

# Predictive Coding in Recurrent Neocortical Circuits

Rajesh Rao

## 1 はじめに

解剖学的に、大脳皮質間に求心性および遠心性の経路が存在すること、皮質の層間に再帰的な結合があることが知られています。しかし、このような構造がどのような機能を担っているのかは明らかではありません。この講義では、このような解剖学的知見に基づく情報処理モデルとして Predictive Coding 仮説を提唱し、それが視覚系の特性を良く説明することを示したいと思います。

## 2 解剖学的知見

マカク猿の視覚系の経路、例えば網膜 (Retina)-外側膝状体 (LGN)-視覚一次野 (V1)-V2-V4-IT には、Feedforward と Feedback の双方向の結合が見られます。また、大脳皮質には 6 層からなる層構造が存在し、層ごとに規則的な投射関係を持ちます。例えば LGN からの入力 V1 の 4 層に入りますが、V1 の 6 層から逆に LGN への出力が存在します。また、V1 の 2,3 層は V2 に投射を送り、V2 の出力が再び V1 の 1 層へ入力します。皮質層内においても再帰的な興奮性結合が存在することが知られています [2][3]。皮質内錐体細胞への入力の 80 分のものとなっています。このように視覚系は階層的なシステムを持ち、それらの間に多数の双方向結合をシステムであると言えます。

## 3 Predictive Coding Hypothesis

- 皮質領域間の相互結合 (reciprocal connection) の役割は何か。
- 皮質領域内の再帰的興奮性結合の役割は何か。

これらの疑問に答える仮説の一つが Predictive Coding 仮説です。  
この仮説は

- Feedback 結合は下位レベルの神経活動の予測値を伝達する。Feedforward 結合は予測値と実際の活動の誤差を伝達する。
- 皮質領域内の再帰的興奮性結合は下位レベルの活動の予測値を生成する。というものです。

図 1 Predictive Coding の概念図 (文献 [1], 図 1)

もう一度、ここで紹介するアイデアをおさらいすると、図 1 - a に示すように入力信号の統計的性質を学習し、フィードバック結合によって入力信号を予測しようとしています。誤差信号はフィードフォワード結合によって次のレベルに、またさらに上位のレベルへと運ばれます。これが一般的なアイデアです。図 1-b, c はそのアイデアを具体化したネットワークの例です。入力としては局所的な画像の集合を与えます。Level 1 のそれぞれの予測器は、入力画像の統計的な性質を学習します。Level 2 では、Level 1 の活動の組み合わせを予測しようとしています。このようにすることで、入力画像の空間的な性質を捉えることができます。

図 2 内部モデルと最適な隠れ状態推定の問題 (文献 [8], 図 1(a))

ここで、カルマンフィルターの概念を簡単に紹介したいと思います。カルマンフィルターで解こうとする問題は図 2 のような問題です。学習したい内部モデル (internal model) は、外部世界 (the world) のシステムです。視覚の問題の場合、外部世界に存在するいろいろな物理的な物体がそうです。しかし、観測によって直接物理的な世界を知ることができません。2次元世界に投影された画像だけが、網膜に与えられます。そのため、私達はこの可視状態 (visible state) から外部世界の隠れ状態 (hidden state) を推定する必要があります。つまり、逆問題を解かなければならないこととなります。このままでは数学的な表現ができません。そこで、数学的に表現可能なある特定の枠組みに対するカルマンフィルターを考えます。

図 3 カルマンフィルターに基づいた推定問題 (文献 [8], 図 1(b))

外部世界とその外部世界を予測する予測器を考えます (図 3)。この外部世界の方の四角に囲まれた箱の中に外部世界のダイナミクスが表現されています (図 3 左半分)。  $I(t)$  が時刻  $t$  で得られる画像信号を表しています。そして、  $U$  や  $V$  が外部世界のパラメーターとなっています。  $U$  が隠れ状態  $r$  から可視状態  $I$  への変換を行うパラメーターで、  $V$  が物体のダイナミクスを表現するパラメーターです。四角の箱の中に書いてあるように、  $r$  は行列  $V$  によって  $r(t-1)$  から  $r(t)$  に状態遷移を繰り返します。このようにそれぞれの時刻  $t$  での状態は過去の状態のみに依存、ここでは 1 ステップ前、時刻  $(t-1)$  の状態に依存して更新されていきます。これは、統計的な手法として用いられている隠れマルコフモデルとみることができます。以上のことから、  $V$  は状態遷移行列とよばれ、  $U$  は観測行列または隠れ状態から可視状態を生成するため生成行列とも呼ばれます。

