

# Imaging and representation learning of solar radio spectrums for classification

Zhuo Chen · Lin Ma · Long Xu · Chengming Tan · Yihua Yan

Received: 30 October 2014 / Revised: 28 January 2015 / Accepted: 22 February 2015  
© Springer Science+Business Media New York 2015

**Abstract** In this paper, the authors make the first attempt to employ the deep learning method for the representation learning of the solar radio spectrums. The original solar radio spectrums are pre-processed, including normalization, enhancement and etc., to generate new images for the next processing. With the expertise of solar radio astronomy for identifying solar radio activity, we build a solar radio activity database, which contains solar radio spectrums as well as their labels indicating the types of solar radio bursts. The employed deep learning network is firstly pre-trained based on the available massive of unlabeled radio solar images. Afterwards, the weights of the network are further fined-tuned based on the labeled data. Experimental results have demonstrated that the employed network can effectively classify the solar radio image into the labeled categories. Moreover, the pre-training process can help improve the classification accuracy.

**Keywords** Deep learning · Solar radio astronomy · Feature learning · Classification

## 1 Introduction

Solar radio astronomy is a burgeoning interdisciplinary of radio astronomy and solar physics which was born in 1940s. The discovery of radio waves from the Sun provided a new window to investigate the solar atmosphere. Thus new information about the sun could be obtained. For example, the properties of the solar corona were much more easily determined at radio wavelengths. Nowadays, solar radio telescopes have improved a lot, so that fine structures in solar radio bursts can be detected. In this study, we use data obtained by Solar Broadband Radio Spectrometer (SBRS) of China [8]. The SBRS is with characteristics of high time resolution, high-frequency resolution, high sensitivity, and wide frequency coverage in the microwave region is described. Its function is to monitor solar radio bursts in the frequency range of 0.7–7.6 GHz with time resolution of 1–10 ms. It consists of five ‘component

---

Z. Chen · L. Xu (✉) · C. Tan · Y. Yan  
Key Laboratory of Solar Activity, National Astronomical Observatories, Chinese Academy of Sciences,  
Beijing, China  
e-mail: lxu@nao.cas.cn

L. Ma  
Huawei Noah's Ark Lab, Hong Kong, Hong Kong

spectrometers' which work in five different wave bands (0.7–1.5, 1.0–2.0, 2.6–3.8, 4.5–7.5, and 5.2–7.6 GHz, respectively). The SBRS monitors the solar radio bursts all day long producing mass of data for researchers to analyze. In the observed data, burst events are rare and always with interference meanwhile. So it seems impossible to identify whether the data containing bursts or not and figure out which type of burst it is by manual operation timely. Thus, classifying the data observed automatically will be quite helpful for solar radio astronomical study.

Nowadays, based on the available mass of data of SBRS, many algorithms have been developed for learning the representation with unsupervised and supervised methods, specifically the deep learning methods. Recent methods based on deep learning [1] have demonstrated state-of-the-art performance in a wide variety of tasks, including visual recognition [12, 17], audio recognition [13, 14], and natural language processing [4]. These techniques are especially powerful because they are capable of learning useful features directly from both unlabeled and labeled data, avoiding the need for hand-engineering, which will be much helpful the automatic of the solar radio spectrum analysis. For analyzing a large volume of data, the simple and widely used method is principal components analysis (PCA), which finds the directions of greatest variance in the data set and represents each data point by its coordinates along each of these directions. However, PCA cannot well learn a good representation of the data for the targeting task, which will be shown in Section IV. 2. 2. Moreover, autoencoder (AE) can also be employed to learn the representation from the mass of available data. AE is an unsupervised learning algorithm that applies backpropagation, setting the target values to be equal to the inputs. The AE tries to learn a function to make the input to be similar with the output of the function. In other words, it is trying to learn an approximation to the identity function, so as to output of the network that is similar to the input. The identity function seems a particularly trivial function to be trying to learn. But by placing constraints on the network, such as by limiting the number of hidden units, interesting structure about the data can be learnt. Therefore, AE is very helpful for representation learning. Also there are many other variations of the AE, such as denoising AE [2], stacked AE (SAE) [3]. Furthermore, in [11], the authors proposed the automatic dimensionality reduction to facilitate the classification, visualization, communication, and storage of high-dimensional data. An adaptive, multilayer "encoder" network to transform the high-dimensional data into a low-dimensional code and a similar "decoder" network to recover the data from the code. With the random weights as the initialization in the two networks, they can be trained together by minimizing the discrepancy between the original data and its reconstruction. Then the representation can be learned in an unsupervised manner. The network can be further named as deep belief network (DBN). With the achievements of these learning methods, we can learn the representations of the solar radio spectrums, which will be employed for further solar radio image analysis, such as clustering, classification, and so on.

In this paper, we make the first attempt to employ the deep learning method, specifically the DBN, to learn the representation of the solar radio spectrum. Based on the representation, we can further classify the solar radio spectrums into different categories automatically. The main contributions of the paper are as following.

- The first attempt is made to employ the deep learning method to automatically learn the representation of solar radio spectrums.
- A solar radio spectrum database is built, which contains the solar radio spectrums as well as their labels.
- A group of image processing methods, including channel normalization, image enhancement, image scale, image normalization and so on, are raised for the following

representation learning and classification tasks.

By evaluating the learnt representation on the built solar radio spectrum database, the experimental results demonstrate that the deep learning method can help to automatically analyze the solar radio spectrum, specifically the classification.

The rest of the paper is organized as following. In [Section II](#), a series of pre-processing methods are proposed for the following representation learning and classification. In [Section III](#), the learning architecture is introduced to learn the representation of the solar radio image. [Section IV](#) gives the experimental results on representation learning and classification. And the final section concludes this paper.

