Tsinghua Science and Technology

Volume 18 | Issue 1 Article 6

2013

C-HMAX: Artificial Cognitive Model Inspired by the Color Vision Mechanism of the Human Brain

Bo Yang
State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

Lipu Zhou
State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

Zhidong Deng
State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

Recommended Citation
Bo Yang, Lipu Zhou, Zhidong Deng. C-HMAX: Artificial Cognitive Model Inspired by the Color Vision Mechanism of the Human Brain. Tsinghua Science and Technology 2013, 18(1): 51-56.

TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 06/12 pp51-56
Volume 18, Number 1, February 2013

C-HMAX: Artificial Cognitive Model Inspired by the Color Vision Mechanism of the Human Brain

Bo Yang, Lipu Zhou, and Zhidong Deng*

Abstract: Artificial cognitive models and computational neuroscience methods have garnered great interest from both neurologist and leading analysts in recent years. Among the cognitive models, HMAX has been widely used in computer vision systems for its robustness shape and texture features inspired by the ventral stream of the human brain. This work presents a Color-HMAX (C-HMAX) model based on the HMAX model which imitates the color vision mechanism of the human brain that the HMAX model does not include. C-HMAX is then applied to the German Traffic Sign Recognition Benchmark (GTSRB) which has 43 categories and 51 840 sample traffic signs with an accuracy of 98.41%, higher than most other models including linear discriminant analysis and multi-scale convolutional neural network.

Key words: artificial cognitive model; machine learning; traffic sign recognition

1 Introduction

Hierarchical cortical based models have achieved great success in most computer vision tasks. Indeed, artificial cognitive models and computational neuroscience have a long history of imitating the human brain in this manner. Partial analogs of the brains ventral stream have been used in many computing methods in canonical computer vision. However, models that only imitate the ventral stream can not provide semantic representation[1,2].

HMAX and Convolutional Neural Network (CNN) which extract only shape and texture features in observations are limited by the lack of color features. These models are only partial analogs of the brains ventral stream that imitate only the hierarchical structures and shape-texture-extracting pathway of the brain. Improvements will need to include color-feature expressions in the computational model with the development of neurophysiology and behavior psychology, the mechanism and the stream of color-feature expression in the color vision pathway of the human brain are now well understood[3-8], which provide great opportunities and strongly physiological knowledge to bring this theory into existing artificial cognitive methods.

As a result, this work introduces an artificial cognitive model to extend HMAX with a color-feature-extraction algorithm inspired by the color vision stream of the human brain that performs very well on the German Traffic Sign Recognition Benchmark (GTSRB) test. The model has two unique properties with a methodology to construct a color-feature-extraction pathway and shape-texture expressions. The methodology is derived from the principles of data fusion theory.

2 Related Work

HMAX was proposed by Riesenhuber and Poggio in 1999 to extract shape and texture features from gray images inspired by the ventral pathway mechanism in the V1, V4, and PIT/IT areas of the brain[9]. In the

• Bo Yang, Lipu Zhou, and Zhidong Deng are with the State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China. E-mail: bo-yang07@mails.tsinghua.edu.cn; zlp09@mails.tsinghua.edu.cn; michael@tsinghua.edu.cn.
∗ To whom correspondence should be addressed.
Manuscript received: 2012-10-15; accepted: 2012-12-19

following years, Serre et al.[10,11] extended the HMAX model by improving the S2 layer with more Gabor filters with different orientations and scales. They also improved the model by adding a lateral inhibition mechanism[4] and a sparse-coding expression[12], which is important in the primary visual cortex. Thus, HMAX was constructed with an input layer where an image is fed into the model, S/C couple layers which calculate the convolution with a response diagram from the lower layer and prototypes at the S layer and the maximum expression at the C layer, and output layers where the features were encoded in some way such as a sparse code.

