Spectra-GANs: A New Automated Denoising Method for Low-S/N Stellar Spectra

MINGLEI WU, YUDE BU, JINGCHANG PAN, ZHENPING YI, AND XIAOMING KONG
1School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai 264209, China
2School of Mathematics and Statistics, Shandong University, Weihai 264209, China
Corresponding author: Yude Bu (buyude@sdu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 11873037, Grant U1931209, Grant 11603012, and Grant 11803016, in part by the Young Scholars Program of Shandong University, Weihai, under Grant 2016WHWLJH09, and in part by the China Postdoctoral Science Foundation under Grant 2015M571124.

ABSTRACT Numerous spectra can be obtained from sky surveys such as the Sloan Digital Sky Survey and the Large Sky Area Multi-Object Fibre Spectroscopic Telescope. However, a considerable fraction of such spectra, which are also valuable for astronomical research, are of low quality, possessing characteristics such as low signal-to-noise ratio (low-S/N). Principal component analysis is widely used to process these low-S/N spectra, but it is not efficient enough to describe the non-linear properties within the spectra. Wavelets are often used to denoise the low-S/N spectra. However, as is well known, the most optimal wavelet basis for each type of spectra needs to be determined; therefore, wavelet analysis is very difficult to use in practice. Restricted Boltzmann machine is a non-linear algorithm that performs poorly when applied to low-S/N spectra. Denoising Convolutional Neural Networks (DnCNN) is a promising denoiser, however, its performance is unsatisfactory due to the lack of suitable noise model. To better exploit the spectra with low-S/N, we propose a new method that can be used to obtain better denoised spectra when compared to those obtained using other methods. A new method called the Spectra Generative Adversarial Nets (Spectra-GANs) is introduced. Spectra-GANs is simply a feedforward neural network that learns the difference between the input vector and the target by minimizing the loss function. It can be used in spectral denoising. The performance of Spectra-GANs is better than those of other methods with regard to denoising the spectra, especially with regard to extremely low-S/N spectral processing. Thus, Spectra-GANs proposed herein is a suitable alternative to previously used methods in spectral denoising.

INDEX TERMS Stellar spectra, machine learning, denoising.

I. INTRODUCTION
Multi-object spectroscopy has substantially improved the effectiveness of observation, and large data sets have been obtained from modern spectroscopic surveys such as the Sloan Digital Sky Survey (SDSS) and the Large Sky Area Multi-Object Fibre Spectroscopic Telescope (LAMOST or Guo Shoujing Telescope) [1], [2]. Such large data sets provide us with a robust basis to conduct astronomical research. However, among them, there are a considerable fraction of low-quality data. These spectra show significant quality defects such as large noise, unobvious spectral line features, low signal to-noise ratio (low-S/N), continuum anomalies, and splicing anomalies, of which low-S/N spectra account for a great part. There are many reasons that result in low-S/N spectra, such as shooting noise, skylight background, cosmic noise, readout noise, dark current noise, and so on. The processing and analysis of these low-S/N spectra are of great significance for the improvement of spectral utilisation, multi-band cross-validation, and the discovery of rare celestial bodies, and so on. Historically, these spectra were inspected by human. However, a systematic processing by experts is too time-consuming to be a practical proposition for largescale surveys. Therefore, the efficient and automatic processing of these low-S/N spectra is one of the main difficulties to be overcomed.

In this paper, we introduce a new method, the Spectra Generative Adversarial Nets (Spectra-GANs), to denoise low-S/N spectra. Generative Adversarial Nets (GANs), as a promising class of generative models, is one of the most popular...
neural networks. Goodfellow et al. first proposed GANs, and its variants have produced remarkable results in various applications, including image synthesis, semantic image editing, image style conversion, and image super-resolution [3]. Since then, several studies on GANs have been conducted. Mirza et al. proposed the Conditional Generative Adversarial Networks (CGANs), which introduced conditional variables in both the generation model G and the discriminant model D [4]. Gulrajani et al. proposed the Wasserstein GAN, which theoretically analysed the reasons for the difficulty of training traditional GANs and formulated the Wasserstein distance [5]. Zhang et al. proposed a Stacked Generative Adversarial Networks (StackGANs) to generate photo-realistic images conditioned on text descriptions [6]. Because convolutional neural networks have substantial advantages with regard to image processing, GANs, such as LAPGAN [7] and DCGAN [8], also use the structure of convolutional neural networks. Similar to LAPGAN, SimGAN can make images more realistic by using real unlabelled images [9]. To better achieve a style transfer between different domains, Zhu et al. presented an approach called the Cycle-consistent Generative Adversarial Nets (Cycle-GANs), which is used to learn how an image can be translated from a source domain to a target domain [10]. In astronomy, CGANs has been used to recover features in astrophysical images of galaxies beyond the deconvolution limit [11]. Stark et al. also used CGANs trained on galaxy images to recover point sources and host galaxy magnitudes with less systematic error and a lower average scattering [12].