## 2 Pre-processing of solar radio spectrums

The raw solar radio data captured by SBRS contains the flux values of radio radiation as well as the observation time. Although the captured data covers all the information of the solar radiation, it is hard for the viewers/researchers to judge or determine whether the solar burst happens or not and in what level the solar burst is. In order to further meet the requirements, the raw data captured by SBRS is firstly converted into images for easy visualization.

### 1. Solar radio spectrum

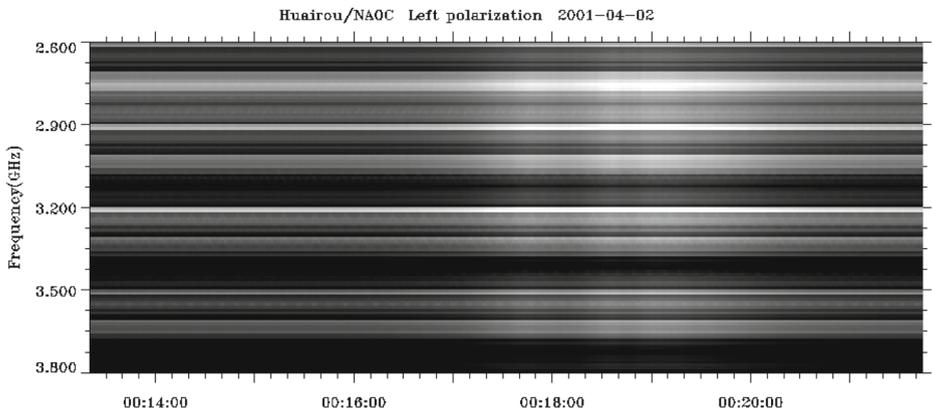
As mentioned before, SBRS contains several channels to monitor the solar burst in different frequencies. Therefore, the signal sensed from each channel will be treated individually. In total, there are 120 channels working toward the solar radio information captured at the same time. Moreover, each captured file contains both left and right circular polarization parts, which should be separated and processed individually for visualization and further processing. We extract the captured data from each channel as a row vector, which is stored according the sensing time. Afterwards, all the vectors from the 120 channels will be assembled together according the frequency values to form a solar radio spectrum, which is used for visualization and further processing. As there are 120 channels and 2520 sensing time points in 8 ms recorded file, the final resolution of the converted image is  $2520 \times 120$ . One sample image is illustrated in [Fig. 1a](#).

### 2. Channel normalization

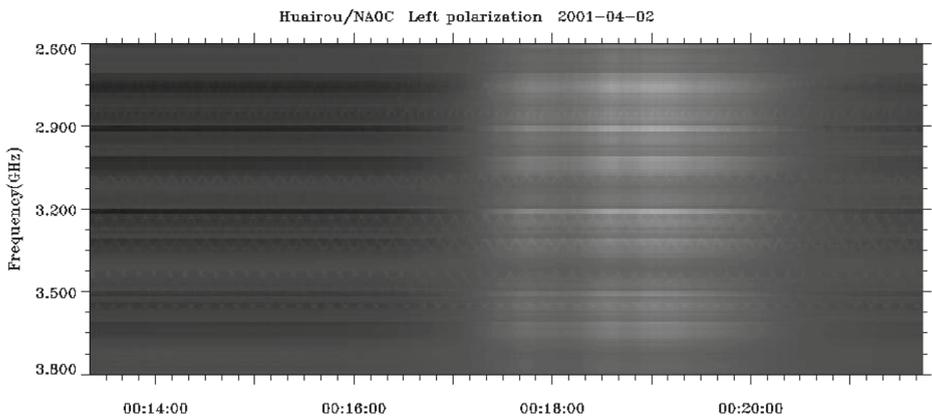
After the conversion process, it can be observed that there are numbers of horizontal-stripes-like interference signal in almost each picture, as illustrated in [Fig. 1a](#). This phenomenon is named as the channel effect in solar radio observation, which is caused by different gains of different channels. It can be observed in [Fig. 1a](#) that each channel produces the signal of the same magnitude. Therefore, clear horizontal lines can be easily detected from the captured solar radio spectrum. The channel effect may disturb the presentation of bursts. In order to eliminate the channel effect, we propose one method for channel normalization, which is formulated as following:

$$g = f - f_{LM} + f_{GM} \quad (1)$$

$f$  is the constructed image,  $g$  is the image after performing the channel normalization,  $f_{LM}$  and  $f_{GM}$  denote the local mean and global mean values, respectively. The local mean  $f_{LM}$  is the mean of the pixels in a frequency channel.  $f_{GM}$  accounts for the mean of a whole image.  $f_{LM}$  is to alleviate the effect of uneven channel gain resulting in horizontal-stripes-like interference, while the  $f_{GM}$  compensates each pixel value by adding a global



(a) Solar radio image before channel normalization



(b) Solar radio image after channel normalization

**Fig. 1** Solar radio image before and after channel normalization. The horizontal axis denotes the sampling time, while the vertical axis indicates the frequency channel. **a** Solar radio image before channel normalization. **b** Solar radio image after channel normalization

background. The solar radio image (for convenience, we call images processed from raw spectrums as solar radio images) after performing the channel normalization is illustrated in Fig. 1b. It can be observed that the channel normalization removes horizontal-stripes-like noise successfully and the solar radio burst can be easier detected.

