



Region-Enhanced Multi-layer Extreme Learning Machine

Xibin Jia¹ · Xiaobo Li² · Ya Jin² · Jun Miao³

Received: 8 October 2016 / Accepted: 19 September 2018 / Published online: 26 September 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

Deep neural networks have made significant achievements in representation learning of traditionally man-made features, especially in terms of complex objects. Over the decades, this learning process has attracted thousands of researchers and has been widely used in the speech, visual, and text recognition fields. One deep network multi-layer extreme learning machine (ML-ELM) achieves a good performance in representation learning while inheriting the advantages of faster learning and the approximating capability of the extreme learning machine (ELM). However, as with most deep networks, the ML-ELM's algorithmic performance largely depends on the probability distribution of the training data. In this paper, we propose an improved ML-ELM made via using the local significant regions at the input end to enhance the contributions of these regions according to the idea of the selective attention mechanism. To avoid involving and exploring the complex principle of the attention system and to focus on the clarification of our local regional enhancement idea, the paper only selects two typical attention regions. One is the geometric central region, which is normally the important region to attract human attention due to the focal attention mechanism. The other is the task-driven interest region, with facial recognition as an example. The comprehensive experiments are done on the three public datasets of MNIST, NORB, and ORL. The comparison experiment results demonstrate that our proposed region-enhanced ML-ELM (RE-ML-ELM) achieves performance increases in important feature learning by utilizing the apriori knowledge of attention and has a higher recognition rate than that of the normal ML-ELM and the basic ELM. Moreover, it benefits from the non-iterative parameter training method of other ELMs, and our proposed algorithm outperforms most state-of-the-art deep networks such as deep belief network (DBN), in the aspects of training efficiency. Furthermore, because of the deep structure with fewer hidden nodes at each layer, our proposed RE-ML-ELM achieves a comparable training efficiency to that of the ML-ELM but has a higher training speed with the basic ELM, which is normally the width single network that has more hidden nodes to obtain the similar recognition accuracy with the deep networks. Based on our idea of combining the apriori knowledge of the human selective attention system with the data learning, our proposed region-enhanced ML-ELM increases the image classification performance. We believe that the idea of intentionally combining psychological knowledge with the most algorithms based on data-driven learning has the potential to improve their cognitive computing ability.

Keywords Local significant region · Selective attention mechanism · Region-enhanced ML-ELM · Sage classification

✉ Jun Miao
jmiao@bistu.edu.cn
Xibin Jia
jiaxibin@bjut.edu.cn
Xiaobo Li
lixiaobo@emails.bjut.edu.cn
Ya Jin
jinya@emails.bjut.edu.cn

¹ Beijing Municipal Key Laboratory of Multimedia and Intelligent Software Technology, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

Introduction

Deep neural networks have led to a revolution in the areas of machine learning. These networks achieve state-of-the-art performance in numerous applications, such as audio

² Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

³ Beijing Key Laboratory of Internet Culture and Digital Dissemination Research, School of Computer Science, Beijing Information Science and Technology University, Beijing 100101, China

analysis and computer vision [1–3]. However, most deep neural networks, such as the deep Boltzmann machine (deep belief network [DBN]) [4], require long training time because of the iterative parameter tuning based on the back-propagation algorithm. Meanwhile, the extreme learning machine proposed by Huang et al. [5] and its variants [6–8] have fast training speeds and good generalization capabilities. Huang also proves that, in contrast to the common knowledge and conventional neural network learning tenets, hidden nodes/neurons do not need to be iteratively tuned for wide types of neural networks and learning models [9]. Therefore, the ELM and ELM variants achieve higher training speeds without iterative parameter tuning. Similar to deep networks, one of extended ELMs, called the multi-layer extreme learning machine (ML-ELM), stacks the extreme learning machine auto encoder (ELM-AE) layer by layer [10]. The ELM-AE learns the deep features faster than the stacked auto encoders. However, as with most deep learning algorithms, the ML-ELM learns the features totally from data and the learning performance largely depends on the probability distribution of data. Actually, some prior knowledge of human cognition may be used to facilitate the discriminative deep feature learning.

