A Hierarchical Visual Recognition Model with Precise-Spike-Driven Synaptic Plasticity

Xiaoliang Xu, Xin Jin, Rui Yan, *Member, IEEE,* and Xun Cao

***Abstract*—Several conventional methods have been implemented in pattern recognition, but few of them have biological plausibility. This paper mimics the hierarchical visual system and uses the precise-spike-driven (PSD) synaptic plasticity rule to learn. The well-known HMAX model imitates the visual cortex and uses Gabor filter and max pooling to extract features. Compared with the traditional HMAX model, our modified model combines with the characteristics of sparse coding, and retains the features of the image in each orientation. In learning layer, it is an effective preparation for the PSD rule that temporal coding conveys precise spatio-temporal information. The PSD rule is simple and efficient in synaptic adaptation, and calculates directly. The results show our scheme provides a powerful approach for handwritten digit recognition in noisy conditions.**

***Index Terms*—Pattern recognition, visual hierarchical system, HMAX model, PSD rule, spike neural network (SNN)**

I. Introduction

IMAGE information is a hierarchical transmission and extraction in the primate visual system. The basic structure and functional unit of the information processing is called Receptive field (RF) in primary visual cortex [1]. Activated by photoreceptor cells, RF disposes the spatio-temporal information through the lateral geniculate nucleus (LGN). It is described as a signal extraction modular with localized, oriented, and bandpass [2], [3].

Researchers imitate the primary characteristics for applications within the visual pattern recognition, and find simple cells in the receptive field of the V1 area are used to deal with external stimuli based on the neural sparse coding principle [4]. Sparse coding is a kind of multidimensional data description method. Image information remains in an active state with only a few components after sparse coding. It is observed that the natural images obtains the over-complete base on the neural sparse coding, which shows the shape like the Gabor wavelet [5]. In order to mimic the hierarchical visual cortex, Poggio [6] first uses Gabor filter and max pooling to propose a HMAX model. It adopts the simple S units and complex C units The simple S units within S1 and S2 combine their inputs with edge filter function to increase selectivity. The complex C units within C1 and C2 pool their inputs through a maximum operation to increase invariance. The HMAX model possesses an extremely efficient on feature extraction.

X. Xu and X. Jin are with the College of Computer Science, Hangzhou Dianzi University, China (e-mail: {xxl@hdu.edu.cn, 2547950843@qq.com}).

R. Yan is with Neuromorphic Computing Research Center, College of Computer Science, Sichuan University, China (e-mail: ryan@scu.edu.cn).

X. Cao is with Department of Statistics, Rice University, US (e-mail: caoxun1013@gmail.com).

Corresponding author: X. Jin.

Pattern recognition not only relies on the extracted image information, but also depends on the classifier. Several conventional classifiers are used to implement pattern recognition, such as naive Bayes classifier, decision trees, support vector machine (SVM), and maximum entropy classifier. However, most of these methods lack biological plausibility. Humans easily distinguish different classes in a very short time. It motivates researchers to find a fast and robust model for pattern recognition from a biological view. Artificial neural network (ANN) is abstracted by structure and function of biological neural system. The spiking neural network (SNN) is generally regarded as the third generation of ANN that has more biological authenticity, and adopts spiking trains to express and process messages [7]. Hebb [8] framed a synaptic plasticity hypothesis: “Cells that fire together, wire together”. This hypothesis emphasizes synergistic activity between presynaptic neuron and postsynaptic neuron. In fact, many learning rules of SNN successfully accomplish the pattern classification by training leaky Integrate-and-Fire (LIF) neuron input synaptic weights, such as Spike Timing-Dependent Plasticity (STDP) [9], tempotron [10], and precise-spike-driven synaptic plasticity (PSD) [11].

Inspired by the Widrow-Hoff rule [12], PSD learning rule trains synaptic weights by connecting an input temporal spike pattern with a desired spike sequence. There are many basic and widely studied schemes of encoding information in these spikes, like rate coding and temporal coding. Rate coding considers the firing rate within a short period and the temporal coding cares for the precise time of spikes [13]. The PSD rule is applied to the association and recognition of temporal spatial pattern, so temporal coding is more suitable for it. Compared the error between the desired and the actual output spikes can drive neuron adaptation: The long-term potentiation (LTP) shows the postsynaptic neuron spikes emerges after the presynaptic neuron pulse; On the contrary, it causes long-term depression (LTD) [14]. The predefined post-synaptic threshold is adjusted to the optimal value in order to achieve the stable state of temporal spike pattern. The modification is proportional to a competency input spiking trace. Studied through experimental simulations in [11], the performance of the PSD rule shows it has computational efficiency and biological rationality.

