，複雑型細胞，超複雑型細胞が存在する．この初期視覚構造を表現するモデルとして，視細胞の受容野を表現する空間分布型の関数が導入される．受容野の有用なモデルにガボール関数 [7–10] がある．ガボール関数はガウス関数と正弦・余弦関数からなる関数であり，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．

著者らは，ガボール関数の受容野構造を導入した視覚系モデル [11] として，TAM(Topographic Attentive Mapping) ネットワーク [4, 12] を提案している．TAMネットワークは，網膜層，外側膝状体 (LGN) 層，特徴マップ層，基盤層，カテゴリー層，出力層の階層構造からなり，上位層は第一次視覚野以降の視覚前野を模擬している．また，網膜層，LGN層，特徴マップ層からなる下位層では，ガボール関数の受容野構造により，対象画像の方位選択成分を抽出して，第一視覚野における特徴マップを構成する．しかし，この受容野構造は単層構造であり，十分な方位選択性を抽出しているとは言えず，対象画像の回転特性に弱い．本論文では，新たに視界の広範囲を網羅する受容野を定義し，受容野構造を多層化して，対象画像の十分な方位選択性を確保する．この方位選択性特性により，対象画像の回転特性が効果的となる．具体的には，受容野の輝度情報からガボール関数によって対象画像の16方位選択成分を抽出し，そのエッジ特徴画像に対して広範囲の方位選択性を抽出して，既存画像との比較で回転特性を推定する．第一視覚野以降では，この回転特性に対応した特徴マップを構成して，TAMネットワークの学習機能により対象画像を認識する．ここでは，簡単な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を検討する．

2. 広範囲受容野構造をもつ TAM ネットワーク

受容野構造をもつTAMネットワークの構成を図1に示す．網膜層やLGN層では，対象画像をガボール関数型受容野で処理し，16方位の方位選択成分を抽出する．得られたエッジ特徴画像に対し，広範囲受容野構造を適用して，より広範囲な方位選択成分を抽出する．この広範囲な方位選択成分から，第一視覚野での16方位の輝度分布の特徴マップを得る．ここで，対象画像の回転特性は広範囲方位選択成分から得られ，第一視覚野での16方位の特徴マップを変更する．得られた特徴マップから，基盤層，カテゴリー層，クラス層での学習機能により視界画像を認識する (図4を参照)．

2次元正弦関数型ガボール関数 $G(x,y)$ は次のように定義される．

$$\begin{aligned} G(x,y) &= Ke^{-\frac{1}{2}(\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2})} \\ &\quad \times \sin(2\pi f_x x\cos\theta + 2\pi f_y y\sin\theta + \phi) \end{aligned}$$

ただし，K は振幅，(μ_x, μ_y) はガボール関数の中心，σ_x と σ_y は標準偏差，f_x と f_y は周波数を表す．図2にガボール関数の例を示す．

次に，方位選択成分の検出過程を図3に示す．いま，視界の原画像の横幅と縦幅をそれぞれ，$R_H\ pixels$ と $R_V\ pixels$ とする．原画像 $O(p,q),\ 0 \le p \le R_H,\ 0 \le q \le R_V$ における受容野の i 番目の方位選択成分 $C_i(x,y),\ \ i = 1,2,\cdots 16$ は，次の畳み込み計算により計算される．

