

# Robust Multi-Layer Hierarchical Model for Digit Character Recognition

Jie Yang<sup>†</sup>, Yadong Sun<sup>\*</sup>, Liangjun Zhang<sup>\*</sup> and Qingnian Zhang<sup>\*\*</sup>

**Abstract** – Although digit character recognition has got a significant improvement in recent years, it is still challenging to achieve satisfied result if the data contains an amount of distracting factors. This paper proposes a novel digit character recognition approach using a multi-layer hierarchical model, Hybrid Restricted Boltzmann Machines (HRBMs), which allows the learning architecture to be robust to background distracting factors. The insight behind the proposed model is that useful high-level features appear more frequently than distracting factors during learning, thus the high-level features can be decompose into hybrid hierarchical structures by using only small label information. In order to extract robust and compact features, a stochastic 0-1 layer is employed, which enables the model's hidden nodes to independently capture the useful character features during training. Experiments on the variations of Mixed National Institute of Standards and Technology (MNIST) dataset show that improvements of the multi-layer hierarchical model can be achieved by the proposed method. Finally, the paper shows the proposed technique which is used in a real-world application, where it is able to identify digit characters under various complex background images.

**Keywords:** Digit character recognition, Multi-layer hierarchical, Neutral networks, Restricted Boltzmann Machine, Computer vision

## 1. Introduction

Digit character recognition is an important research field in computer vision, and it also has been used as a test case for theories of machine learning and pattern recognition algorithm in recent years. Devising model of real-world digit character especially handwritten digit is challenging because characters usually have distortion, variations, and various cluttered backgrounds. One fundamental difficulty in learning the model is to deal with significant noise and clutter background factors. Learning meaningful features of digit character is a crucial step towards building general system that can be easily employed in higher level tasks such as identity verification [1], digital signature [2], and converting subtitle data into text format [3].

Many researches have been concerned with the recognition of digit characters and some high accuracy algorithms have been proposed. Choi *et al.* [4] presented a novel 3D stroke reconstruction algorithm based on a magnetometer-aided inertial measurement unit, which is able to estimate the orientation under both static and dynamic conditions for character recognition. Akhtar *et al.* [5] used a wavelet analysis technique for handwritten digit recognition. Entropy feature was computed in the method. Deselaers *et*

*al.* [6] proposed latent log-linear models which was an extension of log-linear models incorporating latent variables. Sobel features were extracted in this algorithm. These algorithms used low level representation of raw data for digit character recognition. Low level feature representations have great success in many visual recognition problems, yet there has been a growing body of work suggesting that the approach of using only low level feature representations may be insufficient to represent more high-dimensional, complex data [7]. Instead, some researches focused on training deep, multi-layered networks and proposed the algorithms using not only low level feature representation, but also high level feature representation from raw data.

Cecotti *et al.* [8] proposed a radial neural convolutional architecture to extract higher level features for multi-oriented character recognition. The classifier of this method included the Fast Fourier Transform for extracting shift invariant features at the neural network level. Phan *et al.* [9] proposed hierarchical sparse auto-encoder architecture and used linear regression-based features in clustering for handwritten digit recognition. Their algorithm achieved 1.87% error rates on the Mixed National Institute of Standards and Technology (MNIST) handwritten character recognition dataset. Lee *et al.* [10] presented a convolutional deep belief network model and trained with unlabeled data. Another contribution of the paper was probabilistic max-pooling, a novel technique which shrank the representations of higher layers in a probabilistically sound way. The recognition error rate of the model was 0.82% on the MNIST handwritten digits dataset. Ciresan *et al.* [11] proposed convolutional neural network committees for handwritten character recognition. Their algorithm

<sup>†</sup> Corresponding Author: The Key Laboratory of Fiber Optic Sensing Technology and Information Processing, Ministry of Education, Wuhan University of Technology, Wuhan, China. (jieyang509@163.com).

<sup>\*</sup> The Key Laboratory of Fiber Optic Sensing Technology and Information Processing, Ministry of Education, Wuhan University of Technology, Wuhan, China (shunhm\_baby@sohu.com).

<sup>\*\*</sup> School of Transportation, Wuhan University of Technology, Wuhan, China. (616024597@qq.com).

Received: March 31, 2014; Accepted: November 14, 2014

simply adopted average individual committee member outputs and obtained 0.27% error rate on the MNIST dataset. Lately, they proposed multi-column deep neural networks and trained using standard gradient descent method for handwritten digit classification [12]. The recognition error rate record dropped to 0.23% on the MNIST handwriting digit dataset. These algorithms used neural networks model and achieved good performance on the Standard MNIST digit dataset.

Currently, neural networks are among the most suitable architectures for digit recognition which seem to benefit unsupervised learning algorithms applied to the input data. Yet, it is still challenging to correctly recognize digit characters when the data contains an amount of distracting factors.

In this paper, a multi-layer hierarchical model based on neural network is proposed by using Hybrid Restricted Boltzmann Machines (HRBMs) for digit character recognition. The model combines double Restricted Boltzmann Machine (RBM) to learn character features and background clutter and finds robust latent features of both that lead to improvement in recognition performance. The families of Restricted Boltzmann Machine (RBM) model [13] have been shown good results on the speech recognition [14], learning movement patterns [15] and facial expression tasks [16].