Hierarchical visual system in neural networks is applied to solve pattern recognition problems. Masquelier et al. convert pixels into temporal signals and send them into the HMAX model. Their results show this method is a key to understanding the visual system’s notable processing speed [15]. Garrick et al. propose HFirst to enhance the features of dynamic scenes, so that they can navigate with structured surrounds and complex backgrounds [16]. The hierarchical visual system with convolution deep network also does well in pattern recognition [17]. All of them use the visual system to extract features and achieve satisfactory results.

In this paper, we propose a pattern recognition architecture which mimics the visual hierarchical system and uses the PSD rule to learn. We select HMAX model to extract features of images. Inspired by the sparse coding, we just choose a single Gabor filter window in simple cell units. Encoding is the first step in learning, which mimics how information is represented in the retina. A temporal coding is used to emerge precise time spikes, and the PSD rule learns these spikes. Consistent with biological experimental observations, we set the precise timing window on a millisecond level. By calculating the distance between the output spike trains and the predicted spike trains, the model can be applied to classify real data. The experiments show that our approach has biological basis and efficient anti-noise ability.

The organization of this paper can be summarized as follows: Section 2 introduces the proposed framework of the pattern recognition model. Section 3 describes visual hierarchical system and we will make a few alterations on HMAX model. Section 4 describes the temporal learning rule about the PSD rule. Section 5 shows the experimental results and analysis. Finally, Section 6 summarizes this paper, and discusses the advantages and limitations of the proposed method.

II. The System Architecture

In this section, we describe the whole system architecture for visual pattern recognition. It includes three functional parts: the feature extracting part, the learning part, and the output part. Fig. 1 indicates the general architecture of the system. A stimulus is composed of a few components. Some of important components are encoded and they are connected to the spiking learning neurons to generate spike train.

Each part performs different roles. The feature extracting layer analyzes the useful information from external stimuli. The learning layer encodes the features and sends spikes into the neuron network. The output part extracts information based on neural responses. This whole process helps to solve the problem of getting data into and out of SNN, as well as the pattern recognition task.

A. Feature Extracting Layer

The aim of the feature extracting part is to select the important components of the input stimuli. We adopt the HMAX model [18] to extract image features. The Gabor filter intensifies the edge information of each orientation and weakens the effect of noise on the image. Max pooling extracts feature and reduce the amount of computation for the subsequent learning. In this paper, we only choose the first two layer S1, C1, and make some changes to the calculation model (the traditional model is based on four layers S1, C1, S2, C2). We modify the HMAX model to a single filter window inspired by the association between Gabor filter and sparse coding. In order to retain the characteristics of each orientation in C1,

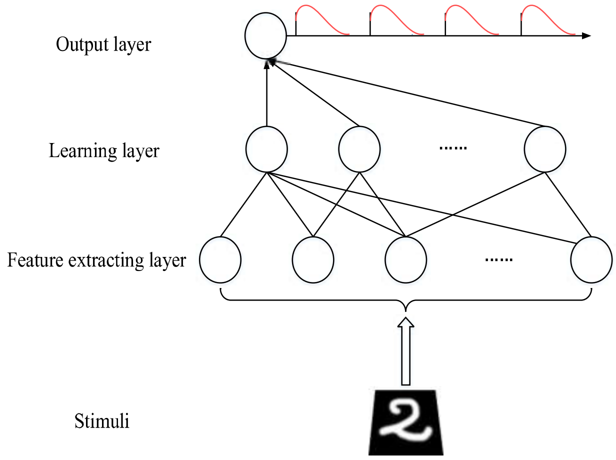

Fig. 1. Architecture of the feedforward computational model for recognition. It includes three functional parts: the feature extracting part, the learning part, and the output part. The feature extracting layer analyzes the useful information from external stimuli. The learning layer encodes the features and sends spikes into the neuron network. The output part extracts information based on neural responses.

we stitch the images of C1 from four orientations to form a single image instead of treating as individual images. Through this modified model, the task of feature extraction could be fulfilled.

B. Learning Layer

The learning part is the most significant layer of our system. In this part, the pixel information is converted to the spiking signals, and then the spiking neural network algorithm is used to learn. So the primary consideration is a coding mechanism of spiking information. The biological neural encoding algorithm of a specific stimulus signals can be divided into three categories: rate coding, temporal coding, and population coding [19]. Rate coding only takes into the effect of high frequency information on spikes, while ignoring the temporal correlation of spike trains. Nevertheless, temporal coding, which makes full use of the information transmission rate, accomplishes conditions that each cell is fired by only one spike [13]. Thereby, we choose the temporal encoding scheme in this paper. Then we send these spikes into the temporal learning rule. Motivated by saving the experimental time, it is better to join a teacher signal to improve the learning speed. In the fourth section, the PSD rule will be introduced. It is a supervised rule and uses the precise time spikes to learn.

C. Output Layer

The purpose of the output layer is to classify stimuli from the responses of spiking neural network. We adopt the van Rossum metric to classify stimuli after learning, which describes clearly in [20]. It measures and analyses the distances between target value and actual value after learning, that is to say, it acts on the result of the spiking neural network. The distances ensures the effectiveness and accuracy of the learning rule.