People can learn new concepts successfully from a single example, while machine learning algorithms typically require tens or hundreds of examples to produce similar accuracies. Artificial intelligence aims to make the machine have the perceptive abilities of humans. To improve the algorithm performance, one of the general methods is to understand the perceptive methods of people and propose a solution that mimics that of humans. According to our understanding, people can quickly focus on the main interesting parts of an image when performing recognition. The interesting parts contain the most significant information to aid cognition, while the other parts of the image normally offer assisting information. Based on this understanding, we suggest modifying the ML-ELM by utilizing the human perception method. From the point of view of mathematics, deep networks are just functions of inputs. In other words, the outputs map the inputs. Therefore, if the important parts of the input are carefully enhanced, the output will carry more information about these parts. Based on this theory, we illustrate our solution of the local region-enhanced ML-ELM in the following parts.

In this work, we have focused on two points. Firstly, find the local significant region of the input. Normally, the significant parts of different data are different. It is necessary to detect the important parts of the input data. Detecting interesting parts is generally a difficult research problem. Before training a model, people usually preprocess the data. The most important part of the data is generally aligned in the middle part of the data. Therefore, we choose the central local region

as the local significant region. Second, train the ML-ELM network with the local significant region. The traditional ML-ELM takes the data as the input. When enhancing the importance of the local significant region in the deep network, the network will contain more information, and thus we take the data and the local significant region as inputs to train the deep network. The results of the experiments show that our model performs better than the other algorithms.

Brief Review of ML-ELM

Extreme Learning Machine

The ELM is a single hidden-layer feedforward neural network that was proposed by Huang et al. [5]. The input data is mapped to an L -dimensional ELM random feature space and the network is given by Eq. 1:

$$f_L(x) = \sum_{i=1}^L \beta_i h_i(x) = h(x)\beta \quad (1)$$

where $\beta = [\beta_1, \dots, \beta_L]^T$ is the matrix of the output weights from the hidden nodes to the output nodes. $h(x) = [h_1(x), \dots, h_L(x)]$ is the row vector representing the outputs of the L hidden nodes with respect to the input x , and $h(x)$ maps the data from the d -dimensional input space to the L -dimensional hidden layer feature space (ELM feature space) H .

The above linear equation can be written in the matrix form.

$$H\beta = T \quad (2)$$

where $T = [t_1, \dots, t_N]^T$ is the target labels and $H = [h^T(x_1), \dots, h^T(x_N)]$ is the output of the hidden layers. The output weights β can be calculated by Eq. 3:

$$\beta = H^{-1}T \quad (3)$$

where H^{-1} is the Moore-Penrose generalized inverse of matrix H . To have a better generalization performance and to make the solution more robust, one can add a regularization term. The output weights β can be calculated by Eq. 4 [11].

$$\beta = \left(\frac{I}{C} + H^T H \right)^{-1} H^T T \quad (4)$$

where I is the identity matrix, and C is a constant.

Multi-layer Extreme Learning Machine

The multi-layer extreme learning machine performs layer by layer unsupervised learning. Each layer is an ELM-AE