In the proposed model, the HRBMs' hidden layer is decomposed into an inheritance hierarchy of two classes based on the scores of each node's activation values and times. By the hierarchical structure, the model can successfully learn the statistical structure of the character's features with the label data. Furthermore, the multiplicative interaction and 0-1 layer are used to induce HRBMs' parameters over the input layer. HRBMs can be seen as an extended RBM. Training the model means adjusting model's parameters such that the probability distribution of the machine representing fits the training data as well as possible. It differs from the existing neural network methods. The model uses a small amount of labeled data and transfers the label information to the data features and 0-1 nodes information through the hierarchical hidden layer, which makes the learned features more compact and efficient to recognize. Finally, the proposed method is evaluated on the Variations of MNIST dataset [17] which contains various cluttered backgrounds in the digit images.

## 2. Model Description

In this section, the HRBMs model is presented. Specifically, Section 2.1 supplies the general idea of the model, Section 2.2 indicates the pre-training process, and Section 2.3 provides how the proposed model works.

### 2.1 Overview of the model

The existing models have been concerned with the

recognition of digit characters. For example, Nair *et al.* [18] proposed implicit mixtures of Restricted Boltzmann Machine, which employed a single component to describe the input data. The key insight of the proposed method is that the mixture model can be cast as a third-order Boltzmann machine. However, it sensitive to the noise in the complex input data.

Vincent *et al.* [19] proposed a denoising autoencoders model, which based on the idea of making the learned representations robust to partial corruption of the input data. The modeling idea means that the proposed model is only robust to small irrelevant changes for input data.

In this paper, the proposed model can be formulated as a generative digit character feature learning model, called HRBMs. Firstly, a standard RBM is trained in an unsupervised learning way, which creates an initial set of hidden nodes. With the pre-training process, these hidden nodes can be seen as hybrid layer, which contains useful character features and background clutter features. Secondly, based on the scores of each hidden node's activation values and times, the hidden nodes are decomposed in hierarchical structures with feature nodes and clutter nodes, which provides an initialization of the HRBMs model. Thirdly, the pixel-wise HRBMs will be trained by using label information and stochastic 0-1 nodes through multiplicative interaction.

The proposed model can be represented in Fig. 1. The structure of the HRBMs can be seen as hybrid graphical model with two combined RBMs for robust feature learning. The model is a multi-layer hierarchical model,

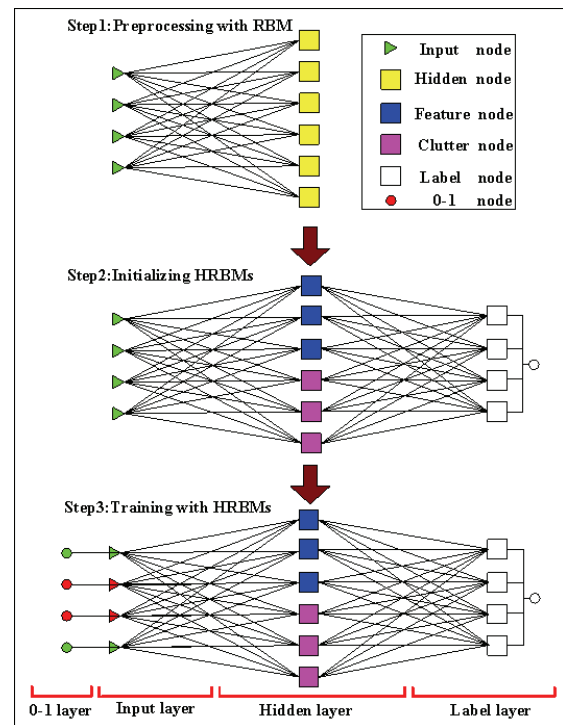


Fig. 1. The graphical framework of HRBMs with the multi-layer hierarchical structure

which includes 0-1 layer, input layer, hidden layer, and label layer. More specifically, the 0-1 layer is used to induce the HRBMs' parameters over the input data; the input layer is used to input the digit character data, which contains amounts of distracting factors; the hidden layer is used to extract features from the input layer, which includes the feature nodes and clutter nodes; the label layer is used to input the label digit character data, which can be used for supervised training.

### 2.2 Preprocessing with RBM

The RBM is an undirected probabilistic graphical model, i.e., a particular form of log-linear Markov Random Field. The nodes of an RBM are usually Bernoulli distributed [20], but if the mean field approximation is employed, the nodes may follow any exponential family distribution. The structure of an RBM is a fully connected bipartite graph and no connections between nodes in the same layer. Thus, one group of nodes (input nodes  $\mathbf{v}$ ) models the data and the other group of nodes (hidden nodes  $\mathbf{h}$ ) models the latent structure of the data.

