A Computational study of Visual Template Identification in the SAIM: A Free Energy Approach

by

KEYVAN YAHYA

A thesis submitted to
The University of Birmingham
for the degree of
Master of Philosophy (MPhil)

School of Psychology
Centre for Computational Neuroscience
and Cognitive Robotics
The University of Birmingham
Nov 2013

Birmingham, 2013

Abstract

This thesis aims to understand how humans could recognize and identify objects. Our main method for doingso is developing a computational model of recgnition/ identification process. This work will account for the process of visual object identification which usually takes place in multiple environments including varioud objects. Since we assumed that visual selective attention is central in disambiguating of objects, the results of our work will include an implementation of what visual selective attention does. Initially this thesis will draw on two successful approaches to human information processing. On one hand, we will base our work on the Selective Attention for Identification model(SAIM). The SAIM combines visual selective attention and object recognition. On the other hand, I will use the "Free Energy" approach proposed by Karl Friston to implement the fundaments of SAIM and expand it by incorporating an identification process. We will then reason for our cliam that holds that perceptual recognition, attention and identification minimizes the "surprise" (prediction error) about incoming sensory signals (Friston, 2006). It will be demonstrated that identification process would lead to an unsupervised extraction of object templates (prior beliefs about the causes of sensory input) from a series of multiple visual scenes to execute a successful object recognition task. At the end of our work, we would test our model by doinga series of computational experiments which are performed in Matlab environment consisted of various neural networks.

In general, this thesis is divided into two main sections. The first section explains our approach that is ging to apply the methods of free energy and information theory to resolve template identification problem. Also it addresses the key concepts of free energy and the architecture of the SAIM. Besides, it compares the capability of the SAIM which benefits from both of top-down and bottom-up streams at the same time to extract the expected object from a shattered scene witht he other models. The second section, envisages the problems our model aims to resolve them through modifying the SAIM by free energy method and then gives a model in which our newversion of free energy methd accounts for the template identification problem. As we will show, repressing the surprise that comes from the environment (herewe we refer to visual information) makes our model provide a new interpretation of the SAIM that augments its efficiency. In other words, we will demonstrate that carrying out the SAIM tasks by a top-down approach is possible and bilogically plausible too.

Acknowledgements

First of all, I would like to thank my supervisor Dr. Dietmar Heinke who genuinely supervised my work by giving useful and thoughtful hints which made this thesis to go through the way we expected . In fact, most of the contents of this thesis emerged from our long debates on the different issues and also from a plenty of brainstorming occasions on which he smartly accessed and criticised all of the ideas I used to come up during the whole period of my study.

Also, my deep appreciation goes to both of my advisers prof.Karl J. Friston who provided me with a great deal of insight and knowledge about the Free energy principle and generally the way we model non-linear complex systems and also to Dr. Max Di Luca who sincerely contributed to improve the quality of this thesis. Although prof. Friston as a leading figure in neuroscience was too busy with his own scientific agenda, he always spent enough time to have a precise look at my results and answer my questions each of them played an important role to accomplish my work.

I would also like to thank my lovely parents Majid Yahya and Mahvash Almassian for their constant and kind support and for whatever they've done to provide me with the bests to go on my way.

Finally, I shall mention some of my friends who helped me somehow to cope with the problems I encountered during my study: Mr.Alireza Miralinaghi, Amir Mohammad Ghasemizadeh, Pouyan Rafeifard,Yashar Mohammadi, Davoud Shahlaei, Saeid Habil and

Ali Bahmani. Also I feel a deep sense of gratitude towards my lovely sisters Arghavan and Armaghan Yahya for their continuou and worthwhile helps.

Contents

1 Introduction **1**
- 1.1 Motivation . . . 1
- 1.2 Background . . . 3
  - 1.2.1 Basics of visual attention . . . 3
  - 1.2.2 Computational models . . . 6
  - 1.2.3 Selective Attention Identification Model(SAIM) . . . 10
- 1.3 Discussion . . . 11

2 Template Identification **14**
- 2.1 Introduction . . . 14
- 2.2 Templates and Recognition . . . 16
- 2.3 Template Identification . . . 22
- 2.4 Connectionist Approach and Identification Problem . . . 26
- 2.5 Discussion . . . 31

3 Free Energy Principle **34**
- 3.1 Introduction . . . 34
- 3.2 Basics of Free Energy . . . 35
- 3.3 Predictive Coding and Attention . . . 40
- 3.4 Discussion . . . 42

4 FR-SAIM **45**
- 4.1 Introduction . . . 45
- 4.2 Identification and the Selective Visual Attention Identification SAIM . . . 46
- 4.3 Free Energy-Based Reconstruction of the SAIM . . . 48
- 4.4 The Generative Model . . . 52
  - 4.4.1 Energy Function for the Content Network . . . 54
- 4.5 Implement . . . 59

5 Outlook **69**

List of References **76**

Chapter 1

Introduction

1.1 Motivation

Computational models of selective visual attention have been widely brought into the center of cognitive studies which attempt to reveal the different concealed mechanism of the brain that overall ruling over visual attention. By doing many cognitive studies, people have started speculating about the possibility of grasping a reciprocal knowledge held between different cognitive functions such as perception and attention. Regarding this possibility, people have come up with some ideas which utterly claimed to show how the different cognitive functions originated from different cognitive levels could interact and affect each other(Merikle & Joordens, 1997).

Within this interdisciplinary field of study, for many years we have faced relatively many difficulties which put into question that how could we approach to figure out the way attention and identification join together. So far the former studies have implied that selective attention and learning have been so tightly linked that one is likely to think of expressing each discipline in terms of the another one. Furthermore, this kind of proximity led many scientists to build up some computational models that not only shed a

light on these phenomena but inspire scholars who are working on image processing in order to offer as efficient as algorithms to recognize the objects appear on the visual field . According to (Posner, 1994), attention and identification are indistinctly coupled somehow: whenever an object appears on the visual field,our complex neural system starts to bring it into attention (consciously or unconsciously) via identifying and recognizing that and this very process occurs through many inter-related mechanisms such as searching, orienting and filtering. Hence, many scientists have gathered a plenty of psychological and computational evidences in favour of the models in which attention and identification would be merged.

Now, one of the most important problem we are dealing with in this respect is *template identification*: a basic preliminary process which is the prerequisite for object recognition and by making more progress in computer science and neuroscience it would have more scientific contribution upon studying attention. We are better to note that from now on by attention we mean selective attention and by template we mean visual template. So, the terms will sound simpler and more straightforward. Given a scattered visual fields consisted of a group of objects, we are going to simulate how the visual information of an attended object enters the brain to be learnt and from there goes up to be identified with the aid of a computational model of attention-identification. Yet, a few scientists have paid attention to the template identification problem and up to this time, no significant study has been done to build a model which benefits from a combination of information theory, non-linear dynamical systems and neural network theory. Up to now, many scholars have put forward a series of important question liaised to this issue, for instance how 'gist perception' takes place in human cognition or how the brain recognized a camouflaged object among other similar objects in a scatter scene (Tononi & Laureys, 2008).

In this section we will introduce the scope of my thesis including its essential ingredients and explain selective visual attention and free energy method which would be elaborated later n. In the next section (1.2.1) we will introduce the computational framework which has been interestingly used to study attention and summarize some of the results along with the relevant implications of the model. Furthermore, we will also give a bundle of basic definitions and terminology that are necessary to follow our discussion at later levels.

1.2 Background

1.2.1 Basics of visual attention

Selective attention is an ubiquitous cognitive process emerged from the complexity of human perception so as to help us efficiently to stand out among a plenty of non-stop incoming information in every instance (Frintrop, 2011). Selective attention plays a fundamental cognitive role which helps the brain to avoid being overloaded by too many information received from the environment. The brain therefore needs to have a mechanism to classify and categorize a sequence of more special and limited information and process this smaller portion of selected information then (Tononi & Laureys, 2008).

Since the the mid-nineteenth century, scientists has begun to address selective visual attention as a famous metaphor that's called 'spotlight' suggested for the first time by Hermann Von Helmholtz (Helmholtz, 1850). According to Hemholtz, selective visual attention could be gained by intentional changing the direction of gaze to focus upon any point-whether peripheral or central- in the visual field. However, among the theories of attention (inspired by this spotlight metaphor) developed by psychologists there came a leading and well established theory called 'Posner Paradigm' which is relied upon the

'biased competition' (Posner, 1994). Before talking about Posner paradigm in the next part, let's brief some essential terms applied to build up this theory.

Given the fact that we always deal with the limited attentional resources, firstly there would be a close competition between stimuli trying to catch the resources and secondly winning the competition strongly depends on the attributes of stimuli and the task of attention (Desimone & Duncan, 1995). Thus, since only one stimulus could be winner and represented by neural mechanisms, a limited capacity would be relocated to the attended stimulus. Seeking for a general framework to explain attention has led the people to take many psychophysical experiments from which some important results emerge. Based on the type of visual search , there would be two kinds of attentional processes, namely *bottom-up* and *top-down* to carry out the task of visual search, that is finding a target among some other objects and distractors. The former(bottom-up) is a stimuli-driven and inductive attentional process while the latter(top-down) is a goal-directed and deductive one(Tononi & Laureys, 2008).

On one hand, bottom-up process, takes into account visual saliency that is a perceptual property of the stimulus and its contrast, for example, popping a pink stimulus out of a gray visual scene including some other gray objects. Saliency is essentially related to stimulus-driven processes and the main property of bottom-up control which does not depend on the attributes of the task and is also very fast and could be influenced by 'figure-ground' effects (Itti & Koch, 2001). In such a process, even if stimuli are task-irrelevant they could catch attention. So, among a scattered visual scene, the visual search that is conducted by a bottom-up process would be biased towards the most salient object.(saliency encompasses various trends like brightness, contrast, geometrical properties and etc.). On the other hand, top-down expectation (prior knowledge) highly emphasizes

on visual task(instead of visual stimulus) and so is a task oriented and biased attentional mechanism. For example, suppose you are seeing a scattered scene in which you are intentionally seeking for a particular object which is camouflaged, now followed by a cue pointing out your target object, the object would be quickly attended and unravelled. In other words, top-down process control the spotlight-mentioned above-by putting that over different objects during visual search.

These two bottom-up and top-down processes do have their own neurological substrates. According to (Itti & Koch, 2001), 'the expression of this top-down attention is most probably controlled from higher areas, including the frontal lobes, which connect back into visual cortex and early visual areas' whereas the bottom-up is triggered 'in a pre-attentive manner across the entire visual field, most probably in terms of hierarchical CENTRESURROUND MECHANISMS'. Finally, they proposed that each time , only one object could be grasped from the visual field and others remain untouched. This process is done by 'inhibition of return'(IOR), that is another import mechanism involved in attentional deployment that prevents already selected location or spot from being selected again (Frintrop et al., 2010).

In neurobiology, through bottom-up processing, selecting the location of attention(where to attend) is primarily controlled by the Dorsal Stream that goes from the primary visual cortex (V1) up to the superior regions of the occipito-parietal cortex. Also, it is worth reminding that object recognition occurs due to the Ventral Stream which affords to affect top-down control. Bottom-up control is usually imposed by ventral stream that goes from V1 down to Inferiotemporal cortex(IT) and from there to the visual cortex (Milner, 2012). In the model we are going to offer, both of the two types of information processing (what and where) would be join and complement together by the virtue of a parallel neural network architecture (Desimone & Duncan, 1995). Also, (Olshausen et al., 1993) has

Figure 1.1: bottom-up vs. top-down. Left: The red T seems to be the first object that quickly draws your attention. This is an example of bottom-up processing, in which your attention is captured by salient sensory information. Right: the second letters of both of the words are cut in half and so look like a same thing like two ladders of same size and shape, but top down processing allows us to read the statement and recognize the disfigured words. adopted from (Mederio et al., 2010).

shown that those features related to identification are involved with cells in inferotemporal cortex, "concerned with representating the properties of known visual shapes". It shall be noticed that the SAIM- a computational model of selective visual attention whereupon we build up our model-is deprived of tuning to retinal position in retinal position throughout the whole model. Besides, "though the templates in SAIM are translation-invariant(Another important property that will be describe later but simply means that it does not matter where stimuli (objects, templates and s on) are going to appear on the visual field), they are sensitive to the spatial positions of parts from a particular vantage point. The SAIM is therefore, sensitive to view angle." (Heinke & Humphreys, 2003).

Finally, we shall point to 'eye movements', which is another important elements that must be considered in modelling of visual attention but because for some reason, the model we build is without eye movement, we will not take that into consideration. Perhaps we are better now just to point out that any model which comprises eye movements is called overt models and any model which does not- that means it explains attention without eye movement-is called covert model(Ryu et al., 2009).

1.2.2 Computational models

Computational studies of visual attention aims to understand the predicted behaviour of primate visual attention and find a proper explanation to elaborate visual perception. To do so, scientists usually exploit a broad range of disciplines including mathematics, physics, computer science and so on to discover as many novel fact as possible. By this account, we could have different architectures to attack the problem from different standpoints but there exists an element that almost every model should take into account, that is, neural information processing and so should we do so too. These models try to insert an information-processing mechanism that controls the visual information going to enter short-term memory (Desimone & Duncan, 1995). As we will show later, we assume an information theory-based approach to build up our computational model.

History reveals that the first attempts at building the computational models of attention were made with the help of the notion of 'saliency' (Koch& Ullman, 1985). Dating back to the 80's decade , (Koch& Ullman, 1985) came up with a over bottom-up model to explain attention. Their model in fact encoded saliency at different locations of the visual field and then took control of visual information processing at the focal point. We ought to note that the vast majority of the computational models have been so far built, tried to gain as much knowledge as possible about bottom-up process and since top-down stands beyond a simple topographical framework, top-down modelling of attention turned out to be a big challenge for neuroscientists. Since top-down approach is involved with higher levels of cognition , one such a low level approach was less likely to answer the challenging questions (Itti & Koch, 2001).

To the best of my knowledge, the majority of the suggested computational models focused on 'space-base attention' which means the target toward which our attentional

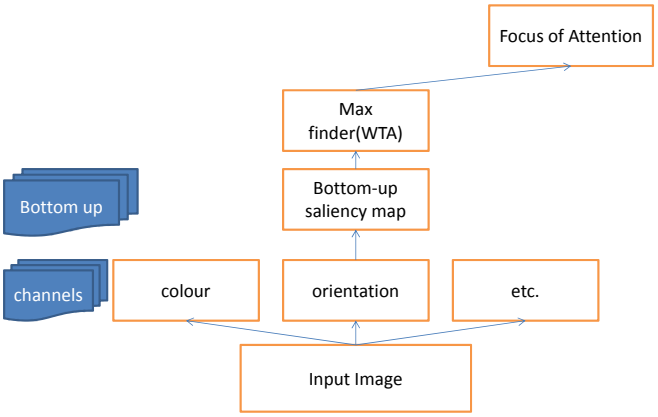

Figure 1.2: A general architecture of a bottom-up model in which information coming into the higher level and a sigmoid function like WTA(winner take all, a neural function which detects the maximum value of what is concerned like saliency) has them summed up to terminates finally into the focus of attention

focus is directed (Frintrop et al., 2010) (Bisley& Goldberg, 2003). Of these all models, we can refer to one of the famous one known as FIT(Feature Integration Theory) proposed by (Treisman & Gelade, 1980). According to FIT, looking over different objects to find the target is done 'serially' such that "different features are registered early, automatically and in parallel across the visual field, while objects are identified separately and only at a later stage, which requires focused attention". Although FIT has shown a remarkable capability when it comes to search the location of a target(where), it fails with doing the same task about identification of the same target(what)(Treisman & Gelade, 1980) (Frintrop et al., 2010). Also, there exist some contradictions that oppose to what the model claims such as *Parallel conjunction search*,*distractor inhomogeneity* and *grouping* which degrade the reliability of the model (Koch, 2000).