III. The Visual Hierarchical System

Visual pattern recognition has always been a hot field of research. In order to simulate nervous system better, neurologists start to apply computing technology in brain researches and cognitive sciences. Poggio has been committed to the bionic research of visual system with his team. They find the visual neural system represents fast and invariantly for human action [21] and adopt this system to deal with object recognition [22], natural videos [23], etc. HMAX works as a recognition model for simulating brain structures. The alternating effect of simple cells and complex cells uses template matching and pooling operation to actualize visual image processing. S1 and C1 layer apply the Gabor filter and the max pooling. S2 filter with N previously gets patches which are in C1 format. Each orientation in the patch is matched to the corresponding orientation in C1 layer. The result is one image per C1 band per patch. C2 values are computed by taking a max over all S2 associated with a given patch. Thus, the C2 response has length N. In this paper, we only choose the first two layer S1, C1, and make some changes to the calculation model. The HMAX of visual cortex model demonstrates advantages in the feature extracting layer.

S1 features are generated to represent the classical simple cells in the primary visual cortex. The Gabor filter uses the template matching of V1 simple units in receptive fields [24]. The V1 area with RF generates a result by sparse coding which extracts features with high-order filter [4]. Many researchers use the 2-D Gabor filter function to establish the computation model of sparse coding [25], [26]. The general form of sparse coding is

$$\begin{array}{rcl} sc_i & = & \sum_{i,j} a_i h_j, \\ SC & = & AH \end{array} \tag{1}$$

where sc_i, a_i, h_i are the element of each patch SC, the basis function A of sparse coding, and the sparse coefficients H,respectively. One popular formulation of A indicates

$$\begin{array}{ll} minimize & \|SC - AH\|_f^2 + \lambda \sum_{j=1}^{k} \|s_j\|_1, \\ subject\ to & \quad \|a_i\|^2 \leqslant, \ \forall i = 1, \cdots, m \end{array} \tag{2}$$

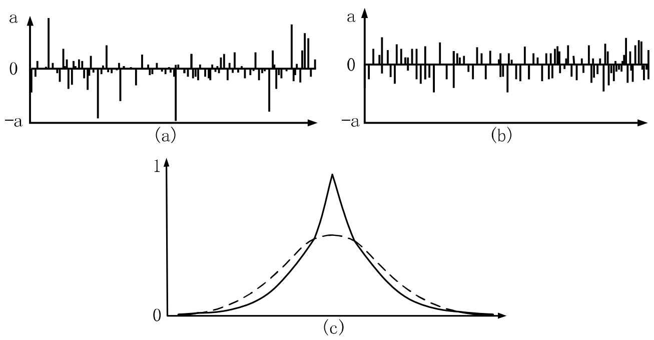

Fig. 2. The relationship between the sparse coding and the Gabor filter. (a) and (b) donates the sparse component and the Gabor kernel function (a kind of Gauss component), respectively. (c) shows the distribution of sparse (solid line) and Gabor kernel function (dotted line).

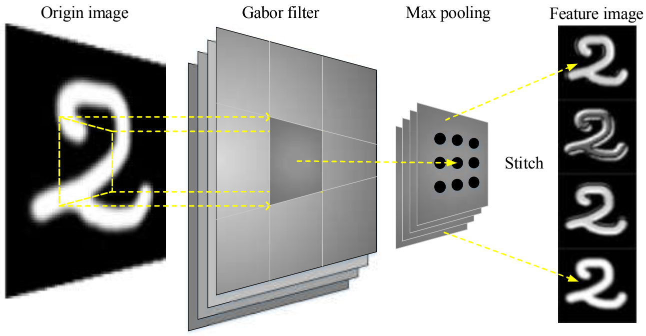

Fig. 3. Illustration of the modified HMAX model. The origin image is the binary image. In S1 layer, we choose a single size of Gabor filter, which extracts spatial features. The max pooling has been applied with C1 layer, which not only ensures the generated image having certain local invariance but also deducts the whole data dimensions.

$\|\cdot\|_f$ is the Frobenius norm and λ is a positive constant. The Gabor response $G(x, y)$ can provide a close form of sparse coding (see Fig. 2) like Eq.(1), which can be described as follows:

$$\begin{array}{rcl} G(x,y) & = & S(x,y) \cdot K(x,y), \\ S(x,y) & = & exp(i(2\pi \frac{x'}{\lambda} + \varphi)), \\ K(x,y) & = & exp(-\frac{x'^2 + \gamma^2 y^2}{2\sigma^2}) \end{array} \tag{3}$$

where $x' = x\cos\theta + y\sin\theta$, and $y' = -x\sin\theta + y\cos\theta$. $S(x, y)$ denotes the complex sine function, its values are not more than 1,and it is satisfied with condition (2). $K(x, y)$ shows the 2-D discrete Gauss function envelope. λ donates the wavelength, which value depends on the pixel of the image. θ determines the direction of Gabor function and φ is the phase deviation. γ donates length-width ratio, and it determines the image's shape after Gabor filter. σ depends on bandwidth. For the sake of simplicity, we only consider the real component of Gabor filter to study features.