翻ってこちら側 (図 3 右半分) の推定器 (The Estimator) はどのようにこのカルマンフィルターの問題を解くかといいますと、アイデアとしては、時刻  $t$  の画像  $I(t)$  が網膜に届いたときに、推定器によって予測された画像  $\bar{I}$  との誤差をとります。この誤差を返すことで  $\hat{r}(t-1)$  が得られます。この  $\hat{r}(t-1)$  から状態遷移行列  $V$  によって次の時刻の隠れ状態  $\bar{r}(t)$  を予測します。観測行列  $U$  を通すことで、予測される画像  $\bar{I}(t)$  を生成します。こうして短期間では、隠れ状態  $r$  を時々刻々に見積もらなければいけないし、長期間ではパラメーター  $U$ 、 $V$  をたくさん画像から学習していかなければならない。以上がカルマンフィルターの流れです。

次にカルマンフィルターの数学を理解するのに役立つ簡単な例を挙げましょう。ここに入力として実数の集合を与えます。

$$\text{Inputs} \quad \{I(1), I(2), I(3), \dots, I(t-1)\} \quad (1)$$

そしてこれらの実数の平均値を求めることを問題とします。平均値は以下の式のようにこれらの実数の和をその実数の要素の数で割ることで求めることができます。

$$\text{Arithmetic Mean} \quad \hat{r}(t-1) = (I(1) + I(2) + I(3) + \dots + I(t-1))/(t-1) \quad (2)$$

ここで新しい実数  $I(t)$  が与えられたとき、これらの実数の平均をどのように計算したらよいでしょうか。これは、次のように以前の平均値の推定値に  $(t-1)/t$  の重みをかけて修正し、新しく得た実数を  $t$  で割った値を加えてやれば求められます。これを書き直せば、新しい推定値が前の推定値に、前の推定値と新たに観測した値との誤差にあるゲインを掛けたものを加えたものとして表現されます。これが一般的な形です。

平均値の逐次更新式:

$$\hat{r}(t) = (t-1/t)\hat{r}(t-1) + I(t)/t = \hat{r}(t-1) + (I(t) - \hat{r}(t-1))/t \quad (3)$$

一般式:

$$\text{New Estimate} = \text{Old Estimate} + \text{Gain} \times \text{Residual Error} \quad (4)$$

このようなイメージで定式化されるのですが、さらにイメージをつかむため学習した時このモデルがどのように働いているかを見てもらいましょう。入力にある画像  $I$  (ある有名人の似顔絵(笑)) を与えます。この画像を水平や垂直のエッジを検出したりするフィルター  $U$  にかけます。そうするとそれぞれのフィルターに対応する活動が生じます。これらフィルター  $U$  とその係数  $r$  となる活動度の積を集めたものが、局所的な入力画像を表します。このように基底関数  $U$  とその係数ベクトル  $r$  の線形和によって画像を生成することになります。それでは数学的な式に戻ります。

空間的生成モデル (Spatial Generative Model):

$$I(t) = U(t)r(t) + n(t) \quad (5)$$

ここで、 $I$ : 入力画像、 $U$ : 生成行列、 $r$ : 神経活動  $n$ : 平均 0、分散  $C_n$  のガウス白色雑音。

入力画像  $I(t)$  は、基底関数  $U$  にその係数ベクトル  $r(t)$  を掛けた積にガウス白色雑音  $n(t)$  を加えたものになります。基底関数  $U$  は、脳内ではフィードバック方向のシナプス荷重に相当します。