### 3. Down-sampling

After performing the conversion and normalization, the image is generated with the resolution of  $2520 \times 120$ . From the solar radio image presented in Fig. 1, it can be observed that the spatial pixel values are highly correlated (neighboring pixels of the solar radio image present similar values), which means that the spatial information of the image contains high redundancy. For the generated image, each row is obtained from each channel containing captured flux value over the time. Therefore, it can be viewed as an observation of a discrete stochastic process. Therefore, each column represents a random variable. In probability theory and statistics, correlation of a stochastic process can be measured by the function, correlation coefficient [6, 15]  $\gamma$ , which is defined in the

following:

$$\gamma(n_1, n_2) = \frac{\varphi[(x(n_1) - \mu_1) \cdot (x(n_2) - \mu_2)]}{(\sigma_1 \cdot \sigma_2)} \tag{2}$$

where  $n_1$  and  $n_2$  are two positions in a certain row,  $x(n_1)$  and  $x(n_2)$  are the values of the pixels lying at the two positions with ensemble average,  $\mu_1$  and  $\mu_2$ , and the standard deviations,  $\sigma_1$  and  $\sigma_2$ , and  $\varphi$  is the ensemble average operator. Due to the variation of local activities of obtained image pixels,  $\gamma(n, n + \tau)$  depends not only on  $\tau$  but also on  $n$  in different positions. In other word, the generated image is not a collection of samples of a wide-sense stationary stochastic process. Consequently, for a given  $\tau$ ,  $\gamma(\tau)$  is a random variable instead of a constant. Moreover, we can treat each row to be a random variable. And the correlation coefficient can be further analyzed. We calculate  $\gamma(\tau)$  with different values of  $\tau$ . The results are illustrated in Table 1.

By referring to the correlation coefficient in Table 1, we can see that the pixel values in the horizontal axis present higher spatial correlations, while the pixel values in vertical axis present lower spatial correlation. Both of them do not vary a lot as  $\tau$  changes. Therefore, the generated image can be further down-sampled to reduce the resolution, which can further remove the redundancies of the image. Another reason is that the image with the resolution  $2520 \times 120$  results in 302,400 pixels. If we feed all the pixels into a neural network, the node will be dramatically large which we cannot afford for learning. Considering this situation, the generated solar radio spectrum is down-sampled into  $75 \times 30$  image with the nearest neighbor sampling method. It can be observed that image characteristic is not sacrificed comparing to the solar radio image in its native resolution, which is illustrated in Fig. 2.

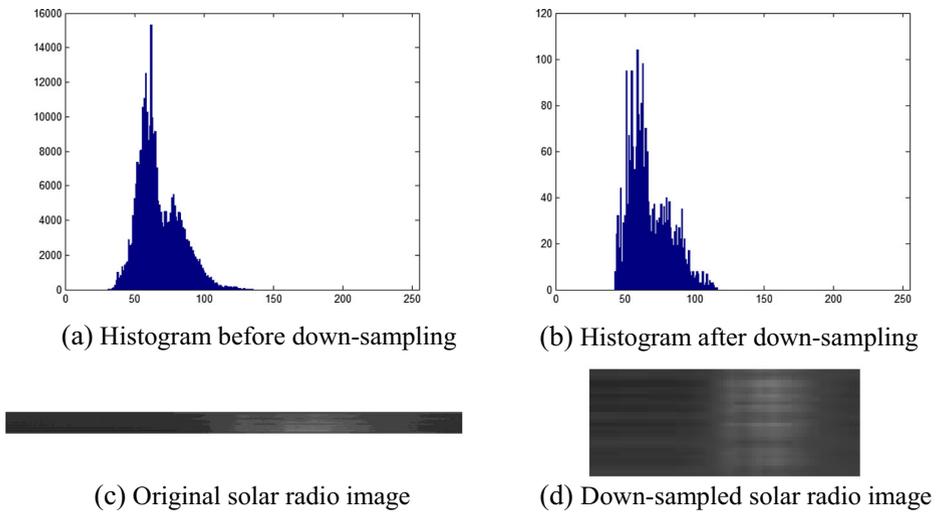
4. Image Normalization & Enhancement

Because of difference of calibration values, the processed images mentioned above are of various average gray levels. It may cause the neural network to learn feature relating to gray scale of images which is not what we want to obtain. To avoid it happens, we employ the image normalization which is formulated as Eq. (1). Different from the channel normalization, the local mean  $f_{LM}$  is calculated by averaging each individual image of a dataset and the global mean  $f_{GM}$  accounts the average value of the whole dataset.

After the image normalization, we propose another process before the data input to the network which is called enhancement. As a matter of experience, the most of effective information is with the value  $\pm 30$  around the mean of image  $\bar{f}$ . In rest of the range  $(0, \bar{f} - 30) \cup (\bar{f} + 30, 255)$ , the gibberishes like noise is in the majority. To make the data more representative, we apply a linear amplification to the pixels in range  $(\bar{f} - 30, \bar{f} + 30)$ .

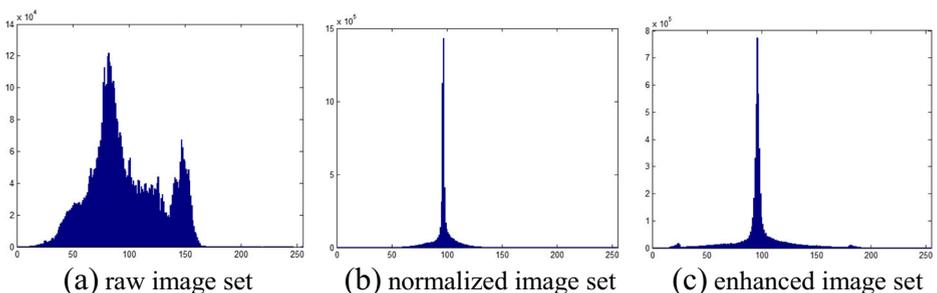
**Table 1** Correlation coefficients of solar radio images

a. $\gamma(\tau)$ in horizontal axis		b. $\gamma(\tau)$ in vertical axis	
$\tau=1$	0.9978	$\tau=1$	0.8343
$\tau=10$	0.9829	$\tau=2$	0.7738
$\tau=20$	0.9771	$\tau=3$	0.7660
$\tau=30$	0.9753	$\tau=4$	0.7096
$\tau=40$	0.9750	$\tau=5$	0.6650



**Fig. 2** The solar radio image and the histogram before and after down-sampling. **a** Histogram before down-sampling. **b** Histogram after down-sampling. **c** Original solar radio image. **d** Down-sampled solar radio image

Examining the solar radio images, the pixel value (the value of flux) should be mostly present a general Gaussian distribution. However, the base values of the calibration cannot be ensured to be the same, which make the average pixel value varies dramatically. The pixel value distributions of the training samples cannot be accurately modeled as the general Gaussian distribution. In this situation, the employed network, as introduced in the following, tends to capture the variations of the mean value changes, which will significant ignore the inherent property of the solar radio image. As discussed with the experts in solar activity, most of the useful information focuses on the area of the mean value. Therefore, in order to handle the aforementioned drawback, a bilinear enhancement is performed to enhance the solar radio image. Figure 3 shows the histograms of the solar radio image set during the processes of image normalization and enhancement. It demonstrates that the distribution of image sets will become more general Gaussian-like when employed image normalization. We can also find that enhanced images act more representative then before through the histograms.