$$g(x, y; \lambda, \theta, \varphi, \sigma, \gamma) = exp(-\frac{x'^2 + \gamma^2 y^2}{2\sigma^2}) \cos(2\pi \frac{x'}{\lambda} + \varphi) \tag{4}$$

For our experiment, the wavelength λ and $\sigma = 0.56\lambda$. For simplicity, φ is set to 0. In [27], the author proposes a sparse HMAX model, which has a single size instead of multiple sizes in S1 layer. So we also use a single scale to template matching with Gabor filter and four orientations (0°, 45°, 90°, 135°).

After previous analysis, C1 unit pool over receptive field organizes images from the previous S1 layer with the same orientation. C1 layer uses max pooling [28] to abstract response from different orientations of the images, and then adds them together. To be more specifically, it divides each response into $m \times m$ size and computes the maximum of each window:

$$R(x, y; \lambda, \theta, \varphi, \sigma, \gamma) = \max_{i=1\cdots m, j=1\cdots m} g(x_i, y_j; \lambda, \theta, \varphi, \sigma, \gamma) \tag{5}$$

The invariant responses in four directions are obtained for each group after data through C1 layer. It should be noted that the selection of maximum is not operated crossing directions, which is the key of HMAX. For object recognition, we extract each class of features from the invariant responses generated by C1 layer and save them as the training results.

In this paper, we don't use the S2 and C2 layer. S2 makes a filter like S1, we have not adopted it to simplify the calculation. C2 values are computed by taking a maximum with different orientations from each peach. However, many research shows the sensitivity of the mammalian visual system with different orientations is different [29]–[31]. So it lacks of biological rationality for C2 layer to select the maximum value. In order to guarantee the characteristic of each direction without distorted, we make a simple stitch with images after C1 layer instead of S2 and C2 operation.

In fact, Fig. 3 indicates the generating procedure of spike pattern with our modified HMAX model, which is made some adjustments in this paper. The procedure can be decomposed as follows: Firstly, the binary image is introduced into the Gabor filter, to generate S1 layer with four different orientations. In this paper, we select a single template size, the other parameter settings are consistent with our previous mention. Secondly, after getting the image in four orientations, the max pooling has been applied with C1 layer. Compared with simple cells within V1, the feature extracting layer uses the HMAX model to imitate the cortical complex cells which tend to have larger receptive fields. In the visual cell simulation system, the template matching operation and scale invariance operation are most significant. S1 layer emphasizes the characteristic of the edge in each orientation. We choose $\theta = 0$, for example, the horizontal stripes are retained, while the other directions are weakened. Throughout the Gabor filter, the unique texture features of the image are obtained. C1 layer uses the scale invariance of image. The max pooling not only extracts key features, but also achieves the purpose of dimension reduction. Therefore, the modified HMAX model is effective for image feature extraction.

IV. Precise Spike Driven Synaptic Plasticity

The temporal learning rules aim to study the spatiotemporal pattern. The spike train $s = \{t^f : 1, \cdots, F\}$ expresses the ordered sequence of the spiking time issued from neuron. The spike train is produced by the following:

$$S(t) = \sum_{f=1}^{F} \delta(t - t^f) \tag{6}$$

where t^f is the $f-$th spike firing time, and $\delta(x)$ is the Dirac delta function. When $x = 0$, $\delta(x) = 1$; otherwise, $\delta(x) = 0$. Although there are differences in various supervised learning algorithms of spiking neural networks, the goal of these methods is consistent: for the input spike train $S_i(t)$ and the target spike train $S_d(t)$, it searches the appropriate synaptic weight matrix W of the spiking neural network to make $S_o(t)$ and $S_d(t)$ as close as possible (see Fig. 4).

Precise-spike-driven synaptic plasticity (PSD) is one of the temporal learning rules, which is mainly to deal with information encoded by precise timing spikes. The foundation of PSD rule is the leaky integrate-and-fire (LIF) model. When the rule modifies the synaptic weights, the trained neuron emits one spike with the pattern corresponding to one category $(P+)$, and emits no spike with the pattern corresponding to another category $(P-)$. It called long term potentiation (LTP) and the long term depression (LTD), respectively. The postsynaptic potentials (PSPs) are the sum of the afferent neural weights from all inputting spikes:

$$V(t) = \sum_i w_i \sum_{t_i} K(t - t_i) + V_{rest} \tag{7}$$

where w_i is the synaptic weights and t_i is the firing time of the $i-$th afferent. V_{rest} is the rest potential of the neuron, and K is the double exponential kernel function and it shows as:

$$K(t - t^f) = V_0 \cdot (exp(\frac{-(t - t^f)}{\tau_s}) - exp(\frac{-(t - t^f)}{\tau_f})) \tag{8}$$