隠れ状態の更新式 (Temporal Dynamics of Hidden State):

$$r(t) = Vr(t-1) + m(t-1) \quad (6)$$

ここで、 $V$ : 状態遷移行列、 $r$ : 神経活動  $m$ : 平均 0、分散  $C_m$  のガウス白色雑音

時間的なダイナミクスは、このように表されます。新しい状態  $r(t)$  は、1 時刻前の状態  $r(t-1)$  に予測行列  $V$  を掛けた積にガウス白色雑音  $m(t-1)$  が加わったものとなります。この簡単なモデルでは、パラメーター  $U$  や  $V$  に内部モデルが符号化されることとなります。では、これからこれらのパラメーター  $r$ 、 $U$ 、 $V$  の学習を考えていきます。一つの考えは、予測誤差を最も小さくするような学習をすることです。もう一つは時間的なダイナミクスから予測される隠れ変数の推定値との差を小さくするような学習です。このため、 $J_1$  のような最適化関数を考えることができます。

$$J_1 = \sum_{i=1}^n (I^i - U^i r)^2 + \sum_i 1^k (r^i - \bar{r}^i)^2 \quad (7)$$

ここで、 $\bar{r}$  が入力画像  $I$  を観測する前の  $r$  の推定値です。この最適化関数を、統計的により一般的な形で書けば、各画像  $I$  について次のように書けます。

$$J = (I - Ur)^T C n^{-1} (I - Ur) + (r - \bar{r}^T M^{-1} (r - \bar{r})) \quad (8)$$

ここで、 $\bar{r}$  は神経活動  $r$  の平均、 $M$  はその分散です。このように最初の項が観測ノイズの影響を表し、2 つ目の項がシステムノイズを表す形となっています。この最適化関数はまた、与えられた入力画像  $I$  の負の対数尤度と  $r$  についての事前確率の負の対数尤度の和でもあります。結局、統計的な立場からみれば、この最適化関数  $J$  の最小化はベイズ推定のフレームワークにおける事後確率の最大化と等価になっています。

では、この最適化関数が与えられたのでここからどのようにして最もよい  $r$  や  $U$  や  $V$  を決めたらよいかと言いますと、 $r$  についての最適値なら、 $J$  を  $r$  で微分して 0 となるときの  $r$  を用いれば最もよい  $r$  を得ることができます。つまり、

$$\frac{\partial J}{\partial r} = 0 \quad (9)$$

とおくと、入力  $I(t)$  を得たときの  $r$  の平均値  $\hat{r}(t)$  は、入力  $I(t)$  を得る前の  $r$  の平均値  $\bar{r}(t)$  をもとに

$$\hat{r}(t) = \bar{r}(t) + N(t) U^T C n(t)^{-1} (I(t) - U \bar{r}(t)) \quad (10)$$

で求められます。ここで、

$$N(t) = (U^T C n(t)^{-1} U + M(t)^{-1})^{-1} \quad (11)$$

$$M(t) = V N(t-1) V^T + C m(t-1) \quad (12)$$

です。こうして得られた  $r$  の平均値  $\hat{r}(t)$  は、次の時刻では

$$\bar{r}(t+1) = V \hat{r}(t) \quad (13)$$

として使われます。

ここでは式の詳しい解説は省きますが、詳しくは optimal control のとてもよい解説書 [4] があるのでそれをごらんになってください。簡単に言えば、その時間その時間で外部世界で起こっていること、つまり  $r$  の分布を推定しています。

その分布の平均値、分散が上式のように表されるのです。それでは、もっと直感的に何が起きているかを見てみましょう。

図 4 カルマンフィルタにおける確率密度関数の推移 (文献 [8], 図 3)

左上にあるのが過去の  $r$  の推定 (Previous Estimator) です。分布は、平均  $\hat{r}(t-1)$ 、分散  $N(t-1)$  のガウス分布をしています。これ (右上) が予測行列 (Prediction Matrix)  $V$  によって予測される時刻  $t$  での分布、平均  $\hat{r}(t)$ 、分散  $M(t)$  のガウス分布です。ふつう、さきほどの分布より分散が大きくなるので裾野が広がった分布となります。ここで、 $I(t)$  を観測しますと、このように再び鋭い分布になります (左下)。以上のように過去の分布からつぎの分布を推定し、それと実際の観測との誤差によって新しい分布に修正するということを繰り返します。

ここで、もう一度ダイアグラムを見てみます。私達はこの予測行列 (Prediction Matrix)  $V$  でシステムのダイナミクスを表現していますが、ここで予測をし、生成行列 (Generative Matrix)  $U$  を通してこちらで入力との差が取られます。それにあるゲイン  $G$  をかけて再び予測器に戻ってきます。ここでゲイン  $G$  は、どれだけ入力信号との誤差が信頼できるか、以前のデータから予測された値との誤差を信頼できるかに依存しています。ここには、トレードオフがあります。もし、入力信号のノイズが大きい場合、 $G$  は小さくなり、かわりに  $r$  の推定をほとんど内部モデルによる予測に頼ることになります。逆に内部モデルによる予測に伴うノイズが大きい場合、 $r$  の推定は入力信号に頼ることになります。こういうわけでノイズの分散が推定時の重み付けを決めていると理解することができます。これまでが数学的なお話でした。

