

Pulsar Candidate Identification with Artificial Intelligence Techniques

Ping Guo,^{1,2*} Fuqing Duan,¹ Pei Wang³, Yao Yao², and Xin Xin²

¹*Image Processing and Pattern Recognition Laboratory, Beijing Normal University, Beijing 100875, China*

²*School of Computer Science, Beijing Institute of Technology, Beijing 100081, China*

³*National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Discovering pulsars is a significant and meaningful research topic in the field of radio astronomy. With the advent of astronomical instruments such as the Five-hundred-meter Aperture Spherical Telescope (FAST) in China, data volumes and data rates are exponentially growing. This fact necessitates a focus on artificial intelligence (AI) technologies that can perform the automatic pulsar candidate identification to mine large astronomical data sets. Automatic pulsar candidate identification can be considered as a task of determining potential candidates for further investigation and eliminating noises of radio frequency interferences or other non-pulsar signals. AI techniques, especially deep convolutional neural networks (DCNNs) had been adopted for pulsar candidate classification in the literature. However, it is very hard to raise the performance of DCNN-based pulsar identification because the limited training samples restrict network structure to be designed deep enough for learning good features as well as the crucial class imbalance problem due to very limited number of real pulsar samples. To address these problems, we proposed a framework which combines deep convolution generative adversarial network (DCGAN) with support vector machine (SVM) to deal with imbalance class problem and to improve pulsar identification accuracy. DCGAN is used as sample generation and feature learning model, and SVM is adopted as the classifier for predicting candidate's labels in the inference stage. The proposed framework is a novel technique which not only can solve imbalance class problem but also can learn discriminative feature representations of pulsar candidates instead of computing hand-crafted features in preprocessing steps too, which makes it more accurate for automatic pulsar candidate selection. Experiments on two pulsar datasets verify the effectiveness and efficiency of our proposed method.

Key words: pulsars:general, methods:statistical, methods: data analysis, methods:numerical

1 INTRODUCTION

Pulsars are highly magnetized rotating neutron stars which emit a beam of electromagnetic radiation. It is compared to a laboratory under extreme physical conditions, and can be used as probes for a wide range of physics and astrophysics researches, such as the equation of state of dense matter, the properties of the interstellar medium, dark matter and dark energy, stellar evolution, the formation and evolution of binary and multiple-star systems. And especially, a spatial array of millisecond-period pulsars can be used as a

gravitational-wave telescope that is sensitive to radiation at nanohertz frequencies. Discovering new pulsars with modern radio telescopes survey projects, such as Parkes Multi-beam Pulsar Survey (PMPS) (Manchester et al. 2001), High Time Resolution Universe (HTRU) (Burke-Spolaor et al. 2011), Pulsar Arecibo L-band Feed Array survey (PALFA) (Deneva et al. 2009) et al., is a very important and meaningful task in astronomical studies. In particular, the Five-hundred-meter Aperture Spherical Telescope (FAST) is expected to revolutionize pulsar astronomy (Nan et al. 2011), and the FAST 19-beams drift-scan pulsar survey is expected to discover ~1500 new normal pulsars and about 200 millisecond pulsars (MSPs) (Smits et al. 2009). The observation data of radiation signals from radio pulsar surveys are processed and folded into diagnostic plots, which are referred as pulsar can-

* E-mail: pguo@bnu.edu.cn (PG); fqduan@bnu.edu.cn (FD); wangpei@nao.cas.cn (PW); bityaoyao@bit.edu.cn; xxin@bit.edu.cn (XX)

didates. Traditional human experts read these plots to determine if the corresponding candidates are from real pulsars or non-pulsar noises, and then perform further verifications on prospective candidates in hope of new discoveries.

Generally, one modern pulsar survey project can produce totally millions of candidates, identification by human's eye cannot meet the necessary for ranking the candidates produced by large surveys, and impossibly accomplish the further requirement of real-time data processing. Automatic scoring or the artificial intelligence related methodology have been successfully applied in radio astronomy study, especially in the pulsar candidate selection. In the past years, [Lee et al. \(2013\)](#) proposed a linear combination of six different quality factors to calculate a score for each candidates, and ranking according to the scores. [Morello et al. \(2014\)](#) developed six hand-crafted features and trained a single-hidden-layer neural network for binary classification on candidates. [Zhu et al. \(2014\)](#) proposed a pulsar image-based classification system, which take candidate plots as inputs, and train single-hidden-layer artificial neural networks (ANN) and support vector machines (SVMs) on histogram plots and convolutional neural networks (CNNs) on two-dimensional image plots. The predictions of all classifiers are then assembled together with a logistic regression classifier.

Artificial intelligence techniques emulate human experts and can automatically learn the mapping relationships between samples and labels from a training dataset labelled by human experts. However, the performance of these methods is generally limited to data incompleteness or sample imbalance. In fact, the positive labeled samples, i.e. real pulsar candidates, are far from enough. If CNN architecture is the LeNet-5 ([Lecun et al. 1998](#)) like, which contains two convolution layers, two max-pooling layers and a fully-connected layer. With very limited true pulsar data samples, this architecture could be not learn good enough features.

In recent years, in the research filed of artificial intelligence, the generative adversarial networks (GANs) ([Goodfellow et al. 2014](#)) framework is very popular and effective for recovering the training data distribution. When learning converges, the generative model (generator, G) will generate new samples which are very similar to training data. And the discriminative model (discriminator, D) will predict the probability that an input sample comes from the training data. Radford *et al.* introduced deep convolutional generative adversarial networks (DCGANs) ([Radford et al. 2015](#)) which adopt CNNs as the generator and discriminator.

As we known, to fully train a deep CNN model, such as AlexNet ([Krizhevsky et al. 2012](#)), will need a large amount of samples. Furthermore, if using class imbalance training data set, we will get a discriminator with prejudice. That means the discriminator tends to recognize most unseen samples as domain class member. Unfortunately, the pulsar candidate sample data just is such a class imbalance data set because the number of positive pulsar candidate samples is very limited compared to huge number of non-pulsar samples. To combat imbalanced training data, one of the tactics is to generate synthetic samples. In this work, we propose a framework that is denoted as DCGAN+L2¹-SVM to deal

with this crucial class imbalance problem in order to improve pulsar recognition accuracy. In our DCGAN+L2-SVM framework, a DCGAN model is trained with whole pulsar samples and equal number of non-pulsar samples. The middle layer activation values of discriminator of the learned DCGAN are regarded as deep features of input samples. A L2-SVM linear classifier is trained on deep features of positive and negative pulsar samples. In inference stage, the label of input candidate is predicted with the trained L2-SVM linear classifier.

The following parts are organized as follows: In section 2, we will briefly introduce about pulsar candidate and the task of automatic pulsar candidate selection, and then give a review of AI-based methods. And our proposed method is summarized in section 3. In section 4, fundamental models including ANNs, CNN, GANs and DCGANs are summarized in short, some training techniques for DCGAN is also described. In section 5, our proposed DCGAN+L2-SVM model is introduced in detail, including network architecture, learning algorithm, and inference procedure. Then experimental results on two pulsar datasets are processed and analyzed in section 6. And in this section, there are some discussions about the proposed framework and feature representations. Conclusions are made and further work consideration is given in section 7.

2 AUTOMATIC PULSAR CANDIDATE SELECTION

In this section, we will briefly introduce about pulsar candidate search, then we formulate the problem of automatic pulsar candidate identification for machine processing as well as machine learning-based selection methods.

2.1 Pulsar Candidates

Modern radio telescope surveys are applied to search periodic broadband signals from universe space that shows signs of dispersion. Analyzing these signals provide a way to discovering new pulsar stars in the universe, and each received signal is recorded and processed in the pipeline of telescope surveys.

As described by [Lyon et al. \(2016\)](#) in their work, a pulsar search involves a number of procedural steps applied to the data set. In general, the first step is to conduct radio frequency interference (RFI) excision via the removal of channels corresponding to known interference frequencies. After these initial steps are complete, processing enters a computationally expensive phase known as de-dispersion. Dispersion by free electrons in the interstellar medium (ISM) causes a frequency dependent delay in radio emission as it propagates through the ISM. This delay temporally smears legitimate pulsar emission reducing the signal to noise ratio (S/N) of their pulses. The amount of dispersive smearing a signal receives is proportional to a quantity called the Dispersion Measure (DM). The degree to which a signal is dispersed for an unknown pulsar cannot be known a priori, thus several dispersion measure tests or "DM trials" must be conducted to determine this value. Periodic signals in de-dispersed time series data, can be found using a Fourier analysis. This step is known as a periodicity search. The

¹ L2 stands for L2 norm.