where t^f is the $f-$th spike delivered from the $i-$th neuron. V_0 is a normalization parameter which makes the value of K less than 1. τ_s refers to the slow decay constant, and τ_f is the fast one. Set $\tau_s/\tau_f = 4$. The postsynaptic current is one of the most important parameter in the PSD rule, it satisfies the following formula:

$$I^i_{PSC} = \sum_{f=1}^{F} (t - t^f_i) H(t - t^f_i) \tag{9}$$

where I^i_{PSC} is the $i - th$ input synaptic current. $H(t)$ is the Heaviside function, which has the value 0 for $x < 0$, 1 for $x > 0$, and 0.5 for $x = 0$. Synaptic weights of PSD rule are derived from the Widrow-Hoff rules:

$$\Delta w_i = \eta[S_d(t) - S_o(t)] I^i_{PSD}(t) \tag{10}$$

As we all known, the change of synaptic weights can also be obtained by differential:

$$\Delta w_i = \frac{dw_i(t)}{dt} \tag{11}$$

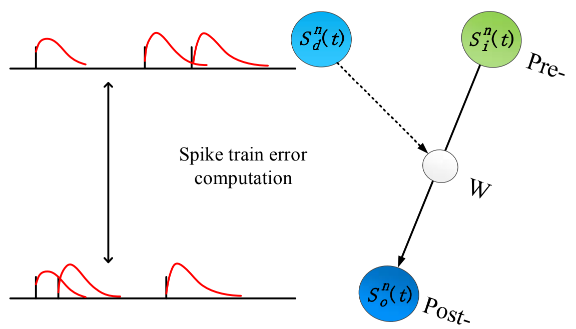

Fig. 4. Basic framework of temporal learning rule. Starting with the random generation of the initial synaptic weight matrix W, the learning process of spiking neural network is divided into three stages: Firstly, the sample data is encoded by a specific method to the spikes $S^n_i(t)$, $n = 1, \cdots, N_i$; Secondly, the spikes is input into the SNN, and obtains the actual output spike train $S^n_o(t)$, $n = 1, \cdots, N_o$ under the learning rule; Then, according to the target spike train $S^n_d(t)$, $n = 1, \cdots, N_o$ calculates errors of the spiking neural network.

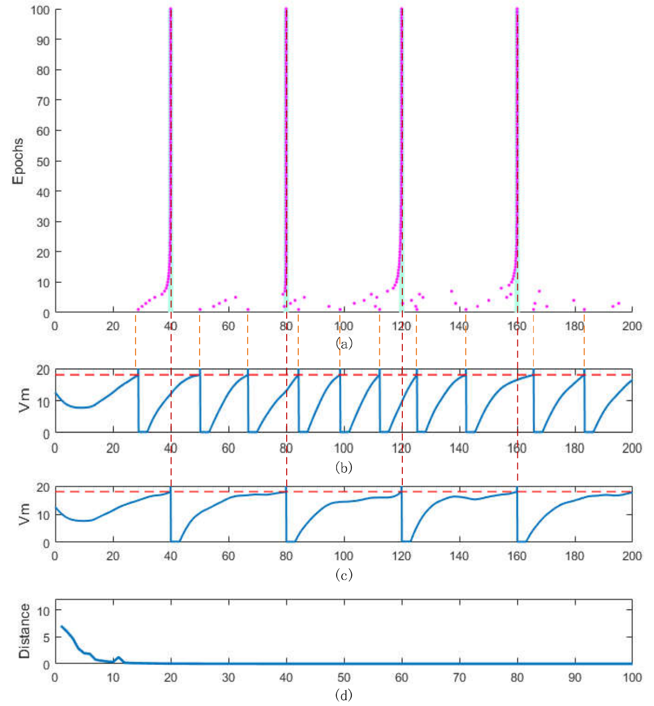

Fig. 5. Typical process of PSD learning. In (a), the red dots are the time of spikes denoted with each epoch. (b) and (c) are the dynamics potential of the neuron before and after learning, respectively. The threshold voltage is shown by the red line. (d) is the distance between the actual output spike train and the target spike train within the whole process.

From formula (6), (10), and (11), the synaptic weights can be described:

$$\Delta w_i = \eta[\sum_{g=1}^{G}\sum_{f=1}^{F} K(t_d^g - t_i^f)H(t_d^g - t_i^f) - \sum_{h=1}^{G}\sum_{f=1}^{F} K(t_o^h - t_i^f)H(t_o^h - t_i^f) \tag{12}$$

V. Experimental Results and Analysis

In this section, we conduct experiments with our visual pattern recognition model, which uses precise-spike-driven synaptic plasticity. The training samples are randomly selected from MNSIT handwritten digital characters dataset. And we randomly add with different degrees of Gauss noise to each selected pictures as our test set. The details are illustrated in the experiment setup.