As time went by, researchers still carried on and took forward their agenda by proposing other theories which even though could not completely put the space-base attention assumption away but tried to find a better explanation for what occurs at higher levels using bottom-up approach. In fact, these new types of models((Treisman & Gelade, 1980), (Bisley& Goldberg, 2003), (Desimone & Duncan, 1995) and many others) have shown more validity in compare to their previous peers. Furthermore, most of the models which take a high level top-down approach are more liaised to our problem than other ones because of sharing a common concern e.g., template matching. Braun's model belongs to this new category who built up a model that could operate simultaneously in serial and parallel manner (Braun, 1994). Braun's model or more precisely "Binary Theory of Attention', proposes that 'attention encompasses two components: a bottom-up, fast, primitive mechanism that selects stimuli based on their saliency (most likely encoded in terms of center-surround mechanisms) and a second, slower, top-down mechanism with variable selection criteria, the spotlight of attention, that is under cognitive, volitional control" (Koch, 2000).

By now, we are gradually getting a bit far away from the bottom-up based theories of visual attention and going towards those models which benefit from a higher level standpoint. 'Posner Paradigm', as such, is another successful framework for selective attention which is not only grounded upon the mechanism we mentioned above namely, bottom-up, saliency and top-down, but involves both central and peripheral cues. Posner paradigm also known as 'cueing paradigm', suggests a three step covert model to carry out an attentional task. Attending to an object usually involves looking at it and placing its image at the fovea (the central area of the retina with highest acuity) (Posner et al., 1978).

So according to Posner paradigm, when a peripheral target appears, subjects would

move towards a central point and start responding as fast as they can. The target is cued with either a central arrow indicating the side it will appear on, or a peripheral box around the targets eventual location (Posner et al., 1978). Posner, also introduces two types of cues and uses both of them in his theory, e.g., *exogenous* and *indigenous*.Exogenous and endogenous cuing fit well with biased competition theory: Exogenous cues are triggered through bottom-up process, "based on the prior expectation that salient events recur in the same part of the visual field" (Frintrop, 2011). Endogenous cues on the contrary, are brought into the visual field by top-down process.

1.2.3 Selective Attention Identification Model(SAIM)

Selective Attention Identification Model (hereafter the SAIM), is a covert model for selective attention proposed for the first time by (Heinke & Humphreys, 2003) and contains a biologically plausible feature extraction property. Because our work are tightly coupled with the SAIM and borrowed some essential features of its structure, now we are going to give to some extent a detailed explanation about that. Although the SAIM has been originally built up to explain selective attention, it gives us further capability to work on higher level functions therein, e.g., Image recognition and template matching and achieve remarkable results. "SAIM was developed to model normal attention and attentional disorders by implementing translation-invariant object recognition in multiple object scenes" (Heinke & Backhaus, 2011). In other words, the SAIM privileges of translation-invariant property it does have. Translation-invariant is a basic property which is necessary to build a well-fitted model for selective attention. It could be formally defined as the following : suppose we are given a curved space X-loosely speaking, a curved space is a vector space consisted of scalars, vectors and a rule like addition that let us to have linear combinations. This space is called curved if every line that links every two elements of it, lies on

the same space-on which some metric function d is defined-metric is a function that takes any two points of the space as input and calculate the distance between them- d is called translation-invariant if and only if :

$$d(x, y) = d(x + a, y + a) \; \forall x,y \in X \tag{1.2.1}$$

Therefore, in visual attention we can redefine this property in that sense that "the contents of any location in the input image can be mapped through to the FOA. The mapping is controlled by the selection network." (Heinke et al., 2008). Translation-invariant in fact, keeps the distance between any two point always the same regardless of any change and thus every thing(points, vectors, objects) in space could be mapped in the same way it was . Besides, the recent achievements in neuroimaging studies confirmed that to carry out the task of the mapping of incoming visual data, our visual system always uses translation-invariant to do so successfully.

The standard version of the SAIM is consisted of several parts and three main neural networks that work simultaneously, namely, the content network , the slection network and the knowledge network which takes on and processes any information appears on the visual field and the focus of attention respectively. At the lowest level, visual information come up to enter both of the content network and the selection network in a parallel manner. The content network , as we said, receives visual information and also makes a translation-invariant mapping from input image to the FOA. To keep a translation-invariant representation, there is a mutual interaction between the selection network and the content network. the Selection network, controls and modifies the units of the content

network and FOA alike by running a competition between it's units , so that each time "input from only one (set of)locations is dominant and mapped into the FOA" (Heinke et al., 2008). This process is called 'Inhibition of Return' (IOR)as we quickly pointed it before. Finally, at the highest level, knowledge network is responsible to store the template and identify the visual information coming up from FOA(identifying and recognizing the objects mapped into FOA). Moreover, the Knowledge network would "modulate the behaviour of the selection network by sending top-down signals down to that".(see fig 1.4)

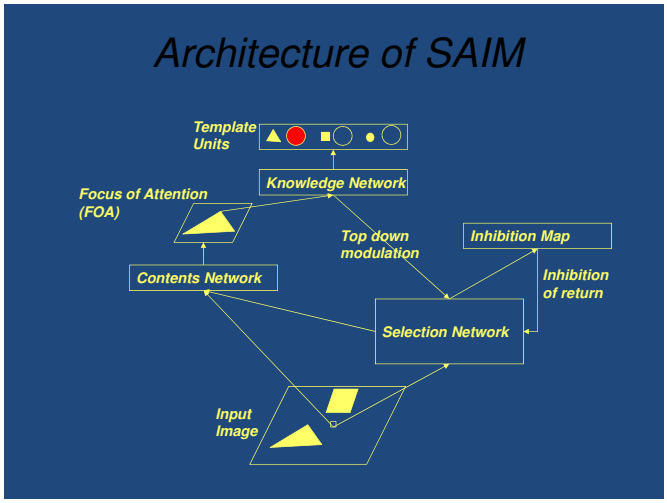

Figure 1.3: General Structure of the SAIM adopted from(Heinke & Humphreys, 2003)

In order to do this job, a translation-invariant representation of an object is formed in the focus of attention (FOA) through a selection process. The contents of the FOA are then processed with a simple template-matching process that implements object recognition. These processing stages are realized by non-linear differential equations often characterized as competitive and cooperative interactions between neurons.

1.3 Discussion

It seems that, up to this time, the models which benefits from a top-down control mechanism have tuned out to be more successful than the other ones whose tasks are mediated by bottom-up process. Moreover, they have shown less capability to demonstrate template identification program. It is believed that being privileged of top-down process inside the model is necessary to explain template matching and template identification processes but as we will depict in the next chapter, it is a necessary but not sufficient condition at all. Since our work is going to end up with building up a computational model for template identification, we will point out some basic issues derived from the topics on visual template matching and template identification.

Having the models which could work at higher levels would bring on this question that whether these models could also been be applied to understand a concealed learning process which is very likely to be adhered to identification process. It is also worth pondering if these models could be reunited under a certain grand unified theory which not only affords to enables them to complement each other and holds a loosely liaison amid, but could shed a light on the important elements that play a decisive role in accomplishing the whole process of template identification.

Identifying objects and classes have been one of the most challenging computational problems for anybody who is interested in the underlying mechanisms of visual attention. Computational studies of visual attention suggested that identification process could not occur without the target being attended. Basically, this hypothesis asserts that the brain should be inevitably endowed with a volitional attention in order to identify objects. As we go further, we would see that having a storage including the basic templates would be

the necessary condition for the brain to identify the objects that appears on the visual scene. The benefit of a storage might collide a learning process, a fluid process by which the brain would be capable of perceiving the objects in a dynamic manner like a child who is going to interact with his world. Here the problem is that how a child can store visual templates during a learning process through which he could both perceive and react to environment simultaneously.

For more than 20 years people have been kept working on computational model either in visual attention or selective visual attention from different standpoints ranging from neuroimaging studies to statistical inferential theories but though almost every model asserts that grasping knowledge about template identification is a necessary step should be taken to accomplish the identification process, no one yet could have provided any considerable account to learn out more about that. That is precisely the way we like to conduct our work. It is nt at all hard to show that we are still deprived of a computational model which take both attention and identification into consideration predicated by learning.

Chapter 2

Template Identification

2.1 Introduction

One of the most important cognitive task that our brain does is taking objective targets out of the environment and then have it then learnt via a complex process including neural feature extraction, neural representation and information processing. We can divide the visual perception function into three different but related categories: *low level* functions(specified to pre-processing), *intermediate level* functions(specified to representation) and *high level* functions(specified to recognition). As we already mentioned, there are no distinct clear-cut boundaries in between to segregate these different level each of them is situated at some different parts of the brain.

By far and large, low level functions are involved with treating incoming sensory information consisted of the processes which require no part of intelligence. In this regard, all the visual pre-processing and receiving tasks(image formation and adaptation) occur at lower level of cognition. When sensory information came into the brain, some other functions would begin to process them afterwards. In the next level, these intermediate functions are dealing with feature extraction and characterizing component (Gonzalez &

Woods, 2002). As long as the brain is operating at lower levels, intelligence wouldn't need to reveal until it gets on higher levels. Intelligence as Hofstadter assigns it to *flexibility* of the human mind that allows the most abstract concepts interact to each other throughout all cognitive levels: an emergent high level cognitive epiphenomenon which can not be seemingly found among mechanical robots at least as much as we could see among humans(Hofstadter, 2000). This kind of flexibility emanate from the enormous number of different rules help us humans to execute intermediate and higher levels procedures. Finally, higher level functions are involved with recognition and interpretation both of which are strongly assimilated by 'intelligent cognition'. The common property that underlies most of the higher level functions is using prior knowledge and expectation to play their important roles in perception and decision making (Gonzalez & Woods, 2002).

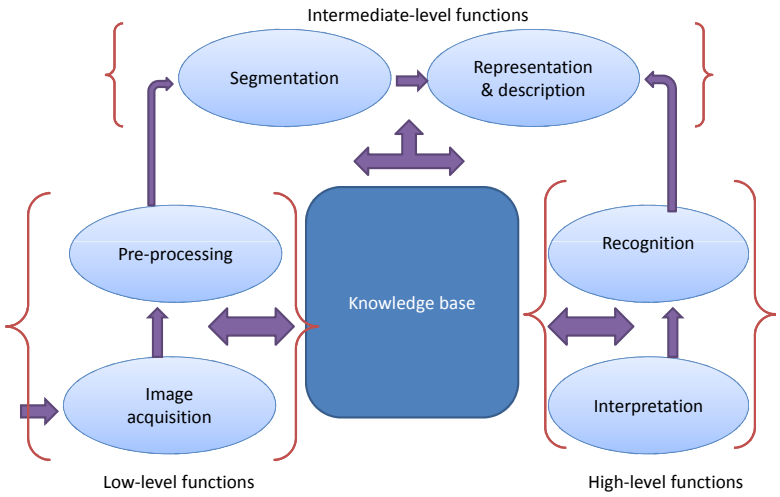

Figure 2.1: Different levels of cognition involved in human visual functions

In this work we only focused on template identification as a simplification of higher level cognitive function. To expand it more than this level, let us begin with a general

review of the problem including its mathematical formalism. First of all, we would like to give a general perspective of template identification and its underlying mechanisms controls that in the brain.

Up to now, lots of models for template identification and pattern recognition have been proposed and since most of them turned out to be successful the field of in digital image processing, we can suspect that they might have enough validity to be generalized so as to depict what really occurs in the brain. As long as pattern recognition within an image processing network is concerned, these models could be quite useful for their abilities to do visual search and grasp the salient object but as we move toward template identification in the brain, it could be no longer as plausible as it was thought. This specific field has been enlarged particularly for specific domains such as face detection and for more general object domains (Rutishauser et al., 2004). Simply and plainly, all of the pattern recognition algorithms in image processing are divided into two different categories: Filtered-base and Differential Equation-based(particularly Partial Differential Equations) approaches but what seems to become their common denominator is the application of convolution to achieve expected result.

2.2 Templates and Recognition

By far and large, there exists a prevalent approach in image processing which suggests to divide the visual scene to a mesh consisted of a finite number of pixels. the more the number of pixels , the higher the resolution would be in a picture. It looks somewhat trivial to assume a scene as a mesh of pixels so that each of them contains a certain degree of intensity. Therefore, we can assume a picture as a function from a spatial space

X domain to the intensity range I, denoted by :

$$u : (x, y) \longrightarrow I \tag{2.2.1}$$

Now, we can propose an easy to do algorithm which aims to match the template already stored in memory with the input visual image by carrying out a serial tasks of search and adaptation. This could be implemented by masking a particular image domain of the template size and starting a visual search that is ensued by adaptation. Let us suppose we have a template denoted by *T(x,y)* and an extracted sample image that has to be searched, denoted by *S(x,y)*. Hence, the algorithm pin the template to the centre of the sample image and then run a sub-procedure to calculate the differences between intensities.(this culd be done by metric functions we briefly noted in previous chapter) We can write the algorithm as follows:

1. put the template T
2. size(T)—calculate the size of the matrix T
3. get a S where size(S)= size(T)
4. translate S to T ($S \longmapsto T$)
5. calculate $D = \sum \sum_{x,y} |\ S(x,y) - T(x,y)|$

As we saw, these methods all emphasize on the notion of saliency and serial visual search belong to bottom-up based models and of course they can extract useful information about the location, size and shape of objects out of a given images (Rutishauser et al., 2004). What comes next should be obtaining a detailed knowledge about the way this bottom-up approach is implemented. As we mentioned above, this approach is usually

accompanied by two efficient methods namely , filters and PDE's. Filter-based methods usually convolve a *kernel*(like Gaussian function N) with the sample image to extract useful information such as location, colour, intensity. Now, we could take 'cross-scale' difference (the nonlinear coupling of picture elements) with regards to these local attributes(l,c,s stand for attributes of each extracted map):

$F_{\mathbf{I},c,s}$=$N|$ I_c(x,y) -I_s(x,y) |

$F_{\theta,c,s}$=$N|$ I_θ(x,y) -I_θ(x,y) |

and eventually to sum over these feature maps:

F_l= $N(\oplus$ $F_{l,c,s}$) where $l \in L_I \cup L_O$ and (L_I , L_O) are the feature maps extracted with regards to intensity I and orientation O. Finally , all the locations start to compete each other to get the highest intensity by a winner-take-all(WTA) function.(WTA is a function which runs a competition between different neurons of a layer till a neuron reaches to its highest activation(the only winner) and makes other neurons turn off) Now, we could easily take the best template to be matched using some different methods such as euclidean distance (Minimum distances), correlators, Bayes classifiers and neural network. We will show it later how to combine Bayes classifiers and neural networks as long as free energy theory is concerned.

Although, many model have been given rise by these computational models, but the human brain is of a high order of complexity and operating in so complex parallel manner that it could perform many recognition, storage and representation tasks in hundreds of a second. So in such a framework, it sounds a bit too irrational to think of the brain as a simple machine in which these enormous tasks are done by a simple serial masking and visual search as in the way mentioned above. That is why the advocates of template theory, feature theory and structural description have been trying for a long time to gather some plausible evidence to cope with theoretical dilemma they have encountered. For exam-

ple, that simple parallel search described above(serial masking) even in a super-computer show far less capability in doing the same kind of cognitive tasks than humans do.

First of all, We Should find out what goes wrong whenever we want to model ur visual system. Daniel Dennett, the American philosopher and cognitive scientist, suggests a mental experiment which can enlighten a paradox that will be revealed through the experiment itself. In his famous book, 'Consciousness Explained', he put forwards a mental experiment which brings up some astonishing results implying that our visual perception is not as enriched as we thought (Dennett, 1991). The experiment is simple: Dennett "asks us to imagine walking into a room papered all over with identical portraits of Marilyn Monroe. We would, he says, see within a few seconds that there were hundreds of identical portraits, and would quickly notice if one had a hat or a silly moustache."