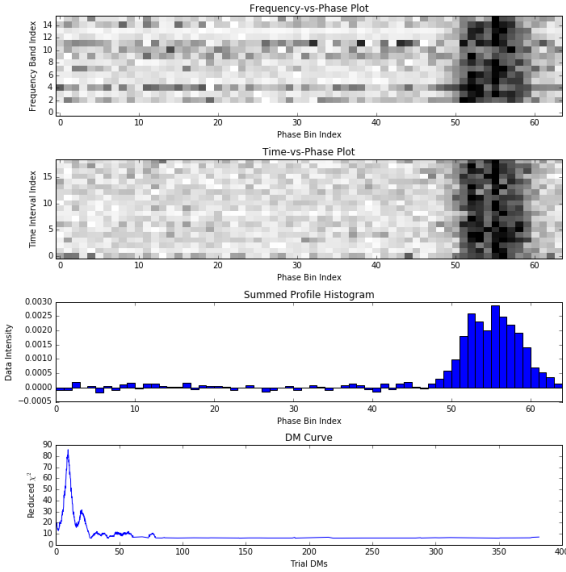


Figure 1. An example of positive (true) candidate from HTRU medlat dataset that is folded from signals of a real pulsar. For a real pulsar, there is a bold vertical stripe on the right side of both frequency-vs-phase and time-vs-phase plots. There is one peak in the summed profile histogram. And DM curve peaks at a non-zero value.

procedure after performing the fast Fourier transformation (FFT) of a periodicity search usually involves filtering the data to remove strong spectral features known as “birdies”. Summing techniques are subsequently applied, which add the amplitudes of harmonically related frequencies to their corresponding fundamentals. Periodic detections with large Fourier amplitudes post summing (above the noise background or a threshold level), are then considered to be “suspect” periods. A further process known as sifting is then applied to the collected suspects, which removes duplicate detections of the same signal at slightly different DMs, along with their related harmonics. A large number of suspects survive the sifting process. Diagnostic plots and summary statistics are computed for each of these remaining suspects forming candidates, which are stored for further analysis. The basic candidate consists of a small collection of characteristic variables. These include the S/N, DM, period, pulse width, and the integrated pulse profile. More detailed candidates also contain data describing how the signal persists throughout the time and frequency domains. Above pulsar search steps are usually implemented via a pipeline software, such as Pulsar Exploration and Search Toolkit (PRESTO)². PRESTO is a pulsar search and analysis software which has helped discovering more than 300 new pulsars up to now.

A folded pulsar candidate is a group of diagnostic plots, which are represented by the two-dimensional matrix. These diagnostic plots can be considered as images to be processed with CNN model. Pulsar candidate diagnostic representations mainly include *summed profile histogram* (SPH), *time-vs-phase plot* (TPP), *frequency-vs-phase plot* (FPP) and *dispersion-measure (DM) curve*.

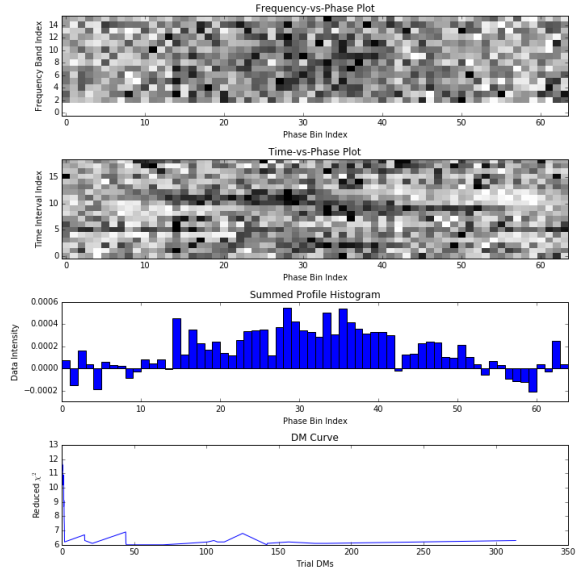


Figure 2. An example of negative (non-pulsar) candidate from HTRU medlat dataset, which is folded from signals of a non-pulsar. For a non-pulsar, there is no obvious vertical stripe in neither frequency-vs-phase nor time-vs-phase plots. There are also no peaks in the summed profile histogram. And DM curve peaks at zero value.

- *Summed profile histogram* is an intensity-vs-phase pulse profile. It is obtained by adding the data over both time intervals and frequencies. The profile of a real pulsar usually contains one or more narrow peaks.

- *Frequency-vs-phase plot* is obtained by summing the data over time intervals. For a real pulsar, there also should be one or more vertical lines, which indicates a broadband signal was observed. If summing the frequency-vs-phase plot over frequency bands, we would obtain summed profile histogram.

- *Time-vs-phase plot* is calculated by summing the data over frequencies. For a real pulsar, there should be one or more vertical stripes which indicates a pulsed signal was observed. When summing the time-vs-phase plot over time intervals, we would also obtain summed profile histogram.

- *DM curve* is a histogram of the trial DMs against the corresponding χ^2 values. For a real pulsar, DM curve peaks at a nonzero value.

Fig.1 shows an example of positive (true) pulsar candidate from HTRU medlat dataset, which comes from a real pulsar. Fig.2 illustrates an example of negative (non-pulsar) candidate from HTRU medlat dataset, which is folded from a non-pulsar signal. From these two figures, we can see that positive (true) and negative (non-pulsar) candidates have different characteristics in diagnostic plots.

2.2 Modeling Automatic Pulsar Candidate Selection

Pulsar candidate selection is the task of determining prospective candidates for further verification and excluding noise candidate which might be radio frequency interferences or other non-pulsar signals. Although for most candi-

² <http://www.cv.nrao.edu/~sransom/presto/>

dates the diagnostic plots show obvious difference between a real pulsar and a non-pulsar, for example Fig.1 is very different to Fig.2 according to human, modern radio observatories produce millions of candidates in a very short time. So automatic pulsar candidate selection with AI techniques is a very meaningful and important method for discovering new pulsars.

Automatic pulsar candidate selection task can be modeled as followings: A pulsar candidate for training is denoted as a tuple of five elements $x = (x_{subints}, x_{subbands}, x_{profile}, x_{dm}, y)$, where $x_{subints}$ refers to a time-vs-phase plot, $x_{subbands}$ refers to a frequency-vs-phase plot, $x_{profile}$ is a summed profile histogram, x_{dm} is the corresponding DM curve and at last y is the ground-truth label of x .

A predicting function f is learned based on the training data.

$$f_* = \underset{f}{\operatorname{argmin}} \frac{1}{n} \sum_{x \in X} L(f(x) - y), \quad (1)$$

where x is a pulsar candidate sample in training dataset X , y is the corresponding label of that candidate, n is the number of training samples in X , and L is a loss function defined in training algorithm, and f_* is a well trained optimal predicting function. In testing stage, $y_{predict} = f_*(x_{test})$ makes the real pulsar (true) or non-pulsar (false) label prediction for input candidate x_{test} .