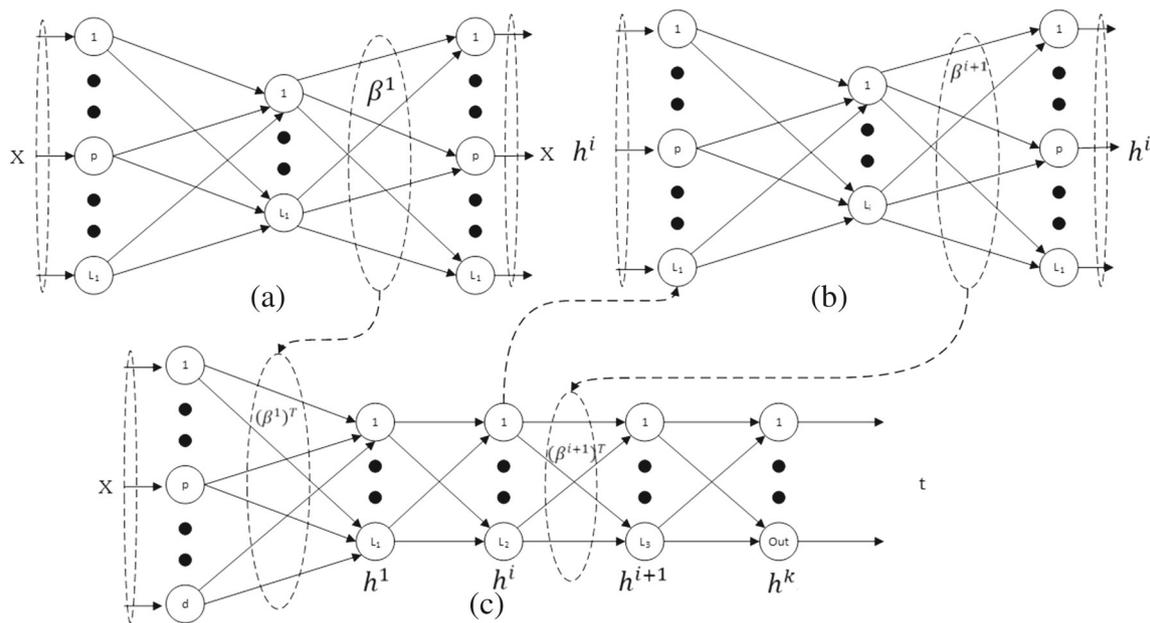


Fig. 1 The network structure of ML-ELM

that represents features based on singular values. The ELM-AE is a modified ELM that performs unsupervised learning, obtains the representation of input X at the coding stage, and recovers the X at the encoding stage. The ELM-AE network is shown in Fig. 1a. The parameters of this ELM-AE are calculated as follows. The input parameters are the weights w and biases b of the hidden respective nodes that are generated randomly and satisfy the orthogonal constraint such that $W^T W = I$ and $b^T b = I$. The orthogonalization of the randomly generated weights and biases tends to improve the generalization performance of the ELM-AE. Under the minimization, the difference between recoding X and inputting X together with the sparse requirement and the output weights of the ELM-AE are calculated as shown in Eq. 5

$$\beta = \left(\frac{I}{C} + H^T H \right)^{-1} H^T X \tag{5}$$

where X is the input data. For the symmetrical ELM-AE, the output weights are calculated by the singular value decomposition (SVD). Kasun et al. [12] proved that the output weights β of the ELM-AE are learned and derived the features of the input data via singular values.

To get the distributed representation, the ML-ELM network is created by stacking the ELM-AE layer by layer, where each layer takes the output from the last layer. Parameter β in each layer is calculated as shown in Eq. 5. Moreover, unlike typical deep networks, such as the DBM, the ELM-AE does not tune the parameters, and it trains the network faster than the other deep networks.

Proposed Methodology

The ELM proposed by Huang was originally inspired by biological learning. They conjectured that some parts of brain systems should have random neurons, and thus they constructed the universal learning structure ELM. All the hidden nodes in the ELM are both independent of the training data and independent of each other [13]. Nevertheless, the final weight calculation still depends on the training data. Furthermore, the selective attention system proves that humans will focus on the local interested area. Inspired by this selective attention mechanism, we make some modifications to the ML-ELM and improve the representation learning ability to mimic the human cognitive method to some extent.

The Main Idea of the Region-Enhanced ML-ELM

In this part, we aim to rectify the ML-ELM by adding additional inputs that are selected from selective attention areas to address the idea of the human visual attention mechanism. The structure of the proposed region-enhanced ML-ELM is illustrated in “Structure of Region-Enhanced ML-ELM”. Our proposed region-enhanced ML-ELM uses the same scheme as most deep networks that learn the distributed representation from the data. However, the original data is not treated equally because the significant areas are repeated and enhanced at the input’s end. This is illustrated in Fig. 2. In this figure, the red rectangle marks the local significant region and this part is multiplexed at the input’s end. The significant areas are selected by referring