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

いま，第 I 番目の方位選択成分が得られたとする．この第 I 番目の方位選択成分に対して，より広範囲な受容野を用いて，第2段目の畳み込み計算を行い，第 I 番目

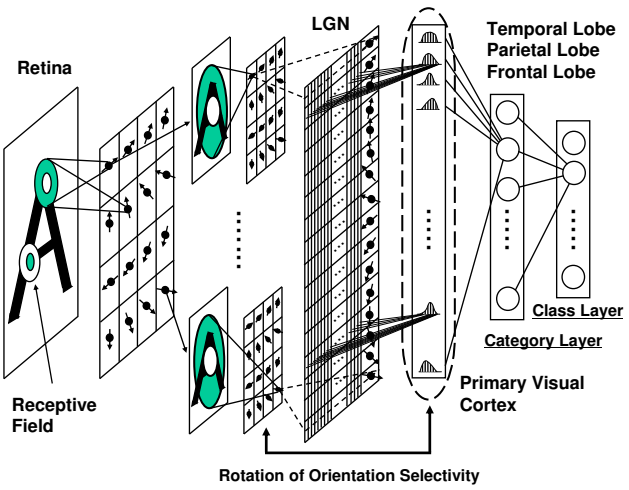

Fig. 1: TAM Network with Extensive Receptive Field

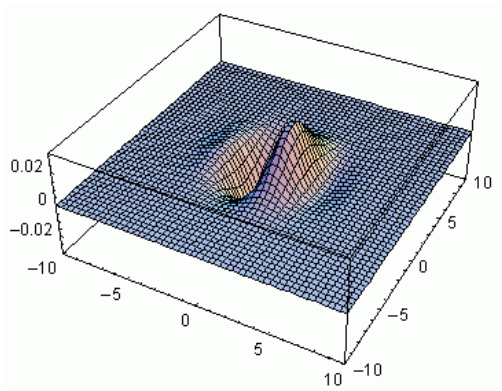

Fig. 2: Gabor Function

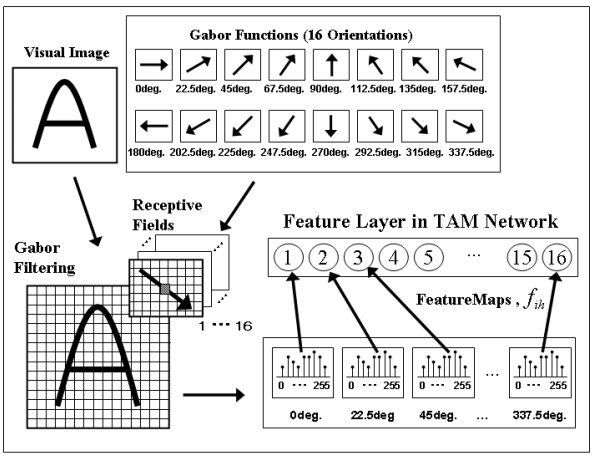

Fig. 3: Gabor Filtering Process

のみの方位選択成分 $C_I^2(x,y)$ を計算する．

$$C_I^2(x,y) = \sum_{q=1}^{R_V}\sum_{p=1}^{R_H} G_I(x-p, y-q) \times C_I(p,q)$$

TAM ネットワークの特徴マップ層での i 番目の特徴マップは，この方位選択成分の輝度を各方位ごとに正規化して構成される．

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

ただし，TAM ネットワークの学習後に，評価用の方位選択成分画像として，$C_{i,new}^2(x,y)$ が得られた場合，その $C_{i,new}^2(x,y)$ の特徴マップ f_{ih} は次式の角度 r により入れ替える．

$$r = \min_i \left(\frac{\sum_x \sum_y C_{i,new}^2(x,y)}{R_H \times R_V} - \frac{\sum_x \sum_y C_I^2(x,y)^2}{R_H \times R_V}\right)$$

カテゴリー層では，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，クラス層への出力 y_j を得る．

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

ただし，ρ はビジランスパラメータであり，b_{ji} は抑制荷重である．

出力値 K は次式より計算される．

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

ただし，p_{jk} は学習荷重である．

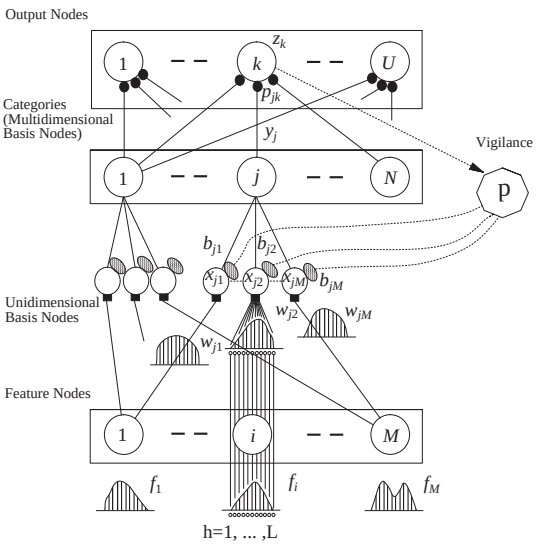

Fig. 4: TAM Network

TAM ネットワークはカテゴリー層のノード増幅機能をもつ．いま，K^* をクラス層での正しい教師値とする．$K \neq K^*$ の場合には，ビジランスパラメータ ρ は初期値

から $z_{K^*}/z_K \geq OC$ の条件が満足されるか最大値になるまで上昇する．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．