2.3 Automatic Pulsar Candidate Selection with Machine Learning Techniques

Recently, scoring and machine learning-based methods show great capabilities in automatic pulsar candidate selection task. Lee et al. (2013) proposed a scoring method called Pulsar Evaluation Algorithm for Candidate Extraction (PEACE), which computes score with six different quality factors. There ranking method helped to find 47 new pulsars. Eatough et al. (2010) hand-crafted 12 features and trained a feed-forward single-hidden-layer artificial neural network. Their work helped to find one new pulsar in PMPS data. Bates et al. (2012) adopted the 12 features from Eatough et al. (2010) and 10 more from (Keith et al. 2009) and trained a feed-forward single-hidden-layer artificial neural network with these 22 features. In their work, 75 new pulsars were discovered. Morello et al. (2014) proposed Straightforward Pulsar Identification using Neural Networks (SPINN), which designs six features and trains a feed-forward single-hidden-layer artificial neural network for candidate binary classification. Their method contributed to 4 new pulsar discoveries. Thornton (2013) also used a single-hidden-layer network as a classifier but he hand-crafted 22 features based on candidate plots and observation parameters. Lyon et al. (2016) designed 8 features from diagnostic histograms to describe pulsar candidate and proposed a fast decision tree model for on-line operation. Their model helped to find 20 new pulsars. In previous machine learning based method, artificial neural networks play an important role, especially single-hidden-layer network. Wang et al. (2017); Guo & Lyu (2004) proposed a pseudoinverse incremental representation learning algorithm for training neural networks and applied it to perform spectra pattern recognition. Zhu et al. (2014) developed a pulsar image-based classification system which

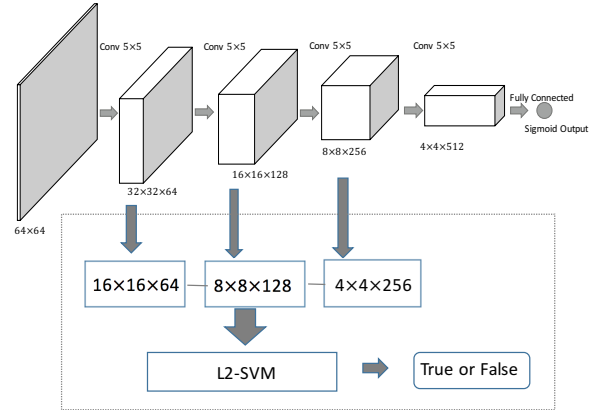


Figure 3. An illustration of our DCGAN+L2-SVM method. The candidate deep features are mapped with the discriminator of the trained DCGAN model, which consist of a 16-by-16-by-64 tensor, a 8-by-8-by-128 tensor, and a 4-by-4-by-256 tensor. A L2-SVM classifier is learned with candidate deep features. In testing stage, the label of input candidate is predicted with previous trained L2-SVM.

take candidate’s histograms and plots as input, and train single-hidden-layer network, SVM and CNN, and combined all the classifiers with logistic regression to form an ensemble net. Among these method, CNNs show good classification performance on two-dimensional diagnostic plots. Compared to other methods, CNNs take the plots as input and realize end-to-end learning, in contrast, other methods require hand-crafted features for pulsar candidate selection. And CNNs performs better than other methods on testing (Zhu et al. 2014). By assembling all the classifiers, accuracy can be further improved over a single classifier. Although CNNs are powerful on two-dimensional image data, training a deep CNN requires more labeled samples. As we known, in real scenario, labeled data is difficult and expensive to obtain. In pulsar candidate selection task, positive samples are very limited because the number of discovered real pulsar are small, and millions of candidates are negative samples. This also lead to a crucial class imbalance problem. So learning a discriminative classification model with limited training candidate and class imbalance samples bring us an important and challenging issue.

3 METHODOLOGY

To tackle class imbalance and small training samples problems, we propose a framework named as DCGAN+L2-SVM for automatic pulsar recognition, an illustration of our proposed framework is shown in Fig.3. In this framework, a DCGAN is first trained with the pulsar dataset. The middle layer activation values of discriminator are treated as candidate’s deep features. A L2-SVM linear classifier is learned with this deep feature representation. In inference stage, we can predict the label of an input pulsar candidate with the learned L2-SVM classifier.

In experiments, a baseline RBF-SVM is trained and tested with pulsar candidate’s PCA features for recognition. For comparison, LeNet-5 like CNNs in Zhu et al. (2014) are also tested and compared. Experiments on

HTRU medlat dataset (Morello et al. 2014) and PMPS-26k dataset collected by National Astronomical Observatories Chinese Academy of Sciences demonstrate that the proposed method, DCGAN+L2-SVM, can improve pulsar candidate recognition accuracy over CNN baselines and traditional kernel SVMs.

The main contributions of our work are summarized as follow:

- A DCGAN-based model, DCGAN+L2-SVM, is proposed to automatically identify pulsar candidates. a DCGAN model can learn the data features and generate new samples from training data, and the middle layer activation values of discriminator are learned as deep features. Then a L2-SVM linear classifier is trained with these features for recognition. In inference stage, DCGAN+L2-SVM predicts the input sample as a pulsar or non-pulsar according to the output of learned L2-SVM classifier.
- Empirical study of RBF-SVM, CNN, and DCGAN-L2-SVM method on both HTRU medlat data and PMPS-26k dataset shows that our proposed method outperforms other methods on pulsar candidate selection task.

4 BASIC MACHINE LEARNING MODELS

In this section, some related basic machine learning models, including ANN, CNN, GAN and DCGAN, are briefly reviewed.

4.1 Feed-forward Artificial Neural Networks

A typical feed-forward fully connected artificial neural network consists of an input layer, several middle hidden layers and an output layer. An example architecture of this kind of ANN is show in Fig. 4. The input layer takes as an input vector $\mathbf{x} \in \mathbb{R}^d$. The elements in former layer are combined with a weight vector and a bias. In Fig. 4, take the i^{th} neuron as an example, the output to the i^{th} neuron in the second layer is

$$h_i = \mathbf{w}_i^T \mathbf{x} + b_i. \quad (2)$$

Then the value h_i is processed with an activation function, such as sigmoid function, which is $g(h_i) = \frac{1}{1+e^{-h_i}}$. After all neuron values in this layer are computed in this way. They are then fed into next layer as an input vector. A similar calculation process is done. In the last output layer, a final value is computed as y .

In training stage, the target is to learn all the weights and biases in all layers. Traditional back propagation learning algorithm (Rumelhart et al. 1988) could be adopted for “shallow” ANNs. Here shallow means the networks only consist of a very small number of hidden layers (one or two hidden layers). When the number of hidden layers becomes larger, usually is more than 5 layers, the ANN model will be a deep neural network. The deep fully connected network contains too many parameter to be learnt, and the gradient in back propagation algorithm may vanish in the deep architecture. Some learning algorithms such as auto-encoders (Bourlard & Kamp 1988; Bengio et al. 2009; Guo & Lyu 2004; Wang et al. 2017; Guo et al. 2017) can be adopted to train a deep neural network. Since single hidden layers ANN

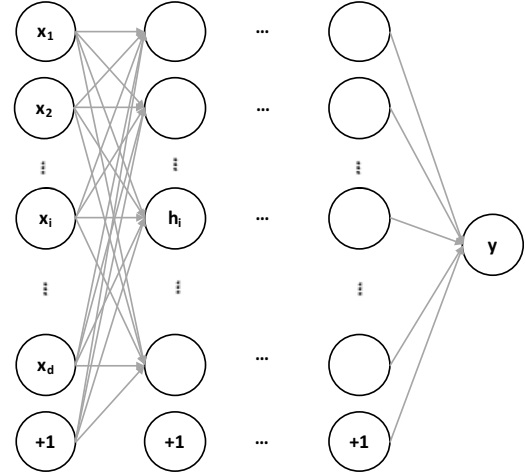


Figure 4. An example of feed-forward fully connected artificial neural network.

has been well studied, in previous automatic pulsar candidate selection works, Eatough et al. (2010); Bates et al. (2012); Thornton (2013); Zhu et al. (2014); Morello et al. (2014), they all adopted this kind feed-forward ANN as the classifier model.

4.2 Convolutional Neural Networks

CNN takes the advantage of local connectivity in receptive field of two-dimensional images, and deep CNN architecture has been widely adopted in the research field of computer vision. Because CNN based neural network architecture has the advantage in deal with images, it can be adopted as a better feature learning model for automatic pulsar candidate selection when we take those diagnostic plots as images. In previous work, Zhu et al. (2014) demonstrated the capability and advantages of CNN for pulsar candidate selection task. Unlike feed-forward ANN where the output of former layer is a linear combination of all input elements, CNN only calculate the output in a two-dimensional receptive field.