さて、このモデルのシミュレーション例を見てみましょう。一番左端の列が学習した物体 (Training Object) の入力画像です。学習後の様子がそれぞれ左から右へ示されています。最初は、予測ができないものの次第にその物体を予測していくようになります。また予測誤差もトレーニングに従って減少し、最後には 0 に近づいています。

図 5 カルマンフィルタによる物体の学習 (文献 [8], 図 5)

図 6 図 5 の実験の学習曲線 (文献 [9], 図 6)

次に、このモデルを用いていくつかのテストを行ってみます。図 6-a が訓練画像です。図 6-b にいくつかの画像でテストした結果を示してあります。図 6-b の上段 2 つ目までは訓練画像を与えたときの、予測画像とその予測誤差ですが、訓練画像に対しては誤差がほとんどなくなっていることがわかります。

図 7 物体の学習とその学習後のカルマンフィルタの動作 (文献 [8], 図 8)

これに対して、上から 3 段目の例は、訓練画像の向きを少し変えたものを入力として与えています。このときは、誤差が少し増えますが、ある程度予測できることがわかります。同じように 4 段目の入力画像に遮蔽物が加わった時も誤差が残ります。そして、興味深いのは 5 段目の訓練集合にない新しい画像を入力として与えたときです。このときは、予測することができず誤差が非常に大きいものとなります。そのため、このモデルが入力された物体が新しい物体であるか、学習済みである物体かを区別することができる可能性を示しています。これによ

て、新しい物体として新たに学習することもできます。

それでは次に階層的な場合を考えましょう。図 7-a にあるようなモデルです。それぞれのレベルでのモジュールは、フィードバックとフィードフォワードの結合をもち、解剖学的には網膜から LGN、V1、V2V4、IT という経路との対応を想定しています。そこでは、3 種類の情報の流れが生じます。1 つ目は入力を受け、それと予測との誤差とり誤差を伝達するもの。2 つ目は、予測器による再帰的な推定。3 つ目はトップダウンによる予測値のフィードバックです。図 8 に皮質内の層間構造をモデル化したものを示します。

図 8 層間構造とそのモデル

解剖学的な構造としては、大雑把に言って 4 層が入力を受け、それが 2、3 層に伝えられ、6 層でフィードバックの出力がなされます。これらがモデルでどう対応付けられるかといいますと、まず入力として下位モジュールからの誤差信号が 4 層に入力され、それを 2、3 層はさらに上位のラシウィールに誤差信号を伝達します。5 層は出力層で、 $\hat{r}$  を出力し、6 層で下位のモジュールへの予測を返します。

それでは、階層性を持つモデルで学習した時のシミュレーション結果を紹介します。訓練画像は、図 8 の上段にあるような自然画像です。Level1 の段で示しているのが、それぞれ Level1 での生成行列  $U$  の行、すなわち個々の  $r$  に対応する基底ベクトルです。これは、この細胞の受容野特性と見ることができます。入力画像は、これら基底ベクトルの線形和で表される事になります。たくさんの画像による学習を行った後では、それぞれの受容野はいろいろな向きのエッジ検出器になりました。このような学習後の受容野の形は、学習する入力画像の統計的性質に依存しています。Level2 ニューロンは、Level1 ニューロンの活動の組み合わせを表現しています。もう一つ、下段にあるのはガウス分布のかわりにスパースな分布を仮定したときの学習後の Level1 ニューロンの受容野です。このときはガボールフィルターやエッジ検出器といったタイプと同じようにもっと限局した受容野が学習されました。

図 9 自然画像学習後の階層的なモデルの受容野 (文献 [1], 図 2)