Our natural conclusion is that we must now have a detailed picture of all those Marilyns in our head. But, says Dennett, this cannot be so. Only the fovea, in the centre of our retina, sees clearly, and our eyes make only about four or five saccades (large eye movements) each second, so we could not possibly have looked clearly at each portrait. Our ability to see so much depends on texture detectors that can see a repeating pattern across the whole room, and dedicated pop-out mechanisms that would draw attention to oddities like a silly moustache or a different colour. So what we see is not a detailed inner picture at all but something more like a guess, or hypothesis, or representation that there are lots of identical portraits. The brain does not need to represent each Marilyn individually in an inner picture, and does not do so. We get the vivid impression that all that detail is inside our heads, but really it remains out there in the world. There is no need to fill in the missing Marilyns and the brain does not do so." So the problem is that when we know that there a few limited light receptors placed in the retina surface,

there wouldn't be wise to assume the brain could achieve as much information as it gets. So, how come we are having with to some some extent little information while we can instantly recognize the disturbed and wrong Marilyn's photo?"(Blackmore, 2005)

Here, of course we shall notice that people have come up with a theory named *sensorimotor* theory which actually dismissed the problem, taking the viewer as an actor and the visions as the actions. Actually we will build our model back to a similar one (free energy method) which shares some assumption in common with the sensorimotor theory. Besides, some authors accounts this theory as an adequate explanation for deploying of human vision. Aaron Sloman and James Gibson each one defends this theory in a similar way somehow: "for organisms the function of vision (more generally perception)is not to describe some objective external reality but to serve biological needs. (Gibbson, 1986) (Sloman, 2011) So now it must explain how actions can become subjective experiences Henceforth, here the question is how this sort of action(seeing) would exploit all the instruction designed for movement and actions sensory motor parts of the brain. Could we make an analogy between action and seeing and further do the same for visual cortex and sensory motor cortex in order to exploit all the achievements and instructions ruling over the Brain.(Blackmore, 2005)

Based on what Dennett suggests we could gather that granting a bottom-up structure of visual search to direct attention and eventually finding the expected object is not a worth doing idea. Roughly speaking , this idea reminds us again the important role that the other attentional mechanism could play, that is, top-down processing through which a control coherent signals relates the focus of attention with other parts of the model that send up information. Given this fact, we could figure out that no matter how rapid a visual search task could be executed it is almost impossible to grasp such subtle changes

Figure 2.2: if you were to enter a room whose walls were papered with identical photos of Marilyn Monroe, you would "instantly" see that this was the case. We know that there are a little visual information are allowed to come into the brain at each instance, so searching and matching is definitely out of question. What makes that moustache Marilyn pops up whenever we take a look at the wall?

(in Marilyn's photo) in a millisecond, nor in image recognition. By now, we could take the attention role specifically top-down process in object recognition into account. (Mozzer & Sitton, 1996) proposed brought one of the first types of the models which claim that selective attention is a necessary element of object recognition. They simply ask us how could we orient our attention in order not to be caught by the most local salient spot in the visual scene. Therefore, putting the recognition task into an attentional framework would endow the model with a mechanism to disambiguate recognition by focusing on one object each time (Itti & Koch, 2001).

Some others have progressively taken the problem from different view that privileges more neural plausibility than the other models we have reviewed. (Schill et al., 2001), offered a model which suggests that recognition task is a matter of 'information gaining' and gives us an explanation as to why the brain prefers some particular object in compare to some other ones. According to their theory, object recognition is dealt with informa-

tion gaining in that sense that attention could be oriented to those parts which entail the most relevant information in compare to other spots with far more less information. In other words, attention would be fixed on the location of object which gains as maximum information as possible. Therefore, given a bundle of potential objects to be identified, attention is to distinct and classify them and finally gain the amount of information each of them contain.

Although we are aware of this fact that object recognition is a cognitive function of high level, but as we mentioned once before, it couldn't be accomplished without being involved with lower level activities. That is why today it has been accepted to impose 'what' and 'where' memory to keep this balance. In most of the model which have been proposed so far, the potential objects are selected and identified via bottom-up approach from the visual field and since then they would be scanned serially until a specific object obtain the highest recognition score which depends on feature analysis (Itti & Koch, 2001).

2.3 Template Identification

Generally speaking , we can address a recognition problem as a process of naming an object in a sense that we tend to identifying objects either as an individual object like a specific 'token' or an object which may belong to a number of a certain class and category('a truck')(Ullman, 1996) (Hofstadter, 1996).

As it is clear, so many different models which have been proposed to account for attention and object recognition, share some key regularities and ideas which holds that, firstly to identify an object the brain does not need a huge restoration of all types of shapes and variations of an object which is supposed to be recognize and secondly, all different varia-

tions of a particular object contain some necessary information carrying similar attributes which put together represent that object. Regarding this kind of information, the brain is going to recognize the object. All the methods we have looked over aim at undermining these regularities which constitute various transformation and variations. Now, we intend to shift incrementally from object recognition to novel object learning problem that of course has many things in common with object recognition but for some reason a few research programs have been devoted to unravel the complexes of this problem.

The vast majority of the models which have been offered to provide an insightful account for object recognition, usually take into account the learning problem in the extent. According to (Ullman, 1996), we can sum up almost all the models offered for object recognition over the three main methods: (i) invariant properties methods, (ii) parts decompositioned method, and (iii) alignment method. the first method says that all the object would remain invariant under all kind of transformation they might get whilst the the second method suggests that to recognize an object, the target object should be decomposed in smaller constituents. Intuitively, this method is fairly straightforward: objects contain their fundamental parts such as face, nose and eyes. These parts could be found and put them together to accomplish the recognition process. Essentially, These methods could also be considered like an inductive bottom-up process. Finally, the (iii) is to "compensate for the transformations separating the viewed object and the corresponding stored model and then compare them." (Ullman, 1996).

Although these three approaches have shown a considerable success in making recognition tasks, deeming that these too would grant success in object learning is far from what is really going on in the brain and we would reason for that. First of all we shall point out that by saying a novel object we mean an object that is being seen for the

first time through visual sense or how could we perceive a previously unknown object not seen before?(Saxena et al., 2006) Also we should notice that even learning a novel object like a triangle or square for a child is quite far from some geometrical transformation in that sense that if the learning process just encompasses sheerly geometrical elements, the robots should have been able to do the same as us human. This question immediately pops out that so why should this very problem be revealed of that high difficulty and then what are the true mechanism undergone to this task?

Perhaps the main problem with template identification is that it is not wholly liaised to the problem of how the brain is going to do decomposition task and feature extraction nor it is precisely a simple confectionist model in which some information are integrated and joint together. If we try to undermine the learning template process by sticking with confectionist approach and information theory, we would probably get into some conclusion which are indeed of great importance to notice. Firstly, we shall notice that as long as we are working on the learning template problem, "there is compelling evidence that different kinds of information" are involved in such that its seems fairly acceptable to assume distinct types of information such as visual knowledge, semantic knowledge and object naming (Eysenck & Keane, 1997).

In order to sort it out , Humphreys and Bruce did some studies comprising both cognitive and lesion ones. One of the most important results of their studies is presumably a hierarchical structure they came up with to explain the different kinds of information and functionalities which are proceeding at the same time. According to (Humphreys & Bruce, 1989), there are several distinct stages of information processing which are involved in the identification problem. Of the stages involved in identification, we could particularly refer to *perceptual classification* and *semantic classification* where the former

"involves matching the visual information extracted from an object with its stored structural description" and the latter "involves the retrieval of information about functions and associates of the object" such as naming (Eysenck & Keane, 1997).

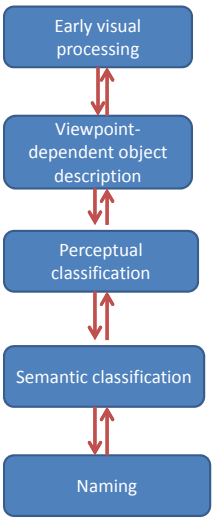

Figure 2.3: The stages involved in object identification, adopted from (Humphreys & Bruce, 1989)

David Marr was one of the greatest leading figures usually known for his contribution on computer vision, developed a theory on vision which got a good deal of acclaim afterwards. According to Marr's theory of vision proposed in 1982 , vision is considered as a proceeding process which maps the two-dimensional retinal contents to a three dimensional description of them as output. From an information processing standpoint this process consisted of three stages as follow:

- "*the primal sketch*, which is mainly concerned with the description of the intensity changes in the image and their local geometry, on the grounds that intensity variations are likely to correspond to physical reality like object boundaries.

- *the 2 1/2 sketch*, which is a viewer-centred description of orientation, contour and depth and other properties of visible surface.

- *the 3-D model*, which is an object-centred representation of three dimensional objects, with the goal of allowing both handling and recognition of the object." (Poggio, 1981)

According to Marr's theory, Objects could be created, restored and represented through a vision process which implies that the process contains some independent stages should concur and get integrated in order to produce the fine representation of the object which has been viewed. Although , the motivation along with Marr'a theory seems quite interesting, but the results show that it rarely could provide an appropriate account for identification (Poggio, 1981).

By far and large, as (**?**) interestingly put in, most of the models(including (Marr, 1982) and (Biederman, 1995)) suggest a bundle of functions involved in recognition and identification task, namely,

- Coding of the edge

- Grouping or encoding into higher-order features

- Matching to stored structural knowledge

- Access to semantic Knowledge

They combined their approaches with a connectionist model that privileges from several neural networks working simultaneously.

2.4 Connectionist Approach and Identification Problem

Connectonism has brought into the attention of neuroscientists since the early years of the 80th decade, although it could alsobe traced back to a century ago (Pinker & Mehler, 1988). Nowadays, connectionst is usually referred to as a set of approaches which aim at modelling functionalities of the mind via producing inter-connected networks of unified neurons(like sigma pi neurons-sigma pi neurons are the neurons who input units separately come in with their corresponding weights, the neuron would fire only if the summation of the weights passes its own threshold).

In general, a connectionist network or a *parallel distributed processing* (PDP) network, is a set of inter-related neurons linked together in some architectural form. To carry out a given task, the connectionist network benefits from "many small, independent units calculating very simple functions in parallel" to which the task is given to be learnt. "These networks are composed of two basic building blocks: idealized neurons (often called units) linked via weighted connections. Each unit has an associated activation value, which can be passed to other units via the links with the connection weights mediating the amount of activation that is passed between units."(Rumelhart & McClelland , 1986) (Blank, 1997).

It is usually claimed that a connectionist processing unit is in some sense taken similar to a biological neural system. A formal connectionist network consists of some layers through which information flow in and out. These layers called *input*, *hidden* and *output*, are to link and gather neurons together via establishing weights. An *activation* value is given to each value which could be conveyed to the other neurons.

Along with some little differences, most of connectionist networks work in a fairly straightforward way as follow: we set out to feed the lowest level of the network by a set of activations(mostly initialized randomly) called input pattern, then activation could be spread over the hidden layers which are to make deduction upon the input activity by summing them up and get them passed through an activation function and finally the new activation for the output layer would be calculated and exposed , called output pattern. (Blank, 1997) In a single neuron model , after getting the incoming activations denoted by a_i the network would calculate its *net input* by summing them up and have them multiplied by an activation function f through the weights as follow:

$$a_n = f(\sigma_i) \tag{2.4.1}$$

$$\sigma_m = \sum a_i w_{i \longrightarrow m} \tag{2.4.2}$$

where the previous units denoted by the index i are linked to unit m, and a_i indicates the activation of i. $w_{i \longrightarrow m}$ is the weight of the connection from unit i to unit m via imposing a logical activation function f (Berkeley, 1997) (Blank, 1997).

Based upon two assumptions, an prevalent interest about connectionist network to the template identification problem has started raising up. The first one is that each template like a circle or triangle has its own activity pattern spreading over the unit which represent them and the other one is that by posing an input activation, the network would be able to make analogy, learn and generalize accordingly. Alongside these two presumption , we could also build up more complex inter-connected networks which gain encoded visual pattern as the input pattern and imposed deduction through activation of *inhibitory* and *exhibitory* connections. And so we can rephrasing the problem and redefine it in terms of such framework.

A connectionist network, by far and large, aims at producing a desirable output which is recognized, categorized, induced and generalized. This all is done by *delta rule* which holds that: given "random weights and feed it a particular input vector from the corpus", activity would propagate "forward to the output layer. Afterwards, for a given unit u at the output layer, the network takes the actual activation of u and its desired activation and modifies weights according to the following rule":

$$\delta_{iu} = \alpha(desired_u - a_u).a_i \tag{2.4.3}$$

where α is the learning rate and a_u and a_i are the correspondent activations of the current actual u and unit i successively. Therefore, after learning occurred, having fed input data, the network would go to take them and produced them in output as such. Moreover, interestingly if just a part of input is exposed, the network could detect and make the correspondent and appropriate values activated. And this all is done by by making adjustment such that the internal input would always remain equal to the total input, as expressed in the above formula.(Eysenck & Keane, 1997)

A major success of connectionist networks is to learning problem is that these models could be regarded as more insightful than the classical learning approach for it allows to the input data to do sophisticated tasks through running different kinds of interaction and competition. They, in general, take input as the encoded representation of the object that is to be learnt and trigger a learning process amongst hidden layers.

Nevertheless, still a lot of problems have not so intimately been raised up, particularly those which twist around validity of the postulates the connectionism encompasses.One

of the vigorous critiques ever imposed belongs to (Block, 1995) who strongly believed that the brain does not work in such a superficial way the connectionism suggests. For instance, he goes to challenge what we are usually dealing with in this area. According to Block, "Connectionist networks have been successful in various pattern recognition tasks, for example discriminating mines from rocks. Of course, even if these networks could be made to do pattern recognition tasks much better than we can, that wouldn't suggest that these networks can provide models of higher cognition. Computers that are programmed to do arithmetic in the classical symbol-crunching mode can do arithmetic much better than we can, but no one would conclude that therefore these computers provide models of higher cognition".(Block, 1995) Connectionist do not aim at simulating a broad range of various tasks of the brain neurons. Furthermore, it is not generally accepted that the prevalent methods in connectionist models such as back-propagation really take place in the brain. (Pinker & Mehler, 1988) Anyhow, according to Block, in general having supposed an exact similarity between connectionist models and what the brain does, looks like a bit superficial.

However, foundations of connectionism reveal that it could not be far too much from the brain functionalities. At the first glance, connectionist models may resemble some sort of what is going on in the brain, that is, a highly complicated inter-connected system consisted of neurons and dendrites. The motivation right on the contrary to classical and sceptical approaches seems intimidatingly outlaw. As Hofstadter points out in his introduction to the new edition of the well-known Ernst Nagel's book, "since the cells of the brain are wired together in certain patterns, and since one can imitate any such pattern in software that is, in a fixed set of directivesa calculating engines power can be harnessed to imitate microscopic brain circuitry and its behavior. Such models been studied now for many years by cognitive scientists, who have found that many patterns of human learning, including error making as an automatic by-product, are faithfully replicated" (Nagel

et al., 2001). But, still a lot of problems have not so intimately been raised up, particularly those which twist around validity of the postulates the connectionism encompasses.

In conclusion, connectionist models benefit from many advantages over the classical *symbol processing* models in which everything is analysed and processed in terms of abstract symbols rather than actual numerical values. Perhaps the most famous aspect by which these models are known, is that giving a few example would be fairly enough to trigger a learning process without resorting to all symbolic representation. Also, as Feldman truly put in, a concept should not necessarily be represented by an unique unit but every concept is mostly exhibited as a pattern of activity which is distributed parallel one the space. Nonetheless, connectionist networks are to map the pathway in which information are mapped from retinal parts to a head-centred coordinate system (Zisper & Anderson, 1998).

2.5 Discussion

Since the capacity of visual system is very limited , a few amount of information could be passed through retina. That is why incoming information starts competing to reach the focus of attention. Making much progress in psychophysics of attention as well as information theory positively has affected ongoing research in competition for attention. It has tentatively been more appealing when neurological experiments affirmed that the firing rate of receptive cells would change the stimuli which is going to be attended. So far, most of the models are determined to establish a paradigm which leans back to information processing taking place in saliency map but again experiments revealed that it is very likely that the brain attend some less salient spot or weak stimulus among a bundle of more salient stimuli. Here, most of the model explained above come to collapse and

the appeal for those models which could enable us to swallow some strange behaviours of our attentional mechanisms would increase.

Biased competition theory have been very favourable to assume a loop of forward/backward feedback signals dispatched from the outer areas of the visual field. to violate the obligation of attention and then conduct it towards some certain stimulus given a loaded scene of different stimuli. Even though, due to the brain-imaging studies that such an outer signal does exist and is projected to an occipital area called "extrastriate visual cortex", the whole theory fails in offering a proper explanation for template identification.