**Fig. 3** Histograms of the solar radio image set. **a** raw image set. **b** normalized image set. **c** enhanced image set

### 3 Representation learning and classification for solar radio images

After pre-processing of solar radio spectrum, we start to learn solar radio image representation by employing DBN in this section. The architecture of DBN is illustrated in Fig. 4. DBN is a multilayer, stochastic generative model which is created by stacking multiple restricted Boltzmann machines (RBMs). Each RBM is trained by taking the hidden activities of the previous RBM as its input data. Each time a new RBM is added to the stack, the new DBN has a better variational lower bound on the log probability of the data than the previous DBN, provided the new RBM is learned in the appropriate way [10].

#### 1. RBM

RBM is a type of graphical model in which the nodes are divided into two sets, specifically, the visible and hidden. Each visible node is only connected to the hidden nodes. It means that there are no intra-visible or intra-hidden connections, which can be illustrated in each layer of Fig. 4. The energy function of an RBM with  $V$  visible units and  $H$  hidden units is defined in the following.

$$E(v, h) = -\sum_{i=1}^V \sum_{j=1}^H v_i h_j \omega_{ij} - \sum_{i=1}^V v_i b_i^v - \sum_{j=1}^H h_j b_j^h \tag{3}$$

where  $v$  is the binary state vector of the visible nodes,  $h$  is the binary state vector of the hidden nodes,  $v_i$  is the state of visible node  $i$ ,  $h_j$  is the state of the hidden node  $j$ ,  $\omega_{ij}$  is the real-valued weight between the visible node  $i$ , the hidden node  $j$ .  $b_i^v$  is the real-valued bias into visible node  $i$ , and  $b_j^h$  is the real-valued bias into hidden node  $j$ . The joint distribution of the visible and hidden nodes is defined in the following:

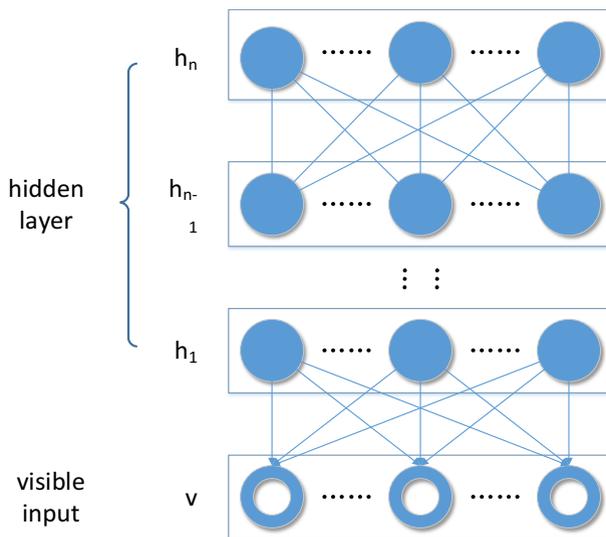


Fig. 4 DBN learning structure

$$p(v, h) = \frac{e^{-E(v, h)}}{\sum_u \sum_g e^{-E(u, g)}} \quad (4)$$

It can be observed that low energy results in high probability and high energy brings is assigned low probability. Also the probability of a visible node turning on is independent of the states of the other visible nodes, given the states of the hidden nodes. Likewise the hidden states are independent of each other given the visible states. The property of the RBM makes the sampling extremely efficient, as one can sample all the hidden nodes simultaneously and then all the visible nodes simultaneously. However, the energy function of the RBM defined in Eq. (3) can only take the binary values as the visible nodes, which is very inconvenient for modeling real-valued data, specifically the pixel values ranging from 0 to 255.

## 2. Gaussian-Bernoulli RBM (G-RBM)

In order to handle the real-valued visible inputs, the energy function in Eq. (3) is modeled as:

$$E(v, h) = -\sum_{i=1}^V \sum_{j=1}^H \frac{v_i h_j \omega_{ij}}{\sigma_i} - \sum_{i=1}^V \frac{(v_i - b_i^v)^2}{2\sigma_i^2} - \sum_{j=1}^H h_j b_j^h \quad (5)$$

where  $v_i$  takes the real-valued activity of the visible node  $v_i$ . Each visible node adds a quadratic offset to the energy function where  $\sigma_i$  controls the corresponding width. Comparing the binary visible and binary hidden node defined in Eq. (3), the G-RBM takes real-valued nodes as the input and output the binary nodes.

## 3. DBN

As mentioned before, each layer of the DBN is composed by an RBM, where the weights in layer  $l$  are trained by keeping all the weights in the lower layers constant and taking as data the activities of the hidden units at layer  $l+1$ . Therefore, the DBN training algorithm trains the layers greedily and sequentially. Layer  $l$  is trained after layer  $l-1$ . If the size of the second hidden layer is the same as the size of the first hidden layer and the weights of the second is borrowed from the weights of the first, it can be proven that training the second hidden layer while keeping the first hidden layer's weights constant improves the log likelihood of the data under the model [16]. Fig. 3 illustrates the multilayer DBN. The probability of the DBN assigns to a visible vector is defined as:

$$p(v) = \sum_{h_1, \dots, h_n} p(h_{n-1}, h_n) \prod_{k=2}^{n-1} p(h_{k-1} | h_k) p(v | h_1) \quad (6)$$

where  $n$  defines the num of hidden layers.