このモデルを用いて終端抑制 (endstopping) ニューロンの動きを説明することができます。さきほどご説明したように、2、3 層のニューロンは誤差を上位モジュールに伝達します。その 2、3 層にあるニューロンの多くは終端抑制の性質を示します。終端抑制とは、刺激としてバーを呈示した時、そのバーの長さが十分小さい時には、ニューロンの応答はバーが長くなるにつれて強くなりますが、バーの長さがそのニューロンの受容野を越えると、バーが長くなればなるほど応答が弱くなり、最後には 0 になるといったものです (図 9-a 右)。これは 2、3 層のニューロンは、トップダウンの予測とボトムアップの入力の誤差を伝達するが、ある長さ以上のバーではトップダウンの予測が正確になるため、誤差が小さくなるからだという風に理解することができます。実際モデルがどのように働くか見てみますと、バーの長さが十分小さければ、バーの長さが長くなるに従って応答は強まりますが、ある長さからは逆に応答が弱くなります (図 9-a 左)。このように実際の視覚野の 2、3 層のニューロンで見られるのによく似た山なりの曲線が得られます。ここでもし、モデルのフィードバック結合を取り払うと、応答は強いまま残るといった結果になります。それに対し、こちらの実際のニューロンも

フィードバック信号を伝達していると考えられる6層が活動できないようにしたときには、2、3層のニューロンの活動は強いまです。

図 10 終端抑制のフィードバックによる影響比較 (文献 [1], 図 5)

ここで終端抑制ニューロンがどうして形成されるのかと言うと、実はその原因は自然画像の中に隠されています。自然画像は、とても高い確率で同じ方向に線分がのびる傾向をもちます。これによって周りの画素から中心の画素を予測することができるのです。モデルは、この統計的性質を学習するので線分が長い時には予測誤差、つまりニューロンの活動が小さくなるのです。

モデルを使った他の実験もみてみましょう。図 10-a は 4 種類の刺激パターンを与えたときのニューロンの活動を表しています。左から 3 番目のように中心の線分が周囲の方向と異なる時が活動が最も強くなっています。逆に左から 2 番目のようにある方向の線分が一面に長く並んでいる場合は、最も活動が弱くなります。これも周囲の領域が中心部分を予測できる形だからです。図 10-b も同様の例です。同じ向きにそろっていると応答が弱い。図 10-c は、時間的な応答パターンを調べたものです。刺激としてテクスチャーが与えられています。このときもやはり、周辺領域が中心と異なる方向のテクスチャーであれば、同じ方向のテクスチャーの時と比べてニューロンの活動は小さくなります。

図 11 モデルにおける古典的受容野の効果 (文献 [1], 図 6)

このシミュレーションに対応するマカク猿、V1 での実験をお見せします。この実験でもモデルの振る舞いと同じように中心と異なる方向のテクスチャーが周辺に与えられたときのほうがニューロンの活動が強くなっています。また、刺激呈示後およそ 80msec でニューロンの活動がほぼ一定値に落ち着くことがわかります。ニューロンの活動を時間軸にそって眺めてみると、図 11 のように書けます。これを見てわかるようにこれらテクスチャーの線分によって違いが生じるのは 80~100msec 後になります。ここから、フィードバック信号が V1 に届くのには 80msec ぐらいかかるのではないかということが予想されます。

図 12 テクスチャーによる古典的受容野の効果 (文献 [11], 図 2)

図 13 いろいろな領野での応答潜時の比較 (文献 [11], 図 10)

最後に、ロバストカルマンフィルターを紹介します。普通のカルマンフィルターと異なるところは、Gating 行列が加わったところです。これは入へと予測との誤差がある閾値より大きくなった時は、その誤差を無視します。このときの最適化関数はどうなっているかと言うと、まず図 14 左が通常のカルマンフィルターの最適化関数です。これに対して図 14 右のロバストカルマンフィルターでは、ある値以上の誤差からは定数となり、それ以上の誤差の増分を無視する形になっています。

図 14 最小 2 乗最適化とロバスト最適化 (文献 [8], 図 4)

図 15 はロバストカルマンフィルターの学習をした例です。図 15-a が訓練画像です。図 15-b のように訓練画像に新しい画像と一緒に与えると、このモデルは訓

練画像だけを予測しつづけ、残りの誤差は大きい誤差なのですが、Gating 行列の働きによってその誤差は無視されることとなります。もう一つは、図 15-c にある訓練画像を 2 つ同時に見せた例です。先ほどと同じように誤差が大きくても無視されてしまうので一方の画像だけが予測されます。しかし、いったん無視していた誤差に対応する画像を予測しようとすれば、もう一方の画像を予測することとなります。これがこのモデルにおける注意の切り替えです。

