Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.12785/amis/081L46

Structural Similarity Sparse Coding

Zhiqing Li[1,*], *Weizhong Zhao*[1] *and Zhixin Li*[2]

[1] Key Laboratory of Intelligent Computing and Information Processing of Ministry of Education, Xiangtan University, 411105 Xiangtan, China
[2] College of Computer Science and Information Technology, Guangxi Normal University, 541004 Guilin, China

Received: 22 May. 2013, Revised: 16 Sep. 2013, Accepted: 17 Sep. 2013
Published online: 1 Apr. 2014

Abstract: Sparse coding theory demonstrates that the neurons in primary visual cortex form a sparse representation of natural scenes in the viewpoint of statistics. In this paper, we propose a novel sparse coding model based on structural similarity for natural image patch feature extraction. The advantage for our model is to be able to preserve structural information from a scene, which human visual perception is highly adapted for. Using the proposed sparse coding model, the validity of image patch feature extraction is testified. Furthermore, compared with standard sparse coding model, the experimental results show that the quality of reconstructed images obtained by our method outperforms standard sparse coding model.

Keywords: Natural image, sparse coding, structural similarity, computational model, biological visual system

1 Introduction

The computation capabilities and limitations of neurons, and the environment in which the organism lives, are two fundamental components driving the evolution and development of human perceptual systems. At the same time, the use of environmental constraints is most clearly evident in sensory systems, where it has long been assumed that neurons are adapted to the signals to which they are exposed. Because not all signals are equally like each other, it is natural to assume that perceptual systems should be able to best process those signals that occur most frequently. Thus, it is the statistical properties of the environment that are relevant for sensory processing.

Efficient coding hypothesis [1] provides a quantitative relationship between environmental statistics and neural processing. Barlow hypothesized that the role of early sensory neurons is to remove statistical redundancy in the sensory input. Furthermore, Olshausen and Field put forward a model, called sparse coding (*SC*), which made the variables (or neurons stimulated by the same stimulus in the neurobiology.) be activated (i.e., significantly non-zero) only rarely [2,3]. Vinje's experimental results validated the sparse properties of neural responses under natural stimuli conditions [4]. Therefore sparse coding theory was broadly investigated [5,6,7,8,9,10,11,12].

Objective methods for assessing perceptual image quality traditionally attempted to quantify the visibility of errors (*differences*) between a reconstructed image and an actual image in SC model and improved models. The simplest and most widely used full-reference quality metric is the mean squared error (*MSE*), computed by averaging the squared intensity differences of reconstructed and actual image pixels, along with the related quantity of peak signal-to-noise ratio (*PSNR*). These are appealing because they are simple to calculate, have clear physical meanings, and are mathematically convenient in the context of optimization. But they are not very well matched to perceived visual quality [13,14].

In this paper, we propose structural similarity sparse coding (*SS_SC*) employing a novel quality assessment method that takes advantage of known characteristics of the human visual system (*HVS*). Under the assumption that human visual perception is highly adapted for extracting structural information from a scene, we introduce structural similarity for quality assessment based on the degradation of structural information.

The rest of this paper is structured as follows. Section 2 describes the SC model and structural similarity measure. In Section 3 we propose SS_SC model and image patch feature extraction approach using SS_SC model. Experiment results are reported and analyzed in Section 4. Finally, we conclude the paper in Section 5.

* Corresponding author e-mail: lizq@xtu.edu.cn

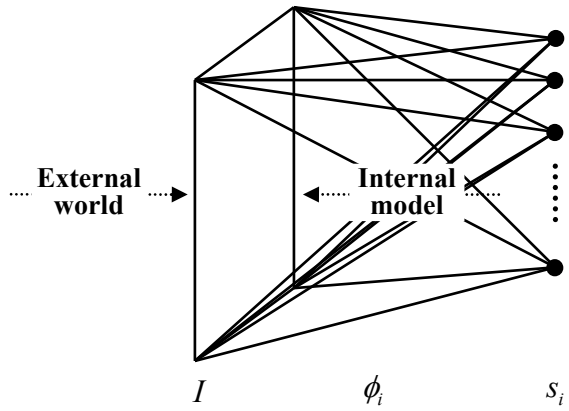


Fig. 1: Linear superposition model

2 Sparse coding model and structural similarity

A perceptual system is exposed to a series of small image patches, drawn from one or more large images, just like the classic receptive field (CRF) of neurons. Imagine that each image patch is represented by the vector I and has been formed by the linear combination of N basis functions. The linear superposition model of image is shown in Figure 1. In this model, images are assumed to be composed of a linear superposition of basis functions, ϕ_i, mixed together with amplitudes s_i.

