2015 IEEE International Conference on Systems, Man, and Cybernetics

Multimodal Learning for Classification of Solar Radio Spectrum

Zhuo Chen
Key Laboratory of Solar Activity,
National Astronomical Observatories,
Chinese Academy of Sciences, Beijing, China
chenzhuo@nao.cas.cn

Lin Ma
Huawei Noah's Ark lab,
Hong Kong
forest.linma@gmail.com

Long Xu
Key Laboratory of Solar Activity,
National Astronomical Observatories,
Chinese Academy of Sciences, Beijing, China
lxu@nao.cas.cn

Ying Weng
School of Computer Science,
Bangor University, Bangor, UK
y.weng@bangor.ac.uk

Yihua Yan
Key Laboratory of Solar Activity,
National Astronomical Observatories,
Chinese Academy of Sciences, Beijing, China
yyh@nao.cas.cn

Abstract—This paper proposes the first attempt to utilize multimodal learning method for the representation learning of the solar radio spectrums. The solar radio signals sensed from different frequency channels, which present different characteristics, are regarded as different modalities. We employ a multimodal neural network to learn the representations of the solar radio spectrum, which can distinguish the differences and learn the interactions between different modalities. The original solar radio spectrums are firstly pre-processed, including normalization, denoising, channel competition and etc., before being fed into the multimodal learning network. Experimental results have demonstrated that the proposed multimodal learning network can learn the representation of the solar radio spectrum more effectively, and improve the classification accuracy.

Keywords-Multimodal learning; solar radio astronomy, solar radio spectrum; classification

I. INTRODUCTION

Solar radio astronomy is an emerging interdisciplinary field of radio astronomy and solar physics. The discovery of radio waves from the Sun provides a new window to investigate the solar atmosphere, and then new information about the Sun can be obtained. For example, the properties of the solar corona are much more easily determined at radio wavelengths. As solar radio telescopes have improved a lot in recent years, fine structures in solar radio bursts can be detected. In this study, we use data obtained by Solar Broadband Radio Spectrometer (SBRS) of China [1].The SBRS is with characteristics of high time resolution, high-frequency resolution, high sensitivity, and wide frequency coverage in the microwave region. Its functionality 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 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 and produces mass data for researchers to analyze. In the observed data, burst events are rare and in the meantime always along with interference. Thus 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. To the end, classifying the observed data automatically will be highly beneficial to the solar radio astronomy 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, especially the deep learning methods. Recent methods based on deep learning [2] have demonstrated state-of-the-art performance in a wide variety of tasks, including visual recognition [3] [4] [17], audio recognition [5] [6], and natural language processing [7] [15]. These techniques are super powerful because they are capable of learning useful features directly from both unlabeled and labeled data to avoid the need of hand-engineering, which will be much helpful to the automatic analysis of the solar radio spectrum. While 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 well. Moreover, autoencoder (AE) can also be employed to learn the representation from the available mass data. AE is an unsupervised learning algorithm that applies back-propagation by setting the target values to be equal to the inputs. AE tries to learn a function which makes the input similar to 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 AE, such as denoising AE [8], stacked AE (SAE) [9]. However, these AEs treat the input equally, which cannot distinguish the characteristics between different input modalities well. Thus the interaction between different modality inputs cannot be well captured. In [10], the authors proposed

978-1-4799-8697-2/15 $31.00 © 2015 IEEE
DOI 10.1109/SMC.2015.187

an automatic dimensionality reduction to facilitate the classification, visualization, communication, and storage of high-dimensional data through 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. Using 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 spectrum analysis, such as clustering, classification, and so on.

In this paper, we make the first attempt to employ the multimodal learning method, specifically the AE with the structured regularization (SR), 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 multimodal learning method to automatically learn the representation of solar radio spectrums.
- A group of pre-processing methods, including channel normalization, denoising, scaling, channel competition, and so on, are raised for the representation learning and classification tasks of solar radio spectrum.

By evaluating the learnt representation on the built solar radio spectrum database, the experimental results demonstrate that the multimodal learning method can help to automatically analyze the solar radio spectrum, specifically the classification. In addition, it is superior to the single modality of deep learning, which was presented in our previous work [11].

The rest of the paper is organized as following. In Section II, the multimodal learning architecture is introduced to learn the representation of the solar radio spectrum as well as the solar radio spectrum pre-processing to be fed into the multimodal neural network. Section III gives the experimental results on representation learning and classification. And the final section concludes the paper.