In this study, we developed a method to denoise low-S/N spectra based on GANs. Our method is inspired by CycleGANs. Similar to Cycle-GANs, Spectra-GANs attempts to capture the special characteristics of one data collection and determine how these characteristics could be translated into another data collection. Therefore, Spectra-GANs can convert a spectrum from one representation to another, such as from a low-S/N representation to a high signal-to-noise ratio (high-S/N) representation. To fairly assess our method for denoising stellar spectra, we compare our method with three typical stellar spectral denoising algorithms, namely, Wavelet, RBM, and PCA. In addition, we also compare our method with a promising supervised denoising method called Denoising Convolutional Neural Networks (DnCNN) [6].

The layout of this paper is as follows. In Section III, we provide a brief introduction to Spectra-GANs. In Section IV, we describe the spectral data used in the experiment. In Section V, we compare the performance of Spectra-GANs with those of PCA, Wavelet, RBM, and DnCNN with regard to spectral denoising. In Section VI and section VII, we discuss the limitations and possible applications of Spectra-GANs. Finally, we conclude the paper in Section VIII.

II. RELATED WORK
Principal Component Analysis (PCA) is among the most widely used techniques for processing astronomical data [13]. Deeming first applied PCA to process astronomical spectra [14]. Since then, PCA has proven to be a viable tool in astronomy, for example, in the classification and detection of atypical spectra by reducing the dimensionality of the original spectral data [15]–[22]. In addition, it is also widely used for processing low-quality spectra such as incomplete data. Connolly et al. used PCA to address the problem of incomplete and noisy galaxy spectra and demonstrated that this method could derive an optimal interpolation that reconstructs the underlying galaxy spectral energy distributions in the regions of the missing data [23]. PCA is also applied to reconstruct the missing regions in the spectra for the gap correction of quasar and galaxy spectra. They first fixed the gaps of the spectra by other methods and then constructed a set of eigenspectra from the gap-repaired spectra. Finally, the gaps in the original spectra were corrected via the linear combination of the eigenspectra [24]. Re Fiorentini et al. applied PCA to accurately estimate stellar atmospheric parameters and found that it could act as an effective filter for denoising and recovering missing fluxes of the spectra [25].

In addition to PCA, many other methods have also been used in the pre-processing of astronomical spectra. Lu et al. used wavelets to extract features for denoising the stellar spectra by removing higher frequency components [26]. Bu et al. proposed a new method called the Restricted Boltzmann machine (RBM) as a substitute to PCA for repairing incomplete spectra, spectral denoising, and spectral dimensionality reduction [27]. To improve the efficiency of spectral pre-processing, Wang et al. applied a deep neural network to recover defective spectra [28].

III. METHODS
A. GANs
To elucidate our method more clearly, we first introduce GANs. GANs are generative models consisting of two neural networks, of which one is the generator G and the other is the discriminator D. We assumed the instances x obtained from the true data distribution \( P_{\text{data}}(x) \) to be real and the candidates \( G(z) \) produced by the generator G from some noise \( z \sim P(z) \) to be fake. The training goal of the generator G was to maximise the discrimination accuracy of the discriminator D. In contrast, the training objective of the discriminator D is to maximise its discriminative accuracy. That is, the discriminator D tries to classify the instances x obtained from the true data distribution \( P_{\text{data}}(x) \) as real and the synthetic data \( G(z) \) as fake, whereas the generator G tries to make the discriminator D to classify \( G(z) \) as real, as shown in Fig. 1. The loss function was formulated as follows:

\[
\min_G \max_D V(D, G) = E_{x \sim P_{\text{data}}(x)}[\log D(x)] + E_{z \sim P(z)}[\log(1 - D(G(z))].
\]