A. Experiment Dataset and Setup

MNIST dataset is the famous digital handwriting digits of 0-9. The sample size of this database is more than 70 thousand about different digital handwriting. It has a wide range of applications in pattern recognition.

In order to classify noisy image processed from the MNSIT handwritten digital characters database, we adopt single layer spiking neural network, which includes 10 neurons. Each origin image is allocated to one neurons, ten neurons represent 1-10 handwritten digital. Each pixel of one image processed by the feature extracting layer represents an afferent neuron (all origin image size is 28×28, and after feature extracting, the size is $12 \times (12 \times 4)$). In learning layer, we also need to avoid that a single weight is too large to make the results over fitting, so we set the value of synaptic weights are not more than $6nA$. The initial synaptic weights are in accordance with the random sequence of Gauss distribution, which the standard deviation is $0.2nA$, and the mean value is $0.5nA$. For the slow decay constants, we set $\tau_s = 10ms$. The neurons are trained for 100 epochs, and the training data set is formed with 10 groups of samples, which stand $0 - 9$ digits. Each group includes ten images, one is randomly selected from MNSIT handwritten digital characters dataset and the other 9 are generated with a random noise level of 0–5%. In the testing date set, we use the same digital images, but the noise level is up to 20%, and each group of samples is up to 100. Finally, we adopt the van Rossum metric to classify stimuli after learning.

Fig. 5 shows a typical process of PSD learning, we choose the digit 2 and the target spike train $[40, 80, 120, 160]$ ms as an example. In the initial 10 generations, although some membrane voltage is up to the threshold voltage, the neurons are fired in different time sequences. (b) is the initial voltage, from which we can hardly find some useful regulation. There are no regular spikes emerging. With the increasing of iterations, the weights of each neuron are constantly changing, and the spike signals emit within the target spike train slowly. In order to achieve the target spike train, we continue to train. This is called the supervision of learning, which makes the results of the previous study as a teacher signal to impact later studying. The emission rate and output spikes of the membrane voltages are also the same as the time of the target spike train by the final training. (d) is the distance between the actual output spike train and the target spike train within the whole process. By the continuing train of PSD rule, the distance is getting smaller and smaller, and finally tends to 0.

B. Experimental Results and Discussion

In this section, we make some experiments to testify the efficiency of our proposed architecture on noise image classification. In this paper, we use three different sets of data to test. The parameter settings for each samples of training sets have been introduced in experiment setup. The pictures in the test groups are processed from each class. We add the Gauss noise to these images, and the noise level is 0–20%. There are 1000 test images for each noise ratio, which need to be classified.

We restore the PSD rule mentioned in [11]. In principle, different digits correspond to different target spike trains. But in [11], each neuron is trained to produce same target spike train for the optical characters. Because the pixel values for each digit are different and the synaptic weights of target spike train for each are also different, the interactions between pixels and weights of different digits don't achieve the target spike train. That is to say, a pattern from the assigned class is presented, and not to spike when patterns from other classes

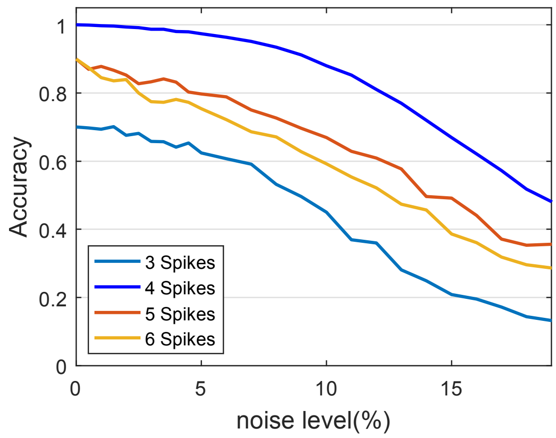

Fig. 6. This is the classification accuracy of different number of spikes in the target spike train.

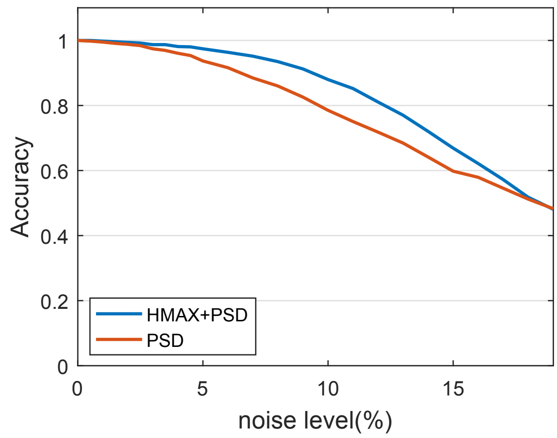

Fig. 7. This is the classification accuracy of different noise level.

