Study of Information Processing Systems Based on Neural Network of Visual Cortex For analysis of Astronomical and Scientific data

ナノシステム研究科 孫哲（Sun Zhe）

指導教員　Ruggero Micheletto

1.General Background Information

With the recent convergence of multi-disciplinary sciences and the necessity of big data processing, the study of the Brain is playing an increasingly important role.

Human beings are borne to have the ability to recognize things around, and this system is working every second to help us deal with things we encounter in daily life. In fact, all Lens and retina are doing is almost like a CCD of non-professional cameras. However, so far, human being still can't even design one camera that can match with this Object Recognition ability, this is due to the fact that we are still far from grasp all the details of the brain functioning. Although we do have gain lots of improvement and achievement in General Object Recognition techniques development, we haven't yet make any breakthrough. Therefore, it have significant meaning to improve our understanding of General Object Recognition by looking into visual stream, whatever it is from theoretical or practical aspect.

For this purpose in this study, we use the computational model of object recognition in the brain cortex (HMAX) to distinguish objects. In particular, we have developed a system able to recognize astronomical object in an astronomical database of data. The HMAX model was proposed by T. Poggio in MIT, it is a feed-forward hierarchical model and extend the classical simple-to-complex cells model. Using bottom up process can speed calculation and promote the general recognition of patterns in a complex database. And in this study, we used support vector machine (SVM) as classifiers. The Sloan Digital Sky Survey (SDSS) is exploited to test our approach with astronomical data.

2.Research Method

2.1Ventral Stream and Hmax model

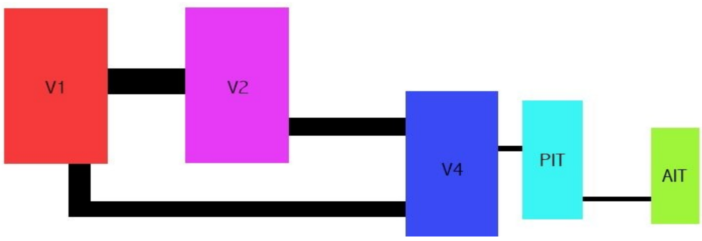

Figure 1. The ventral steam for pattern recognition in our brain

As we know, the visual information work-flow in our brain start from V1, V2, V4 to AIT. Hmax model is a mathematical representation of information processing in the Brain. It contains S1 Layer, C1 layer, S2 Layer, C2 Layer. The S1, S2 layer work like the simple cells and C1, C2 layer was used as the complex cells in visual stream.

2.2 S1 Layer and C1 Layer

We have an initial input image of 256x256 pixels, representing astronomical data, as for example Figure 1:

Figure 2. Images from SDSS database

The <u>S1 Layer</u> is applied on 12 resized versions of this image at different scales (256x256, 214x214, 180x180 down to 38x38 pixels). In higher mammals brain, images are filtered along orientations. A Gabor filter is used to reproduce this functionality. In our HMAX model each scale the Gabor filter is applied at 12 different orientations, for a total of 12x12 set of output data (12 scales by 12 orientations).

The Gabor filters are described by:

$$G(x,y)=\exp\left(-\frac{(X+\gamma^2 Y^2)}{2\sigma^2}\right)\cos\left(\frac{2\pi}{\lambda}\right)$$

Where $X=\cos\theta-y\sin\theta$ and $Y=\sin\theta+y\cos\theta$ .We adjusted the same filter parameters for all of the scales,orientation θ ,and wavelength λ .

After the S1 layer is processed, the 12x12 group of data are received by the C1 layer. In C1, a major rescaling is performed as happens in the hierarchical information flow in the mammal brain. In our algorithm each resized group is compared with the adjacent smallest sized group of data (for example 256x256 compared with 214x214). The comparison is made with normalized XY coordinates on a 3x3 grid. The biggest data value is selected and a single data point is generated. All the image is scanned and a new smaller group of data is generated (for example, 47x47). There are 12 orientations for each group size, so 12 of these resized groups are generated this way. The process continues with all the other sizes (214x214 compared with 180x180 and so on) until a new set of smaller data groups are created. The groups are these sizes (47x47, 39x39, 33x33 down to 5x5). Finally, because of this comparison in couples, one of the sizes is lost and after C1 operations, we have 11 sizes , and 12 orientation, for a total of 11x12=132 groups. This elaboration is done to remove noise in scale-independent pattern recognition processes observed in mammal.

2.1. Layers S2 and C2

In <u>S2 layer</u>, we selected 4075 patches of data. The size of these patches is 4x4x12 or 8x8x12 or 16x16x12 chosen at random and placed at random position. The 12 appears because we have 12 orientations. Because of normalization of coordinates, these patches have the same relative area on each scale. A Gaussian filter of this kind is applied:

$$R(X_i, P_j) = \exp\left(-\frac{(X_i - P_j)^2}{2\sigma^2\alpha}\right)$$

where X_i is a three-dimensional vector representing the data and P_j is the three dimensional patch.

The output of this layer is a new group of data of size AxAx4075, where A is the original size of the group. The 4075 points are the maximum value of the Gaussian filtering above. After this stage we have 11 group of data, each of them AxAx4075 in size, where A is the original group size (47x47, 39x39, 33x33 down to 5x5 as above).

The C2 layer:

These 11 groups are fed to the C2 layer that in the same fashion as C1 layer, it compare different patches on a 3x3 grid. This time the comparison is not made on two adjacent sizes, but is done on 6 of them. For example the group A=47 is compared with A=39, A=33 and so on. A single maximum value is determined. This process is repeated another group of 6 different sizes. At the end two group of data of size A=39 and A=21 is generated. Each group of data has size AxAx4075.

Now spatial dependence is removed. For each AxA sheet data in each group, the maximum value is taken. So each AxAx4075 group is reduced to a one-dimensional vector of size 4075. We do that on the two groups and join the results. So the output of C1 is a single one-dimensional vector of size 8150. This final vector represent the generalization of the original image features, and we call it from now on “the features” embedded in the image.

2.2. Support Vector Machine For Classification

Before the classification, the data should be normalized as preprocessing.

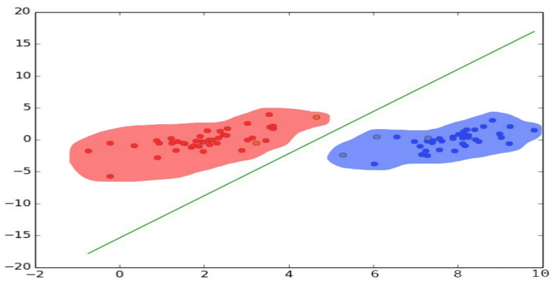

Figure 3. An example for classification

SVM is a classifier for supervised learning which build classification model by analyzing training data. In our case, we used SVM with RBF (Radial Basis function) kernel Model for the classification. We used an external python library called scikit-learn to implement SVM. Scikit-learn harnesses this rich environment to provide state of the art implementations of many well known machine learning algorithms. And it maintains an easy to use interface tightly integrated with python language.

3.Result and Discussion

3.1 The SDSS database

We used the Solar Digital Sky Survey（SDSS）Data Release 8 as our dataset. In this release, all of the imaging data was taken by the SDSS imaging camera which contains totally over 14000 square degree of sky.

In this study , we prepared the dataset with about 500 stars and 500 galaxies for sample training. And we used about 1000 images of objects for prediction.

3.2 Result and Prospect

In this research, we have used Hmax Model to carry Astronomical data and classification. The identifiability have been improved above 90% and in the process, we have simulated the information delivering method of biological brain and achieved effective features. However, in terms of mechanical learning, we have applied traditional SVM. We hope we are able to obtain the category algorithm of brain by means of further studies of brain.

Reference

1.Y. D. Tian, L. Q. Zhang.Relevant Algorithms for General Object Recognition Based on Feature Combination.Shanghai Jiao Tong University Journal,2007.

2.J. Mutch and D. G. Lowe.Object class recognition and localization using sparse features with limited receptive fields. International Journal of Computer Vision. Vol. 56

3.T. Serre, L. Wolf, and T. Poggio. Object recognition with features inspired by visual cortex. In CVPR, San Diego, June 2005.