There is another pathway in the visual cortex that expresses the color features of the scene which is not mentioned in HMAX model. This pathway starts from the V1 area, goes up along the ventral stream, and ends at the V4 area. These cortex areas have some neurons that are compacted together to encode the color information through so called "double opponent" approach[13,14]. These clusters of neurons are also known as blobs[15]. Some also believe that these blobs are related with the color constancy mechanism in human color vision.

3 Design

Color-HMAX (C-HMAX) model is motivated by system work on the human color vision mechanism and the need for a computer vision system with color features not included in HMAX model. The C-HMAX model imitates the human color vision pathway using computer-based methods and forms.

First, the image is encoded in the input layer into three color channels defined in the CIE LAB color space, which imitates the double opponent in the blobs of the human visual cortex. The three channels use L, a, and b to express the opposing color effects of red-green, blue-yellow, and light-dark. The distance between any two colors x and y can then be obtained by calculating the Euclidean distance

$$d = \sqrt{\left(L_x - L_y\right)^2 + \left(a_x - a_y\right)^2 + \left(b_x - b_y\right)^2} \quad (1)$$

Then the hierarchal structures of the C-HMAX models shown in Fig. 1 can be defined as:

(1) The input layer encodes the input image into L, a, and b channels in the CIE LAB color space as three sub-images representing the three separate channels.

(2) The CS/CC couple layers present the convolution and find the maximum for the response diagrams in all

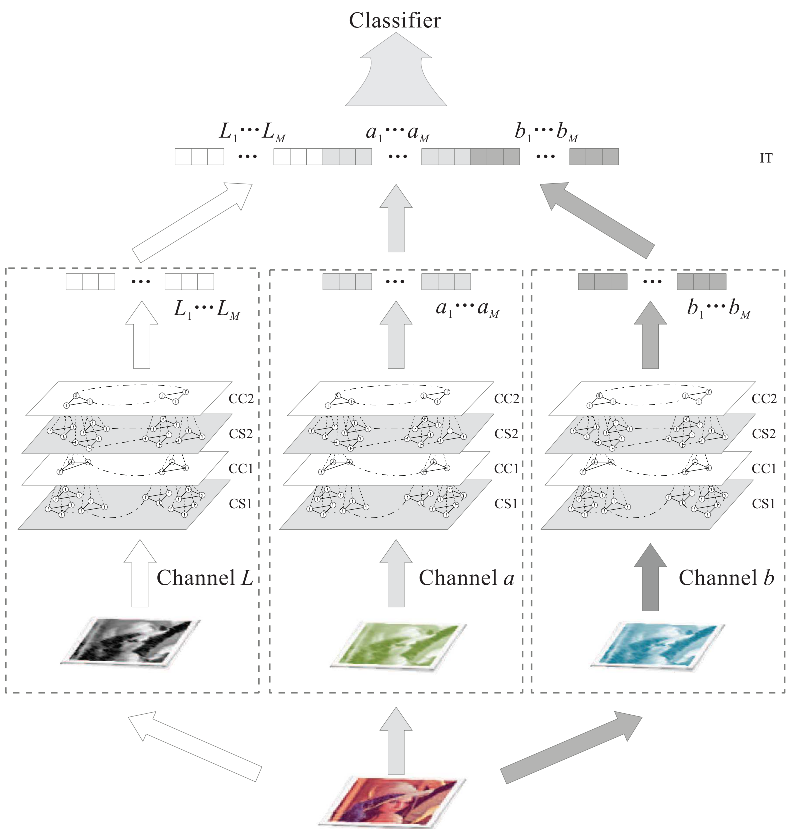

Fig. 1 Hierarchical structure of C-HMAX model.

three channels. The first couple layers use Garbor filters to calculate

$$G(x, y) = \exp\left(-\frac{x_0^2 + \gamma^2 y_0^2}{2\sigma^2}\right) \times \cos\left(\frac{2\pi}{\lambda} x_0\right) \quad (2)$$

where $x_0 = x\cos\theta + y\sin\theta$ and $y_0 = -x\sin\theta + y\cos\theta$, $x, y \in [-5, 5]$, $\theta \in [0, \pi]$, γ is the direction ratio and defines the effective width of the Gabor filter.