We trained D to maximise the probability of assigning the correct label to both training examples and samples from G.
B. CYCLE-GANs

In contrast with GANs, Cycle-GANs combine two GANs in different directions into one. Thus, it has two generators G: \( X \to Y \) and F: \( Y \to X \) as well as discriminators \( D_X \) and \( D_Y \), where \( D_X \) aims to distinguish between images \( [X] \) and translated images \( [F(Y)] \) and \( D_Y \) aims to discriminate between images \( [Y] \) and \( [G(X)] \). The adversarial loss consists of two parts:

\[
L_{GAN}(G, D_Y) = E_{x \sim p_{data}(x)}[\log D_Y(x)] \\
+ E_{y \sim p_{data}(y)}[\log(1 - D_Y(G(x)))] \quad (4)
\]

and

\[
L_{GAN}(F, D_X) = E_{x \sim p_{data}(x)}[\log D_X(x)] \\
+ E_{y \sim p_{data}(y)}[\log(1 - D_X(F(y)))] \quad (5)
\]

Meanwhile, to further regularize the generators, Cycle-GANs introduce two cycle consistency losses that one domain should be back to the initial conditions after experiencing two different generators: (a) forward cycle consistency loss: \( x \to G(x) \to F(G(x)) \approx x \), and (b) backward cycle consistency loss: \( y \to G(y) \to F(G(y)) \approx y \). These two cycle consistency losses can be combined into one function:

\[
L_{cyc}(G, F) = E_{x \sim p_{data}(x)}[\|F(G(x)) - x\|_1] \\
+ E_{y \sim p_{data}(y)}[\|G(F(y)) - y\|_1]. \quad (6)
\]

Then, the full objective function is

\[
L(G, F, D_X, D_Y) = L_{GAN}(G, D_Y) + L_{GAN}(F, D_X) \\
+ \lambda L_{cyc}(G, F) \quad (7)
\]

where \( \lambda \) controls the relative importance of the two objectives.
The generation-consistent loss is shown in Eq. (9). To further reduce the space of possible mapping functions by the generator and to control the forward mapping of the cycle to converge toward the objective function more quickly, we added this generation-consistent loss function (as shown in Fig. 2).

$$L_{gen}(G, F) = E_{x \sim p_{data}(x), y \sim p_{data}(y)}[\|G(x) - y\|_1]$$

(9)

where $x \rightarrow G(x) \approx y$.

Finally, the full loss function is the sum of these three loss functions:

$$L(G, F, D_Y) = L_{GAN}(G, D_Y, X, Y) + \lambda_1 L_{cyc-1}(G, F) + \lambda_2 L_{gen}(G, F)$$

(10)
of which $\lambda_1$ and $\lambda_2$ are hyper-parameters that are equal to each other in this study.

Therefore, compared to Cycle-GANs, the tasks of the discriminator remained unchanged; however, the generator was tasked to not only fool the discriminator but also to be near the ground truth output in an L1 sense.

### IV. DATA

LAMOST is a quasi-meridian reflecting Schmidt telescope located in the Xinglong Station of the National Astronomical Observatories, Chinese Academy of Sciences. It is a special reflecting Schmidt telescope with 4000 fibres and can be used to obtain spectra of celestial objects as well as sky background and calibration sources. The spectra used in this study are from the fifth data release of LAMOST (LAMOST DR5) [30], [31]. Up until July 2017, LAMOST had completed its first five years of regular surveys that began in September 2012. These spectra have been released through LAMOST DR5. After this six-year survey, we obtained a total of 8952297 spectra, including those of 7930178 stars, 152608 galaxies, 50 132 quasars, and 819379 unknown objects. These spectra have different S/N. In LAMOST, the definition of S/N is shown below:

$$S/N = \frac{\sum_{i=1}^{N} Flux \cdot \phi_n}{N}$$

(11)

where $Flux_n$ represents the spectral flux, $\phi_n$ stands for the inverse variance of one flux point and $N$ is the number of the flux points. In LAMOST, the calculation of the inverse variance require error propagation of many variables, which is difficult to be acquired in deep learning, so the S/N is not used as a metric to evaluate the denoising effect in this paper.