4.C. Elting, C. A. L. Bailer-Jones, K. W. Smith. Photometric Classification of Stars, Galaxies and Quasars in the Sloan Digital Sky Survey DR6 Using Support Vector Machines. American Institute of Physics Conference Series. Vol. 1082. pp. 9-14 , 2008

研究概要

宇宙電波望遠鏡の観測データは膨大なため、従来の研究には、単純な数学モデルを利用しているものの、認識の精度は低く、研究者の経験に依存する部分が大きい。これを改善するために、博士後期には、「視覚野神経回路網モデルを用いた情報処理システムの研究」というテーマで研究を展開したい。人間の視覚野神経回路網情報処理モデルと機械学習技術を利用して、観測データを自動的にパターン認識すると分析できるシステムを開発する。「視覚野神経回路網モデル」（HMAX モデル）とは、生理学的実験により判明した、大脳視覚野における情報処理方法を利用し、データの特徴抽出を行うモデルである。具体的には、Ｖ１野とＶ２野の特徴抽出細胞および V4 野を経て IT 野に至るという階層的な処理である。

Study of Information Processing Systems Based on Neural Network of Visual Cortex For Analysis of Astronomical and Scientific Data

平成 26 年度　修士論文

横浜市立大学大学院

生命ナノシステム研究科

ナノシステム科学　専攻

Sun Zhe

Contents

Abstract · 5
1. Introduction · 10
1.1. Research Background · 10
1.2 Research purpose · 11
1.3 Research Method · 11
2.Neural Networks and Artificial Neural Networks · · · · · · · · · · · · · · 13
2.1 Intelligence and machine learning · 13
2.2 The Visual Cortex Structure and Pattern Recognition · · · · · · · · · · 16
3.Standard HMAX model · 22
3.1 General methods(HOG, SIFT) for feature descriptors in computer vision · · 20
3.2 Standard HMAX model · 21
3.3 Support Vector Machine · 26
4.System for astronomical data analysis · · · · · · · · · · · · · · · · · · · 29
4.1 SDSS database · 29
4.2 Astronomical data Features · 30
4.2.1 Initial processing · 30
4.2.2 Process in S1 Layer · 31
4.2.3 Process in C1 Layer · 32
4.2.4 Process in S2 Layer · 33
4.2.5 Process in C2 Layer · 34
4.3 Data Training And Prediction · 34

5. Dataset Prediction · 37
5.1 Test result of Prediction · 37
5.2 Conclusion and Prospect · 37
5.3 Research In the feature · 38
Reference · 40
Japanese Abstract · 42
Acknowledgements · 43
Appendices · 44

Study of Information Processing Systems Based on Neural Network of Visual Cortex For Analysis of Astronomical and Scientific Data

Sun Zhe

主指導員　Ruggero Micheletto

副指導員　木下　郁雄

副指導員　大月 俊也

副指導員　Robert Kanaly

Study of Information Processing Systems Based on Neural Network of Visual Cortex For analysis of Astronomical and Scientific data

ナノシステム研究科 孫哲（Sun Zhe）

指導教員　Ruggero Micheletto

1.General Background Information

With the recent convergence of multi-disciplinary sciences and the necessity of big data processing, the study of the Brain is playing an increasingly important role.

Human beings are borne to have the ability to recognize things around, and this system is working every second to help us deal with things we encounter in daily life. In fact, eye lens and retina work almost like a CCD of non-professional cameras. However, so far, human being still can't even design one camera that can match with the eye and brain object recognition ability, this is due to the fact that we are still far from grasping all the details of the brain functioning. Although we do have done of improvement and achievement in General Object Recognition techniques development, we haven't yet make any breakthrough. Therefore, it have significant meaning to improve our understanding of General Object Recognition by looking into visual stream, whatever it is from theoretical or practical aspect.

For this purpose in this study, we use our original version of the computational model of object recognition in the brain cortex (HMAX) to distinguish objects. In particular, we have developed a system able to recognize astronomical object in an astronomical database of data. The HMAX model was proposed by T. Poggio in MIT, it is a feed-forward hierarchical model and extend the classical simple-to-complex cells model. Using bottom up process can speed calculation and promote the general recognition of patterns in a complex database. And in this study, we used support vector machine (SVM) as classifiers. The Sloan Digital Sky Survey (SDSS) is exploited to test our approach with astronomical data.

2.Research Method

2.1Ventral Stream and Hmax model

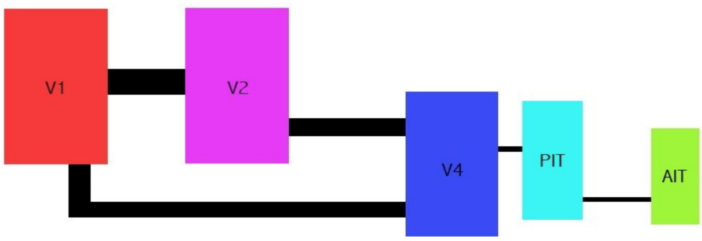

Figure 1. The ventral steam for pattern recognition in our brain

As we know, the visual information work-flow in our brain start from V1, V2, V4 to AIT. Hmax model is a mathematical representation of information processing in the Brain. It contains S1 Layer, C1 layer, S2 Layer, C2 Layer. The S1, S2 layer work like the simple cells and C1, C2 layer was used as the complex cells in visual stream.

2.2 S1 Layer and C1 Layer

We have an initial input image of 256x256 pixels, representing astronomical data, as for example Figure 1:

Figure 2. Images from SDSS database

The <u>S1 Layer</u> is applied on 12 resized versions of this image at different scales (256x256, 214x214, 180x180 down to 38x38 pixels). In higher mammals brain, images are filtered along orientations. In the same fashion, a Gabor filter is used to reproduce this functionality. In our improved HMAX model, each scale the Gabor filter is applied at 12 different orientations, for a total of 12x12 set of output data (12 scales by 12 orientations).

The Gabor filters are described by:

$$G(x, y) = \exp\left(-\frac{(X + \gamma^2 Y^2)}{2\sigma^2}\right)\cos\left(\frac{2\pi}{\lambda}\right) \qquad (1)$$

Where θ is the orientation, $X = \cos\theta - y\sin\theta$ and $Y = \sin\theta + y\cos\theta$. We used the same filter parameters for all of the scales, orientations θ, and wavelength λ.

<u>The C1 layer</u> :

After the S1 layer is processed, the 12x12 group of data are received by the C1 layer. In C1, a major rescaling is performed as happens in the hierarchical information flow in the mammal brain. In our algorithm each resized group is compared with the adjacent smallest sized group of data (for example 256x256 compared with 214x214). The comparison is made with normalized XY coordinates on a 3x3 grid. The biggest data value is selected and a single data point is generated. All the image is scanned and a new smaller group of data is generated. There are 12 orientations for each group size, so 12 of these resized groups are generated this way. The process continues with all the other sizes (214x214 compared with 180x180 and so on) until a new set of smaller data groups are created. Finally, because of this comparison in couples, one of the sizes is lost and after C1 operations, we have 11 sizes, and 12 orientation, for a total of 11x12=132 groups. This elaboration is done to remove noise in scale-independent pattern recognition processes, as observed in mammal.

2.3 Layers S2 and C2

In S2 layer, a maximization process is done along all the 12 orientations. We selected 4075 random patches of data for each group. The size of these patches is 4x4x12 or 8x8x12 or 16x16x12 chosen at random and placed at random positions. The 12 appears because we have 12 orientations. Because of normalization of coordinates, these patches have the same relative area on each scale. A Gaussian filter of this kind is applied:

$$R(X_i, P_j) = \exp\left(-\frac{(X_i - P_j)^2}{2\sigma^2\alpha}\right) \tag{2}$$

where X_i is a three-dimensional vector representing the data and P_j is the three dimensional patch.

The output of this layer is a new group of data of size $A \times A \times 4075$, where A is the original size of the group. The 4075 points are the maximum value of the Gaussian filtering above. After this stage we have 11 group of data, each of them $A \times A \times 4075$ in size, where A is the original group size (214x214, 180x180 … 47x47, 39x39, 33x33 down to 5x5 as in table 1).

Input Layer	S1 Layer	C1 Layer	S2 Layer	C2 Layer
256×256	246×246	47×47	44×44	1×1
214×214	204×204	39×39	36×36	
180×180	170×170	33×33	30×30	
152×152	142×142	27×27	24×24	
128×128	118×118	21×21	18×18	
106×106	96×96	17×17	14×14	
90×90	80×80	15×15	12×12	1×1

Table 1

The C2 layer:

These 11 groups are fed to the C2 layer that in the same fashion as C1 layer, it compare different patches on a 3x3 grid. This time the comparison is not made on two adjacent sizes, but is done on 6 of them. For example the patches of group A=47 are compared with A=39, then A=33 and so on. A single maximum value is determined. This process is repeated another group of 6 different sizes. At the end two group of data of size A=39 and A=21 are generated. Each group of data has size $A \times A \times 4075$.