The higher couple layers do the calculation with prototypes of the three channels. The CS layer uses the Mahalanobis distance to evaluate the similarities for two colors between the response diagram and the prototypes,

$$\begin{aligned} d_L &= \sqrt{(\boldsymbol{L}_A - \boldsymbol{L}_B)^{\mathrm{T}} \boldsymbol{\Omega}_L^{-1} (\boldsymbol{L}_A - \boldsymbol{L}_B)}, \\ d_a &= \sqrt{(\boldsymbol{a}_A - \boldsymbol{a}_B)^{\mathrm{T}} \boldsymbol{\Omega}_a^{-1} (\boldsymbol{a}_A - \boldsymbol{a}_B)}, \\ d_b &= \sqrt{(\boldsymbol{b}_A - \boldsymbol{b}_B)^{\mathrm{T}} \boldsymbol{\Omega}_b^{-1} (\boldsymbol{b}_A - \boldsymbol{b}_B)} \end{aligned} \quad (3)$$

for the three channels separately, where

$$\boldsymbol{\Omega}_z = \mathrm{cov}(\boldsymbol{x}_A, \boldsymbol{x}_B) \quad (4)$$

is the color similarity matrix obtained within the A and B areas of the image, where z represents L, a, and b.

(3) The output layer concatenates the three feature vectors of the three channels. If each channel has a color feature vector of length M, this layer give a $3M$-length vector

$$\boldsymbol{V} = [L_1, \cdots, L_M, a_1, \cdots, a_M, b_1, \cdots, b_M] \quad (5)$$

4 Performance Evaluation on GTSRB

The performance of the C-HMAX model was evaluated using the GTSRB test[16]. GTSRB is a German traffic sign recognition benchmark in 2011 with all the traffic sign samples captured with a Prosilica GC 1380CH Camera on real German streets. The training set contains 43 classes of traffic signs as shown in Fig. 2. The following subsections show the configurations of the datasets and the C-HMAX model. Then, the test results are given to show the accuracy with comparisions with other models.

4.1 Datasets and model configuration

Due to the wide-range of brightness in the GTSRB training set samples, some samples were discarded as "too dark" or "too light" since the color features from these samples were then not obvious. Thus, samples with an average $\bar{L} > 85$ and $\bar{L} < 15$ were excluded from the training set, with those samples numbers listed in Table 1.

The C-HMAX model used three CS/CC couple layers to give six layers plus one input layer and one output layer. The sixth layer of each channel produced an 80-element-color-feature vector obtained from channels a and b. Channel L was then discarded from the finally concatenation in the output layer due to the illumination variation. A 160-element vector containing only the color features from channels a and b was then used for the classification.

The vectors were then fed into an SVM classifier for training and testing. The classifier has a positive integer output value ranging from 1 to 43, which stands for the 43 different categories.

4.2 Test results

The GTSRB testing set had 12 630 samples with no duplicated traffic signs. The C-HMAX model had an accuracy of 98.41% while the HMAX model had an accuracy of 96.53%, as listed in Table 2 with the results of the other models in Ref. [11]. C-HMAX method

Fig. 2 43 categories of traffic signs in the GTSRB training set.