The basis functions form the columns of a fixed matrix, A. The weight of this linear combination is given by a vector, S. Each component of this vector has its own associated basis function, and represents a response value of a neuron in vision system. The linear synthesis model is therefore given by:

$$I = AS = \sum_{i=1}^{M} s_i \phi_i. \tag{1}$$

In a cortical interpretation, the S models the responses of (signed) simple cells, and the column of matrix A closely related to their CRF's.

2.1 Sparse coding model

In an influential paper, Olshausen and Field applied two criteria to seek the optimal basis functions and the coefficients [3]. One of the criteria is how well the code describes the input. It can be measured by the squared error between the input and its reconstruction by the network:

$$Error(A,S) = \sum_{i=1}^{N} [I_i - \sum_{j=1}^{M} s_j \phi_{i,j}]^2. \tag{2}$$

As an additional criteria for sparse coding, Olshausen and Field proposed the "sparseness" cost for seeking sparse codes. The sparseness cost function is given by

$$Sparseness(S) = \sum_{i=1}^{M} \beta(\frac{s_i}{\sigma}) \tag{3}$$

where σ is a scaling constant, and $\beta(x)$ is a nonlinear function such as $|x|$, $exp(-x^2)$, and $log(1+x^2)$. The cost sparseness favors the codes which consist of minimal number of non-zero coefficients. As a result, the network seeks the coefficients which are statistically independent each other over an ensemble of input data. In the case that the data contains some forms of higher-order statistical structure as found in natural images, it can be captured by using this sparseness cost function. So the search for a sparse code can be formulated as an optimization problem by constructing the following cost function to be minimized:

$$E(A,S) = \sum_{i=1}^{N} [I_i - \sum_{j=1}^{M} s_j \phi_{i,j}]^2 + \lambda \sum_{i=1}^{M} \beta(\frac{s_i}{\sigma}). \tag{4}$$

2.2 Structural similarity

Natural image signals are highly structured: their pixels exhibit strong dependencies, especially when they are spatially proximate, and these dependencies carry important information about the structure of the objects in the visual scene [14].

The HVS is highly adapted to extract structural information from the visual scene. Therefore, a measurement of structural similarity should provide a good approximation to perceptual image quality. Wang and Bovik et al. [13] developed a Structural Similarity Index and demonstrate its promise through a set of intuitive examples, as well as comparison to both subjective ratings and state-of-the-art objective methods on a database of images compressed with JPEG and JPEG2000. Shown in Figure 2 is an example. It is comparison of "Boat" images with different types of distortions, all with MSE=210. (a) Original image. (b) Contrast-stretched image, $SSIM$=0.9168. (c) Mean-shifted image, $SSIM$=0.9900. (d) JPEG compressed image, $SSIM$=0.6949. (e) Blurred image, $SSIM$=0.7052. (f) Salt-pepper impulsive noise contaminated image, $SSIM$=0.7748.