Now spatial dependence is removed. For each AxA sheet data in each group, the maximum value is taken. So each $A \times A \times 4075$ group is reduced to a one-dimensional vector of size 4075. We do that on the two groups and join the results. So the output of C1 is a single one-dimensional vector of size 8150. This final vector represent the generalization of the original image features, and we call it from now on “the features” embedded in the image.

2.4 Support Vector Machine For Classification

Before the classification, the data should be preprocessed and normalized.

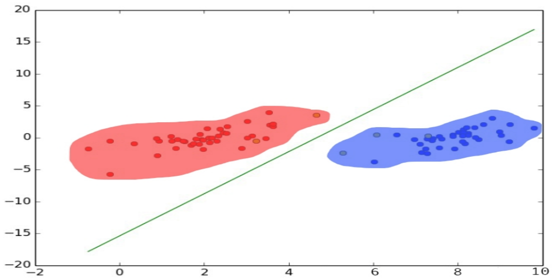

Figure 3. An example for classification

SVM is a classifier for supervised learning which build classification model by analyzing training data. In our case, we used SVM with RBF (Radial Basis function) kernel Model for the classification. We used an external python library called scikit-learn to implement SVM.

3.Result and Discussion

3.1 The SDSS database

We used the Solar Digital Sky Survey（SDSS）Data Release 8 as our dataset. In this release, all of the imaging data was taken by the SDSS imaging camera which contains totally over 14000 square degree of sky.

In this study , we prepared the dataset with about 200 stars and 200 galaxies for sample training. And we used about 1000 images of objects for prediction.

3.2 Result and Prospect

In this research, we have used Hmax Model to carry Astronomical data and classification. The accuracy of detection have been improved to 80%, and in the process, we have simulated the information delivering method of biological brain and achieved effective features. However, in terms of mechanical learning, we have applied traditional SVM. We hope we are able to obtain the category algorithm of brain by means of further studies of brain.

研究概要（日本語）

視覚野ニューロンネットワーク情報処理に基づいた、天文データの識別システムの研究　研究概要

人間の視覚野は、目から得たデータを処理し、学習と記憶に基づき、パターン認識をしている。脳の働き方を模倣した、学習機能を持つアルゴリズムは、一般的なデータパターン（実験データ等）の認識率が高い。従来の研究では、輝度の勾配方向と勾配強度を利用しているが、認識の精度は低い。これは、宇宙電波望遠鏡の観測データはノイズが多く複雑かつ、情報量が膨大なためである。これを改善するために、人間の視覚野神経回路網情報処理モデルと機械学習技術を利用して、観測データを精度良くパターン認識できるシステムを開発する。本研究では、MIT の Poggio 教授によって発明された HMAX モデルを用いた。HMAX モデルは、生理学的実験に基づいた大脳視覚野の情報処理方法を利用し、データの特徴抽出を行うモデルである。具体的には、Ｖ１野とＶ２野の特徴抽出細胞および V4 野を経て IT 野に至る階層的な処理である。このモデルを利用し、視覚野単純型細胞と複雑型細胞の情報処理プロセスを次のような４層で行う。まず、S1 層で処理データと各方向のガボールフィルタを畳み込み計算を行って、次に S2 層で S1 層の直近データを統合する。そして、S2 層でデータパッチを出力し、C2 層で全部のデータを統合する。最後に、C2 層処理後のデータは、サポートベクターマシン（SVM）でパターンの学習と予測を行う。

1. Introduction

1.1. Research Background

The importance of sensory and communicational functions comes from the fact that the human body is an open biosystem , in a permanent exchange of energy, substances and information within its surrounding environment. The human body receives information from the Environment around it: 1% by taste, 1 to 5% by touch, 3 to 5% by smell, 11% by hearing and 83% is by sight. And the vision of human outperforms machine vision systems with respect to almost any measure, and emulating the information system in visual cortex has always been an attractive research topic.

The vision plays an important role in all of the perceptions. Comparing with the other perception information systems, the visual system is also the most complex. The computer can realize the capability of human visual recognition by researching the information system in ventral stream, and clarifying the mechanisms of its cognition and target recognition, which is significant to carry out the pattern recognition for a variety of data of the research, life and industry [1].

The biological vision systems has a completely different mechanism comparing to the image processing system of modern computer, but its capacity of the perception, cognition and the target recognition and the extremely strong ability to adapt the complex environment, all of which the computer system still can’t match, even nowadays with the highly developed computing technology.

Under the suggestion of prof. T. Poggio, professor at MIT, we developed an improved implementation of HMAX model, written in python interpreted language for the purpose of obtaining a pattern recognition method which is much more similar to that of human brain. Besides, as a joint study with JAXA, we used the improved system to process astronomical data to test the effectiveness of our new model. Final goal is to use actual JAXA data for automatic recognition of celestial structures.

1.2 Research purpose

The dataset from the CCD photometric system of Sloan Digital Sky Survey used six sets of CCD to measure the five wave brands (u,g,r,i) of celestial bodies simultaneously and the photometric parameters include color, outline and size. SDSS could obtain the photometric data of more than 100 million celestial bodies. In this study, we studied, identified and classified the stars and the galaxies in the SDSS of the database by using the HMAX model.

Fig 1-1 The star and galaxy image from SDSS database

1.3 Research Method

We will focus on simulating the V1 area of the Ventral Stream by using the HMAX model. The features of the V1 area as follows:

- The lateral geniculate nucleus(LGN which is the primary relay center for visual information) and V1 are both divided into six layers.
- The pyramidal cell and stellate cell project inside the V1.
- The simple cell and complex cell process the information of the lines.
- The complex cell will be abstract (the combination of the straight lines with same angle from different positions).
- For the cells which react to the straight lines of different angles, here we arrange them in columns, as it happens in V1.
- At least, there are three paths for vision: shape recognition, motion analysis and color processing. [2][3]

To help researchers implement pattern recognition of astronomical data more efficiently, we have made improvements to HMAX model in following aspects:

1. We used an improved new method to select patches.

2. We used more orientations (12) for the Gabor filter in order to acquire responses from more different directions. The same parameters are applied to different scales in order to reduce the number of variables and avoid instabilities.

3. We used SCI-KIT as our classification tools, to maximize speed and better integration with python.

2.Neural Networks and Artificial Neural Networks

2.1 Intelligence and machine learning

The human cerebral cortex is a neurons square of approximately 1000 cm^2 and 2mm thickness. For visualizing, please imagine a cloth napkins, which is probably the similar size and thickness of the neocortex. The neocortex is divided into many functional areas: the visual area, auditory area, language area and so on. Assume we observe them under a microscope, and the physical characteristics of these different regions are almost the same. There are some organizational structures in each region throughout the whole cerebral cortex.

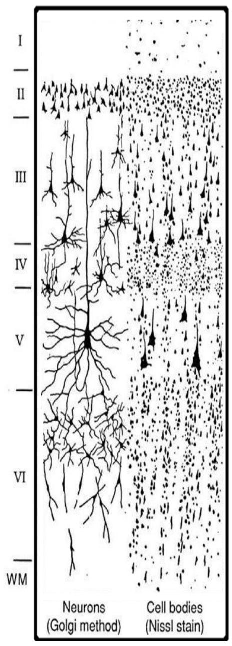

Fig 2-1 The six layers in neocortex

The first organizational structure is called the layer. Overall, the neocortex has six layers, wherein 5 layers include cells while 1 layer is composed almost by only connections and no neurons[3]. The second functional organizational structure is called the column. When the scientists explored what made the neurons become active by using the probe, they found that the neurons crossed the different levels, by responding to the same input by the mode of the vertical distribution.

According to the current study of the brain in biology, we believe that the essence of the intelligence is the memory and the prediction. And the intelligence is not an sudden extraordinary ability coming from above, but an action force running through the entire evolutionary history of organisms.

There are three stages so far throughout the evolution of intelligence. The first stage is from the single cell to an organism appearing the primary neural network. The so called intelligence does not exist in someone biological body, but performs through the DNA, RNA or other forms of genetic evolution. However, in this process, it does exhibit the memory (the DNA heredity), which is one of the natural property of the intelligence.