Since an RBM is a special case of a Markov Random Field, the joint distribution over all nodes is given by a Boltzmann distribution that is specified by the energy function  $E(\mathbf{v}, \mathbf{h}; \boldsymbol{\theta}_R)$ . The most common choice for the energy function is a linear function of the states of the input and hidden nodes. The energy function is given by:

$$E(\mathbf{v}, \mathbf{h}; \boldsymbol{\theta}_R) = -\sum_{i,j} v_i W_{ij} h_j - \sum_i b_i v_i - \sum_j c_j h_j \quad (1)$$

where  $\boldsymbol{\theta}_R = \{\mathbf{W}, \mathbf{b}, \mathbf{c}\}$  are the model parameters, and  $v_i \in \{0,1\}$ ,  $h_j \in \{0,1\}$ . The probability distribution of the configuration  $P(\mathbf{v}, \mathbf{h}; \boldsymbol{\theta}_R)$  is defined as:

$$P(\mathbf{v}, \mathbf{h}; \boldsymbol{\theta}_R) = \exp[-E(\mathbf{v}, \mathbf{h}; \boldsymbol{\theta}_R)] / Z(\boldsymbol{\theta}_R) \quad (2)$$

$$Z(\boldsymbol{\theta}_R) = \sum_{\mathbf{v}, \mathbf{h}} \exp[-E(\mathbf{v}, \mathbf{h}; \boldsymbol{\theta}_R)] \quad (3)$$

where  $Z(\boldsymbol{\theta}_R)$  is the partition function which ensures that the distribution is valid. Noting that the states of the input nodes are conditionally independent given the states of the hidden nodes and vice versa, it can easily be seen that the linear energy function leads to conditional probabilities  $P(v_i = 1 | \mathbf{h})$  and  $P(h_j = 1 | \mathbf{v})$  that are given by the sigmoid function of the input into a node:

$$P(v_i = 1 | \mathbf{h}) = \frac{1}{1 + \exp(-b_i - \sum_{j=1}^n W_{ij} h_j)} \quad (4)$$

$$P(h_j = 1 | \mathbf{v}) = \frac{1}{1 + \exp(-c_j - \sum_{i=1}^m v_i W_{ij})} \quad (5)$$

The detail process of the derivation can be found in

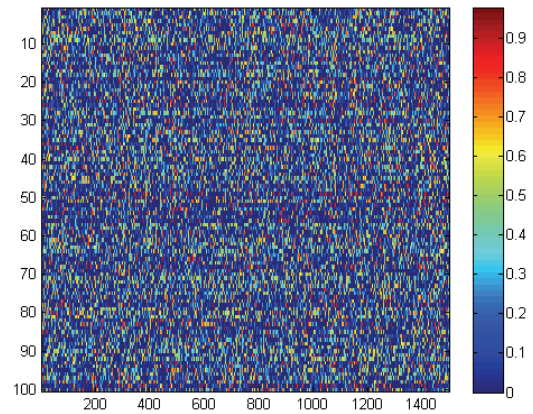


Fig. 2. Hidden nodes activations of RBM

the reference [13]. The basic idea underlying RBM architecture is that the hidden nodes of a trained RBM can be viewed as learned features of the input nodes. From this perspective, posterior  $P(h_j = 1 | \mathbf{v})$  is interpreted as a relevant representation for the input nodes. To avoid computing these distributions, Gibbs operator is employed to sample from the posterior. Fig. 2 shows the activations of 100 samples data on 1500 RBM hidden nodes. In general, when training RBM, data visualization is needed to show what the nodes compute in a layer space. The goal of data visualization is to have an efficient way of adjusting model's parameters, and to make the model as intuitive as possible.

### 2.3 Inferring HRBMs

In order to effectively deal with complex cluttered backgrounds, it is desirable for a digit recognition algorithm to extract the useful digit characters information from the background clutter. To address this problem, the HRBMs' hidden layer is decomposed with two classes of RBM's hidden layer, where each class can extract the corresponding character or clutter feature.

The main process can be described as:

- 1) The hidden nodes are divided into the feature nodes ( $\mathbf{h}^1$ ) and clutter nodes ( $\mathbf{h}^2$ ), and each RBM defines a distinct distribution over the input nodes.
- 2) The 0-1 nodes are determined by the input nodes and the two RBMs' hidden nodes.
- 3) The type of input pixel information (character or clutter) is determined by the 0-1 nodes.

The HRBMs' hidden layer is decomposed into two classes based on the scores of each node's activation values and times after a standard RBM is trained. The score functions of HRBMs hidden nodes are defined as:

$$z_{jk} = \frac{1}{1 + \exp(-c_j - \sum_{i=1}^m v_i W_{ij})} \quad (6)$$

$$W_j = \begin{cases} w & \text{if } z_{jk} \rightarrow y_k \\ \tilde{w} & \text{otherwise} \end{cases} \quad (7)$$

$$\text{score}(W_j) = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (8)$$

where  $W_{ij}$  is the HRBMs weights,  $y_k$  is label information. Eq. (6) is used to compute the activation values for each hidden node. Eq. (7) is used to determine the node whether it is activated. In (7),  $W_j$  records each hidden node's activated ( $w$ ) and non-activated ( $\tilde{w}$ ) state information, which includes the respective means ( $\bar{x}_1, \bar{x}_2$ ) and standard deviations ( $\sigma_1, \sigma_2$ ) corresponding to the activated and non-activated state. Eq. (8) is used to compute the score for each hidden node according to the activated and non-activated values and times. Thus, according to the score of each hidden node, the HRBMs hidden layer is initialized by the feature nodes and clutter nodes.