Fig. 5 presents an illustration of convolutional layer. Denote the i^{th} block of input two-dimensional image as \mathbf{x}_i , the k^{th} filter as \mathbf{w}_k . The output value h_i of convolution operation corresponding to \mathbf{x}_i is

$$h_i = \sigma(\mathbf{w}_k^T \mathbf{x}_i + b_k), \quad (3)$$

where \mathbf{w}_k and \mathbf{x}_i are both flatten into vectors, b_k is a bias and σ is an activation function, typically $\sigma(t) = \max(0, t)$. In Zhu et al. (2014), hyperbolic tangent activation function is adopted. The output image of a convolutional layer is called a feature map. In Fig. 5, the k^{th} feature map is obtained by applying the k^{th} filter on every block of the input image divided by a sliding window strategy. Obviously, the size of input image, stride of the sliding window and convolutional filter size all together determine the size of output feature map.

After a convolutional layer, there usually is a pooling layer. A pooling layer down-samples the feature maps into a smaller map and brings in some invariant characteristics. An illustration of a pooling layer is demonstrated in Fig. 6, where four values in the input feature map are pooled into

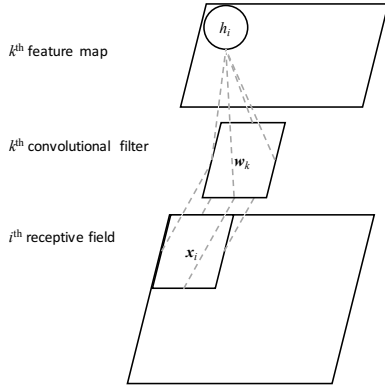


Figure 5. An illustration of stride convolutional layer. \mathbf{x}_i is an image block called i^{th} receptive field on a sliding window of input image. \mathbf{w}_k is the k^{th} convolution filter. \mathbf{x}_i and \mathbf{w}_k have the same size. h_i is the convolution output value corresponding to \mathbf{x}_i .

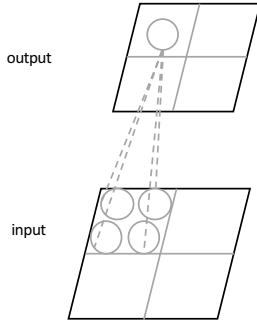


Figure 6. An illustration of max pooling layer. Four values in a neighborhood are pooled into one value in the output.

one in the output. Typically, the pooling method is a maximum or averaging function. In [Zhu et al. \(2014\)](#), max pooling is adopted, and the CNN architecture is a LeNet-5 ([Lecun et al. 1998](#)) like network: The first and third layers are convolutional layers, the second and fourth layers are max pooling layers, and the last layer is a fully connected layer with a sigmoid output.

4.3 Generative Adversarial Networks

Generative Adversarial Network (GAN) is a new framework proposed by Goodfellow et al. ([Goodfellow et al. 2014](#); [Goodfellow 2016](#)). GAN originally aims to estimate generative models via an adversarial learning process. In the process, two models are learned: a generative model G called generator, which captures the data distribution of training samples, and a discriminative model D called discriminator, which predicts the probability of an input sample coming of real training samples rather than fake data generated by G . For generator G the training is to maximize the probability of discriminator D making mistakes. For discriminator D the training is to minimize the probability of making mistakes. The whole training is proceeded in a way of minimax two-player game, final training converges to a stationary Nash equilibrium.

Fig. 7 shows the process. Noise variables \mathbf{z} has a prior

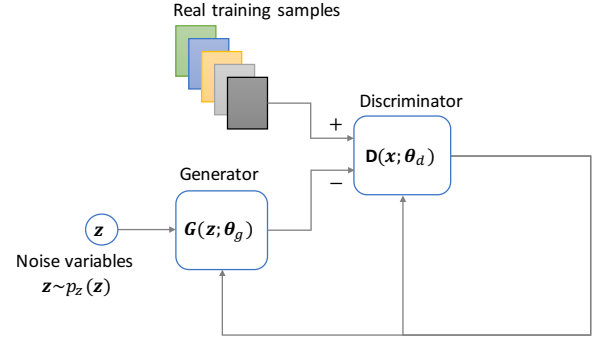


Figure 7. Generative adversarial network framework. Noise prior distribution $\mathbf{z} \sim p_z(\mathbf{z})$. G is the generator. D is the discriminator. Fix G , D is trained with real data samples and synthesized fake data by minimizing the probability of taking fake data as real ones. Fix D , G is trained with some synthesized fake data by maximizing the probability of taking fake data as real ones.

distribution $p_z(\mathbf{z})$. \mathbf{z} is then mapped into the training data space with the generator $G(\mathbf{z}; \theta_g)$. The generated fake data samples and real training data are used together to train the discriminator $D(\mathbf{x}; \theta_d)$. The training of G and D are processed alternatively. D and G plays a two-player minimax game with the following value function,

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{\text{training_data}}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]. \quad (4)$$

where $p_{\text{training_data}}(\mathbf{x})$ refers to the training data distribution, and $\mathbb{E}[\cdot]$ denotes expectation value which in practice is a averaged value on a mini-batch.

The learning algorithm for a GAN framework is summarized as follows:

A) Fix generator G , train discriminator D , repeat k times:

- 1) from noise prior $p_z(\mathbf{z})$, sample a mini-batch of m noise $\{\mathbf{z}^{(1)} \dots \mathbf{z}^{(m)}\}$;
- 2) from real training data, sample a mini-batch of m samples $\{\mathbf{x}^{(1)} \dots \mathbf{x}^{(m)}\}$;
- 3) update discriminator parameter θ_d by ascending its stochastic gradient:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m [\log D(\mathbf{x}^{(i)}) + \log(1 - D(G(\mathbf{z}^{(i)})))] \quad (5)$$

B) From noise prior $p_z(\mathbf{z})$, sample a mini-batch of m noise $\{\mathbf{z}^{(1)} \dots \mathbf{z}^{(m)}\}$;

C) Fix discriminator D , update generator parameter θ_g by descending its stochastic gradient:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m [\log(1 - D(G(\mathbf{z}^{(i)})))] \quad (6)$$

The steps from **A)** to **C)** are updated iteratively until the maximum number of training iterations is reached.

5 DCGAN-BASED AUTOMATIC PULSAR CANDIDATE SELECTION

In this section, the proposed DCGAN-based model shown in Fig 3, DCGAN+L2-SVM, is described. More details, in-

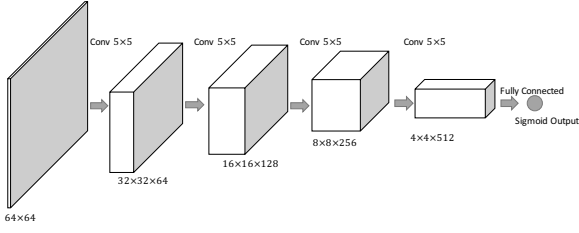


Figure 8. DCGAN discriminator, D , trained in DCGAN+L2-SVM. The input samples are 64-by-64 gray images. There are four stride convolution layers. The final 4-by-4-by-512 features are fed into a single sigmoid output, which provides the probability of input image coming from real dataset.

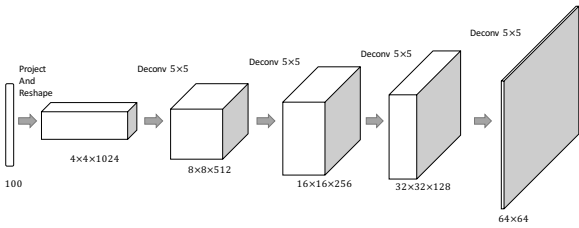


Figure 9. DCGAN generator, G , trained in DCGAN+L2-SVM. The input is a 100 dimensional uniform noise distribution Z . Input is randomly projected and then reshaped into a 4-by-4-by-1024 tensor. There are four fractional-stride convolution layers (or deconvolution layers). The output, $G(z)$, is a generated 64-by-64 pixel gray image.

cluding generator and discriminator neural network architectures, learning and inference techniques, are given in Appendix A.

5.1 DCGAN+L2-SVM

DCGAN defines a family of GAN architectures which are able to perform stable training and to learn good image feature representations in an unsupervised learning manner. In this work, we take the time-vs-phase and frequency-vs-phase two-dimensional plots as the input of DCGAN+L2-SVM, and a L2-SVM linear classifier is trained for predicting labels of new input samples. An illustration of the framework of DCGAN+L2-SVM is shown in Fig.3.