The second stage started from the appearing of the primary neural network to the stage of higher neural network. Through biological evolution, finally, there had a set of lower level of neural networks, such as the mollusks, the plants and so on. After having such set of neural networks, they had been greatly improved in the adaptation to the environment against the enemy and other related aspects.

The third stage started from the advanced neural network to nowadays. In the second stage, the biological neural network is not subject to change with different environments at any time, but only by the DNA genetic generation which has too low efficiency. So up to this stage, it began to appear the advanced neural network which can adjust in real-time in different environments. Needless to say, humans are very good at this stage. However, the footsteps of the intelligence evolution will not stop. There will be even more efficient and advanced intelligence in the future.

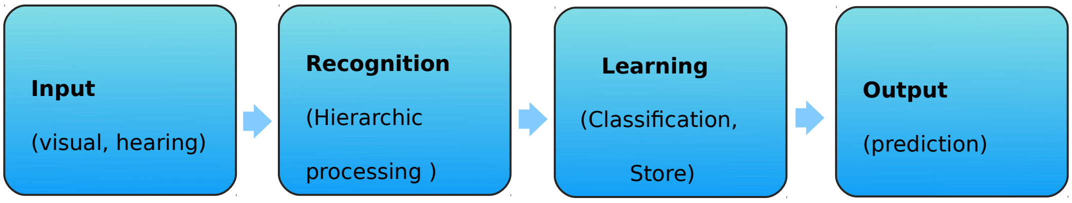

Fig 2-2 The flow chat of perception process

We believe that the Artificial Neural Networks established based on the abstraction and simulation of some basic properties of human brain and Natural Neural Network should have the following characteristics:

1. Self-learning function: for instance, in order to realize image recognition, we can just input the different image templates and corresponding recognition results into the artificial neural network, so that the network can learn to recognize similar images through the self-learning function.

2. Prediction storage function: this kind of prediction can be achieved by the feedback artificial neural network.

3. Capable of finding out the optimal solution quickly: finding out the optimal solution to a complex problem often needs a large amount of computation. However, using the feedback Artificial Neural Networks specially designed for a problem can give full

play to the high-speed computing ability of the computer and find out the optimal solution quickly.

2.2 The Visual Cortex Structure and Pattern Recognition

Fig 2-3 Our Brain can separate pictures of objects belonging to many categories

After the light enters the eye, it will be absorbed by the photo receptor on the retina, which drives a series of chemical reactions into nerve signals, and then they begin to be forwarded. The retina can actually be divided into more detailed layers, which can be used as the integration processing on the visual signals of the first stage. After the photo-stimulation of the retina converted into nerve signals, they will be transmitted along the optic nerves, then the part of the optic nerves of the left and right eyes have the first cross, then they go into an area of the brain below which is the one piece of the thalamus, the English abbreviation for it is the LGN (lateral geniculate nucleus). The LGN is equivalent to a relay station for visual signals from the retina to the visual cortex V1. It mainly has a function of attention, as well as some regulating functions for the feedback signals from some other regions of brain. In addition, until the visual signals arrive the LGN, all the images (to which each nerve cell responses) have circle shapes, which have luminance contrasts between the middle and outer of the ring (light inside and dark outside, or the dark outside and light inside), only by which it will cause the responses from the retina or the LGN cell.

The LGN can also be initially distinguished by two paths: the M Pathway and the P Pathway as the textbooks usually calls[4]. These signals of two paths are respectively related to the "move" and the "shape color". These two signals will also be finally transmitted to be processed in different areas. Then, as the previous picture of the human brain, the visual signals start from the retina, then go through the LGN (as relay station), and then arrive the hind brain spoon, which is also the first station of the visual cortex called the V1. The biggest difference between V1 and LGN is that you need a straight line to make a cell response, rather than an inner and outer circle with a comparison of brightness. The LGN cells in the ventral Stream(There are two streams in the visual cortex: the “dorsal stream” and the “ventral stream”) will only react to the shape of circles, and the V1 cells will only react to the straight lines. The V1 can actually been divided as: the simple cell and the complex cell.

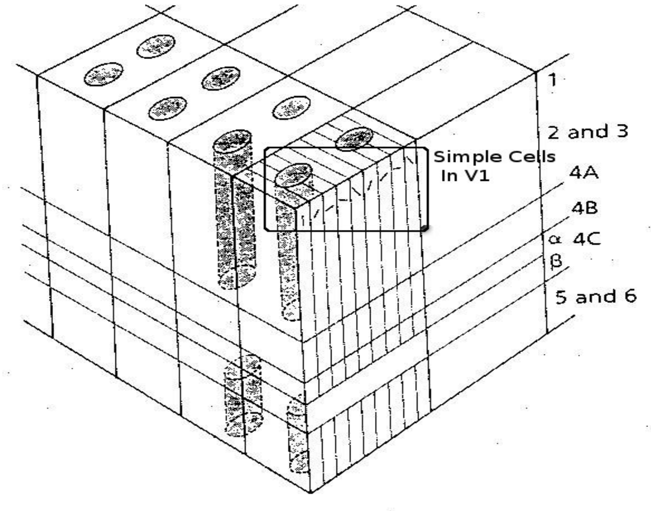

Fig2-4 Simple cells in V1 area of visual cortex

In addition, in the V1 area of visual cortex, there is an model with six layers as shown in above map which describes the columns in V1 cell, and each column is on behalf of a group of cells which react to straight lines in someone direction. The cells that have similar directions and angles in the brain are also close. However, the later studies found that the pinwheel model should be more suitable.

After the study by Hubel and Wiesel, etc., the visual cortex cell in V1 is found to be further divided into simple cells and complex cells. Just like mentioned, the simple cell reacts to the straight lines at an angle (straight line, horizontal line, and the oblique line with an angle). The complex cells can capture spatially displaced occurrences of similarly oriented features. So inside the visual cortex V1, it is the simple cell and the complex cell to make the appropriate processing on the visual signals, so that we finally recognize the square, the rectangular or the more complex visual information. Up to now, it is not entirely clear what are the detailed mechanisms after completing the process in the visual cortex V1. Semir Zeki [5] is the first author to propose that the visual signals are processed in parallel. The current hypothesis propose that there are at least three or more visual paths which process the images of motion, shape or form of the object, and the color. The paths from the M channel and P channel of LGN to the six layers inside V1 for different projections, and then the path from the V1 to the other cortical areas for processing, all of which include above three main visual paths.

Tomaso Poggio (the professor at MIT) proposed one most preliminary visual model. In the neuro-science, he hoped to transform those visual research findings to the Algorithm used in the information science, which can make the computer be closer to the biological model to help people complete some work of the image classification. We can get characteristics with high invariance and selectiveness through the simulation and modeling of visual cortex tissue structure. These characteristics can help us to learn from a small number of training samples, so as to complete the pattern recognition in complex scenes.

3.Standard HMAX model

3.1 General methods (HOG, SIFT) for feature descriptors in computer vision

For feature extraction, Histogram of Oriented Gradient(HOG) method and Scale-Invariant Feature(SIFT) method was the common method in computer vision.

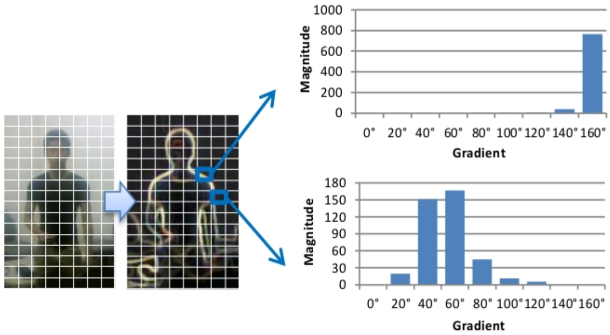

Fig 3-1 The feature vectors from HOG method

The following is the HOG method for feature extraction from images :

A. Image gradation (Images was seen as 3-dimensional(R,G,B) data).

B. Dividing the data in into several cells.

C. Calculating the gradient(orientation) of each pixel in every cell.

D. For each cell, the statistics of the gradient histogram is the descriptor of features.

[6]

Fig 3-2 The feature vectors from the SIFT method

The following is the Scale-Invariant Feature (SIFT) method for feature extraction from

A. Using scale-space filter for edge detection.

B. Finding the all positions of all of the key points.

C. Calculating the gradient magnitude and direction in each region.

Comparison with the HMAX model, the HOG features has no Scale-invariant and Rotation-invariant. Change of the direction or size will change the detection rate significantly. SIFT method has Scale-invariant and Rotation-invariant, but to get features with this method is over dependence on the selection of the principal directions.