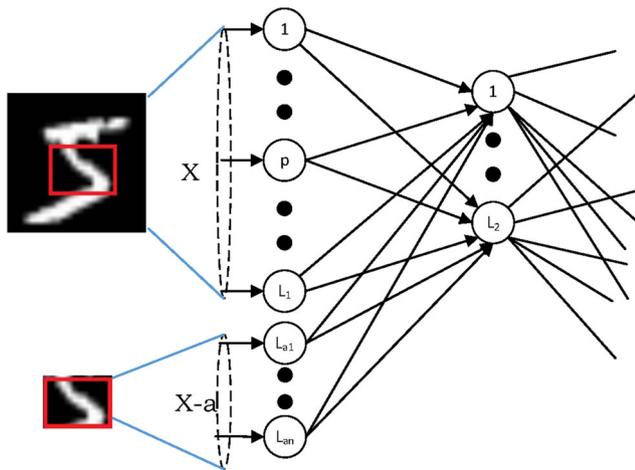


Fig. 2 Illustration of the region-enhanced ML-ELM input

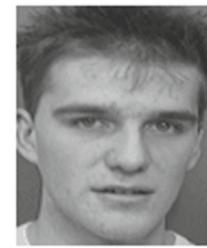
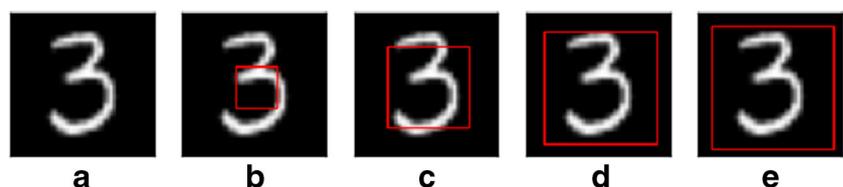
to the human being's visual selective attention mechanism. Nevertheless, the human selective attention system is complex. To simplify the attention simulation problem and focus on the attention-aided learning algorithm, we use only two typical selective attention areas according to our understanding. One is the center area. The other is the task-driven interested area.

Definition of our Local Significant Regions Based on the Attention Knowledge

In cognitive psychology, there are some models that describe how visual attention operates. The first of these models to appear in the literature is the spotlight model. In this spotlight model, attention is described as having a focus, a margin, and a fringe [14]. The focus is the area that extracts information from the visual scene with a high resolution, where its geometric center is where the visual attention is directed. Surrounding the focus is the fringe of attention, which extracts the information with a low resolution. This fringe extends out to a specified area, and the cutoff is called the margin (<https://en.wikipedia.org/wiki/Attention>). The measurements of the strength and direction of the illusory motion at increasing separations from the cue reveal an attentional "perceptive field" with an excitatory center at the cued locus and an inhibitory surround subtending the remaining visual field [15].

To explain the center attention scheme, we use the following example. Here are five example images of the handwritten digit "3" shown in Fig. 3.

Fig. 3 Illustration of the center-based local significant region



(a) whole image



(b) eyes



(c) nose



(d) mouth

Fig. 4 Illustration of task-driven local significant regions

Image (a) is the original image. Image (b) to image (e) were each marked with a different sized red rectangle. It is easy to know the digit in the image (a) is three. However, if you just see the red rectangle blocks from image (b) to image (e), you will understand the information of the image increasingly better. The image showing the digit three is the knowledge that people obtain from their eyes. When the red rectangle block becomes larger, people can see it better. However, from image (b) to image (e), we can see that, the red rectangle block of image (e) shows enough information to allow people to see which number is shown in the image. It seems that we do not need the other part of the image's information. In other words, the part of the red rectangle provides more important information than the remaining part. Combining the spotlight model of the visual attention, we extract the center region from the whole image and add them as multiplexing input nodes to increase the resolution of this region. In images with simple scenes, the geometric center is utilized as we do in this paper for digital and object recognition in pre-processed databases of MNIST and NORB. For an image with a complex background, the excitatory center at the locus must be detected at a certain attention system or a cognition task.

Some attention theories placed a new emphasis on the separation of visual attention tasks alone and they are mediated by supplementary cognitive processes. Duncan et al. state that there is an initial pre-attentive parallel phase of the perceptual segmentation and analysis that encompasses all visual items present in a scene [16]. Being task-driven is a possible stimulus for perceptual segmentation. Based on this understanding, we select several segments as local significant regions for a certain cognition task. When we