The data used in this experiment were divided into three groups: D1, which consisted of spectra with $10 \leq S/N \leq 15$ and with $S/N \geq 50$; D2, which consisted of spectra with $5 \leq S/N \leq 10$ and with $S/N \geq 50$; and D3, which consisted of spectra with $2 \leq S/N \leq 5$ and with $S/N \geq 50$ (see Table 1). Note that each group consisted of two parts, the first part consisted of high-S/N spectra and the second part consisted of low-S/N spectra. For each high-S/N spectrum in the first part, there was a corresponding low-S/N spectrum in the second part. These were the spectra for the same celestial body observed by LAMOST at different times (see e.g. Table 2 and Fig. 3). The difference between D1, D2, and D3 was that the spectra in the second part covered a different S/N range. The performance of our method on different datasets demonstrated the ability of our method to process spectra with different S/Ns.

The resolution of the LAMOST spectra was $R \sim 1800$, and the wavelength coverage was $3700 \sim 9100$ Å. We only used the wavelength range between $4000$ Å and $8096$ Å, which covered important spectra lines that were necessary for our analysis. We did not take any further steps such as continuum subtraction to pre-process the spectra except normalising the fluxes of each spectrum to $[0, 1]$ using the following equation

$$x_{norm} = \frac{x - \min(x)}{\max(x) - \min(x)}$$

(12)

where $x = [x_1, x_2, \ldots, x_n]$ represents a spectrum.

Each of the three aforementioned groups were divided into two subsets: training sets and test sets. The training sets of D1, D2, and D3 contained 6000, 7000, and 5000 spectra, respectively, and all the testing sets consisted of 1000 spectra. During training, the epochs were set to 20000 and the batch size was 5.

### V. EXPERIMENTS AND RESULTS

In the experiment, the reconstructed low-S/N spectra were compared with the corresponding high-S/N ones to verify the denoising effect. Spectra with $S/N \geq 50$ were used as the high-S/N ones, and the corresponding ones with $2 \leq S/N \leq 5$, $5 \leq S/N \leq 10$, and $10 \leq S/N \leq 15$ were used as the low-S/N ones. Each group of data was processed by Spectra-GANs, and as a means of comparison, we also used PCA, Wavelet, RBM, and DnCNN to denoise the spectra. For PCA, we obtained eigenspectra from the test spectra in D1, D2, and D3. Considering the denoising effect, the final eigenspectra used were the first 3 eigenspectra of D1, first 8 eigenspectra of D2, and first 83 eigenspectra of D3, which yielded a relevant variance contribution rate greater than 99% [19]. With regard to the Wavelet, the Haar wavelet is a simple and effective method that is often used in denoising the astronomical spectra [26]. Therefore, we used Haar as the basis of the wavelet. We experimentally found that the 2nd decomposition can well retain the spectral line features and reduce the noise. Therefore, we reconstructed the spectrum with the coefficients on the 2nd decomposition. For the RBM (see Appendix A), the hidden layer vectors were similar to the PC vectors given by PCA. We experimentally set the size of the
hidden layer of RBM to 500 and the number of epochs to 1000 [27]. We set the DnCNN (see Appendix B) parameters similar to those described in [32]. In addition, DnCNN needs pure data and pure noise to train itself, so we use Kurucz synthetic spectra [33] as the pure data and the Gaussian White Noise (WGN) as the pure noise (details are shown in Appendix B).

We applied these five methods on the low-S/N spectra and compared how well the reconstructed spectra matched the original high-S/N spectra both quantitatively and qualitatively.

A. QUALITATIVE COMPARISON

Figure 4, Figure 5, and Figure 6 show the denoised results of Spectra-GANs, PCA, DnCNN, RBM, and Wavelet on the spectra selected from D1, D2, and D3, respectively. In these figures, the five colour spectra represent the denoised results of the five denoising methods. Among these, PCA, Wavelet, and RBM are unsupervised methods, while Spectra-GANs and DnCNN are supervised methods. These figures indicate that our method (the results of which are shown in orange) obtains better denoised results compared to the other methods.