DBN is demonstrated to be very helpful for hand-written digital numbers recognition, as well as unsupervised training [11]. In this paper, we also employed such neural network to learn the representation and perform the classification of the solar radio images. The detailed information about our constructed network is introduced in the following.

## 4. Neural network for solar radio image classification

Based on the learning architecture in previous section, we propose a simple network for representation learning and classification of solar radio images. A classification layer with three output nodes is added on top of one RBM layer, which takes learned representation as input and outputs the classification results for each type of the solar radio image. For

each type, the classification layer will determine the possibility about how the inputs will result in the specific type.

The depth of the neural network depends on the problem and the size of the training set. Overfitting will occur with high probabilities if the training samples are insufficient, as the network requires a larger number of parameters. In this case, due to the limit number of solar radio images, only one hidden layer is employed. Then, we propose the *I-H-C* structure network for the experiment, as illustrated in Fig. 5. *C* stands for the classification, which is defined to give the prediction that which is the most possible type the input is. *I* indicating the number nodes of the input layer is set as 2250 which is the number of dimensions of preprocessed data. *H* stood for hidden is defined as 100 nodes of hidden layer. The bottom layer of the employed network is the RBM and the top layer is a softmax layer for classification. In order to realize the non-linear mapping function for the classification, the object of the learning network is defined as following:

$$\hat{o} = \arg \max_o p(o|x; \Theta) \tag{7}$$

Where  $\Theta$  include all the parameters in RBM and softmax layers. In order to make the inference, we need to obtain the parameters of the constructed network, specifically the parameters of RBM and softmax layer, respectively. For the parameters in the RBM layer, the standard contrastive divergence learning procedure is employed for pretraining. Detailed information about the pretraining method can found in [9].

With the process of pretraining, the constructed network can effectively avoid the risk of trapping in poor local optima. After the pre-training process, the fine-tuning process needs to be further performed to make the network more suitable for solar radio spectrum classification. Thereby, a log-likelihood function is employed as the object function for further training the parameters in the softmax layers and fine-tuning the parameters in the RBM layer:

$$\Theta^* = \arg \max_{\Theta} \sum_{l=1}^k \log P(\hat{L} = L|x; \Theta) \tag{8}$$

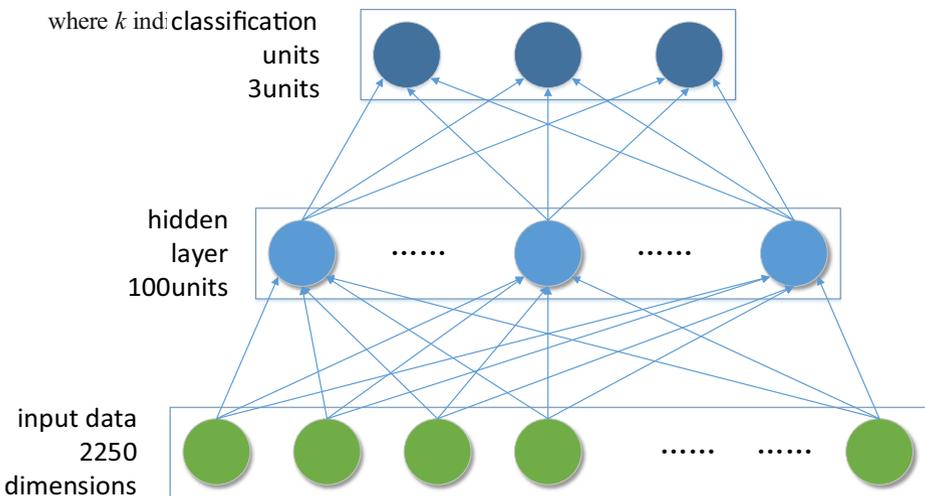


Fig. 5 Neural network for solar radio image classification

inputs, and  $\hat{L}$  represents the outputs of the network. For the parameter training, traditional back-propagation (BP) [5] is employed to fine-tune parameters of the constructed deep network. This algorithm is first proposed by Rumelhart and McClelland, the essence of which is to minimize the mean squared error between actual output and desired output based on gradient descent. BP algorithm is especially powerful because it can extract regular knowledge from input data and memory on the weights in the network automatically. Furthermore, in order to prevent over-fitting in training neural network, drop-out is introduced. Typically the outputs of neurons are set to zero with a probability of  $p$  in the training stage and multiplied with  $1 - p$  in the test stage. By randomly masking out the neurons, dropout is an efficient approximation of training many different networks with shared weights. In our experiments, we applied the dropout to all the layers and the probability is set as  $p=0.2$ .

## 4 Experimental results

To evaluate the proposed representation learning and classification of solar radio spectrums, a solar radio spectrum database is established firstly. Then, the representation learning and classification of solar radio spectrums are tested on this database.

### 1. Solar Radio Spectrum Database

The solar broadband radio spectrometer (SBRS) of China [8] is designed to acquire dynamic spectrograms of solar microwave bursts with the combination of wide frequency coverage (0.7–7.6 GHz), high temporal resolution, high spectral resolution, and high sensitivity. It consists of five “component spectrometers” which operate in five different wave bands (0.7–1.5, 1.0–2.0, 2.6–3.8, 4.5–7.5, and 5.2–7.6 GHz, respectively). The time resolution for sensing the solar microwaves varies for different wave bands. For example, the time interval for the wave band with frequency covering 2.6–3.8 GHz is 0.2 s. All the five ‘component spectrometers’ work simultaneously to make a full view of the solar microwave bursts from the perspective of sensing frequencies. Detailed information about SBRS can be referred to [8].