For each input node  $v_i$ , HRBMs contain a binary 0-1 node, denoted  $s_i \in \{0,1\}$ . Furthermore, HRBMs use pixel-wise multiplicative interaction between the 0-1 nodes and input nodes during training. In this case, the energy function and the joint distribution of the HRBMs are defined as:

$$\begin{aligned} E(\mathbf{v}, \mathbf{s}, \mathbf{h}; \boldsymbol{\theta}_R) &= E(\mathbf{v}, \mathbf{s}, \mathbf{h}^1; \boldsymbol{\theta}_R) + E(\mathbf{v}, \mathbf{s}, \mathbf{h}^2; \boldsymbol{\theta}_R) \\ &= -\sum_{i,j} s_i v_i W_{ij}^1 h_j^1 - \sum_i s_i v_i b_i^1 - \sum_j c_j^1 h_j^1 \\ &\quad - \sum_{i,j} (1-s_i) v_i W_{ij}^2 h_j^2 - \sum_i (1-s_i) v_i b_i^2 \\ &\quad - \sum_j c_j^2 h_j^2 \end{aligned} \quad (9)$$

$$P(\mathbf{v}, \mathbf{s}, \mathbf{h}; \boldsymbol{\theta}_R) = \exp[-E(\mathbf{v}, \mathbf{s}, \mathbf{h}; \boldsymbol{\theta}_R)] / Z(\boldsymbol{\theta}_R) \quad (10)$$

Since the model can be interpreted as a stochastic neural network, the input nodes, hybrid hidden nodes, and 0-1 nodes are conditionally independent given the other two types of nodes. The hidden layer conditional probability distributions can be independently taken the form:

$$P(h_j^1 = 1 | \mathbf{v}, \mathbf{s}) = \frac{1}{1 + \exp(-c_j^1 - \sum_i s_i \cdot v_i W_{ij}^1)} \quad (11)$$

$$P(h_j^2 = 1 | \mathbf{v}, \mathbf{s}) = \frac{1}{1 + \exp(-c_j^2 - \sum_{i=1}^m (1-s_i) \cdot v_i W_{ij}^2)} \quad (12)$$

The conditional probability distribution over the 0-1 nodes can be determined as:

$$\begin{aligned} P(s_i = 1 | \mathbf{v}, \mathbf{h}^1, \mathbf{h}^2) \\ = \frac{1}{1 + \exp(v_i (-b_i^1 - b_i^2 - \sum_{j=1}^r W_{ij}^1 h_j^1 - \sum_{j=1}^r W_{ij}^2 h_j^2))} \end{aligned} \quad (13)$$

Eq. (13) means that  $s_i$  is achieved by the alternating competition between the two hybrid RBMs' hidden nodes based on the matching between input nodes and the reconstructed input nodes. The reconstructed input nodes can be computed as:

$$\begin{aligned} P(v_i = 1 | \mathbf{s}, \mathbf{h}^1, \mathbf{h}^2) \\ = \frac{1}{1 + \exp(s_i (-b_i^1 - \sum_{j=1}^r W_{ij}^1 h_j^1) + (1-s_i) (-b_i^2 - \sum_{j=1}^r W_{ij}^2 h_j^2))} \end{aligned} \quad (14)$$

For classification, a non-linear classifier is trained by using the two RBMs' hidden nodes and label information as inputs. Since the computation of (11) to (14) is intractable, the alternating Gibbs sampling is employed to the above conditional distributions, due to the alternating Gibbs sampling is much more efficient than standard Gibbs sampling. The detail process of the algorithm can be described as follows.

In positive learning phase, the model iterates over (11) to (13) and samples  $\mathbf{h}^1$ ,  $\mathbf{h}^2$ , and  $\mathbf{s}$ . In the negative learning phase, it iterates over (11) to (14) and samples reconstruction  $\hat{\mathbf{v}}$ ,  $\hat{\mathbf{h}}^1$ ,  $\hat{\mathbf{h}}^2$ , and  $\mathbf{s}'$ . When computing the gradient of the log-likelihood of training data, the contrastive divergence approximation is used to approximate the gradient. The algorithm for HRBMs is outlined in Algorithm 1.

---

**Algorithm 1:** Inference in the HRBMs

---

- 1: Pretrain a Standard RBM with training data and initialize  $\{\mathbf{W}, \mathbf{b}, \mathbf{c}\}$ .
  - 2: Decompose RBM hidden nodes into HRBMs' using (6) to (8).
  - 3: Initialize HRBMs parameters:  
*epochs* : the number of iterations.  
*kCD* : the number of contrastive divergence steps.
  - 4: Initialize 0-1 nodes  $\mathbf{s} \leftarrow 1$ .  
**for**  $m=1$ : number *epochs* **do**
  - 5: Infer  $\mathbf{h}^1, \mathbf{h}^2$  give  $\mathbf{v}$  and  $\mathbf{s}$ , using(11) and (12).
  - 6: Infer  $\mathbf{s}'$  give  $\mathbf{h}^1, \mathbf{h}^2$  and  $\mathbf{v}$ , using (13).  
**for**  $n=1$ : number *kCD* **do**
  - 7: Infer  $\hat{\mathbf{v}}$  using (14).
  - 8: Infer  $\hat{\mathbf{h}}^1, \hat{\mathbf{h}}^2$  give  $\hat{\mathbf{v}}$  and  $\mathbf{s}'$ , using (11) and (12).
  - 9: Infer  $\mathbf{s}'$  give  $\hat{\mathbf{h}}^1, \hat{\mathbf{h}}^2$  and  $\hat{\mathbf{v}}$ , using (13).
  - 10: Calculate gradient using contrastive divergence algorithm.
  - 11: Update parameters using stochastic gradient descent.  
**end for**
  - 12: Calculate classification error and select the best  $\mathbf{h}^1, \mathbf{h}^2$  and  $\mathbf{s}$ .  
**end for**
-