II. Learning for Classification of Solar Radio Spectrum

A. Multimodal Learning Architecture

The proposed learning architecture is illustrated in Figure 1, which takes different numbers and types of modalities as the input. The output will be the recognition results, which not only considers each modality property but also accounts for the interactions between different modalities. The proposed multimodal learning architecture is built by stacking the softmax layer on top of the layer of AE with structured regularization.

Eq. 1 represents the global function of the proposed method. $\hat{L}$ is the label output by the multimodal network. $f_{SR}(\cdot)$ is the function calculating the weights mapping from the visual layer

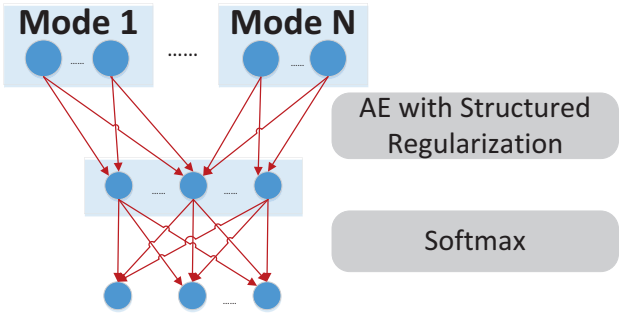


Figure 1. The framework of the multimodal learning.

to the first hidden layer. The algorithm that implements $f_{SR}(\cdot)$ is AE with SR. AE is a simple learning circuit aims to transform inputs into outputs with the least possible amount of distortion. However, the correlations between different modalities cannot be well learned and represented by AE. To overcome this limitation, we employ SR on AE, which allows the network to distinguish different modalities for individual treatments. SR employed on AE could help to distinguish and learn the representation from different multimodal inputs. With SR on AE, we get the joint representations of the input signals, which is the input of the function $\varphi(\cdot)$. In our framework, softmax is employed as the classification function $\varphi(\cdot)$ for our multimodal learning architecture.

$$\hat{L} = \varphi\left(f_{SR}(x_1, x_2, \cdots, x_m)\right) \tag{1}$$

1) Autoencoder (AE)

AE consists of two components, specifically the encoder and decoder. The encoder $e(\cdot)$ encodes the input $x \in R^d$ to some hidden representation $e(x) \in R^{d_h}$, while a decoder $d(h)$ decodes the obtained hidden representation back to a reconstructed version of x, to make the reconstructed signal to be as close as possible to the input. Therefore, the encoder process can be viewed as a single mapping function f:

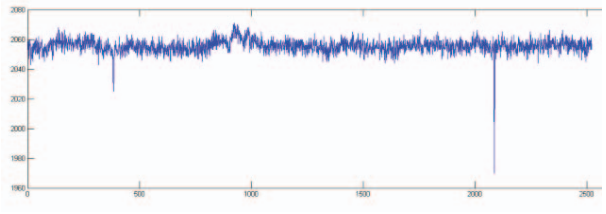
$$y_i = f(x_i) = \sigma(Wx_i + b) \tag{2}$$

where y_i represents the encoder output and x_i is the input of the encoder. W and b are the mapping weight and encoder bias, respectively. σ is a non-linear function, which can employ sigmoid, tanh, and rectified linear unit (ReLU) [16] function.

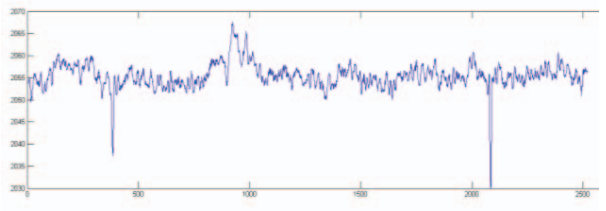
2) Structured Regularization (SR)

SR function is employed for AE with multimodal inputs inspired by [12] [13]. Suppose $S_{r,i}$ as a $K \times N$ modality binary matrix, where K denotes the number of modalities and N indicates the number of units in corresponding modality. For SR, each modality will be used as a regularization group separately for each hidden unit, applied in a manner similar to the group regularization, compared with the traditional regularization that treats each input unit equally and ignores the relationship and correlation between different modalities. SR is defined as:

$$SR\left(W^{[1]}\right) = \sum_{j=1}^{M}\sum_{k=1}^{N} f_B\left(\max_i\left(S_{k,i}\left|W_{i,j}^{[1]}\right|\right) > 0\right) \tag{3}$$



(a) Flux values in a channel before denoising



(b) Flux values in a channel after denoising

Figure 2. Flux values in a channel before and after denoising,

where f_B indicates a Boolean function that takes a value of 1 if its variable is true, and 0 otherwise. The regularization function in Eq. (3) performs a direct penalty on the number of modalities used for each weight without further constraining the weights of modalities with nonzero maxima.