The statistics of solar radio data shows that there are only a small portion of solar radio bursts in all captured data. There are in total millions of microwaves captured by the end of 2001. However, there are only hundreds of them are labeled as burst as shown in Table. 2. It can be observed that the burst microwaves captured in the 2.6–3.8 GHz frequency range are more easily detected by the human viewers. It means that the captured microwaves in the frequency range are more representative to indicate whether the spectrums contain bursts or not. Therefore, the most representative solar microwaves in the frequency range are employed to build the dataset for our experimental results.

In this dataset, 4408 observational data files are labeled by the experts into six categories (0=no burst or hard to identify, 1= weak burst, 2=moderate burst, 3= large

**Table 2** The number of bursts observed with each component spectrometer of SBRS by the end of 2001

Freq. range (GHz)	0.5–1.5	1.0–2.0	2.6–3.8	4.5–7.5	5.2–7.6
Num. of bursts	108	526	921	233	550

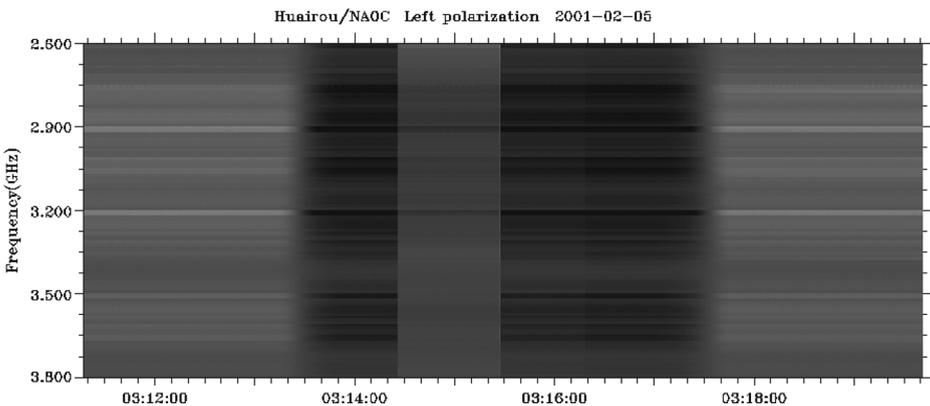
burst, 4=data with interference, 5=calibration). Since the objective of our experiment is to distinguish the bursts from others, the solar radio image in the dataset has been selected and relabeled to form a new database for the experiment. Three coarse categories, specifically the ‘bursts’, ‘non-burst’, and ‘calibrations’ are included in the database. The files of the ‘burst’ category contain at least one solar radio burst and the ‘non-burst’ stands for files not containing an identifiable burst (the spectrum shown in Fig. 1 is a typical one of ‘burst’). The ‘calibration’ type means files with calibration signal which is used to make sure the value obtained by the solar radio telescopes is effective. As shown in Fig. 6, in calibration images, the variation of the gray level values of the image pixel is non-continuous along the horizontal direction.

As introduced before, by performing the imaging process, each observational data file can be converted to two images with the size  $2520 \times 120$  pixel arrays with 8-bit gray scale. The  $Y$ -axis denotes the frequency for capturing the microwave, while the  $X$ -axis indicates the sensing time of the microwave. Furthermore, the channel normalization, down-sampling, enhancement processes are performed, which generate an  $75 \times 30$  image. With the expertise of solar radio activity, the solar radio images are labeled as ‘bursts’, ‘non-burst’, and ‘calibrations’. The detailed information about the labeled data in the built database is illustrated in Table 3.

## 2. Performance Comparisons

As introduced, there are 8816 total labeled solar radio images in the built database. However, there are massive of unlabeled solar radio images. In this part, not only the labeled data are employed to learn the representation for classification, but also the unlabeled data are employed to help representation learning. From the experimental results in the following, it can be observed that the unlabeled data with proper pretraining can help build the network for representation learning. As illustrated in Table 3, there are 1158 ‘burst’, 6670 ‘non-burst’, 988 ‘calibration’ solar radio image through the observed data files. We randomly select 900 ‘burst’, 800 ‘non-burst’, 800 ‘calibration’ for training. The rest labeled data is employed for testing. Also the other unlabeled data is employed for pretraining. The number of the unlabeled solar radio images is 2500.

The experiments are performed in the following. Firstly, the classification effect will be shown after training the network. Then, a PCA+SVM experiment will be performed to prove that PCA is not suitable to learn a good representation of solar radio images. Afterwards, the network will be trained without pre-training process in order to examine



**Fig. 6** A typical solar radio image of ‘calibration’

**Table 3** The details of the database. 0=no burst or hard to identify, 1= weak burst, 2=moderate burst, 3= large burst, 4=data with interference, 5=calibration

Categories	0	1	2	3	4	5	total
Image Number	6670	618	268	272	570	988	8816

whether the pre-training is helpful for final classification. Finally, different numbers of nodes in the hidden layer will be applied to prove network structure suitable.

#### 4.1 The performance of the network

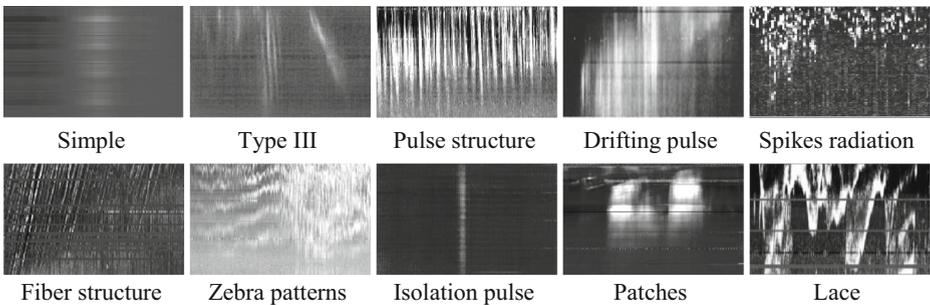
After preprocessing, we input the training set data to the network as batches. The hidden layer is first pretrained to initialize the parameters in an unsupervised way. Then both the hidden layer and the classification layer are fine-tuned with labeled data. After that, preprocessed testing set data will be inputted sequentially and the network will output the classification results in possibilities that how likely the input data belongs to each category respectively. The model classify a solar radio image successfully when the category with highest possibility output by the algorithm matches the labeled category of the file input. The classification results can be found in Table 4.