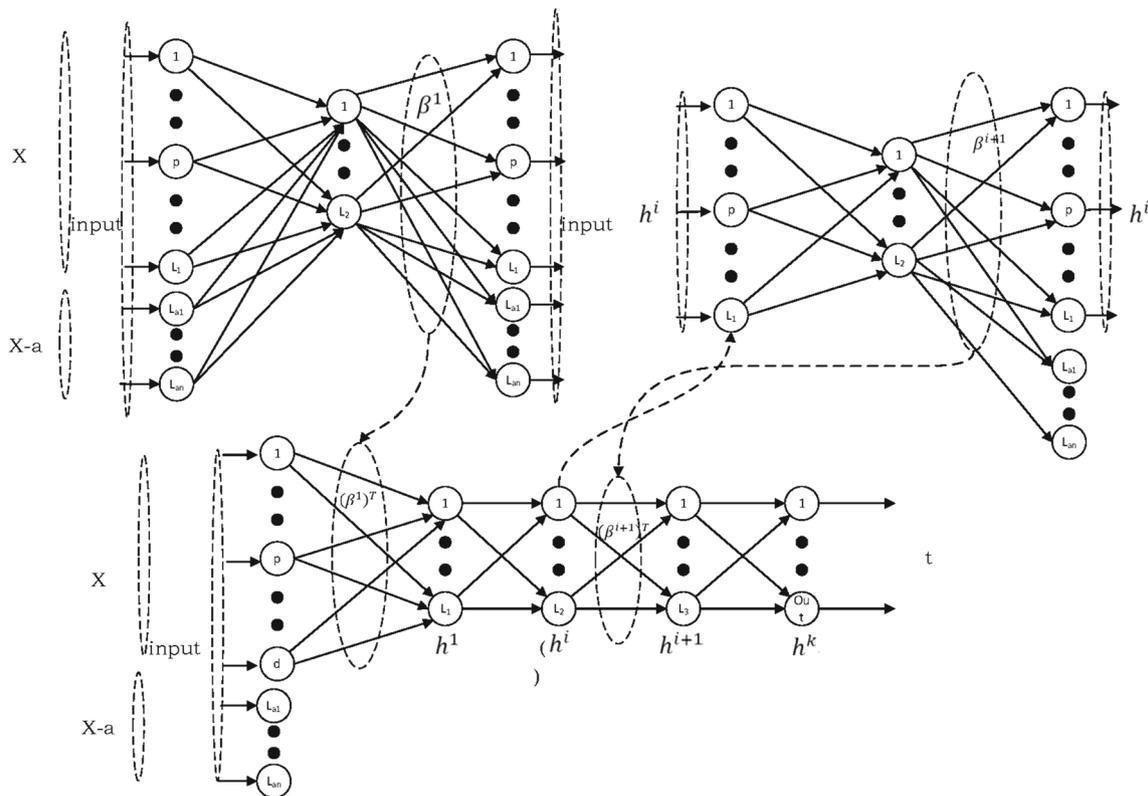


Fig. 5 The structure and parameter-training diagram of region-enhanced ML-ELM

perform motion recognition, we focus on the hands and feet parts of an image. When we perform face recognition, our attention will be concentrated on the eyes, nose, and mouth of the image. When we perform facial expression recognition, we care about the eyes, mouth, eyebrows, etc.

In this paper, we take the face recognition as the task to extract several relative significant regions, and then add them as multiplexing inputs in our improved ML-ELM.

In the task of face recognition, we take the eyes, mouth, and nose parts of image as the local significant regions. In Fig. 4, we use some algorithms that are embedded in OpenCV to detect the eyes (b), nose (c), and mouth (d) from

the image (a). After we get the significant regions, we will use them in the next step.

Implementation of Region-Enhanced ELM-AE

An auto encoder (AE) is an artificial neural network used for learning efficient coding. The aim of the auto encoder is to learn a representation for a set of data. An auto encoder is a feedforward, non-recurrent neural network, that is similar to the multilayer perceptron (MLP), with an

Fig. 6 Source and reconstruction images



(a). Source image and Central regions of the source images



(b). Reconstructed images and Central regions of the source images with region-enhanced ML-ELM



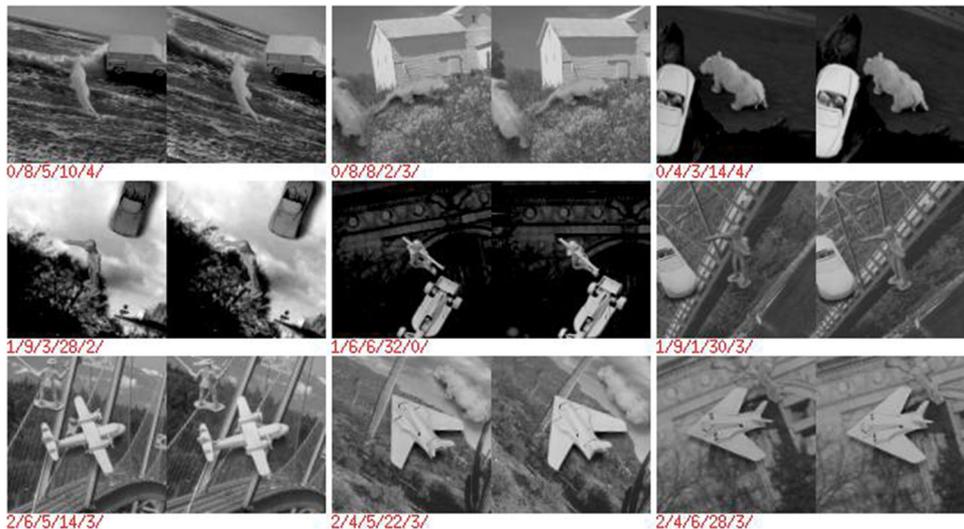
(c). Reconstructed images of the source image with ML-ELM