Table 1 Sample size of each category in the GTSRB training set

Category	Sample size		Category	Sample size	
	Original	After pre-processing		Original	After pre-processing
00	7×30=210	187	22	13×30=390	377
01	74×30=2220	2032	23	17×30=510	452
02	75×30=2250	2170	24	9×30=270	224
03	47×30=1410	1398	25	50×30=1500	1428
04	66×30=1980	1876	26	20×30=600	572
05	62×30=1860	1750	27	8×30=240	223
06	14×30=420	402	28	18×30=540	510
07	48×30=1440	1411	29	9×30=270	252
08	47×30=1410	1388	30	15×30=450	418
09	49×30=1470	1455	31	26×30=780	702
10	67×30=2010	2001	32	8×30=240	210
11	44×30=1320	1230	33	23×30=690	628
12	70×30=2100	2066	34	14×30=420	409
13	72×30=2160	2109	35	40×30=1200	1173
14	26×30=780	746	36	13×30=390	325
15	21×30=630	615	37	7×30=210	198
16	14×30=420	405	38	69×30=2070	2006
17	37×30=1110	1097	39	10×30=300	238
18	40×30=1200	1173	40	12×30=360	319
19	7×30=210	198	41	8×30=240	198
20	12×30=360	351	42	8×30=240	232
21	11×30=330	319			

Table 2 Accuracies of various models

Model	Accuracy (%)
Committee of CNNs	99.46
Human	98.84
C-HMAX	**98.41**
Multi-scale CNN	98.31
HMAX	**96.53**
Random Forests	96.14
Linear Discriminant Analysis (LDA)	95.68

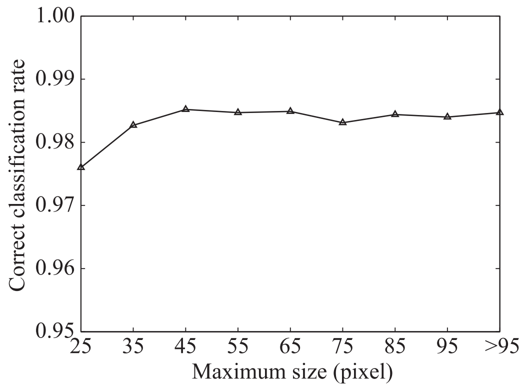

Fig. 3 C-HMAX accuracy in the GTSRB test for various size input images.

Fig. 4 Various samples for class No. 35.

was not as accurate as the Committee of CNNs, but better than H-MAX and multi-scale CNN, with nearly the same accuracy as the human tests.

The recognition accuracy was also evaluated for images scaled from smaller than 25 pixels to greater than 95 pixels. Figure 3 shows the accuracy changed with the maximum image size, which shows that the C-HMAX model can achieve an average accuracy of greater than 98% regardless of the image size. Thus, the C-HMAX model can accurately extract robust scale invariant features from the image as humans do.

Figure 4 shows some samples with various illumination levels for one class of traffic signs, which was distinguished from other classes by the C-HMAX classifier by the color features from the C-HMAX model. The results show that regardless of the brightness, the classifier still gives the correct category information. Thus the C-HMAX model can accurately extract robust color features from input images.

5 Conclusions and Future Work

The C-HMAX model demonstrates that the CS/CC couple layers can extract color features from the input image as in the color vision mechanism of the human brain. Application of the C-HMAX model to classify the GTSRB traffic signs shows that the C-HMAX model is not only scale invariant but also illumination invariant, so it can accurately extract robust color and shape-texture features from visual objects.

More tests will be conducted with the C-HMAX models to further verify the invariant properties of the model. In addition, the complexity of computing the response diagram for each CS/CC couple layer will be studied to simplify the process for practical usage.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (Nos. 90820305, 60775040, and 61005085) and Aeronautical Science Foundation of China (No. 2010ZD01003).

References

[1] D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield, *Analysis of Visual Behavior*. Boston, America: MIT Press, 1982.

[2] M. A. Goodale and A. D. Milner, Separate visual pathways for perception and action, *Trends in Neuroscience,* vol. 15, no. 1, pp.20-25, Jan. 1992.

[3] J. Jin, Y. Wang, H. A. Swadlow, and J. M. Alonso, Population receptive fields of ON and OFF thalamic inputs to an orientation column in visual cortex, *Nat. Neurosci.,* vol. 14, no. 2, pp.232-238, Jan. 2011.

[4] M. S. Gazzaniga, R. B. Ivry, and G. R. Mangun, *Cognitive Neuroscience*, 2nd Ed. New York, America: W W Norton & Co Inc (Np), 2002.