It is worth mentioning that the receiver operating characteristic (ROC) [7] analysis is used to display experimental results, which utilizes the true positive rate (TPR) and false positive rate (FPR) to judge the performance of a classification model. TPR is defined as the number of correct classification among all positive samples available during the test. The larger the TPR value, the better the performance. FPR, on the other hand, measures the number of negative samples which are wrong classified into positive category during the test. Consequently, the smaller the FPR value, the better the performance.

From Table 4, it is easy to notice that the TPR of category ‘burst’ is much lower than the other two. The ‘burst’ data contains numbers of complex subcategories (Fig. 7). Each one differs from others with unique textures, sizes, shapes and other characteristics. Moreover, some subcategories are hard to define clearly in solar radio astronomy. Furthermore, the ‘burst’ type itself is consisted of three sets with different intensity level. So, it is much more difficult to extract features of ‘burst’ images than others, which make the classification performance of burst solar radio images to be lower than the other two types. Another reason is that the number of burst solar radio images is very small. With such small number of images and such complex patterns in the burst images, the neural network cannot effectively learn the inherent property of the burst images. Therefore, in the future, we may include more samples with different types

**Table 4** Performance of DBN and PCA+SVM

	DBN		PCA+SVM model	
	TPR	FPR	TPR	FPR
Burst	67.4 %	13.2 %	52.7 %	26.6 %
Non-burst	86.4 %	14.1 %	0.1 %	16.6 %
Calibration	95.7 %	0.4 %	38.3 %	72.2 %



**Fig. 7** Different burst types of solar radio spectrum

of burst images. The neural network is expected to effectively learn the presentations, which will result in a better performance.

Additionally, we applied a PCA+SVM model to classify the solar radio images. Principal components analysis (PCA) is a classical one of traditional algorithms to reduce dimensionality of data. We employed the PCA algorithm to learn features of solar radio images and classified them with SVMs. This experiment had the same experimental settings with the prior one (with the same training and testing samples). And the dimension of PCA is the same as the length of feature vector generated from our employed DBN. Table 4 shows the comparison of DBN and the PCA+SVM model.

From the table, we can notice that classification result of PCA+SVM model performs poorly. The performance at ‘non-burst’ and ‘calibration’ categories is worse than random. Especially, the TPR at ‘non-burst’ is almost zero. This might because the PCA learned a global representation of SRIs by minimize the reconstruction error, which cannot effective preserve the information for SRI classification. On the contrary, our employed DBN learns a specific representation from both unsupervised and supervised training, which automatically learn the features for the SRI classification task. Therefore, the employed DBN perform better than the PCA+SVM model which is a traditional approach.

#### 4.2 Training the network without pre-training

Table 5 shows the comparison of algorithms with and without pretraining. It demonstrated that, despite of slight decreases in classification success rate with the category ‘non-burst’, the overall performance is improved. The fine-tuning process employing BP algorithm is based on local gradient descent, and usually starts with random initial points, which may incur poor local optima. If pretraining is added to initialize the parameters, the network could be fine-tuned on parameters that more likely to approach the global optima. The comparison result proves it and demonstrates the necessity of pretraining.

**Table 5** Comparison of the algorithms with and without pretraining

	Fine-tuning		Pretraining + Fine-tuning	
	TPR	FPR	TPR	FPR
burst	53.1 %	6.8 %	67.4 %	13.2 %
non-burst	93.1 %	25.1 %	86.4 %	14.1 %
calibration	95.2 %	0.3 %	95.7 %	0.4 %

**Table 6** Comparison of the algorithms with different numbers of units in hidden layer

	50 units		100 units		125 units	
	TPR	FPR	TPR	FPR	TPR	FPR
burst	58.5 %	11.8 %	67.4 %	13.2 %	56.6 %	11.2 %
non-burst	87.8 %	17.0 %	86.4 %	14.1 %	88.5 %	19.0 %
calibration	94.7 %	0.7 %	95.7 %	0.4 %	92.0 %	0.6 %

In Table 5, it is interesting that the TPR of ‘non-burst’ category presents an opposite trend relative to the other two. It may be attributed to the different solar radio features presented by these three different categories. A ‘non-burst’ image is with kind of calmness that lack of fluctuations or variations on the gray background with constant level. Unlike the ‘non-burst’ images, the ‘calibration’ images have dramatic fluctuations of the gray scale and saltus steps covering the whole column of the image. The ‘burst’ images seem to be more complex. The fluctuations on it have diversiform shapes, various sizes and different intensity. A ‘weak-burst’ one may be like a ‘non-burst’ data that it hard for an inexperienced person to identify. On the contrary, a ‘large-burst’ image may looks like a ‘calibration’ one. So the representation learned by the network is more suitable for identify the category ‘burst’. It can also classify some ‘non-burst’ or ‘calibration’ images as ‘burst’ type by mistake. In other words, it may sacrifice the TPR of ‘non-burst’ or ‘calibration’ type to improve which of the category ‘burst’ – and vice versa.

#### 4.3 Hidden layer node numbers

The number of the nodes in hidden layer is the hyper parameter of the network, which affects the representation of the data. So it is very important for the algorithm’s performance. In order to figure out the best parameter, we do a simple test. Table 6 reveals the experimental results. From the table, 100 units tend to give the best performance. Fewer units cannot guarantee the quality of representation with the 2250 input nodes. On the other hand, if the number of units in the hidden layer increases, there will be much more parameters employed to map the inputs from visible layer to hidden layer. However, in our experiment, labeled data is not enough to afford a network with such parameters. Hence, in this paper, we simply selected 100 nodes in hidden layer for the neural network. In future, with more labeled data, we can increase the number of the hidden layer nodes to learn a better representations of the solar radio images.