TABLE I
CLASSIFICATION PERFORMANCE OF HMAX_PSD

Noise Level(%)	Accuracy			
	Testing1	Testing2	Testing3	Average
0	1	1	1	1
0.5	0.999	0.999	1	0.999
1	0.997	0.996	0.998	0.997
1.5	0.996	0.994	0.996	0.995
2	0.996	0.991	0.994	0.994
2.5	0.995	0.987	0.995	0.992
3	0.987	0.984	0.991	0.987
4	0.987	0.985	0.986	0.986
5	0.979	0.977	0.982	0.979
6	0.983	0.976	0.976	0.978
7	0.978	0.971	0.974	0.974
8	0.963	0.962	0.960	0.961
9	0.942	0.954	0.952	0.950
10	0.928	0.942	0.937	0.936

TABLE II
THE TIME OF HMAX+PSD RULE AND PSD RULE

Time	HMAX+PSD Rule	PSD Rule [11]
Training Time	1362.806	1341.673
Testing Time	4751.336	5260.804
Whole Time	6120.410	6607.617

are presented. But which number of spikes in the target spike train need be considered. Fig. 6 shows the result of different number of spikes in the target spike train. The target spike train includes four spikes will perform better. In order to obtain higher accuracy, we adopt the target spike train $[40, 80, 120, 160]$ ms in the following experiments.

With the HMAX model, PSD rule performs well in noisy image recognition. The classification accuracy of each test set is shown in Table I and the blue line of Fig. 7 (the green line donates the result of PSD rule). In the absence of noise, the classification efficiency of the PSD rule based on HMAX is 100%. It has a high accuracy in the case of a small noise, which is up to 99% with the noise level 2.5%. Since the key features are extracted by HMAX model and the unimportant information is ignored, the anti-noise ability of images is significantly improved. However, with the increasing of the noise, the recognition rate of the algorithm is gradually reduced, because the noise interference is too strong to make the images lost the original features. But when the noise level is less than 10%, its accuracy still reaches more than 90%. It is reveal that the PSD rule with modified HMAX model is more useful than just using PSD rule.

Our visual pattern recognition model is effective not only in accuracy, but also in time (see Tab. II). There is that $wholetime = trainingtime + Testingtime + others$. Others is some functions that take a little time. Although in training period, our scenario spends more time, the total experimental time is still shorter than the PSD rule. During the training within our scheme, the images are input into the HMAX model, it will generate extra time; besides, the max pooling achieves the purpose of dimensionality reduction. Because the data of training set is relatively small, the advantage is not obvious, the time is more than the PSD rule. When starting to process the test set, the large number of data after dimensionality reduction will save much time.

Therefore, based on the above reason, our paper combines the advantage of the modified HMAX model and PSD learning rule. Experiments prove that our scheme has a good biological basis and shows the superiority in noise image recognition.

VI. CONCLUSION

A pattern recognition model mimics the visual hierarchical system and uses the PSD learning rule has been presented. The HMAX model adopts template matching and max pooling to possess the features. In this paper, inspired by the association between Gabor filter and sparse coding, we set a single size in the S1 layer. Since C2 selects maximum by crossing directions, which is lack of biological rationality, we don't choose the "patches" in C2. We stitch the images after C1 from four orientation to retain the characteristics of each orientation.

The spike pattern generated by temporal coding conveys significant spatiotemporal information about the input data. PSD rule is designed to process precise timing spikes, where one afferent neuron fires a single spike during the whole time window. It is simple and efficient in synaptic adaptation, which reduces the number of signal sources to calculate directly. PSD rule changes the synaptic weights by the error computation between the real spike train and the target spike train. Notably, the weight modification just relies on the currently deviations and does not affect future results. The PSD rule is terminated when the real spike train is close unlimitedly the target spike train. Such practical ways might be useful for noisy image recognition.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China under Grant No. 61603119 and Zhejiang Provincial Natural Science Foundation of China under Grant No. LY17F020028 and No. LQ14F020013.

REFERENCES

[1] C. D. Gilbert and T. N. Wiesel, "Receptive field dynamics in adult primary visual cortex." *Nature*, vol. 356, no. 6365, pp. 150–152, 1992.

[2] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *Journal of Physiology*, vol. 160, no. 1, pp. 106–154, 1962.

[3] A. C. Bovik, M. Clark, and W. S. Geisler, "Multichannel texture analysis using localized spatial filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 55–73, 1990.

[4] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?" *Vision Research*, vol. 37, no. 23, pp. 3311–3325, 1997.

[5] E. Doi and M. S. Lewicki, "Sparse coding of natural images using an overcomplete set of limited capacity units," *Nips*, pp. 377–384, 2005.

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

[7] J. Brea, W. Senn, and J. P. Pfister, "Matching recall and storage in sequence learning with spiking neural networks," *Journal of Neuroscience*, vol. 33, no. 33, pp. 9565–9575, 2013.

[8] D. O. Hebb, "The organization of behavior; a neuropsychological theory,"*John Wiley , Chapman and Hall*, 1949.