Roughly speaking, most of the models considered above are not going far too much from some sort of locality in that sense that their attempt at indicating a detailed tiny aspect of selective attention backed up by task experiments, would not be able to shed a light on some other vague aspects of the problem which require a different paradigm to be resolved. These models all share a trend that is they usually turn out successful in explaining something and terribly unsuccessful when it comes to explaining else. Hence, we are dealing with the models which could not grant any final explanation. Since each one benefits from different strong points , they are thought to be brought together to get more flourished view in hand. But we shall note that it would be pretty misleading if we suppose these models and theories are essentially to complement each other towards finding a ultimate explanation none of them solemnly entails.

In philosophy of science, a prolong controversially discussion has been kept fresh via the people who push forward the question of the way science do make progress. It used to be a prevalent belief that science keeps going forward its way due to all the theories have so far come into existence and put together provide us with a great deal of knowledge on

the issue they all take that into account. On the other hand and right on contrary to this theory which seems fairly acceptable by common sense, a decade after the mid of the previous century there came another theory which strongly shook the basis of the old theory. The new theory by challenging the cumulative essence of science, turn around its pivotal theme that science is not working in the way we used to think that is by collecting theories and have them summed up to complement each other, rather according to Kuhen, science proceeds through scientific revolutions which come to abrogate the paradigms upon which normal science leans back (Kuhn, 1996). As long as we rest on the Kuhn's theory, it is very natural to encounter a cluster of theories on some specific issue that some or many of them come to contradict each other. Hence, to assume these theories are to gather to demise either a larger perspective or a new more general theory seems a bit absurd. In the next chapter, we will discuss how a new paradigm named "free energy theory of the brain" come into being along with many consequences of great import which not only cover most of the issue have so far been offered, but it could truly unite them in a very subtle manner.

Chapter 3

Free Energy Principle

3.1 Introduction

Connectionist and symbolic approach(that constitutes that the brain operates solely based on symbols like a Turing machine) have been competing each other for a long time and hereby, every now and then, some people tried to bring up some novel approaches which could grant the other aspects have been neglected so far. Like the other disciplines, coming up with a grand unified theory of the brain has been a big motivation. Having inspired by Helmholtz's theory in statistical physics, *free energy principle* indicates that "what cause our sensory inputs and learning causal regularities in the sensorium can be resolved".(Friston & Klaas, 2007) In this chapter , we will give an introduction to the free energy principle and its reminiscent seeds spreading across the cognitive models which attempt at finding action/perception regularities.

The second law of thermodynamics is a firm physical theory which deduces a trend, that is, differences in temperature, pressure, and chemical potential equilibrate in an isolated physical system or generally the *entropy* of the system would increase, as time goes bye. in statistical mechanics entropy is defined a bit differently from what information

theory is based on, commonly known as the Shannon entropy. Friston comes to build up his theory upon this concept of entropy makes an analogy between thermodynamics and human perception in the brain.

Prior to free energy theory, Bayesian theory of the brain and some of its variations went towards the centre of cognitive science. Most of these theories regard the brain as an inference machine that adopts to the rules to deduce and infer over the information obtained by sampling data with our senses. Naturally as long as we are dealing with the outer environment whereupon we have no control, uncertainty would be an undistinguished and big part of the life. Bayesian models, therefore try to provide an explanation for how the brain goes to cope with uncertainty. Given the hierarchical deployment of cortical areas and also existence of forward and backward connections, Bayesian theory of the Brain aims at explaining the essence of this hierarchy and the functional asymmetries in these connections by statistical methods.

So far, many aspects of the cognitive problem have been models by Bayesian theory of the Brain. For instance, in psychophysics: some of the problems regarding human perceptual or motor behaviour have been worked out as such. Moreover, in neural information processing, a theory called Hierarchical Temporal Memory came out to describe how information could be categorized and processed in the brain given Bayesian network of Markov chains. Also, it has been a big help to obligate the studies upon the representation of probabilities in the nervous system.

3.2 Basics of Free Energy

Actually, free energy principle would account for any self organizing system which wants to retain within its state and avoids being distracted and such systems ranging from cellular organism to the large networks. As evolutionary biology asserts, biological systems(such as animals or brains) usually tend to run away from disorder through interaction with the environment which is always changing (Friston, 2006). Such an interaction with a changing environment is called *homoestasia* that is defined a process whereby a system(open or close) regulates its internal environment to maintain within a stable and constant condition (Cannon, 1929).

Again we like to ask what is free energy? Free-energy is a mathematical model based on a modified version of information theory and inverse Bayesian theory that puts a bound on the surprise the system is gaining from its environment. In other words, the range of the states in which a biological system retains is limited and mathematically it means that the probability of these limited state is of a low *entropy*. Entropy in simply a tool to measure uncertainty, technically it defined as the average surprise of outcomes sampled from a probability distribution or density. It is computed as the negative log of the information content of input data:

$$H(x) = E(-lnP(X)) = \sum P(X)ln(P(X)) \tag{3.2.1}$$

where X is a discrete random variable, p is the probability distribution function and E is mathematical expectation. Intuitively, entropy H tells us how much information does the sampled data X contain citepRe61. It is easy to do to show that if the probability of occurrence is zero, the entropy would also be zero and for the probability of 1, the same is held too. It points out that, if we are perfectly certain about occurrence of

something, let's say X, it has no information to get and also if there is no probability of occurrence of X, still no information would be passed out. Therefore, we say entropy is a measure of uncertainty which is dealt with the amount of surprise. the less probably something, the more surprise it would have. For example, a living fish out of water would be surprise. In conclusion, what free energy does is to gather more evidence for existing of something(sensory data) by putting a bound on surprise(to be more certain of it) via violating its internal states.

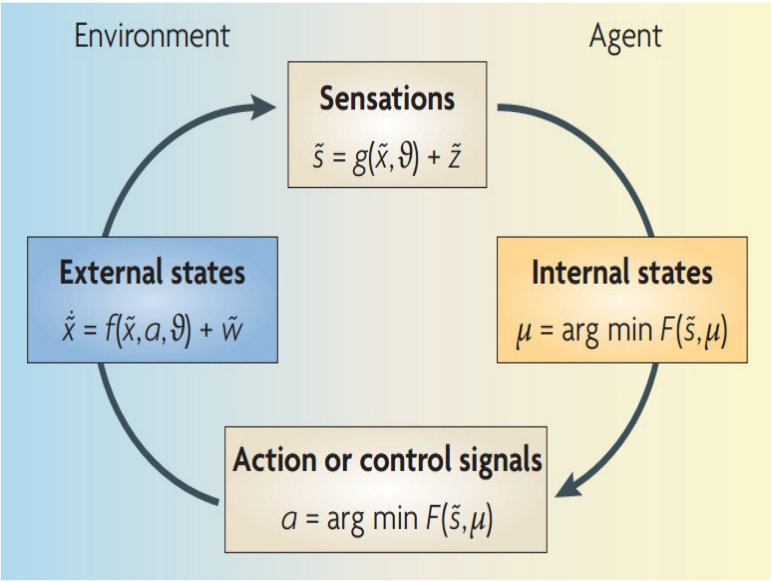

Figure 3.1: Different levels of cognition involved in human visual functions

In fact, according to free energy principle what a biological creature does is to make prediction about what is going on in the environment, given a prediction error , it would go up to improve it by making its internal states change. Simply free energy could be defined as a joint coincidence of causes and its causation and "is a function of sensory data and brain states" (Cannon, 1929). It is known from information theory that surprise couldn't be directly computed and the bound on surprise actually allow us to cope with

this obstacle because as such it is a function of sensory input and internal states.

Although mathematical formalism of free energy looks to some extent sophisticated, but as such, its pivotal claim is simple and plain: given a generative model, by suppressing and minimizing surprise the brain gets from its environment, the brain could gather more evidence for its existence. This simple sentence has hitherto been the core of free energy. Intuitively, the more the brain minimize surprise, the more it gathers evidence for its existence in that sense that having been provided with a generative model, the brain would go to suppress surprise by either changing sensory sample input. Perhaps before going through giving mathematical formula at this disposal, we'd better to give short explanations about some technical terms upon which those formulas would be written.

The first one we shall talks a little bit about is *generative model* that is a fundamental requirement to build up our framework. Basically, generative model is probabilistic structure which tell us how data and causes are dependant to each other in terms of of the likelihood of data, given their causes (parameters of a model) and priors (initial expectations or the probability distribution function of the causes gives the brain some beliefs about those causes before observing the data.) on the causes (Friston, 2006). To us, the second priority goes to *recognition density* which is mentioned very often in almost every model that has been built up based on free energy principle. Recognition density is actually a probability distribution function of the causes of the input (sensory) data; some sort of *conditional density* of the causes given the internal states of the brain. This conditional density that is also called posterior density is the probability distribution function of causes which corresponds causes to observed data. Eventually, maybe we should give a short explanation about the term *Kullback-Leibler divergence*. As we said a couple of lines above, in information theory, entropy is taken as equal to the negative log of the

probability of a random variable. This surprise is mostly known as "self-information" in that sense that it says how much information(in terms of bit) does it contain but when it comes to pass up information from one thing to the other , we would be dealing with "mutual information" or *joint entropy* in which there is a "source" to send up information and a recipient that is called "target". Likewise, the same formula (3.1.1) with a little difference would be revealed. Given an ordered couple of (X, Y) which are our source and target respectively, the amount of information passing between these two denoted by $H(X, Y)$, could be computed as follow:

$$H(X,Y) = E(-lnP(X,Y)) = -\sum_{x,y} P(x,y)ln(P(x,y)) \tag{3.2.2}$$

and when it comes to deal with information transformation or *transinformation*, we would be willing to find about information gain.

This term is applied very often in neuroscience and relevant topics which have something to with information gaining or loosing such the problem of neural coding (Dimitrov & Miller, 2001). However, when we have two random variables which lay into a trnas-information state, we could measure the amount of information via comparing the the probability distribution of each one. Here, Kullback-Leibler divergence comes to measure this amount by subtracting these two probability distribution. Suppose we have a *posterior* probability distribution $P(X)$ and a *posterior* probability distribution $Q(X)$ that is to predict the content of the prior $P(X)$. Kullback-Leibler divergence denoted by $D_{KL}(P(X) \parallel Q(X))$, basically says how close the model $Q(X)$ is to the true probability

distribution $P(X)$:

$$D_{KL}(P(X) \parallel Q(X)) = -\sum_{x} P(x)ln(P(x)) - xQ(x)ln(Q(x))) = -\sum_{x} P(x)ln\frac{P(x)}{Q(x)} \tag{3.2.3}$$

presumably, the potent the Kullback-Leibler divergence is endowed with, has persuaded the people who keen on to know more about something through observing something else in that sense that given an expectation, we could find out how remarkable and reliable it would be.

3.3 Predictive Coding and Attention

Now, time to build up the foundations of our free energy based version of the SAIM. As such, these formulas might seem a bit sophisticated though having a key theme upon which they lean back, could make them more convenient to swallow. To make life easier, we'd better to begin with reminding that free energy is formulated in terms of a non-linear dynamical system and its pivotal ingredient lies upon the notion of entropy and surprise as we described above except roughly speaking, on contrary to surprise, free energy has the privilege of being easily computed for it depends on the brain states and sensory data.

Suppose, Given the brain internal states μ(t) and brain action a(t), we do have sensory signal s=[s,s',s",...] and its causes ϑ of sensory input(that is a function of hidden states and some other parameters such as precision). According to free energy principle , the brain minimizes free energy *F(s,μ)* by taking action on environment or changing its internal states. the free energy could be written as follow:

$$F = - < ln(P(s, \mu) \mid m) >_q + < ln(q(\vartheta) \mid \mu) >_q \tag{3.3.1}$$

$$F = D(q(\vartheta \mid \mu) \parallel P(\vartheta)) - < ln(P(s) \mid \vartheta, \mu) >_q \tag{3.3.2}$$

$$F = D(q(\vartheta \mid \mu) \parallel P(\vartheta \mid s)) - < ln(P(s) \mid m) >_q \tag{3.3.3}$$

where, a=*argmax Accuracy* and a=*argmax Divergance*.

These equation put together describe the condition in which free energy is repressed. Since free energy could be considered as the difference between two p.d.fs namely, conditional and recognition density, by taking action a, by minimizing free energy the brain adjusts its internal states and optimizes recognition density as the a-posterior model to predict that is p.d.f of causes given generative model-for conditional density(prior p.d.f of causes).Again it's wroth asserting that free energy is nothing but the difference between "energy" and "entropy". (Friston, 2009)It could be simply realized that free energy attempts to derive causes from sensory input data. The action the brain takes is tightly liaised with accuracy and changes the way the brain samples sensory data.

Relatively, environment could affect us by giving sensory information and we could act on it by changing the way we sample sensory states. Henceforth, action could minimize free energy by changing sensory input and and perception could suppress free energy by violating predictions. This is usually known as active inference. Actually, the last two equations above (3.2.5) and (3.2.6)-which could be coupled-indicate that the brain ought to infer causes based on their correspondent sensory input data it figures out how sensations are caused.(Friston, 2012)

The term $\mathbf{L}(t) = -ln(p(s, \vartheta \mid m))$ is called Gibb's energy and shows the surprise coming from the joint coincident of sensory data and its causes. As such, free energy could put a measurable bound on this surprise. To compute and minimize surprise we could rewrite recognition density in terms of a hierarchical forms comprising prediction

units and error units. To model sensory information in a hierarchical structure, we bring about some equations which represent our state-space consisted of sensory states:

$$s = g^{\nu}(x^1, v^1, \theta^1) + \omega^{\nu} : \omega^{\nu} \sim N(0, \Sigma^{\nu}(x, \nu, \gamma)) \tag{3.3.4}$$

$$x = f^{\nu}(x^1, v^1, \theta^1) + \omega^{x} : \omega^{x} \sim N(0, \Sigma^{\nu}(x, \nu, \gamma)) \tag{3.3.5}$$

where, f, g are nonlinear functions to which map hidden and causal states which are parametrized by parameter θ. the causal states ν which are mediated by hidden states x through which the hierarchical states link together and provide some kind of memory for the system and establish a dynamic over time. ω^{ν}, ω^{x} are random fluctuations which are produced along with observation.In such a structure, there are two kinds of units, namely those forward connections putative units that convey prediction error and backward connection units which bring up predictions. It could be put in this was during a series of the forward and backward interaction, prediction error would be minimized via imposing a gradient descent on free energy. (Friston, 2012) , (Friston, 2006)

Looking at the fig. 13 reveals that predictions are encoded from the same level and level below whilst prediction error messages are conveyed through the same level and the level above. Fortunately, we can rewrite recognition density totally in terms of prediction error and so what remains is likelihood of prediction errors on the causal and hidden states, that is :

$$\xi^{i} = 1/2 \prod^{i} (\varepsilon)^{i} \tag{3.3.6}$$

this interaction between the state units and error units, trigger a top-down process to lead conditional expectation μ^{i} towards making better prediction. These top-down expectations are indicated by $f(\mu^{i})$ and $g(\mu^{i})$.

3.4 Discussion

In conclusion, what free energy points out could be summed up in a simple term: by optimizing of synaptic gain or mutual information between states, the brain tries to get in hand what caused its sensory input. This is simply done by a series of forward and backward neural coding which put a gradient descent on free energy and make the prediction gets better and the precision to increase.

Free energy has taken many research areas into consideration and provoked questions coming out mostly from prolong discussion have not yet been resolved. Intuitively, its well-defined equations and pivotal notions which came to reconcile information and Bayesian brain theory has, could provide account for those connectionist models which particularly deal with action/perception. Even though its initial masterminds were to give an explanation for "how we represent the world and come to sample it adaptively", it has gone beyond the Sensory-Oculomotor processes and certainly could be applied in attention and biased competition(Feldman & Friston , 2010), associative plasticity (Mathys et al., 2011), perceptual learning and memory(Chumbley et al., 2008), probabilistic neuronal coding, predictive coding and hierarchical inference(Kilner et al., 2008), the Bayesian brain hypothesis (Kilner et al., 2008), the free-energy principle(Friston, 2006), model selection and evolution (Friston, 2009), computational motor control(David et al., 2005), optimal control and value learning (Friston, 2012) and infomax and the redundancy minimization principle(Friston & Kiebel, 2009). Actually, i has been held that free energy aims managing to gather different theories of the brain together and give them all a common denominator which make them look similar at some higher level. For instance, it would come to claim that any kind of modification in "synaptic activity, connectivity and gain" would affect their counterpart coming out as the brain cognitive functions such as per-

ceptual inference, learning and attention (Friston, 2012).

