

Shape Modeling Based on Convolutional Restricted Boltzmann Machines

Xi-Li Wang^{1,*}, and *Fen Chen*¹

¹ School of Computer Science, ShaanXi Normal University, 710119 Xi'an, ShaanXi ,China

Abstract. This paper proposes a kind of shape model based on convolutional restricted Boltzmann machines(CRBM), which can be used to assist the task of image target detection and classification. The CRBM is a generative model that can model shapes through the generative capabilities of the model. This paper presents the visual representation, construction process and training method of the model construction. This paper does experiments on the Weizmann Horse dataset. The results show that, compared with RBM, although the training time of this model is slightly longer, the test time of the model is similar, and it can better shape modeling, modeling of the details of the shape can be well expressed. The samples generated from CRBM look more realistic. The difference between the shape and the original shape generated by Euclidean distance measurement shows that the model has a strong ability to model shapes.

1 Introduction

Modeling the target shape in the image, the resulting shape can be a complete expression of the target contour, for the subsequent image segmentation, classification, target detection and other tasks to lay the foundation. Such as when the shape is applied to the field of image segmentation^[1], because the shape can express the outline of the target, so the shape and target alignment, the target can be a good cut away from the background, to achieve the ideal segmentation results.

At present, there are many ways to represent the shape. A statistical activity shape model represents shapes by means of a set of labeled points^[2]. It selects the edge of the target shape as a set of marked points, through the model deformation to achieve consistent with the target shape. The advantage of the activity shape model is that it can automatically extract the shape features, and the convergence speed is faster, but it is easy to fall into the local optimum, and requires the selection of the appropriate marks for all shapes of the target. Level set method represents the shape by means of symbolic distance functions^[3]. It defines the energy function of the model based on the distance between the shape of the current shape and the shape represented by the model. It controls the evolution of the curve by optimizing the defined energy function, and it can express the shape of the target when the energy function reaches the minimum value. But in the process of curve evolution, the curve is prone to shock, so it can't establish the shape model for the multi class target.

Recently, deep learning is being developed rapidly^[4]. There are many models, such as: restricted Boltzmann machines(RBM)^[5], deep belief networks(DBN)^[6], and these models are applied to many fields, and show a

strong learning ability in a large number data. RBM was proposed by Smolensky. As the model pulls the input image into a one-dimensional vector as an input, ignoring the two-dimensional structure information among images, so when using the model to generate the shape, the sampled shape is blurred.

CNN's network mechanism is local receptive fields and weight sharing, and it has been successfully applied to many areas, such as handwritten digital classification, target recognition^[7-9]. Using CNN's network mechanism^[10-11], Guillaume Desjardins and Yoshua Bengio proposed convolution restricted Boltzmann machine(CRBM)^[12] which brought in convolution into RBM in 2008. But unlike RBM, the input image is input as a two-dimensional structure. When the model is used to represent and generate the shape, not only can the shape of the target be generated, but also the detailed features of the object can be well expressed.

This paper introduces the intuitive expression, construction process and training method of CRBM. Different from the RBM model modeling, it can better extract the target two-dimensional spatial structure information and reduce the number of parameters to be trained in the model. This paper does experiments on the Weizmann Horse dataset, which indicate that CRBM is more efficient than RBM in modeling the shape.

2 Restricted Boltzmann Machines

The RBM is a two-layer, undirected graphical model with a set of visible units \mathbf{v} and a set of hidden units \mathbf{h} . Fig. 1 is the structure of RBM model.

* Corresponding author: wangxili@snnu.edu.cn

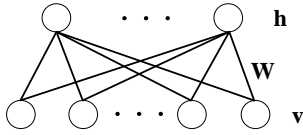


Fig. 1. Structure of RBM model.

The RBM is a generative model. And the key characteristic of the RBM is that its visible units and hidden units are fully connected, but no connection among the visible or hidden units. That is visible units are also conditionally independent and its state is also only affected by the hidden units.

There, we assume the visible units are binary valued. The vector $\mathbf{v}=(v_1, v_2, \dots, v_n)$ and $\mathbf{h}=(h_1, h_2, \dots, h_m)$ respectively represent the state vectors of the visible layer unit and hidden layer unit. We assumption that the visible units are binary valued, that is $\forall i, j, v_i \in \{0, 1\}, h_j \in \{0, 1\}$. So for a given set of states (\mathbf{v}, \mathbf{h}) , the binary state h_j and v_i is set to 1 with probability

$$P(h_j = 1 | \mathbf{v}) = \delta(\sum_{i=1}^n W_{ij} v_i + b_j) \quad (1)$$

$$P(v_i = 1 | \mathbf{h}) = \delta(\sum_{j=1}^m W_{ij} h_j + a_i) \quad (2)$$

Where $\delta(x) = \frac{1}{1 + \exp(-x)}$ is the logistic sigmoid function.