Suppose x and y are two nonnegative image signals, which have been aligned with each other. The structural similarity between signals x and y is given by

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{5}$$

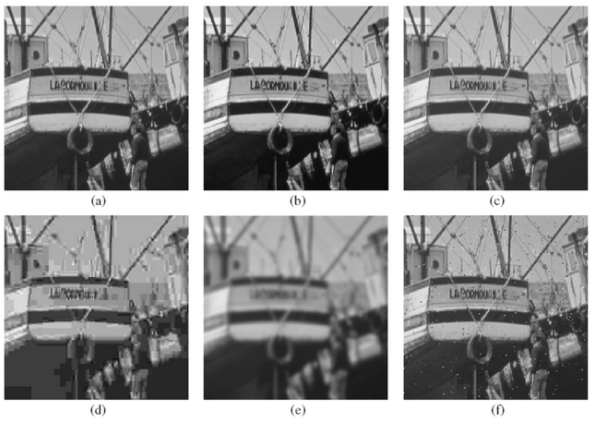


Fig. 2: Boat images with different distortions with different types of distortions, all with *MSE*=210

where

$$\mu_x = \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i,\ \mu_y = \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i,$$

$$\sigma_x = [\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)^2]^{1/2},$$

$$\sigma_y = [\frac{1}{N-1}\sum_{i=1}^{N}(y_i-\mu_y)^2]^{1/2},$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)(y_i-\mu_y),$$

$$and\ 0 < C_1,\ C_2 \ll 1.$$

3 Structural similarity sparse coding

On the basis of the Olshausen's SC model, we propose a novel sparse coding model based on structural similarity called SS_SC model. Here, we use the minimum reconstruction error and the sparseness like Olshausen, but structural similarity for quality assessment between reconstructed image and actual image is also considered. As such, the objective function can be constructed as follows:

$$E(A,S) = \lambda_1 \sum_{i=1}^{N}(I_i - \hat{I}_i)^2 + \lambda_2(1 - SSIM(I,\hat{I})) + \lambda_3 \sum_{i=1}^{M}\beta(\frac{s_i}{\sigma}) \quad (6)$$

where I and $\hat{I}$ denotes respectively actual and reconstructed images, $\hat{I} = \sum_{i=1}^{M} s_i\phi_i$, ϕ_i and s_i denotes respectively the ith column vector of A and the ith component of S, $\lambda_1, \lambda_2, \lambda_3 \geq 0$ is respectively the weights of squared error, structural similarity and sparseness.

$E(A,S)$ is the sum of three terms: the first term computes the squared error, which forces the basis functions, A, to span the input space; the second term measures how well the code describes the structural information from the actual image; and the third term incurs a penalty on the coefficient activities, which encourages sparse representation.

3.1 Learning algorithm

The goal of efficient coding is to learn the basis functions that can best account for the structure in images in terms of statistically independent events. Learning is accomplished by minimizing Equation (6). The process for minimizing $E(A,S)$ can be divided into two nested stages. In the inner stage, $E(A,S)$ is minimized with respect to the s_i for a batch of pattern, holding the A fixed. In the outer stage (i.e., on a long timescale, over many image presentations), $E(A,S)$ is minimized with respect to the A.

Let $B_1 = \sum_{i=1}^{N}(I_i - \hat{I}_i)^2$, $B_{21} = 2\mu_I\mu_{\hat{I}} + C_1$, $B_{22} = 2\sigma_{I\hat{I}} + C_2$, $B_{23} = \mu_I^2 + \mu_{\hat{I}}^2 + C_1$, $B_{24} = \sigma_I^2 + \sigma_{\hat{I}}^2 + C_2$, $B_3 = \sum_{i=1}^{M}\beta(\frac{s_i}{\sigma})$, then

$$\frac{\partial SSIM(I,\hat{I})}{\partial s_i} = \frac{B_{21} * B_{22}}{B_{23} * B_{24}}(\frac{1}{B_{21}}\frac{\partial B_{21}}{\partial s_i} + \frac{1}{B_{22}}\frac{\partial B_{22}}{\partial s_i} - \frac{1}{B_{23}}\frac{\partial B_{23}}{\partial s_i} - \frac{1}{B_{24}}\frac{\partial B_{24}}{\partial s_i})$$

The inner stage minimization over the s_i can be performed by conjugate gradient method, so the s_i is determined by the differential equation:

$$\frac{\partial E(A,S)}{\partial s_i} = \lambda_1\frac{\partial B_1}{\partial s_i} - \lambda_2\frac{\partial SSIM(I,\hat{I})}{\partial s_i} + \lambda_3\frac{\partial B_3}{\partial s_i} \quad (7)$$