The proposed model has several key advantages:

- 1) The model optimizes both characters' features and clutter property simultaneously by decomposing the hidden layer nodes, which makes the learned features better discriminative ability.
- 2) Stochastic 0-1 nodes allow the HRBMs' hidden nodes to independently extract the important features which are observed within the corresponding feature type, leading to the learned features more compact.

These advantages benefit from the hierarchical structure and make the model having a good adaptive ability to robustly extract useful features from significant noise and clutter background factors.

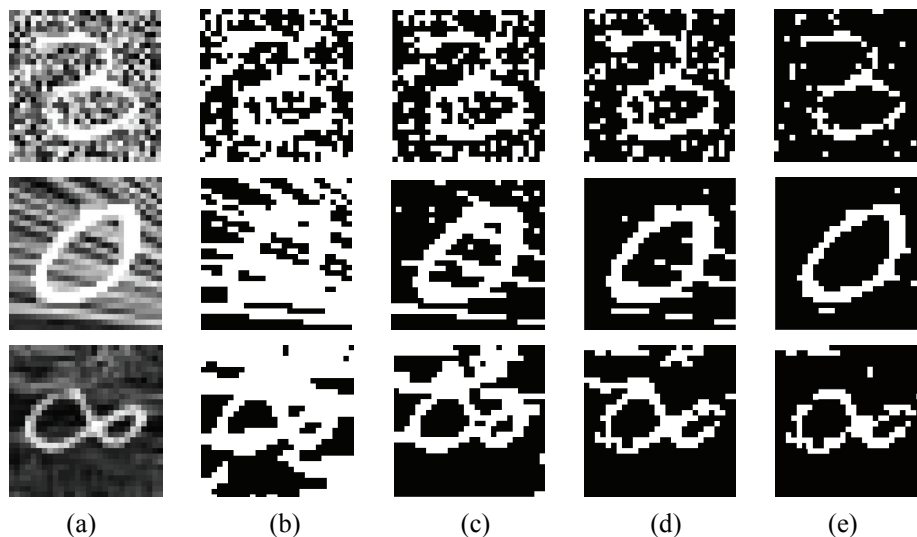
### 3. Experiments and Results

The effectiveness of the HRBMs is evaluated on the Variations of MNIST dataset. The dataset includes 3 types

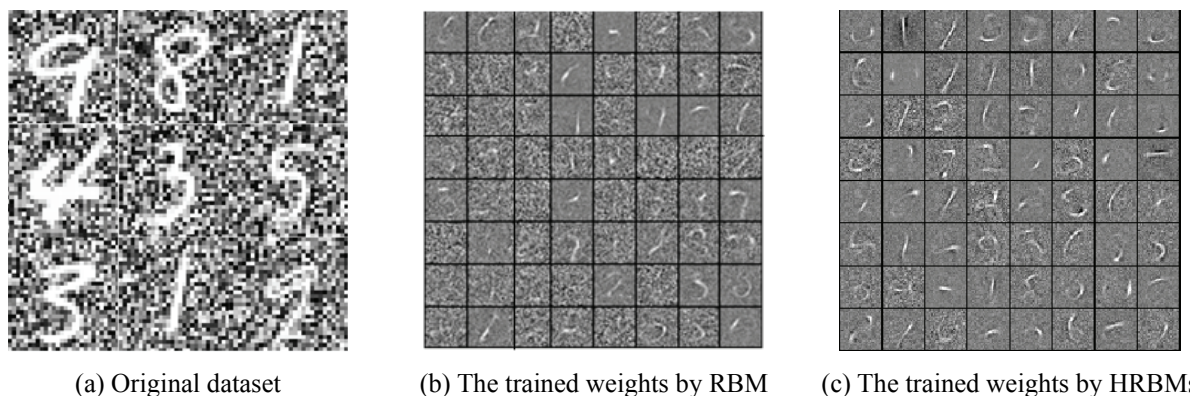
of images. More specifically, the first one is mnist-back-rand dataset, in which a random background was inserted in the digit image; the second one is mnist-back-image dataset, in which a patch from a black and white image was used as the background for the digit image; and the last one is mnist-rot-back-image dataset, in which the perturbations used in mnist-rot and mnist-back-image were combined. Each dataset has 10000 training images, 2000 validation images, and 50000 testing images. Fig. 4(a), Fig. 5(a), and Fig. 6(a) show the three types noisy images, respectively.

HRBMs are trained with two groups of 750 hidden nodes, and initialized with the pre-trained Standard RBM using 1500 hidden nodes. The alternating Gibbs sampling and contrastive divergence are used for posterior inference and approximate gradient respectively in all experiments.