We will show in the next chapter that how free energy could be imposed in a connectionist framework(here the SAIM) to explain template identification process such that it retains both bottom-up and top-down process. This could in the extent compensate what has been repeatedly pointed out as lacking of the top-down models in unravelling selective visual attention.

Information theory have nowadays been a pretty appealing disciplinary to neuroscientists who are favourable to know more about how neural information are encoded and perceived in different distributed parts of the brain.

Chapter 4

FR-SAIM

4.1 Introduction

As described above, free energy principle came to study the integration of Bayesian brain theory, action/perception, neural coding and optimal control in various domains. A pivotal distinction between free energy and other theories of the brain is that it embodies identification as well. We want to indicate that free energy not only could explain those cognitive functions and incorporates it as a direct result coming from suppression of energy.

Now the question is given a scene consisting of two objects, how the visual system is going to represent them on focus of attention and eventually identify them as the best match with template stored in knowledge network.(described at section 1.2.3) As we noted, there is a serious discussion over what kind of mechanisms underlay learning process through selective attention. We aim at showing how energy suppression, that is finding causes based on incoming data, have liaison with identification and object recognition through which we would come across some critical concepts derived originally from game theory and reinforcement learning.

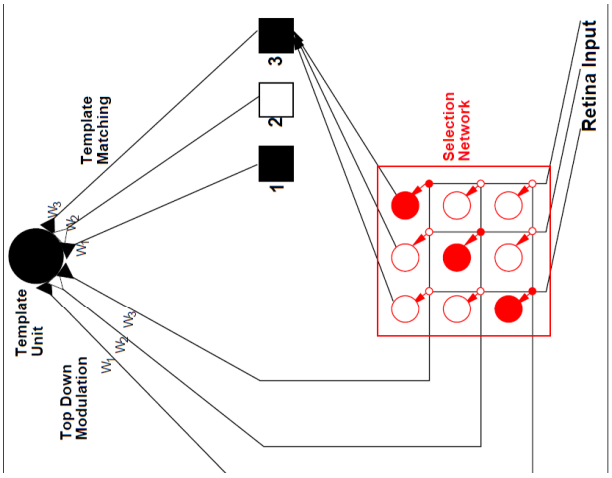

Figure 4.1: reciprocal relation between the content network and the selection network in the SAIM(Heinke & Humphreys, 2003)

4.2 Identification and the Selective Visual Attention Identification SAIM

Identification and representation are two appealing cognitive aspects every successful model shall incorporate them, as such. In this section, first, we will put forward and amplify the problem and then show that how the new version SAIM working based on free energy come to encompass it. As we showed before, selective attention identification model(SAIM) is a successful model of visual attention consisted of three neural networkds working in parallel, namely, the content network, the selection network and the knowledge network. The SAIM came initially to study a cognitive impairment called visual neglect but later on it reveals it is technically strong enough to explain most of the important attentional functions. Its functionality seems fairly straightforward : given an object in visual fields, the SAIM that is based on a Hopfield network (a recurrent neural network that is formed by a pair of binary nodes which get either 1 or -1 values and the whole nodes like a closed graph are related by corresponding weights) is to project the object

to focus of attention.

We know that feed-forward structures are considered as a strong tool to represent and interpret objects via mapping and projecting them through modifying the correspondent neural weights. According to classical theories, it's just enough to find a way so that the network(known as *universal approximator*) could find such a suitable functions which are to modify the weights. But this processing of changing the weights called *adaptation* has encountered important problems that could not be easily resolved. As we said in chapter 2, one of the most appealing for such neural network coming from the fact that the network could learn its weight values from a bundle of examples it is given. But the problem arises when the network is fed by a loaded scene in which objects are not in a rest mood but they run a competition process and besides metaphorically what happens if there is no teacher(supervisor) to conduct the student(network) about how to approximate proper functions based on examples which have been arbitrarily distributed? Albeit there came people like Kohonen who suggested an unsupervised learning structure(Kohonen, 1990) which is devoted to find more about self organizing feature maps but the problem still seems fresh for the efficacy of these methods have been challenged by those problems comprising both state of supervision and self organization.

The problem have been drawn in different scales, for instance recognition of an visual template has gone to be a key theme in image processing and visual science. There exist many procedures and algorithm in amongst textbooks but their efficacy and success in representing an object as it really looks like, is in the extent limited. (Trappenberg, 2009)A crude part of this difficulty goes back to the theories which push forward different structures and assign different values to the network neurons to represent a visual scene. We believe that most of this difficulty gets strength from a theoretical mistake that is

image recognition is not a lower level task and couldn't be treated as in the same way as simple process have been so far. In the chapter 1, we showed how people came to built up models which take recognition and identification into account by assuming them a bottom-up low level process(Itti & Koch, 2001) and we also showed to what extent their models works naively. A crucial common mistake that many of them shared is that they take image representation as a homogeneously distributed process which could be modelled by assigning a feed-forward linear network whilst brain-imaging findings and psychophysics experiments tell another story.(Trappenberg, 2009) Here the SAIM came to say first, how the visual inputs are internally represented and then backed by free energy principle go them up to be projected on focus of attention and eventually identified.

4.3 Free Energy-Based Reconstruction of the SAIM

As the structure of the SAIM suggests, information from visual fields come to be mapped to the focus of attention through two parallel networks, namely, content network and selection network working in a reciprocal manner. Again we shall assert that what is important here is that the SAIM is a translate-invariant model of selective visual attention and this property is gained via mapping contents of visual field to focus of attention which make the model to be capable of identifying retinal inputs. When multiple objects appear in retinal spot, the model would select one and only one object in order to prevent them from being overlapped in focus of attention and this inhibition mechanism is imposed by selection network. At the same time, content network would rectify the objects already selected. This mechanism is divided in the two interconnected phases: one carries out the mapping process which go from retinal parts to the FOA and the other one that takes the mapping process under control.

In this models , two different streams go up and carry the visual information harassed from the retinal parts. Every neuron "in the contents network represents a correspondence between the retina and the FOA" and "the selection network determines which correspondences are instantiated". We should note that units in content network are "simga-pi" nodes. Henceforth, both of the content network and the selection network are to launch two correspondences which put together link the second level of the network dynamic. These following equations imply these correspondences which yield translation invariance through visual data to the FOA: generate translation invariance.:

$$y_{ij}^{FOA} = \sum_{k} \sum l y_{kl}^{VF} y_{SN}^{ikjl} \tag{4.3.1}$$

where, y_{ij}^{FOA} stands for activation of units in FOA, y_{kl}^{VF} stands for activation of units in visual field and y_{SN}^{ikjl} stands for activation of units in selection network. We have also two different spatial indices, namely kl and ij which refer to the visual field and the FOA respectively.

The selection network adjusts the mapping from retinal units to the FOA by imposing some constraints which ensure that each time only one unit in retinal field is selected to be mapped on the FOA. Besides, selection network prevents the network from selecting one unit twice that is mostly called inhibition of return. Finally, selection networks keeps the neighbourhood units that are spread around the selected unit and have them mapped as well. To do this, the standard version of the SAIM applies an energy function called Hopfield network. This energy function puts in a winner-take-all(WTA) to select the the best match unit amongst the input data. This energy function is written as following :

$$y_{WTA}(y_i) = a.(\sum_{i} y_i - 1)^2 - \sum_{i} I i \tag{4.3.2}$$

where, I_i is the input data coming into the visual field and y_i is the output of the units. As soon as y_i is selected, the other units rests in zero and hence the Energy function would be minimized.

Now, times to see how a free energy approach could give rise to the same results of course with a little bit difference. First of all, we have to notice that the standard version of the SAIM is constructed on a bottom-up approach and event though it encompasses a top-down control but there is no such a thing as expectation and prediction error as suggested by the free energy approach. We are going to show that how having been provided with expectation and prediction error could give rise to the same hierarchy that after all converge together.

Free energy approach implements both of feed-forward and backward connections which are to equip the system with memory and expectation. Roughly speaking, we come to claim that identification occurring in the SAIM is i) a high level activation and ii) could be rebuilt based on free energy dynamical structure. Being a high level activity has been shown before particularly the brain-imaging studies have held that identification take place at inferiotemporal cortex(IT) that categorically belongs to the higher levels of the brain. About rebuilding the model, we'd better to say that we are determined to take the action/perception equations and put them all in a prediction-error type. To begin with energy function ought to be derived in order to be used to take its derivatives in respect to each variable.

In fact, as long as the generality is concerned, we prefer to preserve the same structure the SAIM does have and so all the networks and their orders remained the same. Initially, we distinguished the selection network from the location map and wrote different equations

but as the present results strongly suggested, we decided to combine them as consequently the model exhibited a better performance. To do that, we come up with this idea to write an equation comprising both of them. Another important point refers back to the difference that our model is having with the standard models of free energy. Free energy-based models have all a dynamic property and there some hidden state which are responsible to pose the motion of the environment and there exist some noises to make the model to adopt itself with environment. Basically,these dynamical states which is issued from the nature of environment have no part in our model. To our knowledge, up to now, no model has ever offered which leans upon statistic states and works with static states. Therefore, the variables and parameters have been flourished in the extent to let the model works well as long as it has represent a static environment.

We are hitherto determined to push forward the problem from the standard version of the SAIM viewpoint and then go to offer our free energy-based model and draw a comparison between them via showing the results. These results could be compared either intuitively through showing the mapped objects at each level or technically by showing the matrices stand for the objects.

To make a conclusion, we shall make a subtle hint. Strictly speaking, each free energy-based model firstly needs to be turn into a dynamical system and the correspondent equation of energy should be written carefully. Here, energy means the same as what the theory entails, that is the joint coincident of data and its causes. A model of selective visual attention naturally require visual sensation as its incoming input and so data is defined as represented visual data in retina and then lower level of the visual cortex. Further implementation would imply causes restrictedly as the source of data or precisely what causes incoming data which is obviously visual template.

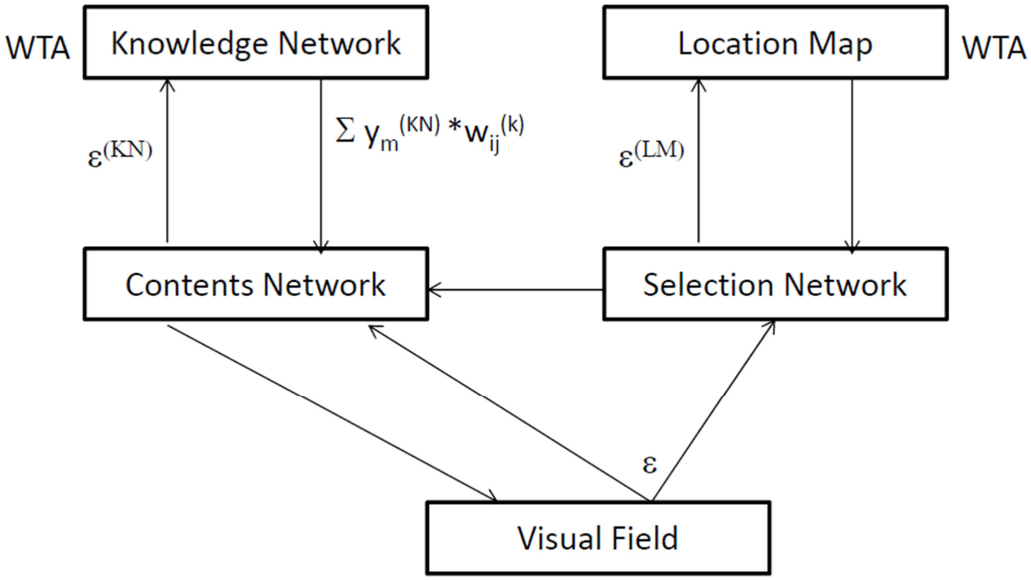

Figure 4.2: The free energy structure of the SAIM

4.4 The Generative Model

First of all, we begin with bringing our generative model to which the model is leaning back. Since we want to write all equations in such a form that is adaptable with 'the neural message passing' system to let predictive coding embarked.

As such, the SAIM is not working as in the way the usual free energy-based models suggest that is environment should be defined based on the equations describing motion. Therefore, we'd better to consider our model as kind of steady state version of free energy.

As we said , we intend to sort out all formulas in terms of prediction error and as we will see, error terms would lye in all of the equations. We are going to write energy function for each part and then integrate them up and we'd also come along with some explanations to interpret what the equations assert.

Since the hidden states have vanished and recombined to causal states, the generative model bears a predictive scheme to map causes to consequences. This scheme has only two levels , the first one is to gauge the spatial representation of the object and the second one is their neural weights in selection network. However, the generative model could be written in terms of these two ingredients and we should point out one of the most important results of free energy perhaps, that is, the inversion of generative model is taken equivocally as precision.(Friston, 2009) In conclusion, this generative model comes to appear in the following form:

$$s^i = f(x^{(i)}) + w : w \sim N(0, \Sigma^\nu(x, \nu, \gamma)) \tag{4.4.1}$$

and also $P(s, u) = P(s|u)P(u)$ so we would have:

$$x^{(0)} = f(x^{(1)}) \tag{4.4.2}$$

$$x^{(1)} = f(x^{(2)} + U_i) \tag{4.4.3}$$

$$x^{(2)} = f(x^{(1)})\ bottom - up \tag{4.4.4}$$

$$x^{(2)} = f(x^{(3)})\ top - down\ prediction \tag{4.4.5}$$

where $U_{(i)}$ is the action the networks takes to violate the selection process of sensory data

and could be put in this way, $U_i = max[x^{(2)}, x^{(3)}]$. Also, f is a continuous nonlinear function.

This generative model, as shown above, benefits from both bottom-up and top-down streams when it comes to calculate the values of x. The essentials of our problem require to take internal states the same as causal states(i.e. $x = \mu$).Now to launch the two important streams in visual search, namely, "what" and "where", we correspond a backward prediction error signal to each hierarchical level starting airing all over the network, as follow:

$$\varepsilon^i = x_i^1 - f(x_i^1, u) \tag{4.4.6}$$

and moreover the theory says a bit more and goes further. Given such backward connections(giving us synaptic gains) existed between the levels of the network, all x values could be backwardly derived due to these very prediction error signals.

$$x_{i-1}^1 = x_i^1 + \frac{\partial \varepsilon_i}{\partial x_i^1}\varepsilon \tag{4.4.7}$$

4.4.1 Energy Function for the Content Network

It's worth bearing in mind that we could expose the Energy function for the content network follows the free energy approach except a tiny difference. Since we take a reciprocal relation between content network and selection network, we write the content network Energy function so that it could encompass the relevant component coming form the selection network in addition according to the results, it would increase the network efficacy.

$$E^{SCN}(x_{ij}^{CN}, x_{kl}^{SN}) = \frac{b_{CN}}{2}\sum_{ij}(\sum_{kl} x_{ij}^{CN} - y_{kl}^{SN} - \sum_{kl} x_{kl}^{VF} y_{k+i,l+j}^{SN})^2 + \sum_{kl}(y_{SN}^{kl} - 1)^2 \tag{4.4.8}$$

It's conveniently traceable to find two highlighted streams in the above formula. The last term of the formula(i.e.$\sum_{kl}(y_{SN}^{kl}-1)^2$) which has expressed in terms of activation units in the selection network. As such, since our model is actually a statistic steady state model, neither noises nor hidden states have been taken into account. To flourish the equations we decided to write them entirely in terms of causal states and some parameters coming from the SAIM per se and hereafter no hidden states would appear in the equations. As we can see in the above equations, what comes to form the bottom-up control is the term $x_{kl}^{VF}y_{k+i,l+j}^{SN}$ which itself is a *convolution* operation; a prevalent function in vision and image processing to impose controls and filters on raw input data. Here, input data coming out from retinal areas are going to be convolved with the values of the selection units y^{SN} and then after mapped on the content network. On the contrary to the standard version of the SAIM and because of the occurrence of this convolution which adhere selection to data, we take the energy function such that it depends on both units of content network and selection network.