3) Integrating SR with AE

By integrating SR with AE, the objective function for training the framework as presented in Figure 1 is represented as:

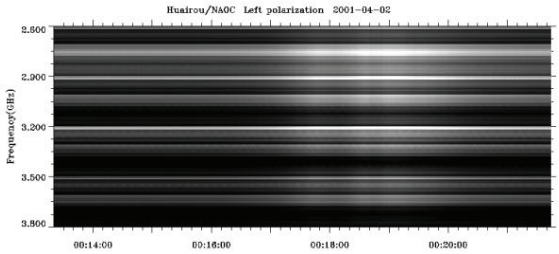
$$W^{[1]} = \underset{W^{[1]}}{\operatorname{argmin}} \sum_{i=1}^{n^{[1]}} \left\| z_i^{[1]} - x_i^{[1]} \right\|_2^2 + \alpha \cdot SR(W^{[1]})$$
$$z_i^{[1]} = \sum_j^{k^{[1]}} \mu_j^{[1]} W_{i,j}^{[1]} \quad (4)$$

where $z_i^{[1]}$ is the signal reconstructed by the decoder of AE. $n^{[1]}$ is the number of the input nodes including all the modality features, and $k^{[1]}$ is the number of the hidden nodes of the multimodal AE. $W_{i,j}^{[1]}$ is the weights of the multimodal AE by introducing SR. α is the parameter to balance the error and the regularization terms.

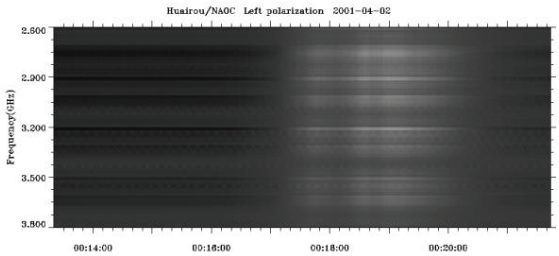
By integrating SR into AE, the obtained representation y_i only connects to partial nodes in the first hidden layer. As Eq. (3) shows, to minimize $SR(W^{[1]})$, the value of $W_{i,j}^{[1]}$ should be 0 as far as possible, leading to some nodes in the first hidden layer connected to only part of the nodes in the visual layer. AE with SR demonstrates that the multimodal network could distinguish different modalities and learn the correlations between them automatically.

B. Pre-processing of Solar Radio Spectrums

Different from previous research [11], we assumed that variations of each channels in a solar radio spectrum can represent more detailed characteristics of solar activities and are more significantly relative to an entire spectrum, which are regarded as different modalities for the representation learning. In this paper, we processed the solar radio spectrums in channel level to form the input vectors.



(a) Solar radio spectrum before channel normalization



(b) Solar radio spectrum after channel normalization

Figure 3. Solar radio spectrums before and after channel normalization. The horizontal axis denotes the sampling time, while the vertical axis indicates the frequency channel.

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 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 to form a matrix, which can be used for visualization and further processing. As there are 120 channels and 2520 sensing time points in 8 ms recorded file, the size of matrixes is 120×2520.

2) Channel Denoising and Normalization

The raw data captured by solar radio antenna is of strong white noise, which may faint the valuable information, as illustrated in Figure 2 (a). To suppress the noise and make the signals more expressive, Gaussian filter is employed for denoising. Each channel is processed by a 1×8 Gaussian kernel with $\sigma = 5$ respectively, the effective signal reveals more obviously as shown in Figure 2 (b).

It can also be observed that there are numbers of horizontal-stripes-like interference signals in almost each picture, as illustrated in Figure 3 (a). This phenomenon is named as the chan-

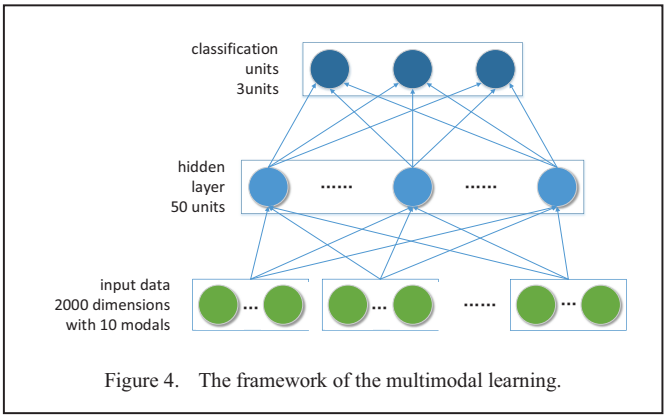


Figure 4. The framework of the multimodal learning.