The architecture of generator and discriminator in DCGAN+L2-SVM follows the guidelines for stable Deep Convolution GANs in Radford et al. (2015), which includes the following items:

- Use stride convolutions (Springenberg et al. 2014) in discriminator to learn to own spatial downsampling. Use fractional-stride convolutions in generator to learn upsampling. No fully-connected or pooling are adopted.
- Use Batch Normalization (Ioffe & Szegedy 2015) in discriminator and generator both.
- ReLU activation function is used (Nair & Hinton 2010) in all layers in generator except output layer, that layer uses Tanh activation function.
- LeakyReLU activation function is used (Xu et al. 2015; Maas et al. 2013) in all layers in discriminator.

With our experience in artificial intelligence research,

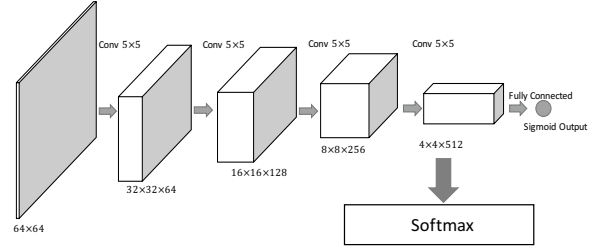


Figure 10. An illustration of observing how discriminative capability of deep feature representation changes.

we design the architecture of both discriminator and the generator with four layers of convolution or deconvolution. Specifically, the structure of the discriminator, where input gray image size is 64-by-64, is shown in Fig.8. The discriminator uses a serial of four strides convolutional layers and predicts the probability with a single sigmoid neuron output. In the first convolutional layer, 64 kernels of size 5×5 are adopted, the stride is 2, and the output feature map is a $32 \times 32 \times 64$ tensor. In the second convolutional layer, 2 kernels of size 5×5 are adopted and the stride is 2. The output feature map is a $16 \times 16 \times 128$ tensor. In the third convolutional layer, 2 kernels of size 5×5 are adopted and the stride is 2. The output feature map is a $8 \times 8 \times 256$ tensor. In the last layer, 2 kernels of size 5×5 are applied, resulting in a $4 \times 4 \times 512$ tensor, which is fed into a sigmoid neuron to obtain final output.

The architecture of the generator is illustrated in Fig.9. The generator produces 64-by-64 gray images by taking a 100 dimensional uniform distribution random noises sampled from $[-1, 1]$ range as inputs. The 100 dimensional noise is projected and reshaped into a $4 \times 4 \times 1024$ tensor with a fully connected layer, four deconvolutional layers follow this fully connected layer. The kernel size is 5×5 , and the stride is 2 for these layers. After each deconvolutional layer, the height and width are doubled and the number of channels are halved. The output sizes of the first three deconvolutional layers are $8 \times 8 \times 512$, $16 \times 16 \times 256$ and $32 \times 32 \times 128$, respectively. The last deconvolutional layer outputs a generated 64×64 gray image, this generated image will be considered as new sample when generator well trained.

The DCGAN model in our DCGAN+L2-SVM framework is trained with 64×64 gray images as input, so in pre-processing stage, all images in datasets are resized to 64×64 and normalized to $[-1, 1]$ range. After the DCGAN model is well trained, the middle layer activation values (or feature maps) of the discriminator is treated as deep features of the corresponding input sample, which will be fed into the discriminator for forward propagation further. This procedure is illustrated in Fig 3. In the figure, the feature maps are $32 \times 32 \times 64$, $16 \times 16 \times 128$ and $8 \times 8 \times 256$ tensors. These feature maps are first max-pooled in 4×4 neighborhood with stride 2, the pooled results are $16 \times 16 \times 64$, $8 \times 8 \times 128$ and $4 \times 4 \times 256$ tensors. All these deep features are reshaped into a long vector of size 28672 as a feature representation for the input sample, then a L2-SVM linear classifier is trained on this feature representation and will be used for future identifying a pulsar candidate to be a real or not pulsar.

For a better observation of how discriminative capability of deep feature representation improves, the discrimina-

tor in DCGAN is checked during the DCGAN training process by extracting deep features with current discriminator for pulsar classification. An illustration for this observation is shown in Fig 10. At a checking point, the outputs of last convolutional layer are fed into a two-class Softmax classifier. And only the Softmax classifier is updated with the training pulsar samples to minimize the training errors, while those layers in DCGAN stay unchanged during checking discriminative capability change situation. Then Softmax classifier is used to predict testing pulsar candidate sample labels, and the generalization performance of a classifier is used to judge the discriminative capability of deep representation, higher classification accuracy reveals stronger discriminative capability of the classifier.

5.2 Training DCGAN on Pulsar Images

The DCGAN training procedure is a two-player competition game. The discriminative network aims to predict the correct label of the input sample which is either from real training data or generative model made data. The generative model aims to produce synthesized samples which can not be correctly classified by the discriminative model. When the training is accomplished, generative network synthesizes samples that are very similar to real pulsar images, and the discriminative network predicts the possibility of input samples coming from real dataset.

All the models can be trained with mini-batch stochastic gradient descent (SGD) with a mini-batch size of M . In order to avoid overfitting, discriminator is optimized k steps, and then generator is updated by one step. Because the activation of generator output layer is Tanh, input images are scaled to $[-1, 1]$, which is the range of Tanh activation function. The more details about description of training algorithm can be found in Goodfellow et al. (2014).

6 RESULTS AND ANALYSIS

In this section, we first introduce some preliminaries, such as dataset information, evaluation metrics, and some parameter settings, then we present and analyze the experiment results on HTRU dataset and PMPS-26k data. Finally, some discussions are given about the advantages and disadvantages of our method.

6.1 Datasets and Evaluation Metrics

Two pulsar candidate datasets are investigated in our experiments: HTRU medlat and PMPS-26k. HTRU medlat dataset is the first public available labeled benchmark dataset released by Morello et al. (2014). HTRU medlat precisely consists of 1,196 positive candidate examples which are sampled from 521 distinct real pulsars, and 89,996 negative candidate examples. PMPS-26k is a pulsar candidate dataset built on the PMPS (Manchester et al. 2001) observation data, which comprises 2,000 positive examples that are obtained from real pulsars, 2,000 negative samples that are obtained from non-pulsar signals, 20,000 negative samples that are obtained from radio frequency interference, and 2,000 samples whose labels are unknown. Table.1 lists the number of examples in both datasets.

Dataset Names	#positive	#negative	#total
HTRU medlat	1,196	89,996	91,192
PMPS-26k	2,000	22,000	26,000

Table 1. The number of examples in HTRU medlat and PMPS-26k datasets. Notably, there are 2,000 examples without labels in PMPS-26k.

Outcomes	Prediction -	Prediction +
Groundtruth -	True Negative	False Positive
Groundtruth +	False Negative	True Positive

Table 2. Binary classification confusion matrix. It defines all the outcomes of prediction, which includes True Negative (TN), False Negative (FN), False Positive (FP), and True Positive (TP).

Evaluation metrics that we adopt for pulsar candidate classification task here include *Precision*, *Recall* and *F-score*. Define binary classification confusion matrix in Table. 2, then

$$Precision = \frac{TP}{TP + FP}, \quad (7)$$

$$Recall = \frac{TP}{TP + FN}, \quad (8)$$

$$Fscore = \frac{2 * Precision * Recall}{Precision + Recall}. \quad (9)$$

6.2 Parameter Settings

The parameter settings basically follow that of DCGAN in Radford et al. (2015). Because the activation function of generator output is Tanh, the images were scaled to the range of Tanh, $[-1, 1]$. The images in both PMPS-26k and HTRU medlat were resized to 64-by-64 pixels. The mini-batch size of SGD was $M = 128$. All weights were initialized by a normal distribution with zero mean value and 0.02 standard deviation. LeakyReLU adopted 0.2 as the slope of the leak. ADAM optimizer (Kingma & Ba 2014) was used for accelerating training, in which the learning rate was set to 0.0002.