To run the program, we feed the network with two visual stimuli, namely, a two(2) and a cross(+), each of them are represented as square matrix of 7 dimension. These two matrices of zeros and ones are fed the network and taken up from visual field. The first level of the network that is the lowest level too, is in charge to receive data and deliver them to the level above to be processes.

Input data are gone up from the lowest level to be sent through the higher levels. In the knowledge network, both visual templates of '2' and '+' have priori been restored. At first and because of competitive underlying mechanisms , both of them try to lay down the FOA and henceforth, an overlapping figure include both of the template would appear right as it is shown in fig.16.

What is really happening here is a generative model comes into being and is prescribed as a neuronal message-passing scheme. Then two different endogenous and exogenous mechanism emerge; "namely, a lateral and top-down modulation of synaptic gain in principal cells that convey sensory information (prediction error) from one cortical level to the next." (Feldman & Friston , 2010) What our free energy-based model would reveal later on is that identification in the usual senses could be corresponded with perpetual inference as is articulated and paraphrased by free energy theory.

Overall, the predictive attribute of our model ensues a precise interaction that is meticulously done in this following way: whenever a stimulus presented, the work and content network begin to make prediction and send them all the way down to the level below(top-down stream) whereas appearance of stimuli suffices to induce the error prediction error which itself evokes the activation of causal states at the levels above.(bottom-up stream). Overall, our model like the standard version of SAIM is privileged to entail both top-down and bottom-up streams simultaneously except in the standard model of the SAIM it is the bottom-up stream which launches the process of identification whereas in the new version of the model we are presenting, this process is originated in the onset of top-down approach.

Energy Function for Knowledge Network

Here the energy function is derived from the original energy function we do have in the standard version of the SAIM.

$$E_{WTA}(y_i) = \frac{a}{2}.(\sum_i y_i - 1)^2 - b.\sum_i y_i.I_i \tag{4.4.9}$$

In fact, despite using a "softmax-function" is commonplace amongst the free energy based

models, here the equation (4.4.9) is applied instead. Of course they do have something important in common and that is both of them are to take the winner neuron regarding a set of input data. As we said, the prediction error

$$\varepsilon_m^{KN} = \sum_{ij}(x_{ij}^{CN} - y_m^{KN}.w_{ij}^l)^2 \tag{4.4.10}$$

dispatches the input data to the knowledge network where $y_m^{KN}.w_{ij}^l$ is top-down stream. Thus, the energy function for the knowledge network could be depicted as follow :

$$E^{KN}(y_m^{KN}, x_{ij}^{CN}) = \frac{a_{KN}}{2}\sum_l(y_l^{KN}-1)^2)^2 + b^{KN}.\sum_l(y_{KN}^l.(\sum_{ij}x_{ij}^{CN} - y_l^{KN}.w_{ij}^l)^2)) \tag{4.4.11}$$

We shall notice that since the error is sent up to the WTA, this WTA should be looser take all and therefore the sign of term $\sum_l$ would be positive.

Energy function for location map

like what we've done for the content network, again we use the energy function from the standard version of the SAIM(the looser take all function). Here again the prediction error

$$\varepsilon_{kl}^{LM} = (x_{kl}^{SN} - y_{kl}^{LM}) \tag{4.4.12}$$

puts the input data into the location map and the energy function for the location map could be written as follow:

$$E^{LM}(y_{kl}^{LM}, x_{kl}^{SN}) = \frac{a_{LM}}{2}\sum_{(}l(y_{kl}^{LM} - 1)^2)^2 + b^{LM}.\sum_l(y_{kl}^{LM}.(x_{kl}^{SN} - y_{kl}^{LM})) \tag{4.4.13}$$

Total energy function

The total energy function for the whole network would be :

$$E^{total}(y_{kl}^{LM}, x_{kl}^{SN}, x_{ij}^{CN}, y_{m}^{KN}) = E^{LM}(y_{kl}^{LM}, x_{kl}^{SN}) + E^{SCN}(x_{ij}^{CN}, x_{kl}^{SN}) + E^{KN}(y_{m}^{KN}, x_{ij}^{CN}) \tag{4.4.14}$$

Gradient Descent

To impose a gradient descent on the WTA-energy functions we've already obtained, we choose the Hopefield method. According to the nature of the energy function(a continuous, differentiable function which could pass the "second derivative test") , there are some minima points distributed across the certain values of y_i. Now a gradient descent method could be applied to find the minima points:

$$\dot{x}_i = -\frac{\partial E(y_i)}{\partial y_i} \tag{4.4.15}$$

In the Hopfield approach x_i, y_i are linked together by the sigmoid function:

$$y_i = \frac{1}{1 + e^{-m.(x_i - s)}} \tag{4.4.16}$$

By using the Euler-approximation the gradient descent would turn into:

$$x_i(t+1) = x_i(t) - \frac{\partial E(y_i)}{\partial y_i} \tag{4.4.17}$$

with regards to the two equations expressed above , the gradient descent is applied to a dynamic, neural-like network, where y_i could be liaised to the output activity of neurons,

x_i the internal activity and $E(y_i)\partial y_i$ gives us the input to the neurons. We by then, ensued the same approach to calculate gradient descent(a linear version of the hoffield version) for the other energy functions, that is

$$x_i(t+1) = x_i(t) - \frac{\partial E(x_i)}{\partial x_i} \tag{4.4.18}$$

We could write down the gradient descent formulas imposed on the energy functions as well.

$$\frac{\partial E^{total}}{\partial y_{kl}^{LM}} = a^{LM}.(\sum_{kl} y_{kl}^{LM} - 1) + b^{LM}.(x_{kl}^{SN} - y_{kl}^{LM}) \tag{4.4.19}$$

$$\frac{\partial E^{total}}{\partial x_{kl}^{SN}} = b^{CN}.\sum_{ij}(x_{ij}^{CN} - x_{kl}^{SN}.x_{k+i,l+j}^{VF}) + b^{LM}.y_{kl}^{LM} \tag{4.4.20}$$

$$\frac{\partial E^{total}}{\partial x_{ij}^{CN}} = b^{KN}.\sum_{ij}(y_l^{KN}.2.(x_{ij}^{CN} - y_l^{KN}.w_{ij}^l) + b^{CN}.(x_{ij}^{CN} - \sum_{kl} x_{kl}^{SN}.x_{k+i,l+j}^{VF}) \tag{4.4.21}$$

$$\frac{\partial E^{total}}{\partial y_l^{KN}} = a^{KN}.(\sum_l (y_l^{KN}-1)+b^{KN}.\sum_{ij}(x_{ij}^{CN}-y_l^{KN}.w_{ij}^l)^2+b^{KN}.y_l^{KN}.2-\sum_{ij}((x_{ij}^{CN}-y_l^{KN}.w_{ij}^l).w_{ij}^l) \tag{4.4.22}$$

Initial values To proceed the algorithm , we set out the following initial values: $y_{kl}^{LM}(0) = \frac{1}{N^2}, x_{kl}^{SN}(0) = \frac{1}{N^2}, x_{ij}^{CN} = \frac{1}{L}.\sum_l w_{ij}^l, y_m^{KN} = \frac{1}{L}$
where N is the size of input image and L is the number of templates.

4.5 Implement

Now everything is ready to have the program run. To do so, let's begin with setting the initial values. We shall remind it that these values have been obtained after a series of trial and errors to find the best initial points which make the program work as good as it gets. We have to notice that since the algorithm has assertively been written with regards

to competition, we render the program when both stimuli as illustrated in the following picture presented together. What follows would first consider the standard version of the SAIM and then our new model called the FR-SAIM.

Results and Discussion

As we could see, the figures following are divided in two parts and each part itself is divided in three counterparts. Here we come to make a comparison between the to version of our model, namely the standard original version of the SAIM(Heinke & Humphreys, 2003) and our Free Energy-based SAIM(FR-SAIM). To do it better, we decided to derive the results out from each network in a point-wise manner in that sense that the results of each network in the one version and whatever reveal, would be compared with its peer in the other version of the model. The following figures are to illustrate what is going on in the selection networks and the content networks of each version. Also, to depict the efficiency of each version, we feed the system with three different input data shall be processed accordingly: the first one , is the multiple incoming data consisted of a '2' and a '+' which are presented together, the second, is a single '+' and finally the third one would be a single '2'.

At the beginning, you can see a multiple scene including incoming input data consisted of a '2' and a '+' as illustrated in the fig. 4.3. Running the both of versions at t=1, give us back some results which are indicated in the fig.4.4 and fig.4.7. As the reader could realized from the caption, the first picture from the left shows the initial input, the middle shows the activity units in the selection network and the righteous shows the activity of units in the content network. From the beginning it is easily detectable that at t=1 the unit activities of the selection network in the FR-SAIM quietly differs with the unit

activities of the same network in the standard version of the SAIM and this difference obviously goes back to the different "metric norms" we have used in our vector space. For the standard version of the SAIM we applied "inner product" norm whilst in the FR-SAIM, we worked out our computations backed by "Euclidean norm". But as time passes by, they try somewhat to converge together and heading the similar results in spite of the different metrics we used and this occurrence could be mathematically proved mostly known as a theory called "strong convergence in Hilbert space " which could be easily found in most of the books on Real Analysis.(Rudin, 1986) For instance, in both firgures 4.8 and 4.9 the template '2' wins the competition and appears robustly at the content network as indicated experimentally in (Heinke & Humphreys, 2003).

The other important thing we have to definitely take into account is the reaction time each experiment shows. Fortunately all the reaction times have turned out in the way we expect from experimental results.(Wolfe, 1998) Whenever the FR-SAIM is fed with a multiple objects, the reaction time to catch the winner('2') is recorded $r_t = 201$ whilst to catch the '+' and '2' templates, the reaction times would be recorded $r_t = 191$ and $r_t = 195$ respectively.These time reaction recordings again affirm what (Wolfe, 1998) elaborates in their publication on how people come to catch the '2' faster than '+'. Albeit running the standard-SAIM gives us back different reaction time which are considerably slower than what we've already illustrated. In the standard-SAIM, r_t for multiple objects, single '+' and single '2' are 198, 169 and 160 all of them are comparatively slower than the FR-SAIM. Therefore, it is obvious that the latter model works faster as long as selecting the templates is concerned and also the reaction times are considerably shorter.

Now, we begin with the results derived from the standard version of the SAIM which is essentially undertaken based on a bottom-up process as we describe above. At the first epoch, these results were achieved in the way that is shown in the figure 4.4. We shall notice to the reaction time each version take to catch up the desired template and get to

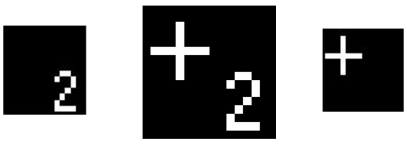

Figure 4.3: the three types of stimuli , namely a '2', a '+' and both '2' and '+' presentd together

fix on it.

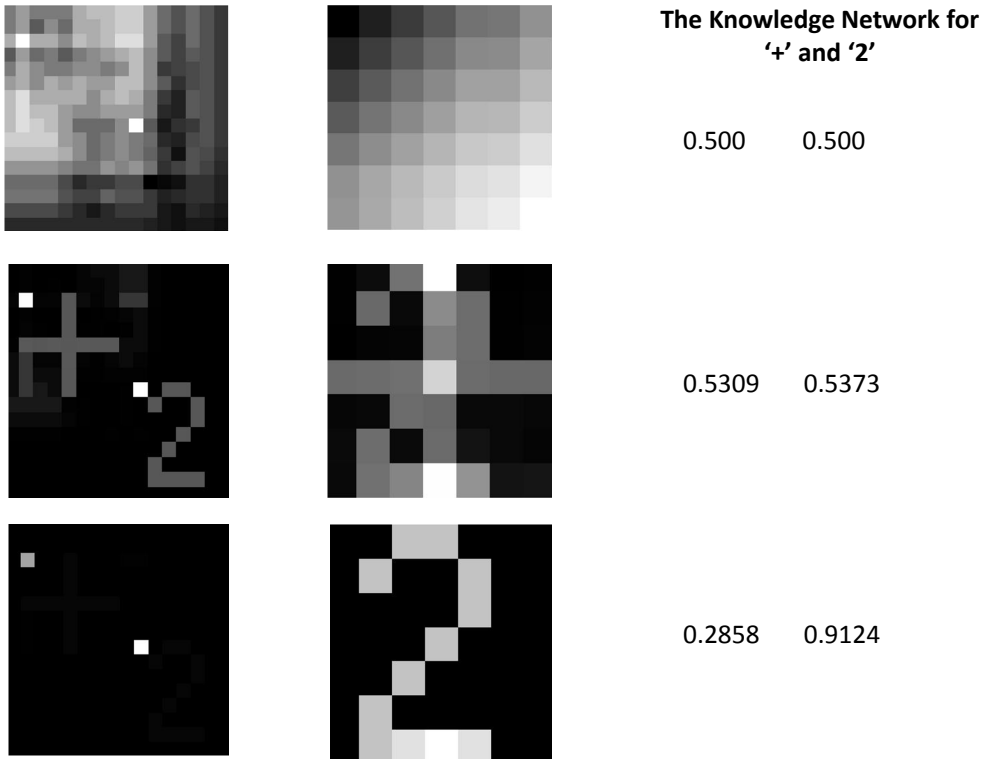

Figure 4.4: This picture shows the input multiple image(both '+' and '2' presented) fed to the standard-SAIM, also activities of the selection network activity and the content network at t=1, t=35 and t=169 at which the template '2' won the competition and has been identified and knowledge network reached to 0.9124

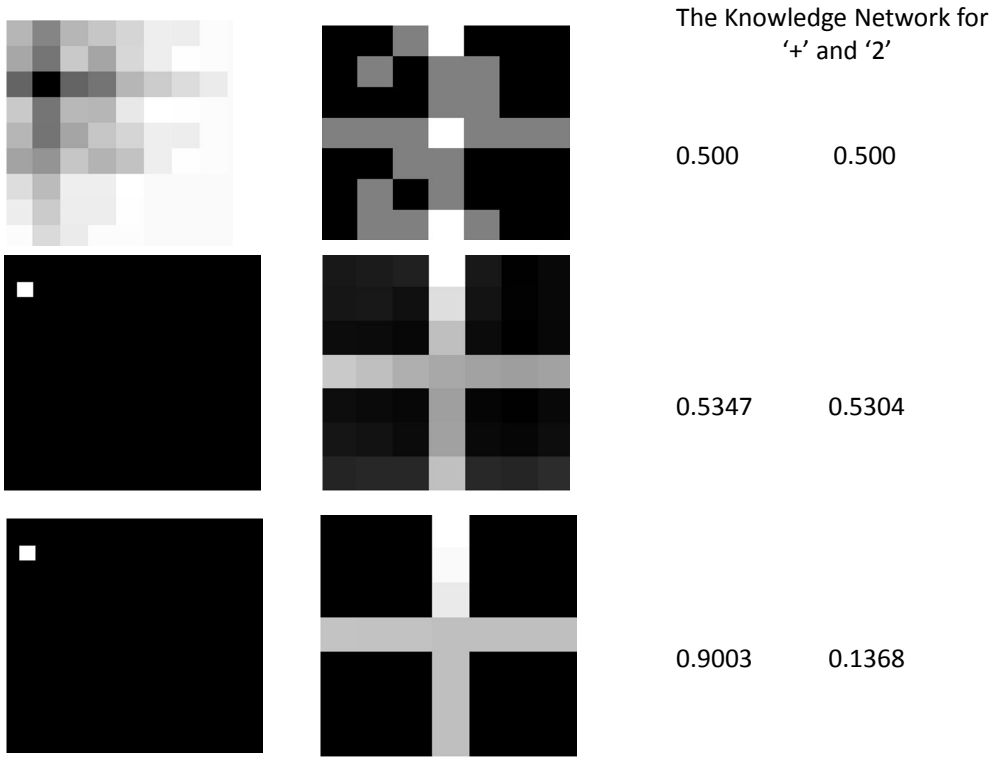

Figure 4.5: This picture shows that a single '+' fed to the standard-SAIM, also activities of the selection network activity and the content network at t=1, t=35 and t=160 at which the template '+' won the competition and has been identified and knowledge network reached to 0.9003