nel effect in solar radio observation, which is caused by different gains of different channels. 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:

$$g = f - f_{LM} + f_{GM} \tag{5}$$

where f is the matrix of a solar radio spectrum, g is the matrix 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 each channel. f_{GM} accounts for the mean of whole matrix. f_{LM} is to alleviate the effect of uneven channel gain, while the f_{GM} compensates each channel by adding a global background. The matrix after performing the channel normalization is illustrated in Figure 3 (b). It can be observed that the channel normalization makes the solar radio burst easier to be detected.

3) Down-sampling with Channel Competition

The radio emission will be greatly enhanced when a solar burst occurs, like a solar flare or CME, which results from a local release of energy in the Sun's low corona. Such process would produce numerous radio spectral structures observed with radio spectrometer. As solar radio bursts occur, the flux values of solar radio spectrum will increase in certain channels. Thus, we defined an activity term to discriminate the difference between channels as:

$$D_i = max(C_i) - mean(C_i) \tag{6}$$

where D_i represents the maximum flux value of channel i minus the mean of it. D_i is assumed to reflect the activities of channels in spectrum, which means we can exploit the inner structural information within a spectrum instead of that at image level.

In this paper, we select 10 channels from 120 channels by ranking D_i, and then downsample each channel with bicubic filter. The original spectrum matrix with the size of 120×2520 is reduced to 10×200. Compared to the previous down-sampling method in [11], the proposed method selects more representative channels and maintains more details in each channel. Furthermore, the dimension of input data to deep learning network is smaller than previous one.

C. Network for Solar Radio Spectrum Classification

We propose a simple network for classification of solar radio spectrum. A softmax layer is added on the top of the multimodal learning architecture, which takes the joint representations as inputs, and outputs the classification results for each spectrum. The classification layer will determine the probabilities of an input spectrum belonging to each given class. The AE with SR is employed to establish the multimodal learning network. Due to limited number of labeled solar radio spectrums, only one hidden layer is employed in this work for avoiding overfitting. Then, we propose an $I - H - C$ structure network as shown in Figure 4. C is defined as the classification nodes which give the probabilities of each input belonging to the given classes (spectrum types). I indicates the size of the data from all the multimodal inputs, which is 2000 in this work. H denotes the size of the hidden layer nodes, which is 50 in this work.

III. Experimental Results

In order to evaluate the proposed method, a solar radio spectrum database we recently built is employed. Firstly, the details of the database are introduced. Then, the multimodal learning for solar radio spectrum classification is tested on the database.

A. Solar Radio Spectrum Database

The SBRS of China [1] 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. The time resolution of 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 0.2s. 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 [1].

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 I. 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 five categories (0=no burst or hard to identify, 1= weak burst, 2=moderate burst, 3= large burst, 4=calibration). Since the objective of our experiment is to distinguish the bursts from others, the solar radio spectrums in the dataset have 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 contains at least one solar radio burst and the 'non-burst' stands for files not containing an identifiable 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. In calibration spectrums, the flux values vary non-continuously within each channel.

TABLE I. 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

As introduced before, by performing the imaging process, each observational data file can be converted to two spectrums with the size of 120×2520. Each row denotes the frequency for capturing the microwave, while the column indicates the sensing time of the microwave. Furthermore, the channel normalization, down-sampling, denoising processes are performed, which generate a 10×200 matrix. With the expertise of solar radio activity, the solar radio spectrums are labeled as 'bursts', 'non-burst', and 'calibrations'. The detailed information about the labeled data in the built database is illustrated in Table II.

TABLE II. THE DETAILS OF THE DATABASE

Categories	*0*	*1*	*2*	*3*	*4*	*total*
Spectrum Number	6670	618	268	272	988	8816

(0=no burst or hard to identify, 1= weak burst, 2=moderate burst, 3= large burst, 4=calibration)

B. Performance Comparisons

The experimental settings are as follows. 900 'burst', 800 'non-burst', 800 'calibration' are randomly selected for training from the dataset. The rest are for testing. After preprocessing, each spectrum is converted into a vector with 10 modals (keep 10 channels). After that, these vectors are imported into the proposed multimodal network. The model classifies a solar radio spectrum successfully when the category with highest possibility output by the algorithm matches the labeled category of the spectrum data input.