Furthermore, compared with the other methods, our method was able to not only preserve the spectral lines consistently, but also effectively correct inaccurate continuum, such as the undesired continuum slope in Fig. 4. From this point of view, our method can perform many other tasks, such as recovering defective spectra, repairing incomplete data, and so on. We plan to further explore the application of Spectra-GANs in processing other types of low-quality spectra in the near future.

As shown in these figures, PCA was able to reconstruct the most important spectral lines and eliminate most sky-lights, but the width and depth of the spectral lines did not match well. Moreover, PCA could not correct the inaccurate continuum, i.e., the continuum of the reconstructed spectrum was not consistent with the original one. For the DnCNN, the denoised effect of the observed spectrum was slightly worse than the other methods. Although the DnCNN is a promising denoiser, it can be seen from our experiments that it tends to produce over-smooth spectra. The most likely reason is that the DnCNN in our experiment is trained using WGN due to the lack of suitable noise model, whereas real noise of the observed spectra is more complex. Therefore, its performance was not as good as expected. The RBM and Wavelet also did not produce satisfactory denoised effects.

To conclude, our method outperforms other methods. The results suggest that our method may have learned the noise model of the spectra and can effectively convert spectra with low-S/N values to those with high-S/N values.
**FIGURE 4.** Comparison of denoised results obtained using Spectra-GANs, PCA, DnCNN, RBM, and Wavelet from D1. The green line represents the original low-S/N spectra, the blue line represents the original high-S/N spectra, the orange line represents the denoised spectra using Spectra-GANs, the red line represents the denoised spectra using PCA, the purple line represents the denoised spectra using DnCNN, the brown line represents the denoised spectra using RBM, and the pink line represents the denoised spectra using Wavelet. The figure shows that all methods can recover the most important lines, whereas only Spectra-GANs can better correct the inaccurate continuum, such as the undesired continuum slope.

**FIGURE 5.** This is the same as Fig. 4, except the results are of spectra from D2. The figure shows that Spectra-GANs can perfectly recover these line intensity and continuum. PCA was able to recover the most important spectral lines, but their width and depth did not match well. DnCNN tended to produce over-smooth spectra. RBM and Wavelet could not denoise the spectra.

**B. QUANTITATIVE COMPARISON**

For the quantitative analysis, we measured the mean absolute error (MAE) of the Lick/IDS index and the full spectrum between the original high-S/N spectrum and the reconstructed high-S/N spectra. The MAE was defined as follows:

\[
MAE = \frac{1}{M} \sum_{m=1}^{M} |e|
\]

wherein, \(e\) is the residual of the corresponding flux (Lick indices) from the original high-S/N spectrum and the reconstructed high-S/N spectra; \(M\) represents the number of flux points in the spectrum. The MAE is the average difference between the original high-S/N spectrum and the reconstructed high-S/N spectra. Hence, when the MAE was high, the reconstructed spectrum was significantly different from the original spectrum, and the performance of the method was unsatisfactory; otherwise, the performance was considered good. Hence, the value of
FIGURE 6. This is the same as Fig. 4, except the results are of spectra from D3. As described in Fig. 5, Spectra-GANs performed better than the other methods in the cases of extremely low-S/N spectra.

FIGURE 7. MAE values of the Lick indices obtained by Spectra-GANs, PCA, DnCNN, RBM, and Wavelet from D1, D2, and D3. The figures show that our method performed significantly better than other methods in terms of Lick index recovery, especially for Balmer lines (Hβ, HδA, HγA, HδF, and HγF); however, PCA performed slightly better than Spectra-GANs for three values (Fe4531, Fe4668, and Fe5709) of spectra with 5 ≤ S/N ≤ 10.

the MAE can be used to measure the performance of our method.