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．

$$
\begin{aligned}
\triangle w_{jih} &= \frac{\alpha y_j^*(1-\lambda^{1/M})(f_{ih}-w_{jih})}{(\alpha-1)\lambda^{1/M}+n_j}, \lambda \in (0,1) \\
\triangle p_{jk} &= \frac{\alpha y_j^*(z_k^*-p_{jk})}{\alpha+n_j} \\
\triangle b_{ji} &= b_j^{(rate)} y_j^*(x_{ji}-b_{ji}) \\
\triangle n_j &= \alpha y_j^*(1-n_j) \\
y_j^* &= \frac{\prod_{i=1}^{M} x_{ji} \times \sum_{k=1}^{U} z_k^* p_{jk}}{\sum_{j'=1}^{N} \prod_{i=1}^{M} x_{j'i} \times \sum_{k=1}^{U} z_k^* p_{j'k}} \\
z_k^* &= 1 \ \text{if} \ \ k=K^*; \quad z_k^*=0 \ \ \text{otherwise}
\end{aligned}
$$

ただし，$b_j^{(rate)}$ は学習定数であり，α と λ は係数，y_j^* はカテゴリー層へのフィードバック信号である．

3. 文字認識による例題

広範囲受容野をもつ TAM ネットワークの有効性を確認するため，アルファベットの文字認識を行った．入力画像として電子タブレットによって作成された画像データを用いる．画像の大きさを $15\ pixels \times 15\ pixels$ とする．ここでは，次の 2 点から TAM ネットワークの有用性を評価した．

- 回転特性をもたない TAM ネットワークを用いて，全 26 文字のアルファベット認識，回転文字認識，斜体文字認識を行う．
- 回転特性をもつ TAM ネットワークを用いて，回転文字認識を行う．

TAM ネットワークの学習パラメータは次の通りである．

L	=	255	ρ^{init}	=	0.0
OC	=	0.8	ρ^{step}	=	0.1
α	=	0.01	ρ^{max}	=	100.0
λ	=	0.33	$b_j^{(rate)}$	=	0.0000001
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	K	=	1.0

まず，回転特性をもたない TAM ネットワークに対する特性を示す．学習用画像を図 5 に示し，全 26 文字のアルファベット認識率の結果を図 6 に示す．認識率はデータ順序を変えた 30 回の平均値を表している．学習回数 (エポック数) を 2 回と 5 回として比較した．どちらのエポック回数においても，ほぼ 60%以上の認識率を示し，高い認識率を確保していることがわかる．しかし，回転文字認識と斜体文字認識に対しては，頑健性が高い文字と低い文字が混在し，回転特性に弱いことがわかった．この原因として，受容野構造が単層であり，十分な方位選択性を確保できず，また，回転特性に対する機能も存在していないことが考えられる．

次に，回転特性をもつ TAM ネットワークに体する特性を示す．ここでは，2 種類のアルファベット 'A' と 'B' の学習画像に対して，回転角度を与えた評価画像を準備し，方位選択性の回転特性に対する依存性を検分した．学習用画像を図 7 に示し，評価用画像を図 8 に示す．回転画像は学習用画像に対して，22.5° ずつの回転を与えた．評価用画像の認識率を表 1 に示し，それぞれの文字の認識率を図 8 に示す．上段の値は回転特性をもたない TAM ネットワークの認識率であり，下段の値は回転角度特性をもつ TAM ネットワークの認識率である．認識率はデータ順序を変えた 30 回の平均値を表している．表 1 と図 8 から，文字 'B' に対して認識率が向上しており，全体の認識率も良くなっていることがわかる．しかし，各文字に対しては，十分に認識できていない文字もあり，まだ検討の余地がある．

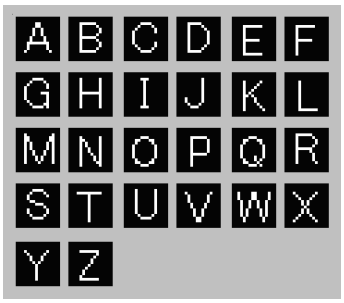

Fig. 5: Training Images

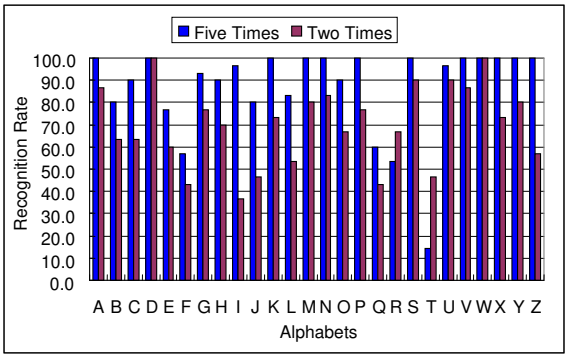

Fig. 6: Results of All Alphabets

Table 1: Recognition Rate

	Char.'A' (%)	Char.'B' (%)	Total (%)
TAM without Rotation	96.04	12.71	54.38
TAM with Rotation	98.33	30.42	64.38