The labeled data were first randomly split into three folds for training, validation and testing by 30%, 30% and 40%, respectively. In this experiment part, three types of classification method are validated and compared: radial basis function kernel SVM (RBF-SVM), CNN and our proposed DCGAN+L2-SVM. RBF-SVM takes 24 principal components obtained from principal component analysis (PCA) of input images as features, and these principal components are adopted to train a RBF kernel SVM classifier. Here, CNN denotes the network architecture used in Zhu et al. (2014). In their CNN method, the images were firstly resized to 48×48 instead of 64×64 pixels.

For our model of DCGAN+L2-SVM, in which a DCGAN is trained with the total labeled samples, and the max-pooled activation values of middle layers in discriminator are regarded as deep features for each sample. A L2-loss linear SVM (Chang & Lin 2011) is then trained with these feature

Method	Fscore	Recall	Precision
RBF-SVM-1	0.868	0.854	0.882
RBF-SVM-2	0.866	0.847	0.886
CNN-1	0.953	0.956	0.950
CNN-2	0.952	0.953	0.951
DCGAN-L2-SVM-1	0.963	0.966	0.961
DCGAN-L2-SVM-2	0.964	0.963	0.965

Table 3. Evaluations of different methods on *HTRU Medlat Dataset*. Three types of classifiers were investigated: RBF-SVM, CNN and DCGAN+L2-SVM. Suffix ‘1’ means the model is trained on time-vs-phase plot data, and ‘2’ means the model is learned on frequency-vs-phase plots.

representations. One thing should be noted that the parameters in RBF kernel and SVMs are validated on the validation data to choose the best to the highest classification accuracy.

For simplicity in later reference, the method names with a ‘1’ suffix refer to classifiers trained with time-vs-phase plots, and the method names with a ‘2’ suffix refer to classifiers learned with frequency-vs-phase plots.

6.3 Evaluations on HTRU Medlat Dataset

The performance metrics of pulsar candidate classification on HTRU medlat dataset are demonstrated in Table.3. Traditional methods, like RBF-SVM-1 and RBF-SVM-2, achieve about 0.86 *Fscore*. CNN methods have better performance than RBF-SVMs. Both baseline methods, CNN-1 and CNN-2 achieved good performance according to *Fscore* values, which are 0.953 and 0.952. Specifically, *Fscore* of CNN-1, 0.953 is larger than that of RBF-SVM-1, which is 0.868. *Fscore* of CNN-2, 0.952 is larger than that of RBF-SVM-2, 0.866. The reason that performance of CNN baselines is good may be two folds: one is that the input data is two-dimensional plots which convey a lot of discriminative information for classification task, the other is that CNNs are able to model the spatial features very well especially in two-dimensional images. After training a DCGAN with whole dataset, the learned discriminator in DCGAN can be used as the hierarchical feature extractors. In the framework of DCGAN+L2-SVM, these hierarchical deep features are used to train a L2-SVM linear classifier. The performance of DCGAN+L2-SVM is improved over baseline CNNs by about 1% on *Fscore*. DCGAN+L2-SVM-1 improves 0.6% over CNN-1. DCGAN+L2-SVM-1 improves 0.7% over CNN-2. These results validate that DCGAN can provide a good way of deep discriminative feature extracting, and the discriminative feature representation can help to raise the pulsar identification accuracy.

6.4 Evaluations on PMPS-26k Dataset

The performance of pulsar candidate classification on PMPS-26k dataset is listed in Table.4. The *Fscores* of RBF-SVM-1 and RBF-SVM-2 are 0.820 and 0.813, respectively. From the table we can see that CNN-1 achieves about 0.883 on *Fscore*, which is 6.3% larger than that of RBF-SVM-1. And CNN-2 achieves about 0.879 on *Fscore*, which is 6.6% larger than that of RBF-SVM-2. This fact

Method	Fscore	Recall	Precision
RBF-SVM-1	0.820	0.806	0.835
RBF-SVM-2	0.813	0.798	0.828
CNN-1	0.883	0.886	0.881
CNN-2	0.879	0.879	0.880
DCGAN-L2-SVM-1	0.889	0.895	0.885
DCGAN-L2-SVM-2	0.886	0.891	0.881

Table 4. Evaluations of different methods on *PMPS-26k Dataset*. Three types of classifiers were investigated: RBF-SVM, CNN and DCGAN+L2-SVM. Suffix ‘1’ means the model is trained on time-vs-phase plot data. ‘2’ means the model is learned on frequency-vs-phase plots.

also shows that CNN methods perform better than those RBF-SVM classifiers on pulsar candidate classification task. While DCGAN+L2-SVM-1 improves accuracy about 0.6% over CNN-1 on *Fscore*, and DCGAN+L2-SVM-2 improves accuracy about 0.7% over CNN-2 on *Fscore*. The main improvement of DCGAN+L2-SVM over CNN here is *recall* value. They both have about 0.88 *Fscore*, DCGAN+L2-SVM-1 improves about 0.9% over CNN-1 on *recall* value, and DCGAN+L2-SVM-2 improves about 1.2% over CNN-2 on *recall* value. These results show that the deep feature representation obtained with DCGAN discriminator output indeed can improve the pulsar candidate classification performance further.

6.5 Observation of How Discriminative Capability Changes

In order to show how discriminative capability of deep feature representation changes, we design a checking technique which is shown in Fig 10. During the training process of DCGAN on the whole labeled samples, we set every 200 epoch of iteration as a checking point. In the experiments, PMPS-26k dataset with time-vs-phase plot images are adopted. At each checking point, we train a Softmax classifier with the training data and obtains a classification F1-score on the testing data. Fig 11 shows the *F-score* results of pulsar candidate classification at 10 different checking points. The curve shows that *F-score* increases along with the training process and reaches to 0.889 as the DCGAN training converges to a stationary point. This result illustrates that the discriminative capability of DCGAN deep feature representation is getting stronger with the increase of epoches in DCGAN training process.

6.6 Discussions and further work consideration

The experiment results in Table.3 and Table.4 show that our DCGAN+L2-SVM framework takes the advantages of deep convolutional generative adversarial network and achieves good performance for automatic pulsar candidate identification task. Even when the number of training samples, specifically positive ones, is relatively limited, deep convolutional generative adversarial learning still can provide a good way to learn a strong discriminative feature representation. The principle behind this framework is that generator can produce fake samples which are almost the same as given train-

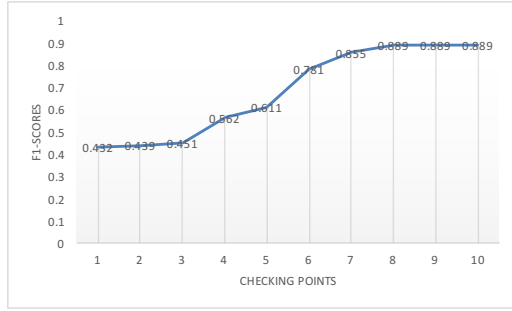


Figure 11. A performance curve showing how the discriminative capability changes over DCGAN training process.

ing samples. For class imbalance problem, one of the tactics is to sample equal number of data point for each class from given dataset. But for pulsar candidate dataset, only very small number of training samples can be drawn because of limited positive pulsar samples, for example, for HTRU medlat dataset only 1,196 training sample pairs, for PMPS-26k dataset only 2,000 training sample pairs can be obtained in this way. It is well known to train a deep neural network (DNN) a large number of labelled samples is needed, if using a small sample dataset to training DNN, we cannot expect to obtain a good generalization performance because overfitting will occur. This can be revealed by comparing the CNN model performance on HTRU medlat and PMPS data. To deal with this dilemma, we propose to adopt the tactic of generating synthetic samples. During training process, generator in DCGAN will generate fake samples, and discriminator will be trained with true and fake samples. After training converges to a stationary Nash equilibrium, we can obtain not only a good generator, but also a good discriminator. With this framework, we can solve class imbalance and small training sample problem simultaneously. Another advantage of DCGAN+L2-SVM framework is that makes it is possible to eliminate the hard-work of designing good hand-crafted features for classification by learning an end-to-end model with the original input data. The results on both datasets show DCGAN+L2-SVM has shown the capability of learning stronger discriminative features for pulsar classification than that traditional CNN model does. Obviously, as a complicated deep neural network model, DCGAN+L2-SVM can perform better with limited labeled training data under the training strategy of the GAN, in this point it outperforms CNN models. It is believed that with more new labeled samples, especially positive real pulsar candidate data, DCGAN+L2-SVM and CNN models both could improve pulsar candidate classification performance further.