3.2 Standard HMAX model

As mentioned above, the brain and the intelligence, as the most mysterious part of human body, has been drawing the attention of neurologists for hundreds of years. The enormous number of neurons and experimental data makes it quite difficult for people to understand the operating principle of the brain. Biologists have put forward various theories to explain the cells behavior of V1 area. However, few studies have been made on a higher level of neuronal behavior, such as Contour Linking and Feature Grouping. In recent years, there has been a lot of research hypotheses proposed by researchers, but these hypotheses just limited to the description of cells behavior. Few researches have been made on how to realize object recognition—the ultimate objective.

In the calculation model developed by Poggio based on the imitation of human brain

(HMAX), a multilayer model which is biologically similar to human brain is given to realize object recognition. This model proved a great success and made a stir in the research field of machine vision.

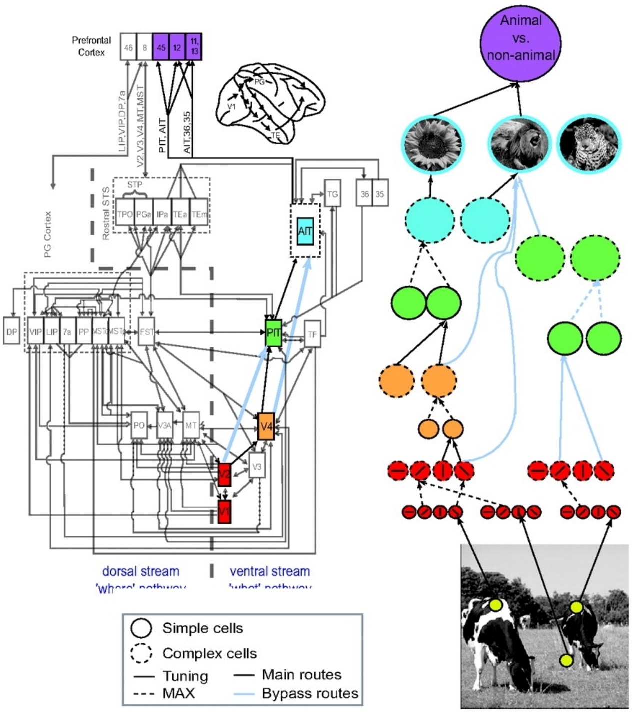

Fig 3-3 Standard HMAX model

The above model is what proposed by the professor Poggio from MIT. The standard model contains 4 layers, “simple unit” S1 Layer, “complex unit” C1 Layer, “simple unit” S2 Layer, “complex unit” C2 Layer[7].

HMAX uses scale-space and a spatial hierarchy to perform processing on small sub-components of the image in multiple image scales/sizes. The hierarchy consists of multiple Levels, with lower levels feeding data into the higher levels, one level to the next. The bottom-most level takes in the original input image while the top-most level performs the final classification decision. Each level will consist of one or more

Layers. Each layer is intended to handle differently scaled sizes of the original input image in order to perform a sort of parallel processing of multiple object sizes. As we ascend the hierarchy we slowly combine the scale layers by selecting the best responses between 2 neighboring scales resulting in an overall best response across many sizes. This gives an increased invariance to size when performing classification.

In S1 Layer, an input gray-scale image is densely filtered by a battery of Gabor filters at each scale and orientation. And at each pixel, all of the filters are centered.
The Gabor filters are described by:

$$G(x, y) = \exp\left(-\frac{(X + \gamma^2 Y^2)}{2\sigma^2}\right)\cos\left(\frac{2\pi}{\lambda}\right) \qquad (1)$$

Where $X = \cos\theta - y\sin\theta$ and $Y = \sin\theta + y\cos\theta$. λ represents the wavelength of the sinusoidal factor, θ is the orientation of the Gabor function, γ is the spatial aspect ratio.

The parameters of filters with 4 orientation and 16 scales as in the table below:

S1 Layer			
Scale Band S	Filter Size s	Gabor parameter σ	Gabor parameter λ
Band 1	7×7 9×9	2.8 3.6	3.5 4.6

S1 Layer			
Band 2	11×11 13×13	4.5 5.4	5.6 6.8
Band 3	15×15 17×17	6.3 7.3	7.9 9.1
Band 4	19×19 21×21	8.2 9.2	10.3 11.5
Band 5	23×23 25×25	10.2 11.3	12.7 14.1
Band 6	27×27 29×29	12.3 13.4	15.4 16.8
Band 7	31×31 33×33	14.6 15.8	18.2 19.7
Band 8	35×35 37×37	17.0 18.2	21.2 22.8

Table 3-1 The Gabor filter parameters in S1 Layer[7]

The results from S2 Layer with the parameters conform to the biological experimental datasets. But the defining too many parameters destroys the simplicity of the model. In C1 layer, for each scale the model gets the local maximum value between 2 neighboring scales . The S2 layer selects N patches of data randomly. And then, each patch is filtered by a RBF filter. The values of C2 Layer comes from the traversing all of 8 groups of the scales, to find the maximum of the responses.

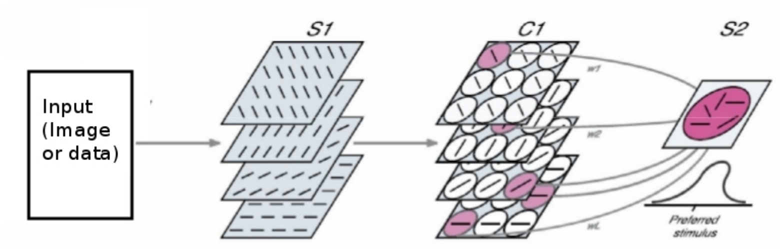

Fig 3-4 The S2 layer select parts of the input as the synaptic weights

However, some problems are what this model needs to modify. Firstly, regarding the information on moving objects, the Aperture Problem involves that how to connect the information in different areas to each other to be overall, which is called the Binding Problem. For example, the neuron A reacts to the images moving to the left, while the neuron B reacts to the image moving to the right. But if a big image moves up as a whole, it is moving to the left or to the right from the regional point of view, for which there is still no appropriate model. However, some research results, e.g., the experimental results of IT (Inferior temporal cortex)from Keiji Tanaka found that the IT neuron will only respond to some particular shapes, rather than simple geometric shapes such as the round or the square, etc.

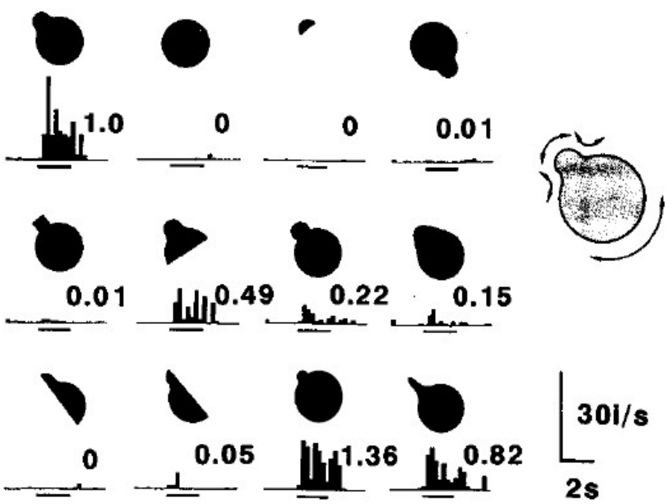

Fig 3-5 The experimental results of IT neuron from Keiji Tanaka

Take above picture as an example, it shows a certain neuron reaction in IT [3]. The reaction at the upper left corner of the picture is the biggest, which is a big circle with a small protrusion. The reaction is 1.0. However, if we turn 180 degrees, then the remaining reaction becomes to be 0.01. If the small protrusion is changed into the square, the remaining reaction is also 0.01. If there is no small protrusion, but only with a large circle in the middle, the reaction is actually 0.

Therefore, the processing mode of IT for handling object shape is still uncertain, because it is not so obvious like V1. A group of neurons are reactive for straight lines in a certain direction. Each neuron of IT only reacts to a certain kind of special shape, while some features are important but some features can be taken away and will not affect. Also because there still remains parts which needs continuing efforts from the researcher. Therefore, we focus to simulate the S and C cells in V1 area which have been biologically and clearly researched and passed in HMAX model.

