EXTRACTING THE BRAIN-LIKE REPRESENTATION BY AN IMPROVED SELF-ORGANIZING MAP FOR IMAGE CLASSIFICATION

Jiahong Zhang,^{1,2} Lihong Cao, ^{1,2} Moning Zhang, ^{1,2} Wenlong Fu^{1,2}

¹ State Key Laboratory of Media Convergence and Communication, Communication University of China, ² Neuroscience and Intelligent Media Institute, Communication University of China, Beijing,China

ABSTRACT

Backpropagation-based supervised learning has achieved great success in computer vision tasks. However, its biological plausibility is always controversial. Recently, the bioinspired Hebbian learning rule (HLR) has received extensive attention. Self-Organizing Map (SOM) uses the competitive HLR to establish connections between neurons, obtaining visual features in an unsupervised way. Although the representation of SOM neurons shows some brain-like characteristics, it is still quite different from the neuron representation in the human visual cortex. This paper proposes an improved SOM with multi-winner, multi-code, and local receptive field, named mISOM. We observe that the neuron representation of mISOM is similar to the human visual cortex. Furthermore, mISOM shows a sparse distributed representation of objects, which has also been found in the human inferior temporal area. In addition, experiments show that mISOM achieves better classification accuracy than the original SOM and other state-of-the-art HLR-based methods. The code is accessible at https://github.com/JiaHongZ/mlSOM.

Index Terms— Self Organizing Maps, Unsupervised Learning, Image classification

1. INTRODUCTION

Backpropagation-based supervised learning has been extensively studied in recent years. For image classification, the adoption of backpropagation enables convolutional neural networks (CNNs) to extract features effectively [1, 2, 3], thus improving classification performance. However, the biological plausibility of the backpropagation is always controversial [4].

The Hebbian learning rule (HLR) is a biologically plausible unsupervised learning mechanism and has been proposed for a long time [6, 7], which suggests that: "Neurons that fire together wire together." In a broad sense, the HLR refers to a family of methods based on the idea of Hebbian. Unfortunately, vanilla HLR does not guarantee high performance



Fig. 1. Visualization of the neuron representations: (a) Hebb-Net, (b) Backpropagation, (c) SOM for handwritten digits. (d) TE for objects (Image Source: [5]). Colored bars in the top (red), middle (green) and bottom graphs in (d) are penetration sites inside the active spots of the stimulus.

for image classification. Recently, several methods have been proposed to improve the classification accuracy for HLR [8, 9, 10]. Self-Organizing Map (SOM) uses the Winner-Takes-All competition HLR to establish connections between neurons [11, 12], which achieves high classification performance. However, they failed to obtain brain-like neuron representation. In human visual cortex, the representation of an object presents a topological structure [5]. For example, Fig. 1 (d) shows the representation of a complex object and its parts in the anterior part of the IT cortex (architectonically defined as area TE). By comparison, it can be found that the neuron representations of existing HLR methods (Fig. 1 (b) and (c)) and the backpropagation method (Fig. 1 (a)) lack the object parts.

This paper proposes an improved SOM, mlSOM. Compared with the original SOM, three modifications are made in mlSOM: from global receptive field (GRF) to local receptive

The corresponding author is Lihong Cao (lihong.cao@cuc.edu.cn).



Fig. 2. The architecture of mISOM. For a given input image, mISOM first divide it into patches through a sliding window. Then for each patch, its WNs in the hidden layer are calculated in the SOM phase. In the Coding phase, the neuron states of the hidden layer will be coded to a feature map, in which 1 denotes the WN. The feature map is then send to the linear classifier to obtain the classification result.

Table 1. The hyper-parameters of mlSOM.

Datasets	Hyper-parameters							
	hidden neurons	w	S	n	σ	k	lr	
MNIST	44×44	14×14	7	5	2	20	0.3	
CIFAR-10	44×44	16×16	4	5	2	100	0.3	

field (LRF), from single-winner to multi-winner, and from single-code to multi-code. The main contributions of this work contain at least three key advantages. Firstly, we propose to improve the representation of SOM from a brain-like perspective, which may contribute to a promising future research direction for SOM. Secondly, we improve the original SOM in three ways inspired by the human brain and demonstrate their effectiveness for classification. Thirdly, the proposed mlSOM shows brain-like representation and gets high classification performance compared with other state-of-theart Hebbian learning-based methods.

2. RELATED WORK

Self-Organizing Map (SOM) is a kind of HLR-based neural network. For image classification, current studies for SOM mainly focus on improving classification accuracy. Supervised SOM was proposed in [13] and got good classification results. Some work presented that deep SOM would get higher classification performance than the single-layer SOM [14, 15]. Combining SOM and CNN to obtain both accuracy and biological plausibility has also attracted widespread interest [16, 17, 18, 19]. This paper presents a new idea to improve SOM from the neuron representation perspective.