In this work, we mainly evaluated the performance of methods based on deep neural networks, and only take two-dimensional diagnostic plots as network's inputs. In later work, we will design a multi-modal input DNN based pulsar recognition model, which will incorporate one dimensional SPH and DM curve, two dimensional TPP and FPP, and other hand-craft features together to make the pulsar identification more accurate. In addition, some traditional machine learning models, which use hand-crafted features of SPH or DM curve to train classifiers, still provides com-

plementary discriminating capabilities. We will investigate these one-dimensional input with stacked auto-encoder deep neural network architecture, and develop the fast learning algorithm (Guo & Lyu 2004; Wang et al. 2017; Guo et al. 2017) for stream data analysis. Because ensemble neural networks can also improve pulsar candidate classification abilities, to assemble deep models and traditional models together to reach more better pulsar identification performance will be studied in our later work.

Our future work will consider to build an artificial intelligence system to deal with pulsar data processing problem, to reinvent the study of the pulsar search procedure, especial to speed up the DM parameter search processing time, and apply DCGAN+L2-SVM model to change the traditional pipeline of pulsar searching by taking the raw data as input for classification before the time when candidates are folded.

7 CONCLUSION

In this work, we have proposed a DCGAN-based automatic pulsar candidate identification framework, called DCGAN+L2-SVM. In this DCGAN+L2-SVM framework, a DCGAN model is trained on the whole labelled balance class samples, and the proposed framework can learn strong discriminative features for pulsar candidate selection task. The max-pooled middle layer activation values of the discriminator in DCGAN is regarded as deep feature representation, this can be considered as unsupervised learning stage. When DCGAN is well trained, those middle layer activation values of the discriminator will be taken as training sample vectors to supervised train a L2-SVM classifier, and this trained L2-SVM classifier will be used for identifying new input of pulsar candidate to be a real or not pulsar. Experimental results on HTRU medlat dataset and our self-collected PMPS-26k dataset show that DCGAN+L2-SVM outperforms previously used CNN as well as RBF-SVM models.

ACKNOWLEDGEMENTS

This work was fully supported by grants from the National Natural Science Foundation of China (61375045), and the Joint Research Fund in Astronomy (U1531242) under cooperative agreement between the National Natural Science Foundation of China (NSFC) and the Chinese Academy of Sciences (CAS).

REFERENCES

- Arjovsky M., Chintala S., Bottou L., 2017, arXiv preprint arXiv:1701.07875
- Bates S., et al., 2012, Monthly Notices of the Royal Astronomical Society, 427, 1052
- Bengio Y., et al., 2009, Foundations and trends® in Machine Learning, 2, 1
- Bourlard H., Kamp Y., 1988, Biological cybernetics, 59, 291
- Burke-Spolaor S., et al., 2011, Monthly Notices of the Royal Astronomical Society, 416, 2465
- Chang C.-C., Lin C.-J., 2011, ACM Transactions on Intelligent Systems and Technology (TIST), 2, 27
- Che T., Li Y., Jacob A. P., Bengio Y., Li W., 2016, arXiv preprint arXiv:1612.02136

- Che T., Li Y., Zhang R., Hjelm R. D., Li W., Song Y., Bengio Y., 2017, arXiv preprint arXiv:1702.07983
- Deneva J., et al., 2009, *The Astrophysical Journal*, 703, 2259
- Dumoulin V., Visin F., 2016, arXiv preprint arXiv:1603.07285
- Eatough R. P., Molkenthin N., Kramer M., Noutsos A., Keith M., Stappers B., Lyne A., 2010, *Monthly Notices of the Royal Astronomical Society*, 407, 2443
- Goodfellow I., 2016, arXiv preprint arXiv:1701.00160
- Goodfellow I. J., Pougetabadie J., Mirza M., Xu B., Wardefarley D., Ozair S., Courville A., Bengio Y., 2014, pp 2672–2680
- Guo P., Lyu M. R., 2004, *Neurocomputing*, 56, 101
- Guo P., et al., 2017, in *Proceedings of the 2017 International Conference on Systems, Man and Cybernetics*. pp 245–250
- Hjelm R. D., Jacob A. P., Che T., Cho K., Bengio Y., 2017, arXiv preprint arXiv:1702.08431
- Ioffe S., Szegedy C., 2015, arXiv preprint arXiv:1502.03167
- Isola P., Zhu J.-Y., Zhou T., Efros A. A., 2016, arXiv preprint arXiv:1611.07004
- Keith M., Eatough R., Lyne A., Kramer M., Possenti A., Camilo F., Manchester R., 2009, *Monthly Notices of the Royal Astronomical Society*, 395, 837
- Kingma D., Ba J., 2014, arXiv preprint arXiv:1412.6980
- Krizhevsky A., Sutskever I., Hinton G. E., 2012, pp 1097–1105
- Lecun Y., Bottou L., Bengio Y., Haffner P., 1998, *Proceedings of the IEEE*, 86, 2278
- Ledig C., et al., 2016, arXiv preprint arXiv:1609.04802
- Lee K., et al., 2013, *Monthly Notices of the Royal Astronomical Society*, p. stt758
- Luc P., Couprie C., Chintala S., Verbeek J., 2016, arXiv preprint arXiv:1611.08408
- Lyon R., Stappers B., Cooper S., Brooke J., Knowles J., 2016, *Monthly Notices of the Royal Astronomical Society*, 459, 1104
- Maas A. L., Hannun A. Y., Ng A. Y., 2013, in *Proc. ICML*.
- Manchester R. N., et al., 2001, *Monthly Notices of the Royal Astronomical Society*, 328, 17
- Mao X., Li Q., Xie H., Lau R. Y., Wang Z., 2016, arXiv preprint arXiv:1611.04076
- Morello V., Barr E., Bailes M., Flynn C., Keane E., van Straten W., 2014, *Monthly Notices of the Royal Astronomical Society*, 443, 1651
- Nair V., Hinton G. E., 2010, in *Proceedings of the 27th international conference on machine learning (ICML-10)*. pp 807–814
- Nan R., et al., 2011, *International Journal of Modern Physics D*, 20, 989
- Nowozin S., Cseke B., Tomioka R., 2016, in *Advances in Neural Information Processing Systems*. pp 271–279
- Qi G.-J., 2017, arXiv preprint arXiv:1701.06264
- Radford A., Metz L., Chintala S., 2015, arXiv preprint arXiv:1511.06434
- Rumelhart D. E., Hinton G. E., Williams R. J., 1988, *Cognitive modeling*, 5, 1
- Salimans T., Goodfellow I., Zaremba W., Cheung V., Radford A., Chen X., 2016, in *Advances in Neural Information Processing Systems*. pp 2226–2234
- Schawinski K., Zhang C., Zhang H., Fowler L., Santhanam G. K., 2017, arXiv preprint arXiv:1702.00403
- Smits R., Lorimer D., Kramer M., Manchester R., Stappers B., Jin C., Nan R., Li D., 2009, *Astronomy & Astrophysics*, 505, 919
- Springenberg J. T., Dosovitskiy A., Brox T., Riedmiller M., 2014, arXiv preprint arXiv:1412.6806
- Thornton D., 2013, PhD thesis, University of Manchester, Manchester, UK
- Wang K., Guo P., Luo A., et al., 2017, *Monthly Notices of the Royal Astronomical Society*, 465, 4311
- Xu B., Wang N., Chen T., Li M., 2015, arXiv preprint arXiv:1505.00853
- Zhu W., et al., 2014, *The Astrophysical Journal*, 781, 117

APPENDIX A: SOME DETAILS ABOUT DCGAN

Here we describe some technical details of the deep convolutional generative adversarial networks.