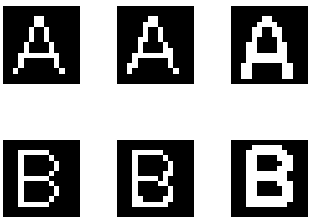

Fig. 7: Training Images of ‘A’ and ‘B’

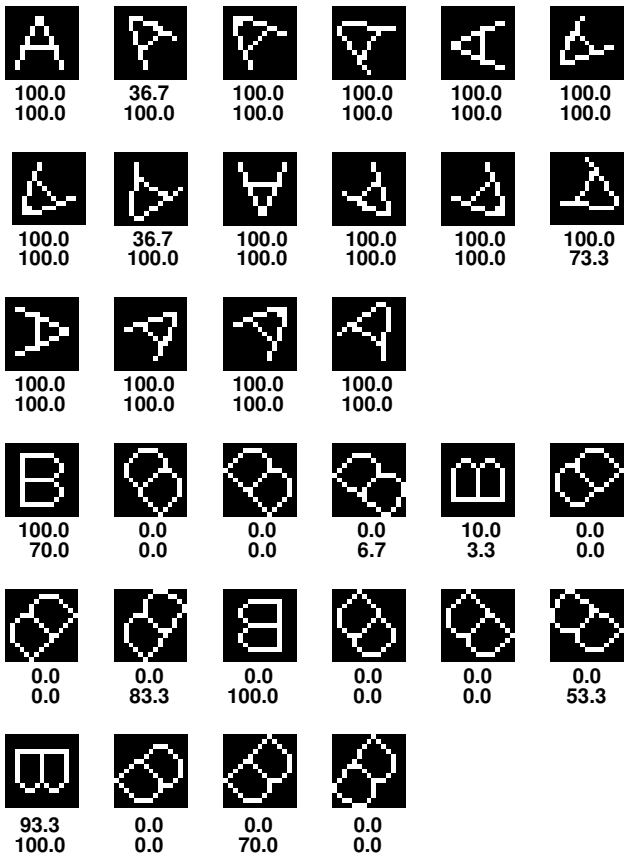

Fig. 8: Checking Images

4. おわりに

ここでは，ガボール関数を用いた広範囲受容野入力構造を提案し，その有用性を文字認識の例で検討した．

参考文献

[1] 松田: 知覚心理学の基礎, 培風館 (2000)

[2] 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)

[3] H.Neumann and W.Sepp: Recurrent V1-V2 interaction in early visual boundary processing, *Biological Cybernetics*, Vol.81, pp.425-444 (1999)

[4] J.R.Williamson: Self-organization of topographic mixture networks using attentional feedback, *Neural Computation*, Vol.13, pp.563-593 (2001)

[5] K.Fukushima: Neocognitron: A hierarchival neural network capable of visual pattern recognition, *Neural Networks*, Vol.1, No.2, pp.119-130 (1988)

[6] K.Fukushima: Use of non-uniform spatial blur for image comparison: Symmetry axis extraction, *Neural Networks*, Vol.18, No.1, pp.23-32 (2005)

[7] S.Marčelja: Mathematical description of the responses of simple cortical cells, *Optical Society of America*, Vol.70, No.11, pp.1297-1300 (1980)

[8] J.Daugman: Uncertainty relation for resolution in space, spatial frequency, and orientaion optimized by two-dimensional visual cortical filters, *Optical Society of America*, Vol.2, No.7, pp.1160-1169 (1985)

[9] 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)

[10] D.C.Lee: Adaptive processing for feature extraction: Application of two-dimensional Gabor function, *Remote Sensing*, Vol.17, No.4, pp.319-334 (2001)

[11] 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)

[12] 林，Williamson: TAM Network のプ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)

[連絡先]
林　　勲，　関西大学 総合情報学部
〒 **569-1095** 大阪府高槻市霊仙寺町 **2-1-1**
tel. 072-690-2448
e.mail ihaya@kcn.res.kutc.kansai-u.ac.jp