3. PROPOSED METHOD

The proposed mlSOM is based on the unsupervised SOM algorithm mentioned in [14]. We first revisit it.

3.1. The original SOM

SOM is a classic unsupervised learning algorithm using the "winner-take-all" learning rule, which gets a non-linear projection of high-dimensional data over a small space. Each neuron in SOM consists of a trainable vector. SOM computes Euclidean distances between the input pattern and each neuron. The neuron that has the least distance is the winner neuron (WN). WN and its neighbors will be updated to be closer to the input pattern during training. The update value decreases as the distance between neurons and WN increases.

3.2. mlSOM

Fig. 2 shows the architecture of mlSOM whose hyperparameters are illustrated in Table 1. mlSOM is based on SOM, and the three modifications in mlSOM are as follows.

1) From GRF to LRF. SOM computes the distance between the whole input image and hidden layer neurons, leading to huge neuron vectors. Inspired by the human eye movement, we propose to use LRF, which can be realized by a sliding window. As shown in Fig. 2, the size of the input image is (H, W) and the window size is set to (w, w) with stride (s, s). The size of the neuron vector in mlSOM is w^2 , which is $(\frac{w^2}{H \times W})$ of the original SOM.

2) From single-winner to multi-winner. The original SOM uses the "winner-take-all" learning rule. It computes

_	8.8.
Inpu	it:
1	raining set of images and labels (\mathcal{X}, Y)
Out	put: Trained mISOM
1: i	nitialize the model parameter W with the standard normal dis-
t	ribution and hyper-parameters in Table 1
l	r, epochs \leftarrow initialize the learning rate, epochs
2: 9	% SOM Phase (unsupervised)
3: f	or $epo \in epochs do$
4:	for $x \in \mathscr{X}$ do
5:	$x_p \leftarrow Sliding(x) \ \%$ image patches obtained by the slid-
	ing window;
6:	for $x_{p_i} \in x_p$ do
7:	% distance of x_{p_i} and neurons in mISOM
8:	for $W_j \in W$ do
9:	$d_{ij} \leftarrow \ x_{p_i} - W_j\ $
10:	end for
11:	sort d_{ij} in ascending order;
12:	WNs \leftarrow the top $n W$;
13:	$[(X_{WN}, Y_{WN})] \leftarrow$ the coordinate of the first n WNs;
14:	% update learning rate
15.	$lr \leftarrow lr \times (1 - \frac{epo}{epo})$
16.	$f_{op} W N \subset W N do$
10:	$\begin{array}{c} \text{Ior } W N_i \in W N S \text{ do} \\ \text{for } W \in W J \\ \end{array}$
1/:	IOF $W_j \in W$ do
18:	% compute updating neuron vector decay value;
10	$\frac{\ (X_{WN_i}, Y_{WN_i}) - (X_{W_j}, Y_{W_j})\ _2}{2}$
19:	$aecuy = e \qquad 2\sigma^2 \qquad ;$
20:	$\Delta W = i r_{epo} \times aecay \times (W_i - W_N);$
21:	$W_i \leftarrow W_i + \Delta W$;
22:	
23:	end for
24: 25	
25:	end for
26: e	nd for
27: 9	6 Coding phase (supervised)
28: f	or $epo \in epochs$ do
29:	for $x \in \mathscr{X}$ do
30:	$x_p \leftarrow Sliding(x);$
31:	for $x_{p_i} \in x_p$ do
32:	$G_p \leftarrow$ binary 2D grid with k WNs of the hidden layer
	for x_p ;
33:	$G_{sum} = Binary(\sum G_p);$
34:	feature map $\leftarrow G_{sum}$;
35:	% prediction of the classifier
36:	pre = f(featuremap);
37:	minimize $L(pre, Y)$;
38:	end for
	and for
39:	end for

Algorithm 1 Learning algorithm for mISOM

Euclidean distances between the input image and neurons in the hidden layer and chooses one WN. However, population coding widely exists in the human visual cortex [20]. It motivates us to change the single-winner to multi-winner. As shown in Fig. 2 SOM Phase, for every image patch, the hidden layer of mlSOM obtains one 2D grid with n winners. For every winner, the vector-updating algorithm is similar to the original SOM.

3) From single-code to multi-code. The original SOM and some deep SOMs use the 2D grid with a single WN as the classification feature map. Also inspired by the population coding, mlSOM generates the feature map with multiple WNs, as shown in Fig. 2 Coding Phase. Specifically, mlSOM uses neurons with the first k WNs to achieve the multi-code. Here, k can be different from the multi-winner n. In Fig. 2 Coding phase, the 2D grids of input image patches are transformed to the corresponding binary matrixs, in which 1 denotes the WN. These matrices are summed together and binarized as the feature map of the input image.

mlSOM is an unsupervised learning algorithm. We trained a linear classifier to classify images using their ml-SOM feature maps.