3.3 Support Vector Machine

Support Vector Machine (SVM) is a very popular and well known machine learning technique proposed by Vapnik et. which has been successfully applied to many real-world classification problems from various domains. SVM is a supervised learning model with associated learning algorithms that analyze data and recognize patterns, and is used for classification. Due to its theoretical and practical advantages (such as solid mathematical background, high generalization capability and ability to find global and non-linear classification solutions), SVM has been very popular among the machine learning and data mining area.

The process of SVM is divided into two parts as depicted below. The is called the learning part. Given a sample dataset, an SVM training algorithm builds a model in order to classify the given data. The second part is called the prediction part. Given new test data, it is classified into a category using the model which was built in learning part.

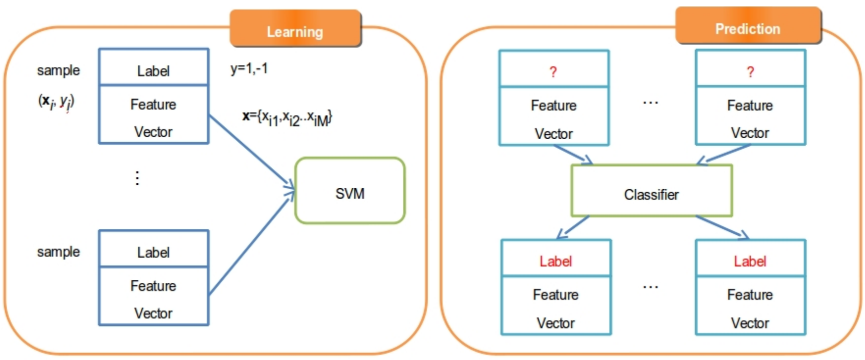

Fig 3-6 The process of Support Vector Machine

The theory of Support Vector Machine initially comes from the disposal for the data classification. For the binary classification of the data, if we apply the neural network, the system randomly generates a hyperplane and then moves it, then we know that the points (in the training set) belonging to different categories are just located at different sides of the hyperplanes. As for the Linear Support Vector Machine, we consider the training samples as { X_i , Y_i }, where the X_i is the example of the input pattern, the Y_i is the corresponding target output. At beginning, we assume that the models represented by the subset $Y_i = -1, +1$ are linearly separable, for separating the hyperplane format to decide the SVM classifier as:

$$f(x)=sign(W^T+b) \qquad (2)$$

and the corresponding hyperplane is

$$W^T+b=0 \qquad (3)$$

Where W is the parameter vector, and b is the biased or offset scalar. Label y of the test data is decided by $f(x)$. When $W^T+b>0$, then $y=+1$ otherwise $y=-1$

4. System for astronomical data analysis

We develop a new improved version of HMAX model. We developed the system in python because of its versatility and available libraries. The differences with the standard model are as mentioned above, an improved method to select patches, use more orientations (12) for the Gabor filter and used SCI-KIT as our classification tools, to maximize speed and better integration with python.

4. 1 SDSS database

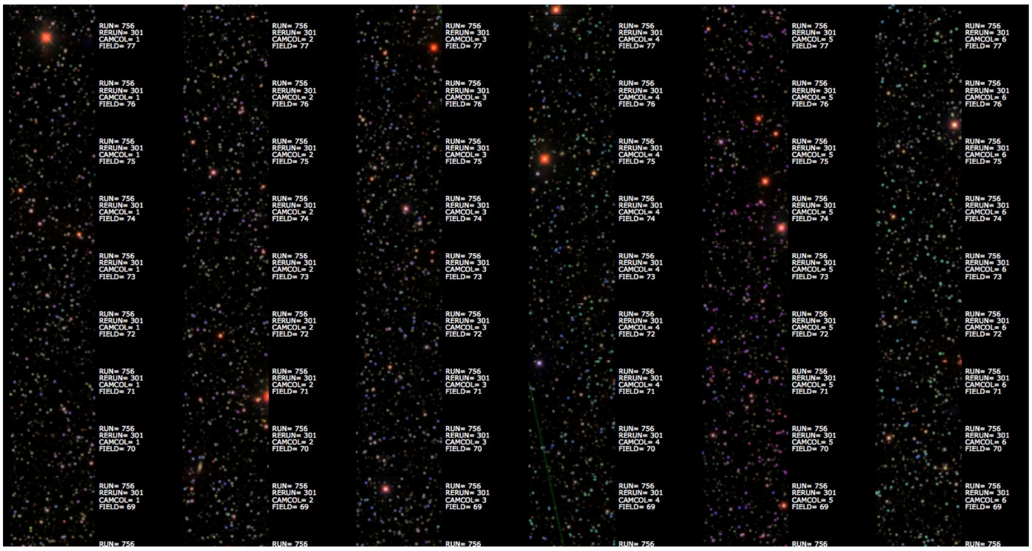

Fig 4-1 The images from Solar Digital Sky Survey（SDSS）Data Release 8

We used the Solar Digital Sky Survey（SDSS）Data Release 8 as our dataset. In this release, all of the imaging data was taken by the SDSS imaging camera which contains totally over 14000 square degree of sky. In this study , we prepared the dataset with about 500 stars and 500 galaxies for sample training. And we used about 5 groups each group contains 60 images for test our model.

4.2 Astronomical data Features

4.2.1 Initial processing

In the initial processing, the program is designed to take an input image and produce 12 resized scales of the image. The intent is that a range of image sizes will help the algorithm become exposed to a greater range of object sizes. Since we keep the Gabor filter size fixed (at say 11x11), as we decrease the image resolution the Gabor filter is effectively covering a larger and larger area of the image. In this way we can detect orientated lines at a range of sizes from rather small to rather large. A single layer then contains a set of data usually to represent a given size scaling. By having multiple layers we can store multiple image scales at once and process then in parallel. As we ascend the hierarchy we slowly combine the scale layers by selecting the best responses between 2 neighboring scales resulting in an overall best response across many sizes. Layers have a concept of both discrete-space and what we call retinal-space. The discrete space is simply the actual indices's corresponding to a numpy 2d matrix that is storing the actual layer values (The pixel values for the image). While retinal-space is a real-valued global space, where (0,0) indicate the center of the image and this location is consistent across all hierarchical levels despite that each level will have discrete matrices of differing dimensions. Using the retinal-space we have a way of ensuring accurate comparisons of locations across different layers in different levels.

We prepared five groups of layers to store the feature vectors in different process, and size of the layers is used as the table following[8]:

Input Layer	S1 Layer	C1 Layer	S2 Layer	C2 Layer
256×256	246×246	47×47	44×44	1×1
214×214	204×204	39×39	36×36	
180×180	170×170	33×33	30×30	
152×152	142×142	27×27	24×24	
128×128	118×118	21×21	18×18	
106×106	96×96	17×17	14×14	
90×90	80×80	15×15	12×12	1×1
76×76	66×66	11×11	8×8	
64×64	54×54	9×9	6×6	
52×52	42×42	7×7	4×4	
44×44	34×34	5×5	2×2	
38×38	28×28			

Table 4-1 The size of layers

4.2.2 Process in S1 Layer

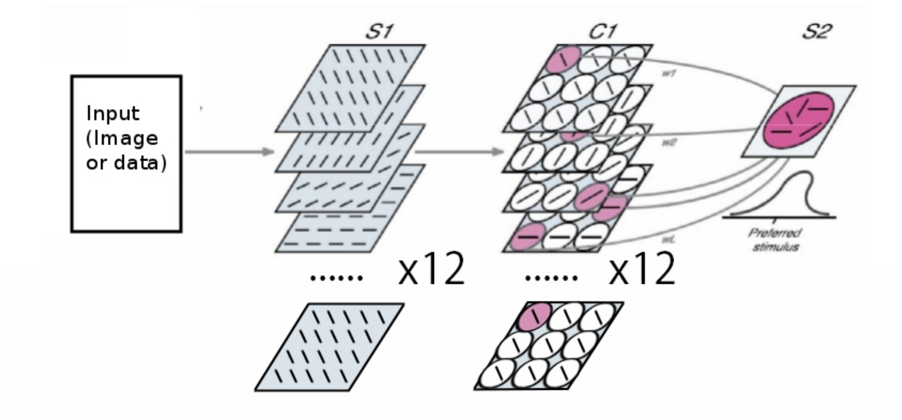

Fig 4-2 Process in our HMAX model

The S1 Layer is applied on 12 resized versions of this image at different scales (256x256, 214x214, 180x180 down to 38x38 pixels). In higher mammals brain, images are filtered along orientations. A Gabor filter is used to reproduce this functionality. In our HMAX model each scale the Gabor filter is applied at 12 different orientations, for a total of 12x12 set of output data (12 scales by 12 orientations).