[5] T. J. Bussey and L. M. Saksida, Memory, perception, and the ventral visual-perirhinal-hippocampal stream: Thinking outside of the boxes, *Hippocampus*, vol. 17, no. 9, pp.898-908, Sep. 2007.

[6] T. M. Preuss, I. Stepniewska, and J. H. Kaas, Movement representation in the dorsal and ventral premotor areas of owl monkeys: A microstimulation study, *J. Comp. Neurol.,* vol. 371, no. 4, pp.649-676, Aug. 1996.

[7] R. Gattas, A. P. Sousa, M. Mishkin, and L. G. Ungerleider, Cortical projections of area V2 in the Macaque, *Cereb. Cortex.,* vol. 7, no. 2, pp.110-129, Jul. 1997.

[8] D. Y. Ts'o, M. Zarella, and G. Burkitt, Whither the hypercolumn, *Journal of Physiology,* vol. 587, no. 12, pp.2791-2805, Jun. 2009.

[9] M. Riesenhuber and T. Poggio, Hierarchical models of object recognition in cortex, *Nature Neuroscience,* vol. 2, no. 11, pp.1019-1025, Feb. 1999.

[10] T. Serre and T. Poggio, Reverse-engineering the brain, *Communications of the Association for Computing Machinery,* vol. 53, no. 10, pp.54-61, Oct. 2010.

[11] T. Serre, L. Wolf, and T. Poggio, Object recognition with features inspired by visual cortex, presented at the 2005 IEEE Conference on Computer Vision and Pattern Recognition, San Diego, USA, 2005.

[12] J. Mutch and D. G. Lowe, Multiclass object recognition with sparse, localized features, presented at the 2006 IEEE Conference on Computer Vision and Pattern Recognition, New York, USA, 2006.

[13] N. W. Daw, Goldfish Retina: Organization for simultaneous color contrast, *Science,* vol. 158, no. 3803, pp. 942-944, Nov. 1967.

[14] B. R. Conway, *Neural Mechanisms of Color Vision: Double-Opponent Cells in the Visual Cortex*. Dordrecht, Netherlands: Kluwer Academic Press, 2002.

[15] J. E. Dowling, *Neurons and Networks: An Introduction to Behavioral Neuroscience*, 2nd Ed. Massachusetts, USA: Belknap Press of Harvard University Press, 2001.

[16] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition, *Neural Networks,* vol. 32, no. 1, pp. 323-332, Aug. 2012.

Bo Yang received his BEng degree from Tsinghua University, Beijing, China, in 2007, where currently, he is working towards the PhD degree at the Department of Computer Science and Technology. He was awarded Excellent Graduate of Department of Computer Science and Technology in 2007. He has published several papers on several journals such as *Science in China* and *Frontiers of Electrical and Electronic Engineering* since 2008. His research interests include computational intelligence, computational neuroscience, machine learning, and unmanned ground vehicle.

Lipu Zhou received his BEng degree from Tianjin University, Tianjin, China, in 2005 and MS degree from Peking University, Beijing, China, in 2009, respectively, both in computer science. Currently, he is working towards the PhD degree at the Department of Computer Science and Technology in Tsinghua University. His research interests include pattern recognition and computer vision.

Zhidong Deng received his BEng degree from Sichuan University, Chengdu, China, in 1986 and PhD degree from Harbin Institute of Technology, Harbin, China, in 1991, respectively, both in computer science and automation. From 1992 to 1994, he was a Postdoctoral Researcher at the Department of Computer Science and Technology, Tsinghua University, Beijing, China, where in 1994, he became an associate professor. From 1996 to 1997, he served as a research associate at Hong Kong Polytechnic University, Hong Kong, China. From 2001 to 2003, he was a visiting professor at the Washington University in St. Louis, USA. Currently, he is a professor at the Tsinghua University. His research areas include computational intelligence, complex network theory, computational biology, computational neuroscience, virtual reality, wireless sensor network, and robotics.