In order to efficiently assess the performance of the proposed algorithm, we compare it with previous DBN method and a PCA+SVM model for classification of solar radio spectrum [11].

It is worth noting that the receiver operating characteristic (ROC) [14] analysis is used to evaluate the performance, which is more general and reliable than recognition accuracy. The classification results can be found in Table III. From the table, we can notice that the proposed multimodal network is better than DBN [11] with respect to the classification of 'burst'. The gain may be that the multimodal learning exploits the differences between different modalities and learn their interactions for final classification. It can be also observed that there is only insignificant loss on 'non-burst'. In the future, we will consider fusing the merits of DBN and multimodal learning to develop better representation learning networks for solar spectrum classification.

TABLE III. PERFORMANCE OF MULTIMODAL, DBN AND PCA+SVM

	Multimodal		*DBN*		*PCA+SVM model*	
	TPR	*FPR*	*TPR*	*FPR*	*TPR*	*FPR*
Burst	70.9%	15.6%	67.4%	13.2%	52.7%	26.6%
Non-burst	80.9%	13.9%	86.4%	14.1%	0.1%	16.6%
Calibration	96.8%	3.2%	95.7%	0.4%	38.3%	72.2%

IV. CONCOLUTIONS

The paper makes the first attempt to employ multimodal learning network for classification of the solar radio spectrums. The solar radio spectrums are pre-processed to generate different modalities to be fed into the multimodal learning network. The proposed multimodal learning architecture is demonstrated to better exploit the structural information within different modalities, and therefore a better performance of the proposed multimodal network is achieved.

ACKNOWLEDGMENT

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

REFERENCES

[1] Fu Q, Ji H, Qin Z, et al, "A new solar broadband radio spectrometer (SBRS) in China." Solar Physics 222.1 (2004): 167-173.

[2] Bengio, Yoshua. "Learning deep architectures for AI." Foundations and trends® in Machine Learning 2.1 (2009): 1-127.

[3] Le, Quoc V. "Building high-level features using large scale unsupervised learning." Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. IEEE, 2013.

[4] Sohn, Kihyuk, et al. "Efficient learning of sparse, distributed, convolutional feature representations for object recognition." Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011.

[5] Lee, Honglak, et al. "Unsupervised feature learning for audio classification using convolutional deep belief networks." Advances in neural information processing systems. 2009.

[6] Mohamed, Abdel-rahman, George E. Dahl, and Geoffrey Hinton. "Acoustic modeling using deep belief networks." Audio, Speech, and Language Processing, IEEE Transactions on 20.1 (2012): 14-22.

[7] Collobert, Ronan, et al. "Natural language processing (almost) from scratch." The Journal of Machine Learning Research 12 (2011): 2493-2537.

[8] Chen, Minmin, et al. "Marginalized denoising auto-encoders for nonlinear representations." Proceedings of the 31st International Conference on Machine Learning (ICML-14). 2014.

[9] M. Chen, Z. Xu, K. Weinberger, F. Sha, "Marginalized Stacked Denoising Autoencoders for Domain Adaptation", 29th International Conference on Machine Learning (ICML), 2012.

[10] Hinton, Geoffrey E., and Ruslan R. Salakhutdinov. "Reducing the dimensionality of data with neural networks." Science 313.5786 (2006): 504-507.

[11] Chen, Zhuo, et al. "Imaging and representation learning of solar radio spectrums for classification." Multimedia Tools and Applications (2015): 1-17.

[12] Lenz Ian, Honglak Lee, and Ashutosh Saxena. "Deep learning for detecting robotic grasps." The International Journal of Robotics Research 34.4-5 (2015): 705-724.

[13] Jalali Ali, et al. "A dirty model for multi-task learning." Advances in Neural Information Processing Systems. 2010.

[14] Fawcett, Tom. "An introduction to ROC analysis." Pattern recognition letters 27.8 (2006): 861-874

[15] L. Ma, Z. Lu, and H. Li, "Learning to answer questions from image using convolutional neural network," arXiv 1506.00333.

[16] G. E. Dahl, T. N. Sainath, and G. E. Hinton, "Improving deep neural networks for lvcsr using rectified linear units and dropout", ICASSP, 2013.

[17] L. Ma, Z. Lu, L. Shang, and H. Li, "Multimodal convolutional neural networks for matching image and sentence," arXiv 1504.06063.