The Gabor filters are described by[8]:

$$G(x, y) = \exp\left(-\frac{(X+\gamma^2 Y^2)}{2\sigma^2}\right)\cos\left(\frac{2\pi}{\lambda}\right) \qquad (4)$$

Where $X = \cos\theta - ysin\theta$ and $Y = \sin\theta + ycos\theta$. We adjusted the same filter parameters for all of the scales,orientation,and wavelength . And in our model, we make γ =0.3, λ =5.641, σ =4.5128. The result value will be returned and is expected to then be stored in the S1 network layer.

4.2.3 Process in C1 Layer

After the S1 layer is processed, the 12x12 group of data are received by the C1 layer. In C1, a major rescaling is performed as happens in the hierarchical information flow in the mammal brain. In our algorithm each resized group is compared with the adjacent smallest sized group of data (for example 256x256 compared with 214x214). The comparison is made with normalized XY coordinates on a 3x3 grid. The biggest data value is selected and a single data point is generated. All the image is scanned and a new smaller group of data is generated (for example, 47x47). There are 12 orientations for each group size, so 12 of these resized groups are generated this way. The process continues with all the other sizes (214x214 compared with 180x180 and so on) until a new set of smaller data groups are created. The groups are these sizes (47x47, 39x39, 33x33 down to 5x5). Finally, because of this comparison in couples, one of the sizes is lost and after C1 operations, we have 11 sizes , and 12 orientation, for a total of 11x12=132 groups. This elaboration is done to remove noise in scale-independent pattern recognition processes observed in mammal.

4.2.4 Process in S2 Layer

In S2 layer, we selected 4075 patches of data. we only store it as a learned template patch if:

A. At least 35% of the cells are non-zero.

B. The patch has less than 35% similarity to any existing learned patch.

If we have 10 learned patches already, and our similarly threshold is 90%, then we will not accept a new patch unless its RBF value is below 90%similar to all the existing patches. This helps ensure a minimal level of uniqueness/variety among learned patch templates. Once we have enough template patches learned we can run the filter in inference. For inference each patch in the current C1 Composite is compared against all learned template patches and scored using a Gaussian Radial-Basis-Function similarity value. All of these values are sent as output to the next layer C2.

The size of these patches is 4x4x12 or 8x8x12 or 16x16x12 chosen at random and placed at random position. The 12 appears because we have 12 orientations. Because of normalization of coordinates, these patches have the same relative area on each scale. The Gaussian Radial-Basis-Function is applied:

$$R(X_i, P_j) = \exp\left(-\frac{(X_i - P_j)^2}{2\sigma^2\alpha}\right) \quad (5)$$

In our case, we set $\sigma=1.6$ and $\alpha=1$ (α is the normalizing factor for patch sizes can vary), The output of this layer is a new group of data of size $A \times A \times 4075$, where A is the original size of the group. The 4075 points are the maximum value of the Gaussian filtering above. After this stage we have 11 group of data, each of them $A \times A \times 4075$ in size, where A is the original group size (47x47, 39x39, 33x33 down to 5x5 as above).

4.2.5 Process in C2 Layer

These 11 groups are fed to the C2 layer that in the same fashion as C1 layer, it compare different patches on a 3x3 grid. This time the comparison is not made on two adjacent sizes, but is done on 6 of them. For example the group A=47 is compared with A=39, A=33 and so on. A single maximum value is determined. This process is repeated another group of 6 different sizes. At the end two group of data of size A=39 and A=21 is generated. Each group of data has size $A \times A \times 4075$.
Now spatial dependence is removed. For each AxA sheet data in each group, the maximum value is taken. So each $A \times A \times 4075$ group is reduced to a one-dimensional vector of size 4075. We do that on the two groups and join the results. So the output of C1 is a single one-dimensional vector of size 8150. This final vector represent the generalization of the original image features, and we call it from now on “the features” embedded in the image.

4.3 Data Training And Prediction

We built the learning and prediction of vectors process by Using python. We used numpy module to store vectors and scikit-learn module for machine learning process. In our model training and prediction was used as following:
A. Normalization in the input space. The same axle of all the sample features should

be normalized [9].

B. We set the type of Support Vector machine to “C-SVC”, the type of kernel function to radial basis function[10].

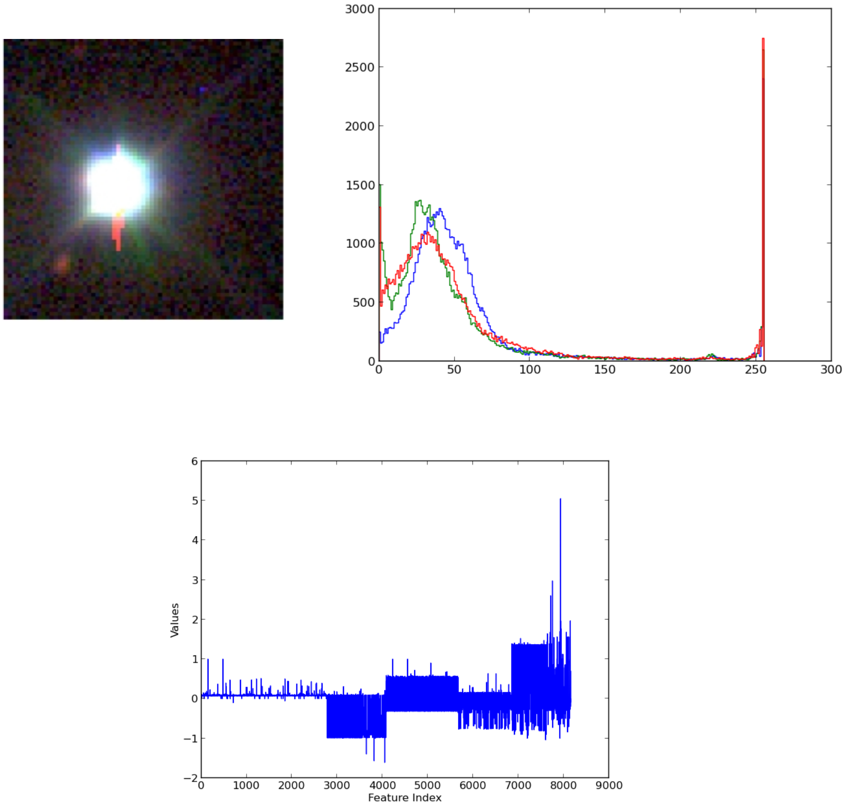

Fig 4-2 Sample A(Star), Histogram of Sample A, C2 features of Sample A

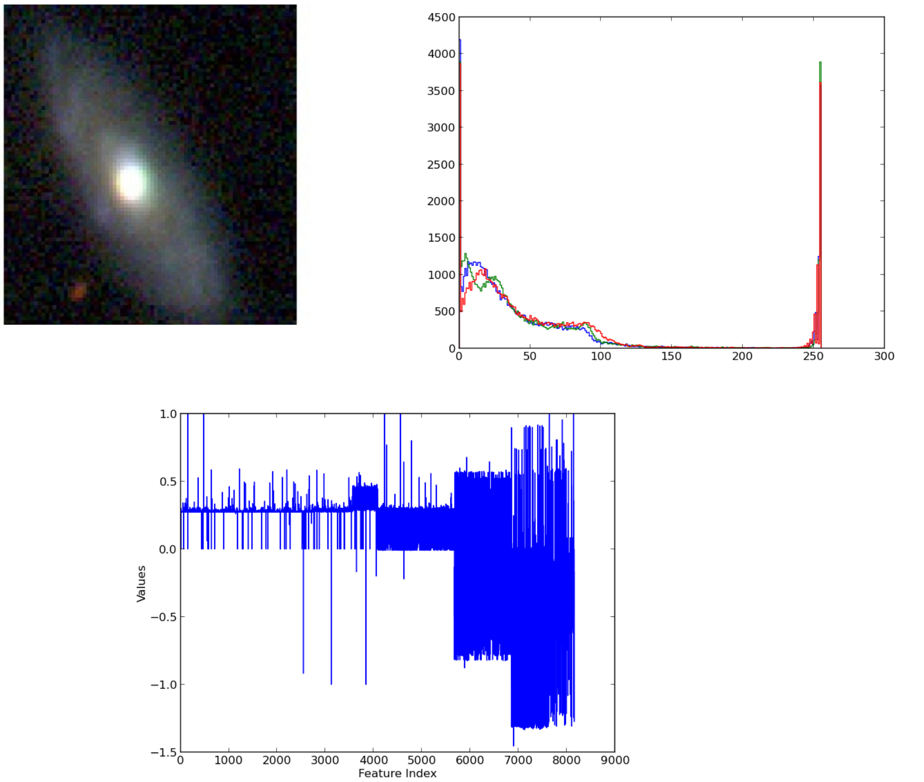

Fig 4-3 Sample B(Galaxy), Histogram of Sample B, C2 features of Sample B

5. Dataset Prediction

5.1 Test result of Prediction

The test dataset consists of 5 groups, each group has 200 images from SDSS. The accuracy rate for our model is in the table following:

Testing dataset	Learning process time for each sample(average)	Prediction process time for each sample(average)	Accuracy
Group1	1 Hour 20 Minutes	0.17 seconds	86%
Group2		0.17 seconds	91%
Group3		0.16 seconds	72.5%
Group4		0.17 seconds	89.5%
Group5		0.19 seconds	84%

In this research, we have used Hmax Model to carry Astronomical data and classification. The Accuracy have been improved to 80% and in the process, we have simulated the information delivering method of biological brain and achieved effective features. We also found the problem of long timescales, Now we are trying to improve this problem by using admixture programming with C Language and python.

5.2 Conclusion and Prospect

By improving HMAX model, this study has established an astronomical image recognition platform based on Python. We improved the selection method of patches and increased the number of Gabor filter orientation. We selected positive and negative samples from SDSS database randomly, tested our model and achieved fruitful results. HMAX model is an object recognition model in brain resembling structure which works very well. It verifies Gabor filter and multilayer feature combination based on biological features and extract the features with invariance and selectiveness by maximizing the space scale filters. The classification of feature vectors of each image or each group of data can be completed by using machine learning. However, Support Vector Machine has removed its biological consistency to some extent at the final stage.

5.3 Research in the future

In the present study, this system is based on the characteristics of static image or the data identification. We use the Support Vector Machine which doesn't consider time sequence. But in reality, most of the image and data are often accompanied by the time. If we assume we do not use the time sequence, we almost can not infer anything from the dynamic images or audio information. Human eyes can understand the static scene. Therefore, the visual processing doesn't always require inputting of the temporal information. However, in the usual visual scene, we are constantly moving our eyes, head and body, and there are too many moving objects around us. After many generations human can recognize everything in this world, quickly. For the general conditions of vision, audition and thigmesthesia (touch sense) [11], the reasoning always requires the input of temporal change. So in the next step of study, we hope to join the dynamic timing of the data into the existing algorithms. Simply speaking, it will be a set of data flow of the spatial characteristic data varying with the time. This new algorithm still needs a lot of data for training. Just like that if you want to identify different varieties of ships by observing many different types of vessels instead of just looking at someone type. In future, we want to realize an algorithm will learn the time series of the input stream, for example, you can build a model of one pattern following another. The difficulty lies in this algorithm is that you don't know when to start the sequence and when to end, and the possible overlap at the same time point.

Reference

1. Y. D. Tian, L. Q. Zhang.Relevant, Algorithms for General Object Recognition Based on Feature Combination.Shanghai Jiao Tong University Journal,2007.

2. J. Hawkins, S. Blakeslee, On Intelligence, 2004.

3. H. Kolb, R. Nelson, E. Fernandez, B. Jones, The Organization of the Retina and Visual System,1995.

4. Fang F, He S., Cortical responses to invisible objects in the human dorsal and ventral pathways. Nature neuroscience, 2005.

5. Zeki S., A Vision of the Brain[M]. Blackwell scientific publications, 1993.

6. Kishino T, Zhe S, Micheletto R. A fast and precise HOG-Adaboost based based visual support system capable to recognize Pedestrian and estimate their distance,2013

7.T. Serre, L. Wolf, and T. Poggio. Object recognition with features inspired by visual cortex. In CVPR, San Diego, June 2005.

8.J. Mutch and D. G. Lowe.Object class recognition and localization using sparse features with limited receptive fields. International Journal of Computer Vision. Vol. 56

9. Graf A B A, Borer S. Normalization in support vector machines, Pattern Recognition, 2001.

10.C. Elting, C. A. L. Bailer-Jones, K. W. Smith. Photometric Classification of Stars, Galaxies and Quasars in the Sloan Digital Sky Survey DR6 Using Support Vector Machines. American Institute of Physics Conference Series. Vol. 1082. pp. 9-14 , 2008

11. Numenta org, Hierarchical temporal memory including HTM Cortical Learning Algorithms ,2011.

研究概要（日本語）

視覚野ニューロンネットワーク情報処理に基づいた、天文データの識別システムの研究

概要: 人間の視覚野は、目から得たデータを処理し、学習と記憶に基づき、パターン認識をしている。脳の働き方を模倣した、学習機能を持つアルゴリズムは、一般的なデータパターン（実験データ等）の認識率が高い。従来の研究では、輝度の勾配方向と勾配強度を利用しているが、認識の精度は低い。これは、宇宙電波望遠鏡の観測データはノイズが多く複雑かつ、情報量が膨大なためである。これを改善するために、人間の視覚野神経回路網情報処理モデルと機械学習技術を利用して、観測データを精度良くパターン認識できるシステムを開発する。本研究では、MIT の Poggio 教授によって発明された HMAX モデルを用いた。HMAX モデルは、生理学的実験に基づいた大脳視覚野の情報処理方法を利用し、データの特徴抽出を行うモデルである。具体的には、Ｖ１野とＶ２野の特徴抽出細胞および V4 野を経て IT 野に至る階層的な処理である。このモデルを利用し、視覚野単純型細胞と複雑型細胞の情報処理プロセスを次のような４層で行う。まず、S1 層で処理データと各方向のガボールフィルタを畳み込み計算を行って、次に S2 層で S1 層の直近データを統合する。そして、S2 層でデータパッチを出力し、C2 層で全部のデータを統合する。最後に、C2 層処理後のデータは、サポートベクターマシン（SVM）でパターンの学習と予測を行う。

謝辞：

終始熱心なご指導を頂いた生命のシステム研究科の先生達に感謝の意を表します。日本にいる修士の２年間にあたり、終始適切な助言を賜り、また丁寧に指導して下さった木下郁雄教授、大月俊也教授、Robert Knaly 準教授 に感謝します。最後まで一緒に頑張って来たミケレット研究室の同期の皆様，研究を手伝ってくれた後輩に心より感謝しております。学業と生活に Ruggero Micheletto 教授から、いろいろな応援をいただき、本当に有難うございました。それで、NITORI 国際奨学財団から十分な経済面のご支援のおかげで、自分の学業に専念することができました。
今後、博士課程をより一層勤勉に勉強し、いろいろな交流活動にも力を傾けようと思う。さらに、世界一流の研究能力を把握することを目指し、努力していきたいと思う。
来日してから、いろいろなことを経験し、初めてじ自分の成長を実感した。皆のおかげで、勉強も生活両立することができて、研究を順調に終わりました。帰国後、社会に貢献しようと思う。また、皆さんとの縁を銘記し、恩返したいと感じています。

Research Products

Conference Presentation:

Sun Zhe, “A fast and precise HOG-Adaboost based based visual support system capable to recognize Pedestrian and estimate their distance.”, the 17th International Conference on Image Analysis and Processing, 2013.

Sun Zhe, “Study of Information Processing Systems Based on Neural Network of Visual Cortex”， Sensory Substitution Symposium， 2013

Papers in International Journals:

T. Kishino, Sun Zhe, R. Micheletto, “A Fast and Precise HOG-Adaboost Based Visual Support System Capable to Recognize Pedestrian and Estimate Their Distance”, Lecture Notes in Computer Science, Vol. 8158, pp 20-29 (2013)

Submitted to International Journals:

Sun Zhe, Ruggero Micheletto, “Emergence of intelligent behavior from a minimalistic stochastic model for the navigation of autonomous Robots”

T. Kishino, Sun Zhe, R. Micheletto, “Cross-modal codification of images with auditory stimuli: an efficient language for the visually impaired ”