input layer, an output layer, and one or more hidden layers connecting them. The difference between auto encoders and MLPs is that in an auto encoder, the output layer has the same number of nodes as the input layer. Instead of being trained to predict the target value Y given inputs X , auto encoders are trained to reconstruct their own inputs X , as does the ELM-AE. The ELM-AE learns the representative features from data in an unsupervised way, which is based on the assumption that the reconstructed data forms the feature in the encoding stage to approximate the source data to generate that feature in the coding stage. By stacking

the coding network of the ELM-AE, the deep feature is then learned with the ML-ELM. Furthermore, the ML-ELM is modified by making use of some human cognition knowledge to improve the performance of the representation learning. The details of our proposed region-enhanced ML-ELM are elaborated as follows.

Structure of Region-Enhanced ML-ELM

The network’s structure is shown in Fig. 5. Different from the ML-ELM, our proposed region-enhanced ML-ELM



(a)

Four-Legged Animals	Human Figures	Airplanes	Trucks	Cars	Four-Legged Animals	Human Figures	Airplanes	Trucks	Cars

(b)

Fig. 7 Sample images in the NORB dataset

has added several input nodes by multiplexing the local significant region or regions at the input's end. In other words, the input of the proposed region-enhanced ML-ELM includes the original data, and it contains two parts. The first part is the source data (X in Fig. 5), and the second part is the data from the local significant region ($X-a$ in Fig. 5). In the training stage, the parameter is calculated in each layer separately based on the ELM-AE. At the input's end, the ELM-AE takes them as inputs. If the ELM-AE takes more information from the data, the representation learning of the ELM-AE will be improved.

Calculation of Network Parameters

Since our region-enhanced ML-ELM has a similar structure as the ML-ELM, the parameter calculation method is similar, except for the weights in the input layer. Therefore, the hidden and output weights are calculated based on the ML-ELM, as illustrated in section 3. Here, we address how we calculate the weights in the input layer.

The input of ELM-AE with local significant region is given by Eq. 6:

$$X_{\text{region}} = [x, x-a] \quad (6)$$

where the x represents the original input and the $x-a$ denotes the multiplexing input of the local significant region of the input. Thus, the hidden output of the region-enhanced ELM-AE is calculated by Eq. 7:

$$H_{\text{region}} = g(a * x_{\text{region}} + b) \quad (7)$$

where a is the orthogonal weights, b is the orthogonal bias, and the g is the activation function.

The optimization target minimizes the loss function between the estimated input and the original one, as in Eq. 8

$$J_{\text{cost}} = \min |X_{\text{region}} - \tilde{X}_{\text{region}}| \quad (8)$$

where $\tilde{X}_{\text{region}} = H_{\text{region}} * \beta$, the parameter β is the output weight under-computing. Then, β is calculated as in Eq. 9.

$$\beta = H_{\text{region}}^{\dagger} X_{\text{region}} \quad (9)$$

Fig. 8 Sample images in the ORL dataset



Analysis of Reconstructed Results

Figure 6 contains the source sample images and their relative central regions, and Fig. 6b, c contains the corresponding reconstructed sample images with and without the region enhancement from Fig. 6a, respectively. From Fig. 6b, c, we can see that the reconstructed images become slightly cluttered and fuzzy compared with the source images. Nevertheless, the reconstructed results are very acceptable with complete information. Meanwhile, when comparing the reconstruction results with/without the region-enhanced ML-ELM in Fig. 6b, c, the difference is tiny but could still be found. After adding the local region enhancement, the edge is a bit clearer and less fuzzy than that found without using the region enhancement. More noise could be found in the reconstruction results without using the region enhancement than that using our region enhancement. For example, note the extra flaw in the vertical stroke around the bottom-right corner of the center region of the digit “4” image. This, to some extent, demonstrates that our proposed region-enhanced method improves the representation of important parts by increasing the resolution of the relative parts at the input's end as intended by multiplexing the local significant regions. In the future, more experiments will be done with various increasing resolution rate. Of course, considering the random generation of weights between the input and hidden nodes, the strategy needs to make full use of the apriori knowledge of the selectiveness.

Performance Evaluation and Analysis

Image Datasets

The MNIST dataset is a commonly used dataset for testing the performance of deep networks. It contains 28*28 images of handwritten digits with 60,000 training samples and 10,000 testing samples [17].