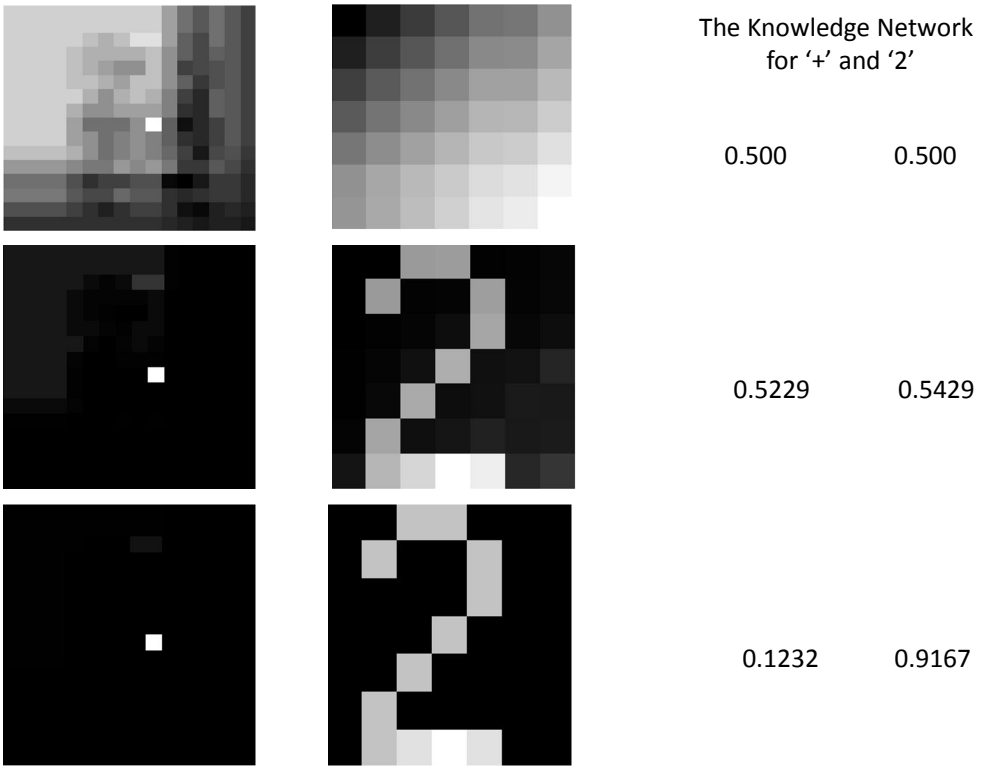

Figure 4.6: This picture shows that a single '2' fed to the standard-SAIM, also activities of the selection network activity and the content network at t=1, t=35 and t=198 at which the template '2' won the competition and has been identified and knowledge network reached to 0.9167

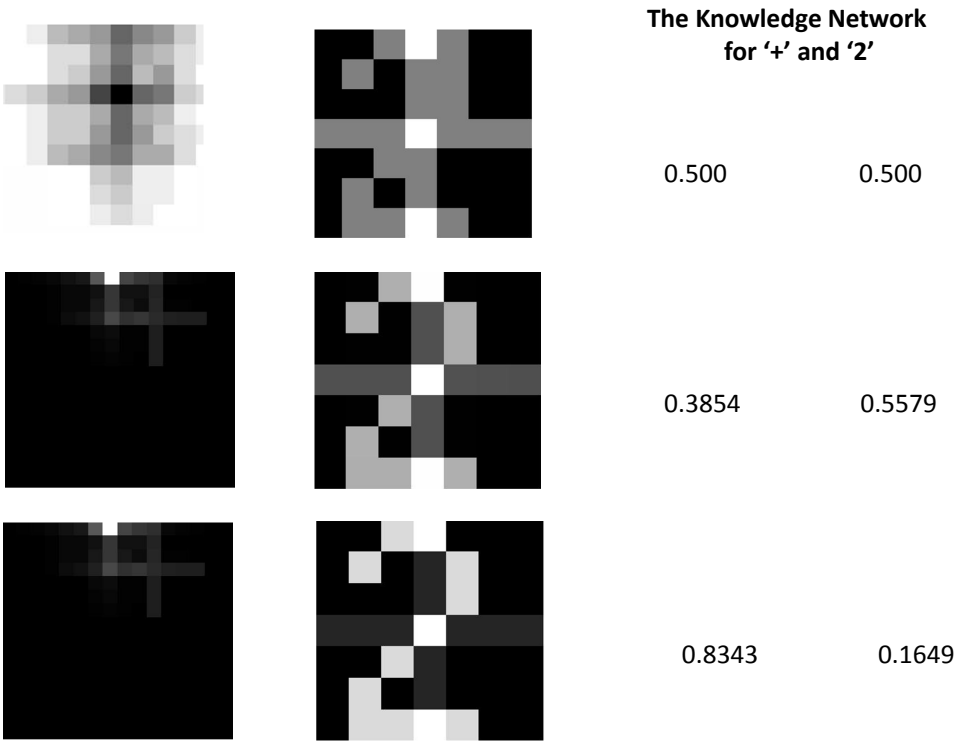

Figure 4.7: This picture shows the input multiple image(both '+' and '2' presented) fed to the FR-SAIM, also activities of the selection network activity and the content network at t=1, t=35 and t=201 at which the template '2' won the competition and has been identified and knowledge network reached to 0.8343

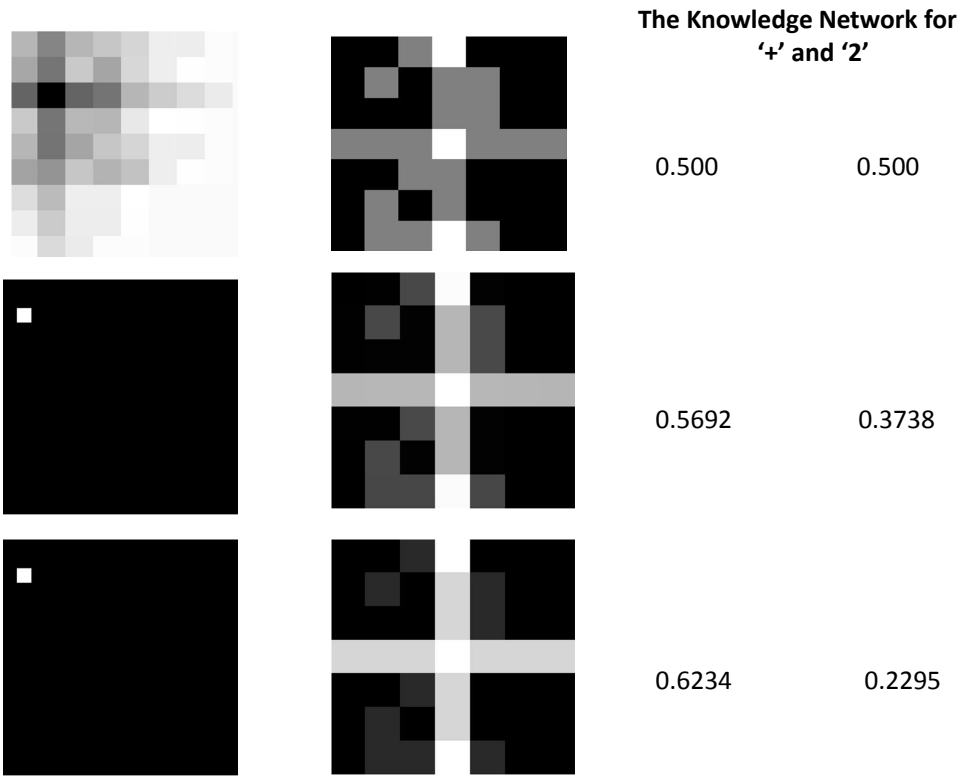

Figure 4.8: This picture shows that a single '+' fed to the FR-SAIM, also activities of the selection network activity and the content network at t=1, t=35 and t=91 at which the template '+' won the competition and has been identified and knowledge network reached to 0.6234

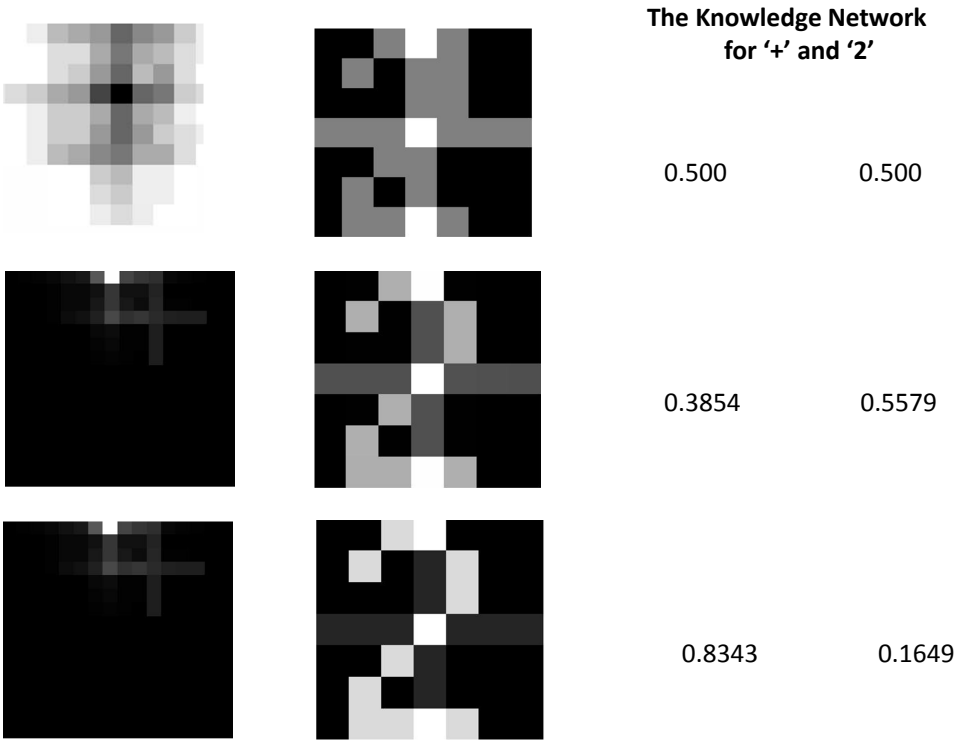

Figure 4.9: This picture shows that a single '2' fed to the FR-SAIM, also activities of the selection network activity and the content network at t=1, t=35 and t=195 at which the template '2' won the competition and has been identified and knowledge network reached to 0.8343

Chapter 5

Outlook

This study has offered a new version of the SAIM which approached a hard problem in cognitive neuroscience (combination of top-down and bottom-up approach to explain identification) that to my knowledge almost no model has ever succeeded in doing that so far. Although the structure of the SAIM (both the standard and FR-based versions) is consistent enough to run a spatial serial and temporary parallel process simultaneously, in this study we only applied the algorithm to single and multiple scene of objects. What might be hitherto considered as a significant progress on this case, will be designing an algorithm which could successfully execute the selection procedure and identification process amongst a multi-objects scattered scene in which the object may overlap or camouflaged.

Quite apart from the problem of identification in single and multi scenes explained above, a future study will embark a novel approach to a well-known problem that is "template learning". Since, we have so far shown how the template can be taken on , identified and recognized within a top-down free energy approach and so in a subsequent project we could develop a model which ends up with learning visual templates. These templates should be learnt throughout the hierarchy of model and stored in the knowledge network.

This problem has been highlighted by a few neuroscientists like (Brady & Kersten, 2003) and (Op Beeck & Baker, 2009) both of them present some evidences as to how humans learn novel objects. Of course many neuroscientists have so far faced this question of how visual objects are basically learnt and stored but it seems paying too much attention to its neural substrate and neglecting computational ideas consequently seems inadequate to resolve it. In general, those aspects which might have contributed to unravel the sophistication of the template learning, did not gain much weight in this current study.

Whenever we hear the word 'learning', an important issue could raise up: firstly, it is important to consider learning a contextual process not an obligatory one. Secondly, what type of learning we are talking about: "non-associative learning" or "associative learning". Determining that which category does engulf our desired learning process is the first step ought to be taken in order to going forward (Wood, 1988) According to the classical definitions which are comparatively agreed upon, it is held that any type of learning is dealing with "habituation" that is a learning response and could be gained via reiterating of a certain stimulus. Normally, this type of learning occurs in instinctive low level behaviours in animal and humans (Wood, 1988). On the other hand association is a vital cognitive capability which yields an association between two stimulus or behaviours attending together usually. The latter type of learning is perhaps one of the most important cognitive mechanisms having many things to do with complexity and neural network. This kind of learning in spite of the former type of learning , reflecting a response that is gained due to either simultaneous presenting(*classical conditioning*) or reward or punishment directing the learning process(*operant conditioning*). Nevertheless, most of the known learning mechanisms have heavily leant back to association whether classical(Pavlovian, Hebbian) or oprant(supervised learning) and further research needs to work details for these questions.

Here we'd better to say that when we attempt to embark this venture this study will revolve around many distinguished whilst inter-related fields like visual attention, information theory, learning, free energy , dynamical systems and etc. To do that we initially design a pilot study given the Kohonen's self organising feature map theory to see whether the model could identify the templates without any supervision and only by the virtue of SOM data classification. We initially came up with a hypothesis on how the standard version of the SAIM could be extended by a learning process. Some of the biological process to which the attentional mechanisms relied inspired us to try *unsupervised learning* algorithm (particularly Self Organising Map)to make it (Hinton et al., 1999). The implementation followed the standard model: A grid containing some randomly initialized nodes is trained with incoming input data, according to Kohonen's SOM algorithm. The shortest distant between the input and the nodes of the grid is computed (based on Euclidean norm) to find the nearest node to input data called best match unit(BMU). Then after, by a simple updating rule,$W_v(t+1) = W_v(t) + \theta(v,t)\alpha(t)(I(t) - W_v(t))$ the BMU neighbour nodes are drawn towards it, where θ is the lattice neighbourhood function between the grid nodes and input data and α is the learning rate. The aim of the learning algorithm was to classify the objects and then have the grid developed the most similar structure with input data. Having fed the network with the same stimulus we used throughout this study namely, '2' and '+' , we obtained this following results with regards to classification of input stimuli.

And as you may find in fig 5.2 a cross has been shaped in within the randomly initialized grid, but neither the classification nor learning the '+' template learning seem fairly successful.

Analogy making is another successful higher level theories backed primarily to the Marvin Minsky's works in artificial intelligence and then has been blossomed particularly

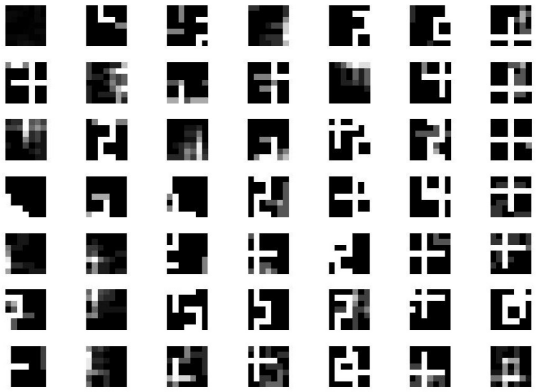

Figure 5.1: scattered data which should be classified after applying the SOM algorithm to the stimuli '2' and '+'

by Douglas Hofsteader and his pupils later on. Even Hofstadter went so fas as to say analogy in the core of human cognition (Hofstadter, 1996). Analogy making is basically "perception of two or more non-identical objects or situations as being the 'same' at some abstract level." (Mitchell, 2001) In other word, in accordance with AI terminology, analogy making is to find out how people can extract *classes* from *instances*.