図 15 ロバストカルマンフィルタによる認識 (文献 [8], 図 13)

このような仕組みを担う生物学的な構造として視床を考えることができます。皮質の異なる領野、V1、V2、V4 などは Pulvinar (視床枕) にフィードバック結合やフィードフォワード結合を持っています。その結合の仕方は V1 → Pulvinar → V2 → Pulvinar → V4 というような形であるので V1 - V2 - V4 - の経路の存在も考えると、三角形の回路をなしていることがわかります。また、これらの皮質と Pulvinar との間には抑制性のニューロン群が存在しています。そのため、このニューロン群で伝えられる誤差信号をカットすることが可能です。そういうわけで Pulvinar になんらかの形で注意を与えることで、物体の認識を切り替えるという機構を考えることができます。

## 4 まとめ

・動物の感覚運動系にとって予測能力は有用である・Predictive Coding の本質: - 大脳皮質の再帰的興奮性結合とフィードバックの機能的役割を示唆する- 終端抑制や古典的受容野外効果などのニューロンレベルの現象を自然画像の性質に基づいて理解できる- 視覚における注意などの認知的性質を説明できる

## 5 質疑応答

- Q1 このモデルは内部状態と出力を同時に推定するという意味で野心的ですが、工学的なアプローチとしてはどのようにとらえられるでしょうか。
- A1 このモデルはオンラインで学習するネットワークになっています。このような問題は工学ではシステム同定 (System Identification) と呼ばれます。詳しくは Ljung の本 [10] を参考にして下さい。
- Q2 フィードフォワード結合を表わすのに  $U^T$  は単純すぎるのではないのでしょうか。
- A2 前のレベルの情報全てが必要なわけではなく、例えばエッジ検出などの結果のみを表現すれば良いので大丈夫です。
- Q3 生物学的な現象として、生まれた時と大人になってからでは可塑性の度合いが異なってきますが、その影響についてはどう思われますか。



A3 網膜などは遺伝的に決定されている要素が多いでしょう。その後細かく調整 (fine tuning) されると考えられます。子猫の実験で Critical Period が存在し特殊な環境に置くと内部モデルが正しくできないことが分かっています。

Q4 心理物理学的な方法で  $U^T$  や  $V$  の変化を知る方法は考えられますか。

A4 神経生理学的な方法では、視覚刺激に対する receptive field の変化を見ることができると考えられます。心理物理学的な実験はより難しいでしょう。しかし応答時間の変化などから推測できるかも知れません。

Q5 新しい内部モデルを学習するにはどの程度の時間がかかりますか。

A5 学習率 (learning rate) やゲインによります。実際の脳ではこれの値は興味や驚きなどの情動的な要素、周囲の環境などで決まると思います。

## 参考文献

- [1] Rao RH, and Ballard DH, "Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects", Nat Neurosci, 79-87, 1999.
- [2] Markram H, et al., "Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex", J Physiol, 409-440, 1997.
- [3] Markram H, "A network of tufted layer 5 pyramidal neurons", Cereb Cortex, 523-533, 1997.
- [4] Bryson AE, "Applied Optimal Control : Optimization, Estimation, and Control", Taylor & Francis, 1981
- [5] Fellmen DJ and van Essen DC, "Distributed Hierarchical processing in the primate cerebral cortex", Cereb Cortex, 1-47, 1991.
- [6] Crick and Koch, "Constraints on cortical and thalamic projections: the no-strong-loops hypothesis", Nature, 245-250, 1998.
- [7] Dan Y, Atick JJ and Reid RC, "Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory", J Neurosci, 3351-3362, 1996.
- [8] Rao RP, "An optimal estimation approach to visual perception and learning", Vision Res, 1963-1989, 1999.
- [9] Rao RP, Ballard DH. "Dynamic model of visual recognition predicts neural response properties in the visual cortex", Neural Comput, 721-763, 1997.
- [10] Ljung L, "System Identification: Theory for the User. Second Edition", Upper Saddle River, N.J.:Prentice Hall, 1999.
- [11] Zipser K, Lamme VA, Schiller PH, "Contextual modulation in primary visual cortex", J Neurosci, 7376-7389, 1996.