3.3. Training method

The training method of mlSOM is shown in Algorithm 1. This training process can be divided into two phases. In the SOM phase, an input image is firstly divided into patches by the sliding window. Then, Euclidean distances of the image patches and the hidden layer neurons are computed. The top n neurons with minimum distance will be selected as WNs. For each neuron of the WNs, the algorithm updating the neuron vectors in mlSOM is similar to the original SOM.

When the SOM phase is finished, we get a trained hidden layer. During the coding phase, for each input image patch, a corresponding 2D grid is obtained from the hidden layer of mlSOM. We convert this 2D grid into a binary coding matrix, in which 1 denotes the WN. Here, the number of 1 in the coding matrix is set to k. Then, we obtain the sum of all the patch grids and binarize it to get the feature map representing a specific object. To verify the classification ability of the feature map, we train a linear classifier with the help of error backpropagation. The training object is minimizing the crossentropy loss:

$$L(x) = \sum_{c=1}^{N} y log(f(x)),$$
 (1)

where x denotes the input image, $f(\cdot)$ denotes the classifier, and y denotes its label.

4. EXPERIMENTS

4.1. Experiment results

We use two datasets MNIST[21], CIFAR-10[22] to evaluate the proposed mISOM. This section compares mISOM to some popular Hebbian learning-based methods. We choose the methods with a single hidden layer for a fair comparison. As shown in Table 2, mISOM performs significantly better classification results than that of HebbNet and SOM, achieving 96.79% test accuracy on MNIST. According to the ablation experiments results in Table 2, the three modifications,



Fig. 3. (a) Neuron representations learnt by the mlSOM after training on MNIST. (b) Visualization of the feature maps and neuron representations of the specific input images. The top row shows the input images and their feature maps. Yellow dots in feature maps represent the WNs. The bottom row shows the neuron vectors of the WNs.

 Table 2. Classification accuracy results and ablation experiments results of mISOM on the MNIST dataset.

Method	Test Accuracy		
Vanilla Hebbian	10.28		
HebbNet[9]	93.25		
SOM	93.07		
DSOM [14]	96.17		
SOM+mult-winner	95.40		
SOM+mult-winner+LRF	96.72		
mlSOM	96.79		

LRF, multi-winner, and multi-code, effectively contribute to better performance. Table 3 shows the results on CIFAR-10. It can be observed that mISOM can achieve competitive results with the state-of-the-art methods. Furthermore, mISOM trades off the classification accuracy and the convergence speed.

4.2. Neuron representations of mISOM

According to Fig. 1 and Fig. 3, the neuron representation obtained by mISOM exhibits brain-like characteristics which are similar to the neuron activity detected in the human TE [5]. Specifically, we found that neurons in mISOM respond to the whole and part of the object. Furthermore, mISOM presents a brain-like distributed coding method. Fig. 3 (b) shows the feature map and neuron representations in mISOM for digits two and eight. Taking the digit two as an example, its representations in mISOM consist of the arcs, slashes, object features, and corresponding parts. These representations are observed in the human visual cortex [23]. It is worth noting that neurons representing slashes are also detected in digit eight. This evidence suggests that neurons in mISOM can respond to fea-

Table 3. Classification accuracy results on the CIFAR-10dataset.

Method	Train Acc	Test Acc	Epochs
Vanilla Hebbian	11.56	15.23	200
BackProp	39.89	41.28	200
Krotov et al. [8]	55.05	50.75	1500
Amato et al. [10]	-	41.78	20
HebbNet [9]	43.13	45.69	200
mlSOM	51.82	43.65	200

ture combinations of different objects, contributing to a larger encoding capacity. It is very similar to the sparse distributed representation in human IT [24, 25, 26].

5. CONCLUSION

This paper proposes an improved SOM method, mlSOM, which achieves better classification accuracy than other Hebbian learning-based methods on MNIST and competitive accuracy on CIFAR-10. mlSOM makes improvements based on the brain-inspired designs, including LRF, multi-winner, and multi-code. Ablation experiments show the effectiveness of these three modifications. The most significant contribution of mlSOM is that it exhibits brain-like neuronal representations and coding. Our research may inspire the design of visual cortex computing model and provide a novel direction for SOM research.

6. ACKNOWLEDGMENT

This paper is supported by supported by the STI 2030—Major Projects (grant No. 2021ZD0200300) and the National Natural Science Foundation of China (grant No. 62176241).