The outer stage minimization over the A may be finished by simple gradient descent method. The learning rule for it is given by

$$\frac{\partial E(A,S)}{\partial \phi_{i,j}} = \lambda_1\frac{\partial B_1}{\partial \phi_{i,j}} - \lambda_2\frac{\partial SSIM(I,\hat{I})}{\partial \phi_{i,j}}, \quad (8)$$

$$\Delta\phi_{i,j} = -\eta\frac{\partial E(A,S)}{\partial \phi_{i,j}} \quad (9)$$

where η is the learning rate.

3.2 Image patch feature extraction using SS_SC

Olshausen and Field forcefully argued that the receptive field is emerged by sparse coding and they applied successfully a sparseness-maximization network to input data to testify their theory. Thus, sparse coding technique can be exploited to perform image patch feature

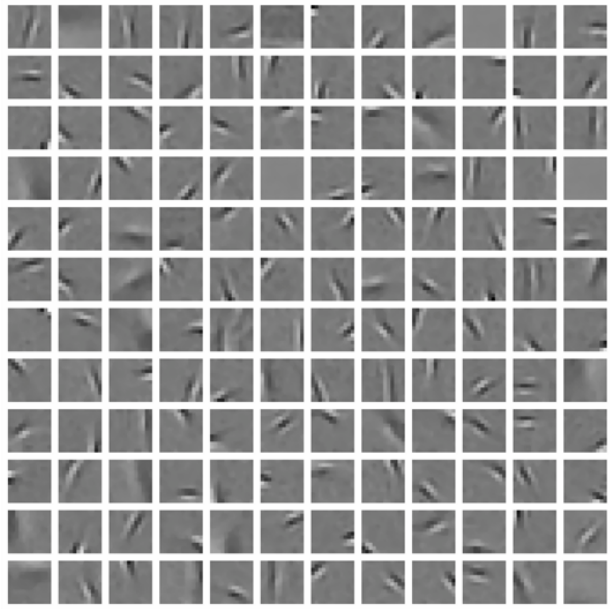


Fig. 3: Basis functions learned after training on natural scenes by SS_SC model

extraction. With learned basis functions, image feature patch extracting is accomplished by minimizing $E(A,S)$ with respect to the s_i for a batch of pattern, holding the A fixed.

4 Experiment results

In order to test the effectiveness of the SS_SC model, we conduct our experiments on a nature image data set. Firstly, we must get the input data matrix. Selecting nature images, which are available on the Internet http://www.cis.hu-t.fi/projects/ica/data/images/, we construct the original images sample set. Then, we sampled randomly subwindows of 12×12 pixels 250000 times from original images, and converted every patch into one column. Thus, the input data set with the size of 144×250000 is acquired, here denoted by matrix X. Consequently, each image patch is represented by a 144 dimensional vector. Secondly, using the updating rules of A and S in turn, we minimized the objective function given in Equation (6).

A stable solution was arrived at after 5000 updates (250000 image presentations) when $\eta = 0.01$, $\sigma = 0.316$, $\lambda_1 = 50$, $\lambda_2 = 10$, $\lambda_3 = 2.2$ and $\beta(x) = log(1+x^2)$. Shown in Figure 3 is a set of 144 basis functions learned after training on 12×12 image patches extracted from natural scenes by the SS_SC model. The learned basis functions simply reflects the fact that natural images contain localized, oriented structures with limited phase alignment across spatial frequency.

4.1 Image patch feature extraction

We sampled randomly subwindows of 12×12 pixels 10000 times from original images, and converted every patch into one column. Using the image patch feature extracting algorithm and 144 learned basis functions image patch features of the input data set with the size of 144×10000 are extracted. For comparison, we also used the SC method to extract image patch features from the same data set, and the experimental results are shown in Table 1. In Table 1, *Avg_Sparseness* denotes averaged sparseness cost of reconstructed image patches; *Avg_Error* denotes averaged squared error between reconstructed image patch and actual image patch; and *Avg_SSim* denotes averaged structural similarity between reconstructed image patch and actual image patch.