[9] S. Song, K. D. Miller, and L. F. Abbott, "Competitive hebbian learning through spike timing-dependent plasticity," *Nature Neuroscience*, vol. 3, no. 9, pp. 919–926, 2000.

[10] R. Gtig and H. Sompolinsky, "The tempotron: a neuron that learns spike timing-based decisions," *Nature Neuroscience*, vol. 9, no. 3, pp. 420–428, 2006.

[11] Q. Yu, H. Tang, K. C. Tan, Li, and Haizhou, "Precise-spike-driven synaptic plasticity: learning hetero-association of spatiotemporal spike patterns," *Plos One*, vol. 8, no. 11, pp. : e78318, 2013.

[12] S. Hui and S. H. Zak, "The widrow-hoff algorithm for mcculloch-pitts type neurons," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 924–9, 1994.

[13] R. V. Rullen and S. J. Thorpe, "Rate coding versus temporal order coding: What the retinal ganglion cells tell the visual cortex," *Neural Computation*, vol. 13, no. 6, pp. 1255–1283, 2001.

[14] N. Caporale and Y. Dan, "Spike timing-dependent plasticity: a hebbian learning rule," *Annual Review of Neuroscience*, vol. 31, no. 1, pp. 25–46, 2008.

[15] T. Masquelier, S. J. Thorpe, "Unsupervised Learning of Visual Features through Spike Timing Dependent Plasticity," *Plos Computational Biology*, vol. 3, no. 2, pp. : 247–257, 2007.

[16] G. Orchard, C. Meyer, R, Etiennecummings, "HFirst: A Temporal Approach to Object Recognition," *IEEE Transactions on Pattern Analysis Machine Intelligence*, vol. 37, no. 10, pp. : 2028–2040, 2015.

[17] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Unsupervised learning of hierarchical representations with convolutional deep belief networks," *Communications of the Acm*, vol. 54, no. 10, pp. 95–103, 2011.

[18] A. W. Bitar, M. M. Mansour, and A. Chehab, "Algorithmic Optimizations in the HMAX Model Targeted for Efficient Object Recognition,"*Springer International Publishing*, vol. 598, pp. 374-395, 2015.

[19] Q. R. Quian, "Principles of neural coding," *Current Biology Cb*, vol. 21, no. 19, pp. 798–799, 2011.

[20] M. C. van Rossum, "A novel spike distance," *Neural Computation*, vol. 13, no. 4, pp. 751–763, 2001.

[21] L. Isik, A. Tacchetti, and T. Poggio, "Fast, invariant representation for human action in the visual system," 2016.

[22] L. Isik, E. M. Meyers, J. Z. Leibo, and T. Poggio, "The dynamics of invariant object recognition in the human visual system," *Journal of Neurophysiology*, vol. 111, no. 1, pp. 91–102, 2013.

[23] Q. Liao, J. Z. Leibo, and T. Poggio, "Unsupervised learning of clutter-resistant visual representations from natural videos," *Computer Science*, 2014.

[24] P. Dayan and L. F. Abbott, "Theoretical neuroscience: Computational and mathematical modeling of neural systems," *Quarterly Review of Biology*, pp. 154–155, 2005.

[25] L. Shang, W. Huai, G. Dai, J. Chen, and J. Du, "Palmprint recognition using 2d-gabor wavelet based sparse coding and rbpnn classifier," in *Advances in Neural Networks - Isnn 2010, International Symposium on Neural Networks, Isnn 2010, Shanghai, China, June 6-9, 2010, Proceedings*, pp. 112–119, 2010.

[26] M. Abdolali and M. Rahmati, "Facial expression recognition using sparse coding," *Machine Vision and Image Processing*, pp. 150–153, 2013.

[27] X. Hu, J. Zhang, J. Li, and B. Zhang, "Sparsity-regularized hmax for visual recognition," *Plos One*, vol. 9, no. 1, pp. e81 813, 2014.

[28] A. Giusti, D. C. Ciresan, and J. Masci, "Fast Image Scanning with Deep Max-Pooling Convolutional Neural Networks," *Computer Sciences*, pp. 4034–4038, 2013.

[29] F. L. Kitterle and R. S. Kaye, "Hemispheric symmetry in contrast and orientation sensitivity," *Perception and Psychophysics*, vol. 37, no. 5, pp. 391–396, 1985.

[30] F. Thoss, B. Bartsch, D. Tellschaft, and M. Thoss, "The light sensitivity of the human visual system depends on the direction of view," *Journal of Comparative Physiology A Neuroethology Sensory Neural and Behavioral Physiology*, vol. 188, no. 3, pp. 235–237, 2002.

[31] A. V. Slavutskaia, G. Niu, and E. S. Mikhailova, "mechanisms of human orientation sensitivity of the vision system. part i. behavioral characteristics of human orientation sensitivity. influence of task type, experimental conditions and gender," *Human Physiology*, vol. 40, no. 40, pp. 660–668, 2014.