Fig. 3 shows the original input digit images containing the three type cluttered backgrounds and the reconstructed input data performed by HRBMs. More specifically, Fig. 3(a) shows the original data images. Fig. 3(b)-(e) show the reconstructed results for the input digit images with sampling 500, 1000, 5000, and 10000, respectively. It can



**Fig. 3.** The original digit images and the reconstructed input data performed by the proposed model: (a) Original images; (b)-(e) Reconstructed input data with sampling 500, 1000, 5000, and 10000.



**Fig. 4.** Trained weights of mnist-back-rand dataset

be seen that the proposed method allows the model to be robust to significant noise and clutter background factors.

Fig. 4(b)-(c), Fig. 5(b)-(c), and Fig. 6(b)-(c) demonstrate the visualization of trained weights of RBM and HRBMs on the Variations of MNIST dataset. More specifically, Fig. 4(b)-(c) show the trained weights from the mnist-back-rand dataset. It can be seen that the RBM extract features not only including meaningful pen stroke features but also having random clutter features. While the HRBMs' hidden nodes extract the features which are mostly can be seen as useful pen stroke features. Obviously, the trained weights of HRBMs are more compact and correlated with each other than RBM. The similar pattern appears in Fig. 5(b)-(c) and Fig. 6(b)-(c). In Fig. 5(b), RBM extracts the features not only including useful pen stroke features but also having patch features. These patch features will reduce the recognition performance. It is worth noting that Fig. 6(b)-(c) visualize the trained weights from the mnist-rot-back-image dataset. Although the digits are rotated by an angle generated uniformly between 0 and  $2\pi$  radians, HRBMs capture the pen stroke features still maintaining compact which consistent with the features of Fig. 4(c) and Fig. 5(c). Particularly, the visualization of trained weights seems curve strokes. This helps higher representations extracted from HRBMs are better than RBM.

These suggest that the HRBMs model can effectively

**Table 1.** The comparison of recognition error rate

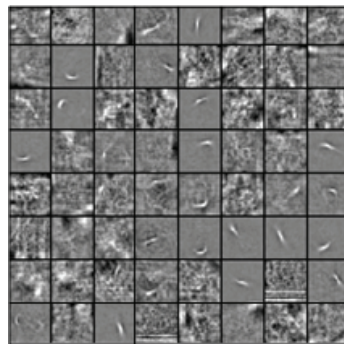
Method	mnist-back-rand	mnist-back-image	mnist-rot-back-image
<b>Proposed</b>	<b>4.45</b>	<b>10.19</b>	<b>40.33</b>
RBM	11.43	15.26	49.58
baseline SVM	14.58	22.61	55.18
DBN-3 [19]	6.73	16.31	47.39
Rifai <i>et al.</i> [21]	10.90	15.50	45.23
Cheung <i>et al.</i> [22]	6.10	10.70	38.88

separate the useful character features from the noisy or distracting backgrounds. For comparison, test recognition error is introduced to evaluate the model quantitatively. Table 1 summarizes the performances of different methods. Our proposed method achieves 4.45%, 10.19%, and 40.33% errors rate on the three types of dataset, respectively. It can be seen that the proposed method obtains the best score in 2 and the second best in 1 out of the 3 type cluttered background datasets. These results show that the proposed HRBMs model is more effective for cluttered background digit character recognition compared to other methods.

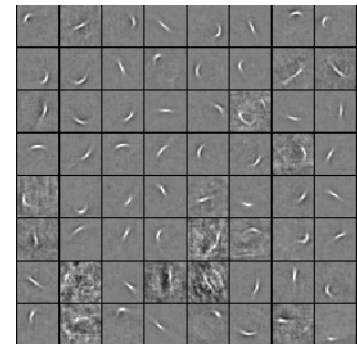
Fig. 7 shows the results of digit character recognition performed by the proposed method on the digit images with various complex cluttered backgrounds. These cluttered backgrounds include rand background, patch background, and rot digit with patch background. It can be seen that all



(a) Original dataset



(b) The trained weights by RBM

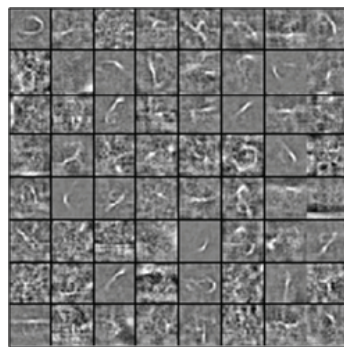


(c) The trained weights by HRBMs

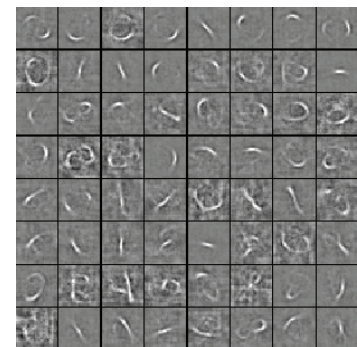
**Fig. 5.** Trained weights of mnist-back-image dataset