Simply, Hofstader tried to explain that people could easily recognize the letter 'A' given a different class of shapes and handwriting styles of 'A' "because of some essential abstract similarity". so far, analogy making could be considered as one of the best well-established theories which would directly address the template identification problem and puts that in a better way to go forward. Although, it has apparently nothing to do with vision problem but has inspired many people working on visual perception. (Hofstadter, 1996)

To my knowledge, a few models have came out to being some knowledge about novel object learning. Amongst model, I can particularly refer to the two models which tried

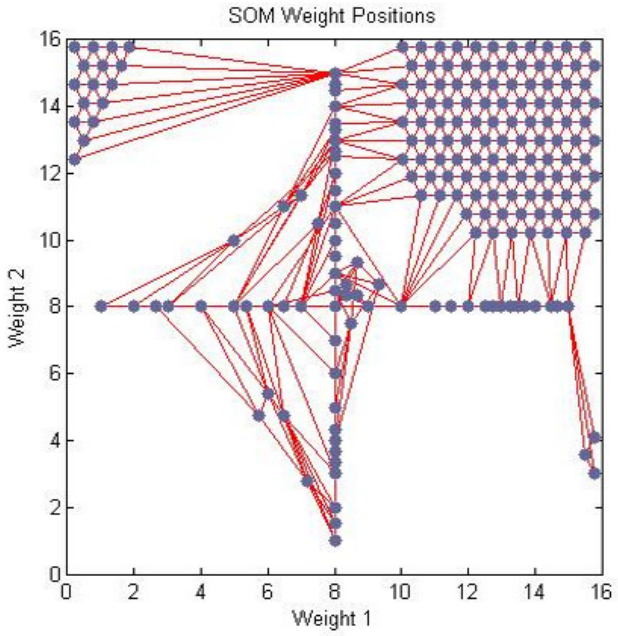

Figure 5.2: The neurons of the SOM grid trying to reshape the template '+' based on the best match unit algorithm.(after 500 epochs)

to resolve the problem by using two different approaches. (Saxena et al., 2006) offered a 'supervised learning' algorithm to which they exposed some novel objects to be learned and then claimed that this learning method named 'logistic regression algorithm' work out as such in an uncluttered visual scene. However, the model Sexana et al. offered is not neuro-biologically plausible. As we've seen, they grounded their model upon a supervised learning paradigm and derived the results through implementing an algorithm which definitely requires a 'feedback error signal'. Perhaps, it's bizarre that there is no such thing as a feedback control in visual system in that sense that we are dealing with in machine learning. Indeed there are some sort of feedback signals in visual system as we will take them into account later, but the point is that the brain operates in a more complex manner than it might look. Even though Sexana et al. themselves don't make any claim about validity of their theory when it comes to be applied in the human visual perception, but a few tried to follow the same supervised approach : For instance ,

Figure 5.3: Analogy making goes to figure out how people can recognize letters of the alphabet, e.g., A, in many different typefaces and handwriting styles. adopted from(Hofstadter, 1996)

instead of emphasizing on attentional resources as is prevalent among cognitive scientists, Dayan et al. "consider statistical and informational aspects of selective attention, divorced from resource constraints" by offering a Reinforcement learning based model which turns the problem into some sort of conditioning and learning one.(Dayan et al., 2000) They suggested an articulated form of TD(0) (Temporal Difference algorithm) along with a more sophisticated 'Rescolar Wagner' update rule.

$$\widehat{w}_i(t+1) = \widehat{w}_i(t) + \alpha_i(t)\delta(t); \quad \delta(t) = r(t) - x(t)\widehat{w}_i(t) \tag{5.0.1}$$

$$\alpha_j(t) = \frac{\sigma_i(t)x_i(t)}{\sum_j \sigma_j(t)x_j(t) + E} \tag{5.0.2}$$

(Brady & Kersten, 2003) also addressed somewhat the same problem but the way they went through amplifying that could not be considered as a learning algorithm we are

having in common sense . They truly point out that to recognize an object , the visual needs to be fed properly by enough knowledge about an object properties and what if the system is deprived of them?

To cope with this dilemma, they began to argue for a specific kind of learning process which might be engaged in and have nothing to do with usual methods like segmentation and decomposition. With regards to a camouflages object amid a scattered scene, their algorithm called 'bootstrapped learning' seems capable to meet the required task. Here, we would like to assert that generally speaking any kind of learning model should have some essential common properties none of the above models do comprise; namely, generalization, interaction, induction and adaptation. We hope in the near prospect we could accommodate template learning in the FR-SAIM based on the basic fact of free energy principle. Perhaps the change in connection strengths that minimise the same free energy that is used to optimise the activity. This usually reduces to some form of Hebbian plasticity that is formally related to back propagation of errors. The connection between the back propagation of errors and free energy minimisation is revealed (intuitively) in the predictive coding formulation of free energy minimisation.

List of References

Friston, K., The free-energy principle: a unified brain theory? , 2010,*Nature Reviews Neuroscience*, 1:1-12

Merikle, P.M. & Joordens, S., Parallels between perception without attention and perception without awareness, *consciousness*,6:219-236

Posner, M.L., Attention: The mechanisms of consciousness , 1994, *Proc Natl Acad Sci USA*, 91:7398-7403

Tononi, G., Laureys, S., *the neurology of consciousness*, 1st edition, Academic Press(2008)

Frintrop, S., Computational visual attention, 2011, *Computer Analysis of Human Behavior*, 69-101

Helmholtz, H.,*On the rate of transmission of the nerve impulse*, Veit & Comp., Berlin(1850)

Rumelhart, D. E., & McClelland, J. L., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*, Cambridge, MA: MIT Press(1986)

Cohen, J. D., & Huston, T. A., *Attention and Performance*, In C. Umilt & M. Moscovitch (Eds.), Cambridge, MA: MIT Press(1994),XV:453-476 Tononi, G., Laureys, S., *the neurology of consciousness*, 1st edition, Academic Press(2008)

Itti, L., Koch, C., Computational modeling of visual attention, 2001, *Nature Reviews Neuroscience*,2:1-11

Medeiro, F., & Verd, B.P.,& Vzquez, A.R. *top-town design of high-performance sigma-delta modulators*, 1st edition, Springer(2010)

Desimone, R., & Duncan, J., Neural mechanisms of selective visual attention, 1995, *Annu. Rev. Neurosci*,193222

Milner, A.D.,: Is visual processing in the dorsal stream accessible to consciousness?, 2012, *The Royal Society*

Ryu, G.G., & Suh, I.H., & Lee, S., Covert Visual Attention by Object-based Selective Visual Features and Their Saliency Map, 2009, *CSREA Press*, 170-173

Braun, J., Visual search among items of different salience: removal of visual attention mimics a lesion in extrastriate area V4., 2009, *J.Neurosci.*, 14: 554-567.

Koch, C. & Ullman, S., Shifts in selective visual attention: towards the underlying neural circuitry, 1985, *Hum. Neurobiol.*,4:219227

Treisman, A., & Gelade, G., A feature integration theory of attention., 1980, *Cognitive Psychology.*,12:97 136

Frintrop, S., & Rome, E.,& I. Christensen, H.I, Computational visual attention systems and their cognitive foundations: A survey 2010, *ACM Transactions on Applied Perception(TAP)*,7(1)

Bisley, J. & Goldber, M., Neuronal activity in the lateral intraparietal area and spatial attention., 2003, *Science 299*,5603:8186

Treisman, A.,& Gelade, G. A feature integration theory of attention, Cognitive Psychology, 2003, *Cognitive Psychology*,12:97-136

Koch, C., Selective Visual Attention and Computational Models, 2000, *CNS/Bi 186:Attention*

Hofstadter, D.R., *Metamagical Themas: Questing for the Essence of Mind and Pattern*, new edition, Basic Books(1996)

Mitchell, M.,: Analogy-Making as a Complex AdaptiveSystem in: *Design Principles for the Immune System and Other Distributed Autonomous Systems*, ed. Lee A. Segel, I. R. C., New York: oxford university press(2001),335-359

Posner, M.I., & Nissen, M.J., & and Ogden, W.C., Attended and unattended processing modes: the role of set for spatial location in *Modes of Perceiving and Processing Information*, 1978,*eds H. L Pick and N. J. Saltzman (Hillsdale, NJ: Lawrence Erlbaum Associates)*

Gonzalez, R.C., & E. Woods, R.E., *Digital Image Processing*, 2nd edition, Penguin(2000)

Hofstadter, D.R., *Godel, Escher, Bach: An Eternal Golden Braid*, 20th Anniversary edition, Penguin(2000)

Heinke, D.G., & Backhaus, A.,Modelling Visual Search with the Selective Attention for Identification Model (VS-SAIM): A Novel Explanation for Visual Search Asymmetries, 2011, *Cognitive Computation*,1:185-205

Heinke, D., & Backhaus, A., & Sun, Y.R., & Humphreys, G.W., The Selective Attention for Identification model SAIM): simulating visual search in natural colour images., 2008, *Paletta L, Rome E (eds) Attention in cognitive systems, lecture notes in computer science* ,4840:141-54

Heinke, D. & Humphreys, G. W.,Computational Models of Visual Selective Attention: A Review, 2003, *Houghton, G., editor, Connectionist Models in Psychology*

Rutishauser, R. , & Walther, D., & Koch, C., & Pietro Perona, P., Is bottom-up attention useful for object recognition?, 2004, *IEEE Conference on Computer Vision and Pattern Recognition*

Dennett, D. C., *consciousness explained*, Little, Brown, 1992 Review by Glenn Branch on Jul 5th 1999 Volume: 3, Penguin(1991)

Blackmore, S., *Consciousness: A Very Short Introduction*, Oxford Umniversity Press(2005), 3:56-57

Mozzer, M., & Sitton, S., in *attention*, (ed. Pashler, H.), University College London(1996)

Mozzer, M., & Sitton, S., in *attention*, (ed. Pashler, H.), University College London(1996)

Gibbson, J. J., *the ecological approach to visual perception*, new edition, Psychology Press(1986)

Sloman, A., Whats vision for, and how does it work?, http://www.cs.bham.ac.uk/research/projects/cogaff/talks/sloman-beyond-gibson.pdf, 2011, Birmingham Vision club

Schill, K. , & Umkeher, E., & Beinlich, S., & Kerieger, G., & Zetzsche, C., Scene analysis with saccadic eye movement:top-down and bottom-up modelling, 2001, *J. Electronic Imaging*

Ullman, S., *high-level vision*, 1st edition, MIT Press(1996)

Saxena, A., & Driemeyer, J., & Kearns, J., & Osondu, C., & Ng, A. Y., Learning to grasp novel objects using vision , 2006, *10th International Symposium of Experimental Robotics (ISER)*

Brady, M. J., & Kersten, D.,: Bootstrapped learning of novel objects , 2003, *Journal of Vision*,3:413-422

Dayan, P., & Kakade, S., & Montague, P. R., :Learning and selective attention , 2000, *nature america*,http://www.nature.com/neurosci/index.html

Eysenck, M. W., & Keane, M. T., *cognitive psychology*, 3rd edition, Psychology Press, UK(1996)

Humphreys, G. W., & Bruce, V., :*visual cognition: Computational experimental and neuro-psychological perspectives* , Hove, Lawrence Erlbaum Associates Ltd, UK(1989)

Poggio, T.,: Marr's Approach to Vision, 1981, *AIM*,645

Poggio, T.,: Marr's Approach to Vision, 1981, *AIM*,645

Humphreys, G. W., & Riddoch, M., J., How to define an object: Evidence from the effects of action on perception and attention, 2007, *Mind and Language*, 22 (5):534547

Marr, D.,: *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, Freeman, New York(1982)

Biederman, I.,: *Visual object recognition*, MIT press, Boston(1995)

Pinker, S., & Mehler, S., : *Connections and Symbols*, MIT press, Cambridge(1988)

Blank, D. S., : *Learning to See Analogies: A Connectionist Exploration*, PhD Thesis, Indiana University, Bloomington(1997)

Berkeley,. I. S. N., *Some Myths of Connectionism*, http://www.ucs.louisiana.edu/ isb9112/dept/phil341/myths/myths.html, The University of Southwestern Louisiana(1997)

Nagel, E., & Newman, J. R., & Hofstadter, D. R., : *Godel's Proof*, NYU Press, Revised edition (October 1, 2001)

Block, N., The Mind as the Software of the Brain., 1995, *An Invitation to Cognitive Science. MIT Press*

Pinker, S., & and Mehler, J. , On Language and Connectionism: Analysis of a Parallel Distributed Processing Model of Language Acquisition, 1988 : *cognition*

Zisper, D., & Anderson, R. A., A back-propagatio programmed network that simulates response properties of a subtle of posterior parietal neurons, 1998, *Nature*,331, 679-684

Friston, K. J., & E. Stephan, K. E.,Free-energy and the brain, 2007, *Synthese*, 159:417458

Cannon, W. B., Organization For Physiological Homeostasis. , 1929,*Physiol Rev.*, 9: 399-431

Reza, F. M., : *An introduction to information theory*, McGraw-Hill, New York(1961)

Dimitrov, A. G., & Miller, J. P., Neural coding and decoding: Communication channels and quantization. , 1929,*Network: Computation in Neural Systems*, 12:441472.

Friston, K. J.,The free-energy principle: a rough guide to the brain?, 2009, *Trends in Cognitive Sciences*, 13:293-301

Friston, K. J.,Policies and Priors , in *Computational Neuroscience of Drug Addiction*, Springer Series in Computational Neuroscience, 10, 3:237-283, 2012

Feldman, H., & Friston, K. J., Attention, uncertainty, and free-energy., 2012, *Hum Neurosci.* , 4:215

Mathys, C., & Jean, D. J., & Friston K. J., & Stephan K. E., A Bayesian foundation for individual learning under uncertainty. , 2011, *Hum Neurosci.* , 5:35 DOI: 10.3389/fnhum.2011.00039

Chumbley, J. R., & Dolan, R. J., & Friston K. J., Attractor models of working memory and their modulation by reward., 2008, *Biol Cybern.* , 98(1):11-8

Kilner, J. M., & Friston K. J., & Frith, C. D., The mirror-neuron system: a Bayesian perspective., 2007, *Neuroreport.* , 18(6):619-23.

David, O., & Harrison, L., & Friston K. J., Modelling event-related responses in the brain. , 2005, *NeuroImage.* , 25(3):756-70.

Friston, K. J., & Kiebel S. J., Predictive coding under the free-energy principle., 2009, *Phil. Trans. R. Soc. B.* , 364:1211-122

Olshausen, B. A., & Anderson, C. H., & Van Essen, D. C. A Neurobiological Model of Visual Attention and Invariant Pattern Recognition Based on Dynamic Routing of Information., 1993, *J. of Neuroscience* , 13(11):47004719.

Heinke, D. G., & Humphreys, W. G. Attention, spatial representation, and visual neglect: simulating emergent attention and spatial memory in the selective attention for identification model (SAIM)., 2003, *Psychol Rev.* , 110(1):29-87.

Op Beeck, H. P. & Baker, C. I., The neural basis of visual object learning., 2009, *Trends in Cognitive Sciences*, 14,1: 22-30

Wood, D.C., Habituation in Stentor produced by mechanoreceptor channel modification., 1988, *Journal of Neuroscience*, 8: 2254

Hinton, G., & Sejnowski, T. J., *Unsupervised Learning: Foundations of Neural Computation.*, MIT Press(1999)

Rudin, W., *Real and Complex Analysis.*, McGraw-Hill Science/Engineering/Math, 3rd edition (1986)

Wolfe, W., *Visual Search: A Review.*, Psychology Press, pp.13-74, (1998) VisualSearchAReview.pdf:/home/axb388/DEVELOP/usr/axb388/BibTexLibrary/ePaper/[201-300]/[221]Wolfe1998-VisualSearchAReview.pdf:PDF

Kuhn, T. S., *The Structure of Scientific Revolutions.*, University of Chicago Press(1996)

Kohonen, T., The self-organizing map., 1990, *IEEE*, 78(9):1464-1480.

Trappenberg, T., *Fundamentals of Computational Neuroscience.*, OUP Oxford(2009)

Leslie Lamport: LaTeX, *A document preparation system*, 2nd edition, Addison-Wesley (Reading, Massachusetts, 1994).

Wettig, T., & Brown, G.E., The evolution of relativistic binary pulsars, 1996, NewA, 1, 17-34.

Elson, R.A.W., Santiago, B.X., & Gilmore, G.F., Halo stars, starbursts, and distant globular clusters: A survey of unresolved objects in the Hubble Deep Field, 1996, NewA, 1, 1-16.

Governato, F., Moore, B., Cen, R., Stadel, J., Lake, G., & Quinn, T., The Local Group as a test of cosmological models, 1997, NewA 2, 91-106.