Table 1: Comparison of coding capability using different models

Models	SC	SS_SC
Avg_Sparseness	8.0153	8.1600
Avg_Error	0.0842	0.0753
Avg_SSim	0.8542	**0.9231**

It is easy to see that the SS_SC model is promising for image patch feature extraction. Furthermore, Table 1 show that SS_SC model preserves more structural information than SC model.

It is more important that structural similarity between reconstructed image patch and actual image patch is stable in SS_SC model. However, in SC model, the fluctuation range of structural similarity between reconstructed image patch and actual image patch is wide, i.e., perceptual quality of some reconstructed image patches is very poor. Shown in Figure 4 is structural similarity of between reconstructed image patch and actual image patch in SS_SC model and SC model. In SC model, it's evident that the fluctuation range of structural similarity is wider than SS_SC.

4.2 Image reconstruction

Four images were selected to reconstruct by SS_SC model, and these images were used widely in the image processing field. Each image is randomly sampled 5000 times with 12×12 pixels to get the data set. Moreover, in order to find the accurate position of any image patch, we must remember the positions of each image patch appeared. Because of sampling randomly, the same pixel might be found in different image patches. Therefore, we averaged all reconstructed pixels' values of one sample pixel, and used the averaged pixel value as the approximation of the original pixel.

Actually, because of the computation capabilities and limitations of neurons, neural responses are sparser than

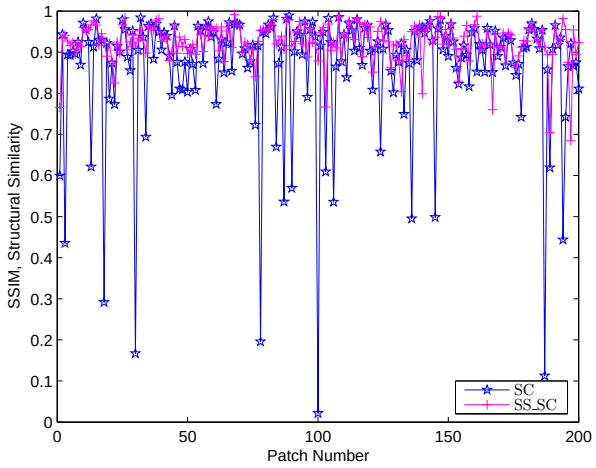


Fig. 4: Structural similarity of between reconstructed image patch and actual image patch

shown in Table 1. We extract sparser image patch feature through increase the weight of sparseness, and then reconstruct images. Shown in Figure 5 are original images and reconstructed images by different model under similar *Avg_Sparseness* when responses are sparser, i.e., *Avg_Sparseness* is smaller. (a-d) Reconstructed images using SC model. (e-h) Reconstructed images using SS_SC model. (i-l) Original images.
(a) *Avg_SSim*=0.3365, *Avg_Sparseness*=6.6830.
(b) *Avg_SSim*=0.3358, *Avg_Sparseness*=6.7997.
(c) *Avg_SSim*=0.2852, *Avg_Sparseness*=6.7820.
(d) *Avg_SSim*=0.3820, *Avg_Sparseness*=7.6822.
(e) *Avg_SSim*=0.7300, *Avg_Sparseness*=6.6190.
(f) *Avg_SSim*=0.7641, *Avg_Sparseness*=6.7566.
(g) *Avg_SSim*=0.7061, *Avg_Sparseness*=6.7628.
(h) *Avg_SSim*=0.7995, *Avg_Sparseness*=7.6775.

Obviously, reconstructed images using SS_SC preserve more structural information than SC model, and the quality of reconstructed images of the former outperforms the latter.

5 Conclusion

In this paper, we proposed a novel sparse coding model based on structural similarity for extracting natural image features. Basis functions obtained by our method much resemble the receptive fields of neurons in primary visual cortex, which behave clearer localized, oriented, bandpass. In order to validate performance of our sparse coding model, we conducted the experiments of image reconstruction. The experimental results showed that SS_SC model can preserve the original image structural information as possible as. Furthermore, the structural similarity of reconstructed image patch using SS_SC model is more stable and greater than using SC model. It is most valuable that the quality of reconstructed images by SS_SC model outperforms SC model under similar *Avg_Sparseness*.