For the RBM, when the image dimensionality is relatively high or the image is relatively large, the size of the model becomes very large, training parameters increased. But CNN network mechanism makes this model very suitable for dealing with natural images. In order to solve the problem of RBM in image application, CNN is introduced into RBM and also proposes the ability to generate of RBM.

3 Convolution Restricted Boltzmann Machines

3.1 Model structure and mathematical expression

The CRBM is an extension of the RBM. It has a visible layer \mathbf{v} and a hidden layer \mathbf{h} . Simultaneously, the CRBM is also a generative model. But unlike RBM, the input layer of CRBM is an image. The characteristic of the CRBM is local receptive fields and weight sharing^[17], that is, the weights of CRBM between the visible layer and the hidden layer are local connection and also shared among all locations. Fig.2 is the structure of CRBM model. Here is just a group showing a hidden layer.

For CRBM, we assume that the input image size is $N_v \times N_v$, so the input layer of CRBM is an $N_v \times N_v$ matrix of two-dimensional. The hidden layer includes K groups, where each graph represents a feature map of the hidden layer and is $N_h \times N_h$ matrix of two-dimensional.

So the number of hidden units is $K \times N_h \times N_h$. Each group is connected to a convolutional filter of size $N_w \times N_w$ ($N_w = N_v - N_h + 1$). In addition, all visible units share the same bias c . All hidden units of one group share the same bias b_k , that is, hidden layer of K group, there are K bias.

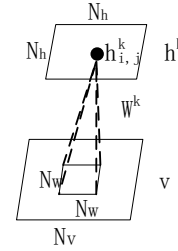


Fig. 2. Structure of CRBM model

Before defining the energy function of the model, we will do same assumptions and symbolic descriptions for the convenience of the latter description. First, we assume that the input to the algorithm is a binary image. We use 1 to represent the target and 0 to represent the background. Second, we use $*$ to denote convolution, and \cdot to denote an element-wise product followed by summation and A to denote flipping the matrix A horizontally and vertically.

We define the CRBM models' energy function as follows:

$$E(\mathbf{v}, \mathbf{h} | \theta) = -\sum_{k=1}^K \sum_{i,j=1}^{N_h} \sum_{r,s=1}^{N_w} h_{ij}^k W_{rs}^k v_{i+r-1, j+s-1} - \sum_{k=1}^K b_k \sum_{i,j=1}^{N_h} h_{ij}^k - c \sum_{i,j=1}^{N_v} v_{ij} \quad (3)$$

Where $\theta = \{W_{rs}^k, b_k, c\}$ is the model parameter.

We define the models' energy function using the previously defined symbols as follows:

$$E(\mathbf{v}, \mathbf{h}) = -\sum_{k=1}^K h^k \cdot (\tilde{W}^k * v) - \sum_{k=1}^K b_k \sum_{i,j=1}^{N_h} h_{i,j}^k - c \sum_{i,j=1}^{N_v} v_{ij} \quad (4)$$

The joint probabilistic distribution for an CRBM is defined by its energy function as follows:

$$P(\mathbf{v}, \mathbf{h}) = \frac{1}{Z(\theta)} \exp(-E(\mathbf{v}, \mathbf{h})) \quad (5)$$

Where $Z(\theta) = \sum_{\mathbf{v}, \mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h} | \theta)}$ is a normalization constant.

The binary state h_{ij}^k and v_{ij} is set to 1 with probability

$$P(h_{ij}^k = 1 | \mathbf{v}) = \delta((\tilde{W}^k * v)_{ij} + b_k) \quad (6)$$

$$P(v_{ij} = 1 | \mathbf{h}) = \delta((\sum_k W^k * h^k)_{ij} + c) \quad (7)$$

Where $\delta(x) = \frac{1}{1 + \exp(-x)}$ is the logistic sigmoid function.

3.2 Model training and sampling

The training of the CRBM model uses the same training algorithm as the RBM—the contrast divergence(CD) algorithm^[13]. Therefore, the gradient update formula for the parameters θ based on the CD algorithm as follows:

$$CD_i(\theta, \mathbf{v}_0) = -\sum_{\mathbf{h}} p(\mathbf{h} | \mathbf{v}_0) \frac{\partial E(\mathbf{v}_0, \mathbf{h})}{\partial \theta} + \sum_{\mathbf{h}} p(\mathbf{h} | \mathbf{v}_k) \frac{\partial E(\mathbf{v}_k, \mathbf{h})}{\partial \theta} \quad (8)$$

After the model has been trained, it's used to generate shapes. The method of generating shapes is multi-step Gibbs sampling. Using a shape as the initial state of the visual layer, perform a multi-step Gibbs sampling, i.e., "1", "2", "3", "4", ..., "n" in Fig.3 gets approximately the distribution defined by CRBM, and finally, the \mathbf{v} obtained from step n sampling is the shape of the model generated. The specific process is shown in Fig. 3.