The NORB dataset is intended for experiments in 3D object recognition from shapes. The NORB dataset contains 24,300 stereo images that are used for training and 24,300

Table 1 Performance comparison of the RE-ML-ELM with the state-of-the-art deep networks using the MNIST dataset

Algorithms	Testing accuracy%	Training time
RE-ML-ELM	99.12	238.9713 s
ML-ELM	98.97	240.7770 s
ELM	97.16	11493 s
DBN	98.87 [20]	20580 s
SAE	98.6 [21]	–

stereo images that are used for testing. It contains images of toys belonging to five generic categories: four-legged animals, human figures, airplanes, trucks, and cars. As shown in Fig. 7a [18], the objects were imaged by 2 cameras under 6 lighting conditions, 9 elevations (30 to 70 degrees every 5 degrees), and 18 azimuths (0 to 340 every 20 degrees). Referring to Prof. Huang's work, we use the preprocessed images with normalized object size and the uniform background of the NORB dataset. The size of the images is 32*32 with stereo pairs. Some samples are shown in Fig. 7b [19].

The ORL dataset is used in the context of a face recognition project. There are ten different images of each of the 40 distinct subjects [6]. To get more subjects, the images were taken at different times with varying lighting conditions, facial expressions, and facial details. All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position. Some sample images of ORL dataset are shown in Fig. 8.

Experiments

The experiments were conducted on a desktop computer with a core Intel Xeon E5-2687 3.10-GHz processor and 32-GB RAM running MATLAB 2013a. The symbol “RE-ML-ELM” marked in Tables 1, 2, and 3 indicates region-enhanced multi-layer extreme learning machine.

The ML-ELM (network structure, 784-700-15000-10 with ridge parameters 10^{-1} for layer 784-700, 10^3 for layer 700-15000, and 10^8 for layer 15000-10) with sigmoidal hidden layer activation function generates an accuracy of

Table 2 Performance comparison of the RE-ML-ELM with the ML-ELM, and ELM using the NORB dataset

Algorithms	Testing accuracy%	Training time
RE-ML-ELM	88.91	144.9884 s
ML-ELM	87.79	163.3013 s
ELM	85.42	112 s

Table 3 Performance comparison of the RE-ML-ELM with the ML-ELM, and ELM using the ORL dataset

Algorithms	Testing accuracy%	Training time
RE-ML-ELM	91.25	14.0666 s
ML-ELM	90.00	14.371 s
ELM	88.75	1.3884 s

98.97% in the MNIST dataset. In addition, the RE-ML-ELM that we propose (network structure, 784-700-15000-10 with ridge parameters 10^{-1} for layer 784-700, 10^4 for layer 700-15000, 10^8 for layer 15000-10, and $2*28$ for the local region size) with the sigmoidal hidden layer activation function generates an accuracy of 99.12% in the MNIST dataset. The ML-ELM uses the same parameters as the RE-ML-ELM. For the ELM, in order to get a higher accuracy using the MNIST dataset, we incorporate significantly more nodes in the hidden layers, and so it costs much more time. Moreover, the two-layer DBM network produces better results than the three-layer DBM network [20]. We achieved the same test results for DBN as those reported in [20]. We also perform experiments using the NORB datasets and the ORL datasets, and the results are shown in Tables 2 and 3.

As shown in Tables 1, 2, and 3, the RE-ML-ELM outperforms the ML-ELM, ELM, DBN, and SAE. It takes less time than the DBN and the other deep networks. The RE-ML-ELM takes more data as inputs. Because the RE-ML-ELM incorporates more information than ML-ELM, the accuracy of the RE-ML-ELM is higher than that of the ML-ELM. There is more useful information could be used in RE-ML-ELM. The experiments show the additional significance information can be utilized to improve the accuracy of classifier.