7. REFERENCES

- Alex Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, no. 2, 2012.
- [2] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.
- [3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
- [4] R. C. O'Reilly and Y. Munakata, "Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain," *MIT Press*, 2000.
- [5] Kazushige Tsunoda, Yukako Yamane, Makoto Nishizaki, and Manabu Tanifuji, "Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns," *Nature neuroscience*, vol. 4, no. 8, pp. 832–838, 2001.
- [6] D Allport, "Distributed memory, modular systems and dysphasia— bibsonomy," *Current Perspectives in Dysphasia*, 1985.
- [7] A Harry Klopf, Brain function and adaptive systems: a heterostatic theory, Number 133. Air Force Cambridge Research Laboratories, Air Force Systems Command, United ..., 1972.
- [8] Dmitry Krotov and John J Hopfield, "Unsupervised learning by competing hidden units," *Proceedings of the National Academy of Sciences*, vol. 116, no. 16, pp. 7723–7731, 2019.
- [9] Manas Gupta, ArulMurugan Ambikapathi, and Savitha Ramasamy, "Hebbnet: A simplified hebbian learning framework to do biologically plausible learning," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* IEEE, 2021, pp. 3115–3119.
- [10] Giuseppe Amato, Fabio Carrara, Fabrizio Falchi, Claudio Gennaro, and Gabriele Lagani, "Hebbian learning meets deep convolutional neural networks," in *International Conference on Image Analysis and Processing*. Springer, 2019, pp. 324–334.
- [11] David E Rumelhart and David Zipser, "Feature discovery by competitive learning," *Cognitive science*, vol. 9, no. 1, pp. 75– 112, 1985.
- [12] C. Gielen, "Neural computation and self-organizing maps, an introduction," *Neurocomputing*, vol. 5, no. 4-5, pp. 243–244, 1992.
- [13] Willem Melssen, Ron Wehrens, and Lutgarde Buydens, "Supervised kohonen networks for classification problems," *Chemometrics and Intelligent Laboratory Systems*, vol. 83, no. 2, pp. 99–113, 2006.
- [14] Nan Liu, Jinjun Wang, and Yihong Gong, "Deep selforganizing map for visual classification," in 2015 international joint conference on neural networks (IJCNN). IEEE, 2015, pp. 1–6.
- [15] Chathurika S Wickramasinghe, Kasun Amarasinghe, and Milos Manic, "Deep self-organizing maps for unsupervised image classification," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 11, pp. 5837–5845, 2019.

- [16] Ehsan Mohebi and Adil Bagirov, "A convolutional recursive modified self organizing map for handwritten digits recognition," *Neural Networks*, vol. 60, pp. 104–118, 2014.
- [17] Mohamed Sakkari and Mourad Zaied, "A convolutional deep self-organizing map feature extraction for machine learning," *Multimedia Tools and Applications*, vol. 79, no. 27, pp. 19451– 19470, 2020.
- [18] Hiroshi Dozono, Gen Niina, and Satoru Araki, "Convolutional self organizing map," in 2016 international conference on computational science and computational intelligence (CSCI). IEEE, 2016, pp. 767–771.
- [19] Saleh Aly and Sultan Almotairi, "Deep convolutional selforganizing map network for robust handwritten digit recognition," *IEEE Access*, vol. 8, pp. 107035–107045, 2020.
- [20] Huw DR Golledge, Stefano Panzeri, Fashan Zheng, Gianni Pola, Jack W Scannell, Dimitrios V Giannikopoulos, Roger J Mason, Martin J Tovée, and Malcolm P Young, "Correlations, feature-binding and population coding in primary visual cortex," *Neuroreport*, vol. 14, no. 7, pp. 1045–1050, 2003.
- [21] Arthur Asuncion and David Newman, "Uci machine learning repository," 2007.
- [22] Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton, "Cifar-10 (canadian institute for advanced research)," URL http://www. cs. toronto. edu/kriz/cifar. html, vol. 5, no. 4, pp. 1, 2010.
- [23] Eric R Kandel, James H Schwartz, Thomas M Jessell, Steven Siegelbaum, A James Hudspeth, Sarah Mack, et al., *Principles* of neural science, vol. 4, McGraw-hill New York, 2000.
- [24] Edmund T Rolls and Martin J Tovee, "Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex," *Journal of neurophysiology*, vol. 73, no. 2, pp. 713– 726, 1995.
- [25] Edmund T Rolls, Alessandro Treves, Martin J Tovee, and Stefano Panzeri, "Information in the neuronal representation of individual stimuli in the primate temporal visual cortex," *Journal of computational neuroscience*, vol. 4, no. 4, pp. 309–333, 1997.
- [26] Leonardo Franco, Edmund T Rolls, Nikolaos C Aggelopoulos, and Jose M Jerez, "Neuronal selectivity, population sparseness, and ergodicity in the inferior temporal visual cortex," *Biological cybernetics*, vol. 96, no. 6, pp. 547–560, 2007.