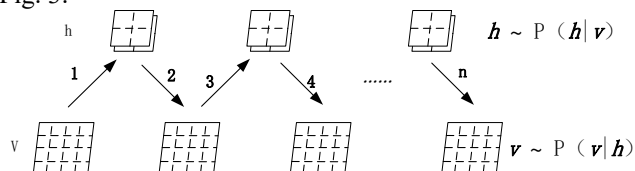


Fig. 3. CRBM model sampling process

4 Experiments

4.1. Experimental data and parameter settings

For our experiments, we used the Weizmann Horse dataset. This dataset consists of three types of images: color, grayscale, binary. In this experiment, we used the binary image in the data set. We select 200 pairs images as training set, 128 pairs images as test set, and all image are normalized to 32×32. Fig.4 is the image in the training set. Form Fig.4, we can see that the trained horse has different shapes, and the head of the horse is directed in one direction.

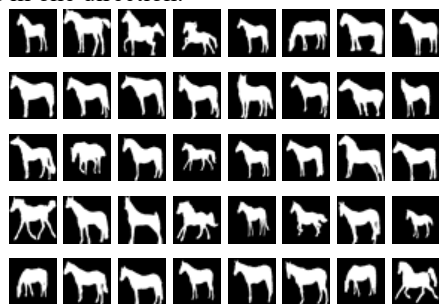


Fig. 4. Training set images

The experimental environment is Matlab R2014a installed under the Win 10 system, and the computer is configured as Intel(R) Xeon(R) CPU E5-2690, 2.6GHz, 256GB RAM. In the experiment, CRBM consisted of 20 groups of 3×3 pixel filters. We used 0.05 as the learning rate and 1000 as iterations. For RBM, the number of units of visual layer and hidden layer is 1024 and 500, and 0.05 as the learning rate and 1000 as iterations.

4.2. Experimental results and analysis

We compare the modeling capabilities of the proposed model CRBM with RBM, and the two models are trained

on the same data set. The training time of CRBM is 564.54s, and the training time of RBM is 303.39s.

After the training of the modeling, two different inputs are used to test the modeling ability of the model. The first is the first input: the full image of the training set and the test set is used as input. Table 1 is the average Euclidean distance measure of the two models.

Table 1. The average Euclidean distance measure of the complete image reconstruction result

	Training set	Test set
RBM	7.3862	7.9314
CRBM	4.1946	4.5842

Fig.5 shows the five sampling result images of training set and test set respectively. In Fig.5, (A) and (B) represent the sampling result from CRBM and RBM of training set and test set respectively, where (a) is the input image, and (b) and (c) are the result of RBM and CRBM sampling to generate the shape respectively.

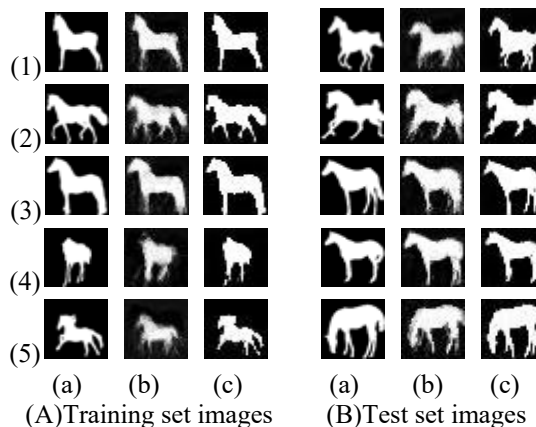


Fig. 5. The results of full images input

From Fig.5, we can see that CRBM sampling to generate the shape is to be more practical than RBM, RBM model to generate the shape is of more vague, easily lack of details, such as the horse's leg, and CRBM is a good retain the details of the horse, the reconstructed horse similar input. The Euclidean distance measure of the image in Fig.5 is shown in Table 2. From Table 2, we can see that whether the training set or the test set, the Euclidean distance measurement shows that the shape of CRBM model is closer to the original target, and the modeling shape is better.

Table 2. CRBM and RBM generate shape Euclidean distance metrics

Num	Training set		Test set	
	RBM	CRBM	RBM	CRBM
(1)	5.6974	3.4495	7.4465	5.0990
(2)	7.4745	5.6458	7.1335	4.0056
(3)	5.8784	3.4641	6.9445	3.4495

(4)	8.4750	4.2426	7.0250	3.9426
(5)	8.8402	4.7990	8.4089	4.3166

In order to further verify the modeling ability of the model, one type of noise is added, whether it is an image in training set or test set. It is added the mean 0 and variance for the Gauss noise 0.04. Table 3 is the average Euclidean distance measure of the two models with Gauss noises.