A1 Development of Generative Adversarial Networks

GAN is a very popular and exciting framework for estimating generative models (data distributions) proposed by Goodfellow et al. (2014); Goodfellow (2016). GAN is learned via an adversarial process, in which two models are trained simultaneously: a generative model that fitting the data distribution and a discriminator model that predicts the probability of a sample coming from the training data rather than a fake synthesized example of the generative model. Since then, it attracted a lot of researchers' attentions. Salimans et al. (2016) applied GAN to semi-supervised learning and generation of images that people find visually realistic. They achieved states-of-the-art results on semi-supervised classification tasks and high quality synthesized images according to a visual Turing test. Ledig et al. (2016) adopted GAN to improve the results of single image super-resolution. Isola et al. (2016) tried to do image-to-image translation where one kind of image is transformed into another kind (such as black-white to color image) with GAN framework. Luc et al. (2016) achieved improved semantic segmentation accuracy on Stanford Background and PASCAL VOC 2012 datasets with adversarial training. They trained a convolutional semantic segmentation network along with an adversarial network that determines whether segmentation maps coming the ground truth or the segmentation network. Schawinski et al. (2017) proposed to recover features of galaxy images via GANs which outperforms conventional deconvolution techniques.

The biggest problem facing GANs is that GANs have been known unstable to train. When that happens, the trained generator will produce nonsensical outputs. Researchers developed a lot of upgraded algorithms to make the training become stable. In classic GAN learning, Jensen-Shannon divergence is used as a default distance measure. Arjovsky et al. (2017) developed the WGAN which take Earth Mover distance. WGAN solved the problem of unstable training, but it is very slower in training to converge. Nowozin et al. (2016) proposed f-GAN which adopts f-divergence to measure distribution distance. Unlike classic GANs choose cross-entropy loss, Mao et al. (2016) proposed Least Square GANs which take mean least square as loss function and Pearson χ^2 divergence as distance measure. Least Square GANs are demonstrated to be faster than WGANs in convergence and more stable than classic GANs. Qi (2017) designed loss-sensitive GANs (LS-GANs) which are trained in a space of Lipschitz continuous functions and take ReLU loss function. Qi also argued that LS-GANs and WGANs can be unified with Generalized LS-GANs (GLS-GANs). They are specific cases of GLS-GANs with different types of loss functions.

Now, more GAN variants have been proposed, such as Hjelm et al. (2017) invented boundary-seeking GANs (BGAN) for adversarial training. At each update iteration, a generator in BGAN is trained to produce samples that is

on the decision boundary of the current discriminator. [Che et al. \(2016\)](#) proposed regularized GAN which can hugely stabilize adversarial training of GANs. [Che et al. \(2017\)](#) designed maximum-likelihood augmented discrete GANs for the application of GANs to discrete settings. [Radford et al. \(2015\)](#) proposed deep convolutional generative adversarial networks (DCGANs) for unsupervised representation learning. They introduced certain architectural constraints and trained DCGANs on image datasets. A hierarchy of representations of images can be obtained in the generator and discriminator of DCGANs, and empirical study demonstrated the feature learning applicability of DCGANs in image classification tasks.

A2 Deep Convolutional Generative Adversarial Networks

GAN provides a framework for adversarial learning. Originally, the specific architectures of generator G and discriminator D are both fully connected feed-forward neural network. In order to build good image representations by training GANs, [Radford et al. \(2015\)](#) proposed deep convolutional generative adversarial networks (DCGANs). After GANs are trained, parts of discriminator and generator are reused as feature extractors for supervised classification task.

The core upgrades of DCGANs over traditional GANs are summarized as follow:

- 1) In discriminator, all neural network layers are strided convolutional layers ([Springenberg et al. 2014](#)). The last convolutional layer is reshaped and fed into a sigmoid output to compute a probability of the input coming from real training data.
- 2) In generator, all neural network layers are fractional -strided convolutional layers ([Springenberg et al. 2014](#)).
- 3) There are no spatial pooling layers in both discriminator and generator. Discriminator learns its own spatial down-sampling with stride convolution layers. Generator learns its own spatial up-sampling with fractional -strided convolution layers.
- 4) There are no fully connected hidden layers in both discriminator and generator.
- 5) In discriminator, LeakyReLU activation function ([Maas et al. 2013](#); [Xu et al. 2015](#)) is used in all strided convolutional layers.
- 6) In generator, ReLU activation function ([Nair & Hinton 2010](#)) is used in all fractional -strided convolutional layers, except that the output layer uses hyperbolic tangent activation function.
- 7) Batch Normalization ([Ioffe & Szegedy 2015](#)) are used in both discriminator and generator to normalize the input of each neuron to have zero mean and unit variance.

A3 Training Techniques for DCGAN

A3.1 Convolutional Layers

Denote an $i \times i$ input two-dimensional image as i^2 -size vector \mathbf{x} by reshaping it from left to right and top to bottom, applying a stride convolutional layer on the input image is equivalent to multiplying a sparse matrix \mathbf{C} of size $o^2 \times i^2$.

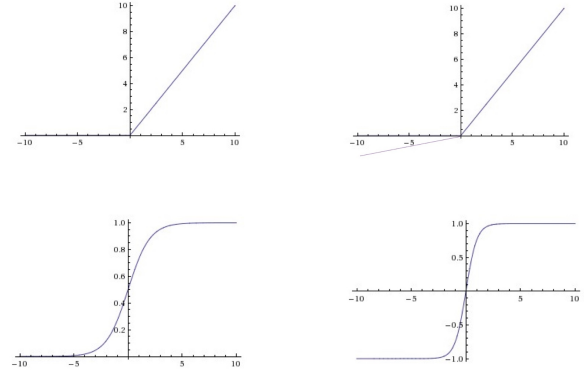


Figure A1. Plots of four related activation functions: ReLU, LeakyReLU, Sigmoid and Tanh (left to right, top to bottom).

Then the output $o \times o$ feature map is reshaped from $\mathbf{y} = \mathbf{C}\mathbf{x}$. The non-zero elements in \mathbf{C} are all from the convolutional filters ([Dumoulin & Visin 2016](#)).

Under this representation, The backward pass is obtained by multiplying \mathbf{C}^T . This backward pass is called a fractional -strided convolutional layer (also named transposed convolutions). The stride convolutional layer learns its own spatial down-sampling. While the fractional -strided convolution layers learns its own spatial up-sampling.

A3.2 Activation Functions

The related activation functions in DCGANs are ReLU, LeakyReLU, Sigmoid and Tanh functions.

ReLU function is adopted in generator's fractional -strided convolutional layers, which is

$$\sigma(x) = \max(0, x). \quad (\text{A1})$$

LeakyReLU function is used in discriminator's strided convolutional layers, which is

$$\sigma(x) = \mathbb{T}(x < 0)(\alpha x) + \mathbb{T}(x \geq 0)(x), \quad (\text{A2})$$

where $\mathbb{T}(\text{true}) = 1$, $\mathbb{T}(\text{false}) = 0$ and the slope, α , is a small constant.

Sigmoid function is used in the output layer of the discriminator to make the predicted probability value in the range of $[0, 1]$:

$$\sigma(x) = \frac{1}{1 + e^{-x}}, \quad (\text{A3})$$

Tanh function is used in the output layer of the generator to make the synthesized data in the range of $[-1, 1]$:

$$\sigma(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \quad (\text{A4})$$

Note that Tanh is a scaled sigmoid function: $\text{Tanh}(x) = 2\text{sigmoid}(2x) - 1$.

A3.3 Batch Normalization

Batch normalization is applied on discriminator and generator layers, except the generator output layer and the discriminator input layer. Batch normalization makes the inputs of

a mini-batch to each neuron have zero mean and unit variance. Denote a mini-batch of activations on some neuron as x_1, \dots, x_m , then they are normalized with,

$$\mu = \frac{1}{m} \sum_{i=1}^m x_i, \quad (\text{A5})$$

$$\sigma^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu)^2, \quad (\text{A6})$$

$$\hat{x}_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \textit{eps}}}, \quad (\text{A7})$$

where *eps* is the least positive real number to ensure the divider larger than zero, \hat{x}_i is the normalized activations in that mini-batch.

According to [Radford et al. \(2015\)](#), batch normalization helps to get deep generators to begin learning and prevent the generator from collapsing all samples to a single data point.

Additional, the source code of implementation of DCGAN can be downloaded at https://github.com/Newmu/dcgan_code. And the source code of implementation of L2-SVM can be downloaded at <http://www.csie.ntu.edu.tw/~cjlin/liblinear/>

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.