(a) Original dataset



(b) The trained weights by RBM



(c) The trained weights by HRBMs

**Fig. 6.** Trained weights of mnist-rot-back-image dataset

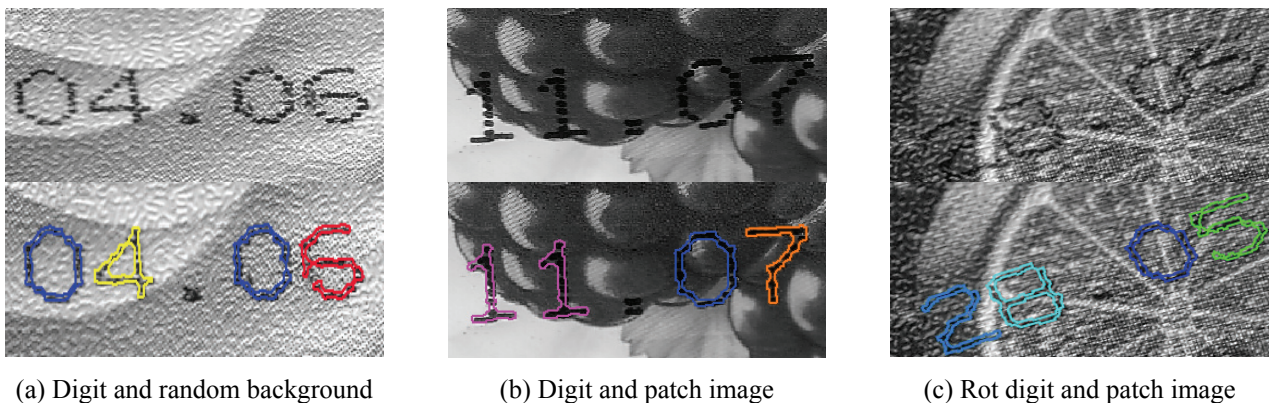


Fig. 7. Digit recognition results based on various background images

the digit characters are correctly recognized by the proposed method.

#### 4. Conclusion

In this paper, a new digit character recognition method is proposed, which can effectively extract useful features from data containing complex distracting factors. In the preprocessing, HRBMs' hidden layer is decomposed with two classes to build a hybrid hierarchical structure, which can induce independently stroke features and clutter property from the input data. Furthermore, HRBMs employ the pixel-wise multiplicative interaction between the 0-1 nodes and input nodes during training, which can be useful in improving the recognition performance. The proposed model is evaluated on the Variations of MNIST dataset and shows superior performance compared to other methods. Results on the digit recognition experiments with three types cluttered backgrounds showed that the efficiency of HRBMs. It is believed that the proposed method will be appealing in building a robust algorithm that can learn from complex data. In the future, the authors will plan to extend HRBMs to deep network, where the stack neutral network on the top of HRBMs will be applied. The authors also plan to explore HRBMs for other tasks, such as segmentation and large image recognition.

#### Acknowledgements

The authors are grateful for the support from the National Science Foundation of China (51479159) and the Soft Science Project of China's Ministry of Transport (2013-322-811-470).

#### References

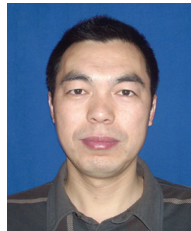
[1] R. G. J. Wijnhoven and P. H. N. de With, "Identity

- Verification Using Computer Vision for Automatic Garage Door Opening," *IEEE Transactions Consumer Electron.*, vol. 57, no. 2, pp. 906-914, May. 2011.
- [2] N. Y. Lee and P. H. Ho, "Digital Signature with a Threshold Subliminal Channel," *IEEE Transactions Consumer Electron.*, vol. 49, no. 4, pp 1240-1242, Nov. 2003.
- [3] K. S. Yildirim, A. Ugur, and A. C. Kinaci, "Design and Implementation of a Software Presenting Information in DVB Subtitles in Various Forms," *IEEE Transactions Consumer Electron.*, vol. 53, no. 4 pp. 1656-1660, Nov. 2007.
- [4] S. D. Choi and S. Y. Lee, "3D Stroke Reconstruction and Cursive Script Recognition with Magnetometer-aided Inertial Measurement Unit", *IEEE Transactions Consumer Electron.*, vol. 58, no. 2, pp. 661-669, May. 2012.
- [5] Muhammad Suhail Akhtar, Hammad A. Qureshi, "Handwritten Digit Recognition through Wavelet Decomposition and Wavelet Packet Decomposition," in *Proc. IEEE International Conference on Digital Information Management*, Islamabad, Pakistan, pp. 143-148, Sep. 2013.
- [6] Thomas Deselaers, Tobias Gass, Georg Heigold, and Hermann Ney, "Latent Log-linear Models for Handwritten Digit Classification," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 34, no. 6, pp. 1105-1117, Jun. 2012.
- [7] R. Mittelman, H. Lee, B. Kuipers, and S. Savarese, "Weakly Supervised Learning of Mid-level Features with Beta-Bernoulli Process Restricted Boltzmann Machines", in *Proc. IEEE International Conference Computer Vision and Pattern Recognition*, Portland, OR, USA, pp. 476-483, Jun. 2013.
- [8] H. Cecotti and S. Vajda, "A Radial Neural Convolutional Layer for Multi-oriented Character Recognition," in *Proc. IEEE International Conference on Document Analysis and Recognition*, Washington DC, USA, pp. 668-672, Aug. 2013.
- [9] H. T. Phan, A. T. Duong, N. D. H. Le, S. T. Tran, "Hierarchical Sparse Autoencoder Using Linear