Acknowledgements

This work was supported by the grants from the National Basic Research Priorities Programme of China (973 Program) (Grant No. 2007CB311004), the National Science Foundation of China (Grant Nos. 61105052, 61165009), the Natural Science Foundation of Hunan Province, China (Grant No. 11JJ4051), the Research Foundation of Education Bureau of Hunan Province, China (GrantNo. 10C1262).
The authors are grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper.

References

[1] H. B. Barlow, Possible principles underlying the transformation of sensory messages, in: W.A. Rosenblith ed., Sensory Communication (MIT Press, Cambridge,) 217-234 (1961).
[2] D. J. Field, What is the goal of sensory coding?, Neural Computation, **6**, 559-601 (1994).
[3] B. A. Olshausen and D. J. Field, Emergence of simple-cell receptive fieldproperties by learning a sparse code for natural images, Nature, **381**, 607-609 (1996).
[4] W. E. Vinje and J. L. Gallant, Sparse coding and decorrelation in primaryvisual cortex during natural vision, Science, **287**, 1273-1276 (2000).
[5] D. B. Grimes and R. P. N. Rao, Bilinear sparse coding for invariantvision, Neural Computation, **17**, 47-73 (2005).
[6] P. O. Hoyer, Non-negative Matrix Factorization with sparsenessconstraints, Journal of Machine Learning Research, **5**, 1457-1469 (2004).
[7] Z. Li, Z. Shi, Z. Li and Z. Shi. Image Classification Using Structural Sparse Coding Model, Proceedings of the International Conference on Natural Computation, 624-628 (2009).
[8] Z. Li, Z. Shi, X. Liu and Z. Shi. A Novel Sparse Coding Model Based On Structural Similarity, Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, 4170-4173 (2010).
[9] W. Liu and N. Zheng, Learning sparse features for classification by mixture models, Pattern Recognition Letters, **25**, 155-161 (2004).
[10] J. Malo, I. Epifanio, R. Navarro and E.P. Simoncelli, Non-Linear Image Representation for Efficient Perceptual Coding, IEEE Transactions on Image Processing, **15**, 68-80 (2006).
[11] L. Shang, D. Huang, C. Zheng and Z. Sun, Noise removal using a novel non-negative sparse coding shrinkage technique, Neurocomputing, **69**, 874-877 (2006).