Conclusion and Future Job

In the paper, we propose an improved ML-ELM with multiplexing applied to the local significant regions at the input end and apply it to image classification. By enhancing the center area and the selected areas, respectively, in the relative recognition tasks, our proposed RE-ML-ELM achieves the best recognition results and a competitive fast training time comparing to that without the region-enhancement ML-ELM, the ELM, and some state-of-art deep networks. This proves our idea of combining the apriori knowledge of attention with data learning to improve the performance of the ML-ELM. Actually, our proposed idea of enhancing the local significant regions can be applied to most data-driven algorithms. However, the definition of significant regions in this paper is not complete and practical for all applications. Like the focus

center, in this paper, we used only geometric center, and we verify that our local region enhanced idea using two datasets with simple backgrounds. In the future, we will extend the region selection with more comprehensive and practical significance referring to the psychological research of the selective attention mechanism. At the least, in the next step, we will employ a more practical focal attention mechanism that detects the location of the excitator and utilizes the center region around the locus instead of only the geometric center. In another perspective, our proposed region-enhanced ML-ELM does not consider the spatial attribution of images such as the CNN and ELM-RLF do. It reconstructs the two-dimensional image into the one-dimension vector. This causes the serious loss of spatial information during the image cognition and accordingly a lower accuracy rate on the NORB dataset comparing to that of the CNN and ELM-RLF given in paper [9]. In the next step, we could consider making use of the spatial information in our RE-ML-ELM to improve the cognition computing performance.

In a word, there are several jobs we will do in the future to increase the performance of our current algorithm. (1) We will extract more comprehensive and practical significant regions based on the psychology research of the attention mechanism, such as the center around excitatory locus instead of the simple geometry center. (2) We will extract the local regions with more complete semantic information. (3) We will improve the algorithm's scheme with the consideration of spatial information.

Funding Information This research was partially sponsored by the National Nature Science Foundation of China (Nos. 61871276, 61672070, and 61672071), the Beijing Municipal Natural Science Foundation (Nos. 7184199 and 4162058), the Research Fund from Beijing Innovation Center for Future Chips (No. KYJJ2018004), and the 2018 Talent-Development Quality Enhancement Project of BISTU (No. 5111823402).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Informed Consent Informed consent was not required as no human or animals were involved.

Human and Animal Rights This article does not contain any studies with human or animal subjects performed by any of the authors.

References

- Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems (NIPS); 2012.
- Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. *IEEE Trans Pattern Anal Mach Intell.* 2013;35(8):1798–828.
- Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw.* 2015;61:85–117.
- Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science.* 2006;313(5786):504–7.
- Huang GB, Zhu QY, Siew CK. Extreme learning machine: theory and applications. *Neurocomputing.* 2006;70:489–501.
- Guo T, Zhang L, Tan X. Neuron pruning-based discriminative extreme learning machine for pattern classification. *Cogn Comput.* 2017;9(4):581–595.
- Liu Y, Zhang L, Deng P, et al. Common subspace learning via cross-domain extreme learning machine. *Cogn Comput.* 2017;9(4):555–563.
- Huang GB. What are extreme learning machines? Filling the gap between Frank Rosenblatt's dream and John von Neumann's puzzle. *Cogn Comput.* 2015;7(3):263–78.
- Huang GB, Bai Z, Kasun LLC. Local receptive fields based extreme learning machine. *IEEE Comput Intell Mag.* 2015;10(2):18–29.
- Kasun LLC, Zhou H, Huang GB, et al. Representational learning with extreme learning machine for big data. *Intell Syst IEEE.* 2013;28(6):31–4.
- Tang J, Deng C, Huang GB. Extreme learning machine for multilayer perceptron. *IEEE Trans Neural Netw Learn Syst.* 2016;27(4):809–21.
- Salakhutdinov R, Larochelle H. Efficient learning of deep boltzmann machines. In: International conference on artificial intelligence and statistics; 2010.
- Huang GB. An insight into extreme learning machines: random neurons, random features and kernels. *Cogn Comput.* 2014;6(3):376–90.
- Eriksen C, Hoffman J. Temporal and spatial characteristics of selective encoding from visual displays. *Percept Psychophys.* 2014;12(2B).
- Steinman BA, Steinman SB, Lehmkuhle S. Visual attention mechanisms show a center-surround organization. *Vision Res.* 1995;35(13):1859–69.
- Raftopoulos A. Cognition and perception. Oxford: Oxford University Press; 2007, pp. 5–7.
- Meier U, Ciresan DC, Gambardella LM, et al. Better digit recognition with a committee of simple neural nets. In: 2011 international conference on document analysis and recognition (ICDAR); 2011. p. 1250–4.
- LeCun Y, Huang FJ, Bottou L. Learning methods for generic object recognition with invariance to pose and lighting. *CVPR.* 2004.
- Zhang Z, Zhao XG, Wang GR. FE-ELM: a new friend recommendation model with extreme learning machine. *Cognitive Computation.* 2017;9(3):1–12.
- Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science.* 2006;313(5786):504–7.
- Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol PA. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *J Mach Learn Res.* 2010;11:3371–408.