Table 3 The average Euclidean distance measure for adding Gauss noise reconstruction result

	Training set		Test set	
	RBM	CRBM	RBM	CRBM
Gauss noise	7.5513	4.5100	8.0654	4.9574

One types of noise added images is fed into the CRBM and RBM. Figure 6 show the five sampling result images of training set and test set respectively. In Fig.6, (A) and (B) represent the sampling result of image to add noise from CRBM and RBM of training set and test set respectively, where (a) is the original image, (b) is the noise image, and (c) and (d) are the result of RBM and CRBM sampling to generate the shape respectively.

From Fig.6, we can see that whether the training set or the test set of images, these two models can effectively remove the impact of noise, but the shape RBM generated blurred, and lost a lot of horse information, CRBM is better to retain the horse's information. The Euclidean distance measure of the image in Fig.6 is shown in Table 4. From Table 4, we can see that the Euclidean distance measurement shows that the shape of CRBM model is more similar to the original target, and is better for modeling shapes.

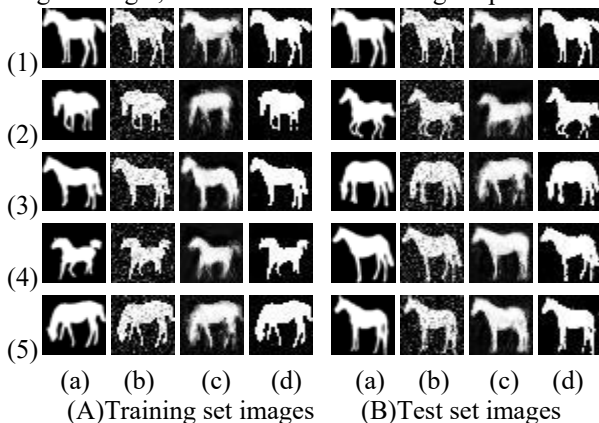


Fig. 6. The results of add Gauss noise images as input

Table 4. CRBM and RBM add Gauss noise generate shape Euclidean distance metrics

Num	Training set		Test set	
	RMB	CRBM	RBM	CRBM
(1)	7.7427	4.5357	8.8935	5.0050
(2)	8.4117	3.7417	7.6836	4.4325
(3)	6.6984	3.5331	7.6712	5.2995
(4)	7.3935	4.2050	6.9919	4.3499
(5)	7.0117	4.9621	7.2027	4.8137

5 Conclusuin

As a further improvement of RBM, CRBM combines RBM and CNN, adding CNN advantages to RBM, while retaining the characteristics of RBM. This paper builds shape models based on the generative capabilities of CRBM and RBM. The experiment was carried out in two aspects. By comparing with the experimental results of RBM, it is proved that the modeling ability of CRBM model can be fully expressed for the target information. The next step is to consider the shape which is applied to image segmentation and shape as priority information constraints to be segmented, improved the segmentation results.

References

- Huang L, Nie J, Wei Z. Human body segmentation based on shape constraint[J]. Machine Vision and Applications, (2007):1-10.
- Yan P, Xu S, Turkbey B, et al. Discrete deformable model guided by partial active shape model for TRUS image segmentation[J]. IEEE Transactions on Biomedical Engineering, **57.5**(2010): 1158-1166.
- Rousson M, Paragios N. Prior knowledge level set represent & visual grouping[J]. International Journal of Computer Vision, **76.3**(2008): 231-243.
- LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, **521.7553**(2015): 436-444.
- Fischer A, Igel C. An introduction to restricted Boltzmann machines[J]. Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, (2012): 14-36.
- Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural computation, **18.7**(2006): 1527-1554.
- He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C], Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016): 770-778.
- Elfwing S, Uchibe E, Doya K. Expected energy-based restricted Boltzmann machine for classification[J]. Neural Networks, **64**(2015): 29-38.
- Zhao Z, Jiao L, Zhao J, et al. Discriminant deep belief network for high-resolution sar image classification[J]. Pattern Recognition, **61**(2017): 686-701.
- Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C], Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015): 1-9.
- Smolensky P. Information processing in dynamical systems: Foundations of harmony theory[R]. Colorado University at Boulder Dept of Computer Science, (1986).
- Guillaume Desjardins, Yoshua Bengio. Empirical Evaluation of Convolutional RBMs for Vision[J], DIRO, Université de Montréal, (2008):1-13.

13. Hinton. A practical guide to training restricted Boltzmann machines. **9.1**(2010): 599-619.
14. Honglak Lee, Roger Grosse, Rajesh Ranganath, et. Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations[C]. Proceedings of the 26th International Conference on Machine Learning, (2009):609-616.