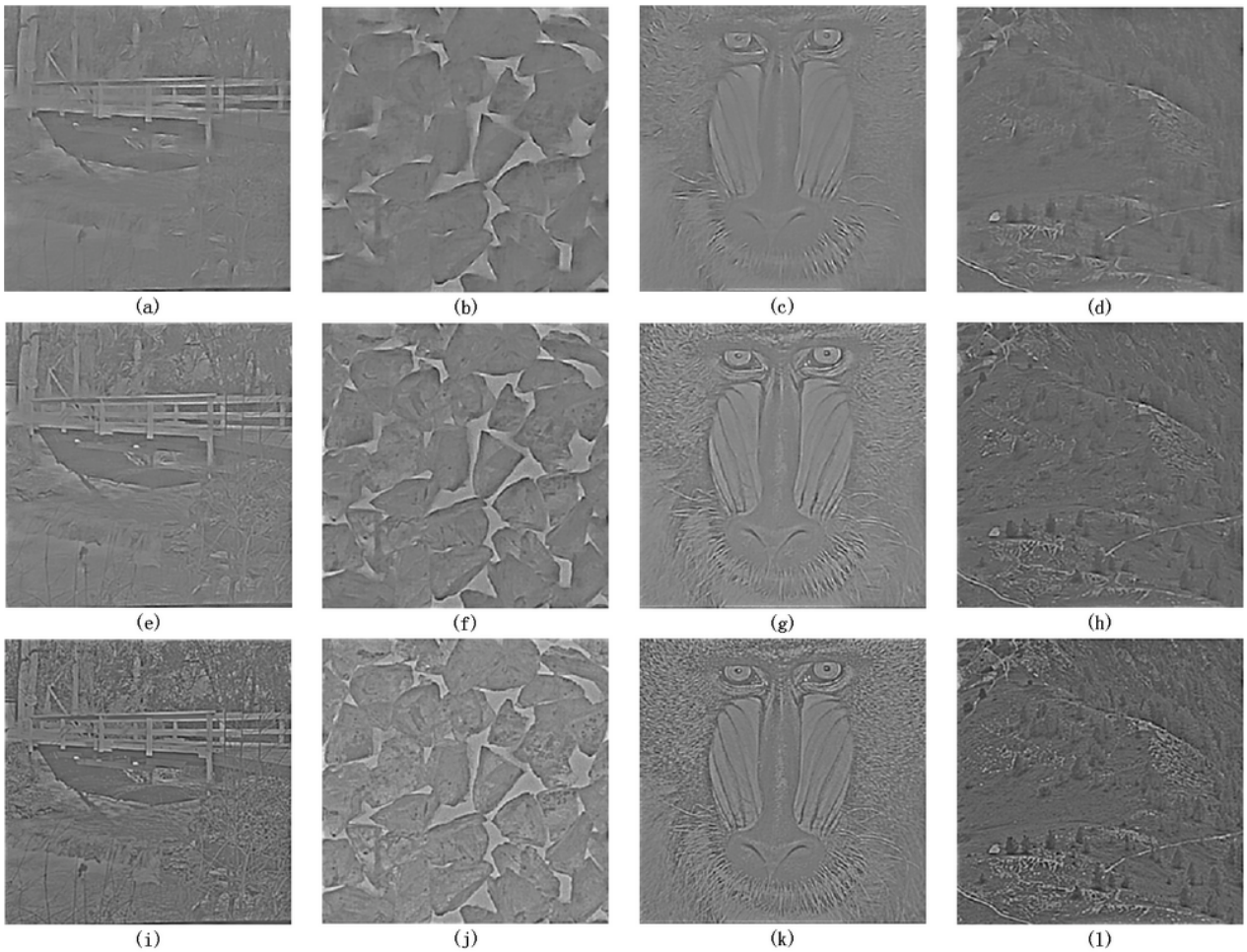


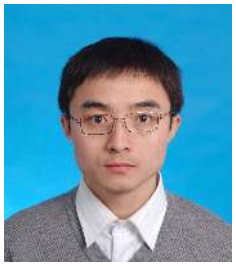
Fig. 5: Original images and reconstructed images by different model under similar *Avg_Sparseness*

[12] A. Balinsky and N. Mohammad, Non-Linear Filter Response Distributions of Natural Images and Applications in Image Processing, Applied Mathematics & Information Sciences, **3**, 367-389 (2011).

[13] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE Transactions on Image Processing, **13**, 600-612 (2004).

[14] Z. Wang, A. C. Bovik and E. P. Simoncelli, Structural approaches to image quality assessment, in: A.C. Bovik ed., Handbook of Image and Video Processing, 2nd Edition (Academic Press, Orlando,) 961-974 (2005).

Zhiqing Li received the Ph.D. degree in computer software and theory from Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), Beijing, China, in 2010. He is currently an associate professor in the College of Information Engineering, Xiangtan University, Xiangtan, China. His research interests include image processing, artificial neural networks, machine learning and cognitive informatics.



Weizhong Zhao received the Ph.D. degree in computer software and theory from Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), Beijing, China, in 2010. He is currently a lecturer in the College of Information Engineering, Xiangtan University, Xiangtan, China. His research interests include Semi-Supervised Learning, pattern recognition and machine learning.

Zhixin Li received the Ph.D. degree in computer software and theory from Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), Beijing, China, in 2010. He is currently an associate professor in the College of Computer Science and Information Technology, Guangxi Normal University, Guilin, China. His research interests include image understanding, machine learning and probabilistic graphical model.