## 5 Conclusion

The paper makes the first attempt for the imaging and representation learning of the solar radio spectrums. The solar radio spectrums are pre-processed to generate images for classification. A solar activity database which consists of the solar radio image and its label is built. Based on the massive of unlabeled and limit number of the labeled solar radio images, we employed network for the solar radio classification. Experimental results demonstrate that the massive of unlabeled data is helpful to initialize the network weight and improve the performance of the classification.

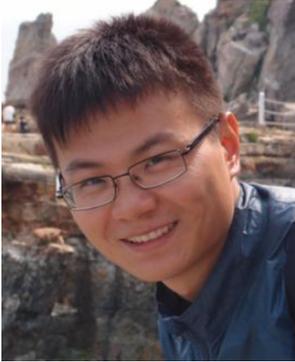
**Acknowledgement** This work was partially supported by the National Natural Science Foundation of China under Grant 61202242, 100-Talents Program of Chinese Academy of Sciences (No. Y434061V01).

## References

1. Bengio Y (2009) Learning deep architectures for AI. *FTML* 2(1):1–127
2. Chen M, Weinberger K, Sha F, Bengio Y (2014) Marginalized denoising autoencoders for nonlinear representation. *ICML*
3. Chen M, Xu Z, Weinberger K, Sha F (2012) Marginalized stacked denoising autoencoders for domain adaptation 29th International Conference on Machine Learning (ICML)
4. Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P (2011) Natural language processing (almost) from scratch. *JMLR* 12:2493–2537
5. Deng L (2012) Three classes of deep learning architectures and their applications: a tutorial survey, *APSIPA transactions on signal and information processing*
6. Dong J, Ngan KN, Fong CK et al (2009) 2-D order-16 integer transforms for HD video coding. *IEEE Trans Circuits Syst Video Technol* 19(10):1462–1474
7. Fawcett T (2006) An introduction to ROC analysis. *Pattern Recogn Lett* 27(8):861–874
8. Fu Q, Ji H, Qin Z et al (2004) A new solar broadband radio spectrometer (SBRs) in China. *Solar Phys* 222(1):167–173
9. Hinton GE (2010) A practical guide to training restricted boltzmann machines. Technical report, University of Toronto
10. Hinton GE, Osindero S, Teh Y (2006) A fast learning algorithm for deep belief nets. *Neural Comput* 18: 1527–1554
11. Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. *Science* 313(5786):504–507
12. Le Q, Ranzato M, Monga R, Devin M, Chen K, Corrado G, Dean J, Ng A (2012) Building high-level features using large scale unsupervised learning. *ICML*
13. Lee H, Largman Y, Pham P, Ng AY (2009) Unsupervised feature learning for audio classification using convolutional deep belief networks. *NIPS*
14. Mohamed AR, Dahl G, Hinton GE (2012) Acoustic modeling using deep belief networks. *IEEE Trans Audio Speech Lang Process* 20(1):14–22
15. Papoulis A, Pillai SU (2002) Probability, random variables, and stochastic processes. McGraw-Hill, Boston
16. Salakhutdinov R, Murray I (2008) On the quantitative analysis of deep belief networks. *ICML*
17. Sohn K, Jung DY, Lee H, Hero A (2011) Efficient learning of sparse, distributed, convolutional feature representations for object recognition. *ICCV*



**Zhuo Chen** received his B.S. degree from the School of Electronic and Information Engineering, Beijing Jiaotong University, China in 2014. He is currently a research assistant at the National Astronomical Observatories, Chinese Academy of Sciences. His research interests include 3D image quality assessment and deep learning.



**Lin Ma** received the B. E., and M. E. degrees from Harbin Institute of Technology, Harbin, China, in 2006 and 2008, respectively, both in computer science. He received his Ph.D. degree in Department of Electronic Engineering at the Chinese University of Hong Kong (CUHK) in 2013. He is now a Researcher with Huawei Noah's Ark Lab, Hong Kong.

He was a Research Intern in Microsoft Research Asia from Oct. 2007 to Mar. 2008. He was a Research Assistant with the Department of Electronic Engineering, CUHK, from Nov. 2008 to Jul. 2009. He was a Visiting Student with the School of Computer Engineering, Nanyang Technological University (NTU), from Jul. 2011 to Sep. 2011. His current research interests include multimodal matching, deep learning, image/video processing, and quality assessment.

He got the best paper award in Pacific-Rim Conference on Multimedia (PCM) 2008. He was awarded the Microsoft Research Asia fellowship in 2011. He was a finalist to HKIS young scientist award in engineering science in 2012.



**Long Xu** received the M.S. degree in applied mathematics from Xidian University, Shanxi, China, in 2002, and the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China.

He was a Senior Research Associate with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong, from August 2009 to July 2011. He was a Post-Doctoral Fellow with the Department of Electronic Engineering, Chinese University of Hong Kong, Shatin, Hong Kong, from July 2011 to July 2012. He is now with National Astronomical Observatories, Chinese Academy of Sciences. His current research interests include image/video coding, wavelet-based image/video coding, and computer vision.



**Chengming Tan** is now with Key Laboratory of Solar Activity, National Astronomical Observatories of Chinese Academy of Sciences. He was born in 1977, Doctor of natural science, Associated Professor, member of Chinese Astronomical Society. His major research is on solar radio astrophysics.



**Yihua Yan** is a professor and chief scientist of the solar radio research, and the director of the CAS Key Laboratory of Solar Activity and the Solar Physics Division, the CAS National Astronomical Observatories. Since 2006 Prof. Yan has served as the Chairman of the Sun & Heliospheric Physics Committee of the Chinese Astronomical Society and the Chairman of the Radio Astronomy Committee of the Chinese Institute of Electronics. He obtained his bachelor's degree in 1982, master's degree in 1985 and Ph.D in 1990. Research interests include solar magnetic fields, solar radio astronomy, space solar physics and radio astronomical methods.