- Regression-based Features in Clustering for Handwritten Digit Recognition,” in *Proc. International Symposium on Image and Signal Processing and Analysis*, Trieste, Italy, pp. 183-188, Sep. 2013.
- [10] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, “Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations,” *In Proc. International Conference on Machine Learning*, Montreal, Canada, pp. 609-616, Jun. 2009.
- [11] D. C. Ciresan, U. Meier, L. M. Gambardella, and J. Schmidhuber, “Convolutional Neural Network Committees for Handwritten Character Classification,” *In Proc. IEEE International Conference on Document Analysis and Recognition*, Beijing, China, pp. 1135-1139, Sep. 2011.
- [12] D. C. Ciresan, U. Meier, and J. Schmidhuber, “Multi-column Deep Neural Networks for Image Classification,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, USA, pp. 3642-3649, Jun. 2012.
- [13] A. Fischer and C. Igel, “An introduction to restricted Boltzmann machines,” *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, Springer Berlin Heidelberg, pp. 14-36. 2012.
- [14] A. R. Mohamed, G. E. Dahl, and G. Hinton, “Acoustic Modeling Using Deep Belief Networks,” *IEEE Trans. Audio, Speech, and Language Processing*, vol. 20, no. 1, pp. 14-22, Jan. 2012.
- [15] G. W. Taylor and G. E. Hinton, “Factored Conditional Restricted Boltzmann Machines for Modeling Motion Style,” *In Proceedings of the 26th Annual International Conference on Machine Learning*, Montreal, Canada, pp. 1025-1032, Jun. 2009.
- [16] M. Ranzato, J. Susskind, V. Mnih, and G. Hinton, “On Deep Generative Models with Applications to Recognition,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Colorado Springs, USA, pp. 2857-2864, Jun. 2011.
- [17] Variations on the MNIST handwritten digit dataset: <http://www.iro.umontreal.ca/~lisa/twiki/bin/view.cgi/Public/MnistVariations>. Oct. 23th 2012.
- [18] Nair Vinod, Geoffrey Hinton, “Implicit mixtures of restricted Boltzmann machines,” *Advances in neural information processing systems*. pp. 1145-1152, 2009.
- [19] P. Vincent, H. Larochelle, Y. Bengio, and P. A. Manzagol, “Extracting and Composing Robust Features with Denoising Autoencoders,” *In Proc. International Conference on Machine learning*, Helsinki, Finland, pp. 1096-1103, Jul. 2008.
- [20] G. E. Hinton, “Training Products of Experts by Minimizing Contrastive Divergence,” *Neural Computation*, vol. 14, no. 8, pp. 1771-1800, Aug. 2002.
- [21] S. Rifai, P. Vincent, X. Muller, X. Glorot, and Y. Bengio, “Contractive Auto-encoders: Explicit Invariance During Feature Extraction,” *In Proc. International Conference on Machine Learning*, Bellevue, Washington, USA, pp. 833-840, Jun. 2011.
- [22] B. Cheung and C. Sable, “Hybrid Evolution of Convolutional Networks,” *In Proc. IEEE International Conference Machine Learning and Applications and Workshops*, Hawaii, USA, vol. 1, pp. 293-297, Dec. 2011.



**Jie Yang** was born in Wuhan, China, in 1960. She received her BS degree in Communication Engineering from XiDian University, China, in 1982 and MS degree in Computer and Automation from Wuhan Transportation University, China, in 1988. She received Ph.D. degree in Electronic Engineering from the Shanghai Jiao Tong University, China, in 1999. She has been a Professor of School of Electronics and Information at Wuhan University of Technology since 1999. Her current research interests include computer vision and pattern recognition, information transmission theory and application.



**Yadong Sun** was born in Shandong, China, in 1977. He received his BS degree in Electronic Information Engineering from Air Force Engineering University, China, in 2000 and MS degree in Signal and Information Processing from Wuhan University of Technology, China, in 2007. He is currently a PHD candidate in Information and Communication Engineering in Wuhan University of Technology, Wuhan, China. His current research includes signal processing on image and computer vision.



**Liangjun Zhang** was born in Jiangsu, China, in 1987. He received his BS degree in Electronic Information Engineering from Wuhan University of Technology, China, in 2009 and MS degree in Signal and Information Processing in 2012. He is currently a PHD candidate in Information and Communication Engineering in Wuhan University of Technology, Wuhan, China. His current research includes signal processing on image and passive radar, and blind signal processing.





**Qingnian Zhang** received his BS degree in Design and manufacture of port machinery from Wuhan Institute of Water Transportation Engineering, China, in 1982 and MS degree in Transportation management engineering from Wuhan Transportation University, China, in 1988. He received

Ph.D. degree in Machinery design and theories from the Wuhan University of Technology, China, in 2002. He has been a Professor of School of Transportation at Wuhan University of Technology since 2002. His current research interests include Traffic and transportation planning, Optimization and decision making of transportation system, Transportation safety management etc.