1) PERFORMANCE ON LICK INDEX
In this subsection, we compare the performance of Spectra-GANs, PCA, DnCNN, RBM, and Wavelet on recovering the Lick index. The Lick/IDS index system is a set of absorption-line indices, defined by the Lick group for low-resolution (R ~ 8) spectra with wavelength coverage between the range of 4000 Å and 6400 Å [34]–[39], which is often used as an approach to measure the atmospheric parameters in stars and galaxies. Barbuy et al. explored the relations between the Lick indices and the stellar parameters Teff, log g, [Fe/H], and [α/Fe] [40]. Thomas et al. derived a rough estimate of [α/Fe] of a galaxy using the relations between Lick indices and [α/Fe] [41]. Graves et al. used Lick indices to measure the individual abundance ratios for unresolved stellar populations by using the different sensitivities of various Lick indices to the elemental abundances [42]. Thomas et al. studied the relationship between the Lick index and [α/Fe] [43].

In this experiment, we first denoised the spectra with Spectra-GANs, PCA, DnCNN, RBM, and Wavelet, and then computed the Lick indices of the denoised spectra. The derived Lick indices were then compared with those derived
from the corresponding high-S/N spectra, and the MAE between these two Lick indices was used to measure the performance of different methods in denoising the spectra. The wavelength definitions of the Lick indices are shown in Table 3. Figure 7 shows the MAEs for the three test sets D1, D2, and D3 derived with Spectra-GANs and the other four methods. The plots show that Spectra-GANs performs significantly better than other methods in terms of Lick index recovery, especially for the Balmer lines (Hβ, HδA, HγA, HδF, and HγF). However, the PCA performs slightly better than Spectra-GANs when recovering Fe4531, Fe4668, and Fe5709 from spectra with $5 \leq S/N \leq 10$. Thus, Spectra-GANs was able to better recover spectral lines than other methods even for extremely low-S/N spectra ($2 \leq S/N \leq 5$).

2) PERFORMANCE ON THE FULL SPECTRUM

In this subsection, we compare the full spectra denoised by Spectra-GANs with those denoised by PCA, DnCNN, RBM, and Wavelet. We first denoised the spectra with different methods and then computed the MAE between the denoised spectra and the corresponding high-S/N spectra. Evidently, the method with smaller MAE values performed better than that with larger MAE values. The results are shown in Fig. 8, Fig. 9, and Fig. 10. As indicated in these figures, our method presents the smallest mean and variance values compared to other methods, and the DnCNN produces the largest ones due to the lack of suitable noise model. Furthermore, the mean and variance values of the other three methods are almost two times larger than those obtained with our method.
M. Wu et al.: Spectra-GANs: New Automated Denoising Method for Low-S/N Stellar Spectra

FIGURE 10. This is the same as Fig. 8, except the results are of spectra from D3.

FIGURE 11. The green line represents the original low-S/N spectra, the blue line represents the original high-S/N spectra, the orange line represents the denoised spectra using Spectra-GANs. Four example spectra on which the Spectra-GANs failed are shown. The reason for the failure in the top panel was that there were too many skylights in some low-S/N spectra. The reason for the failure in the bottom panel was the fault correspondence; that is, the high-S/N spectra and the corresponding low-S/N spectra (as shown in Table 2) were not from the same star.

In general, the MAE obtained by Spectra-GANs was significantly smaller and more concentrated than the MAE obtained by other methods, indicating that the denoised spectrum obtained by Spectra-GANs was closer to the corresponding true high-S/N one in comparison with other methods.

C. ANALYSIS OF TIME COMPLEXITY

In this subsection, we analyze the time complexity of Spectra-GANs, PCA, DnCNN, RBM, and Wavelet. The experiments were carried out in the PyCharm 2019.2.4 environment running on a workstation with an Inter(R) Core(TM) i9-9900X (3.50 GHz) CPU with 64 GB memory and Windows 10 system.

We calculated the denoising time of test sets in D1, D2 and D3 respectively, each of which contains 1000 stellar spectra. From Table 4, we could conclude that the time complexity of our method is almost equal to the DnCNN, inferior than PCA, RBM and Wavelet. The reason is that as the deep learning algorithms, the models of the Spectra-GANs and DnCNN are more complex, so their time complexity is inferior to other three methods. In addition, for the Spectra-GANs and DnCNN, we need to train them before denoising, and the time complexity of training is much higher than that of denoising process. However, in the processing of astronomical spectra,
TABLE 3. Lick index category. In the third column, 1 represents the molecular bands and 0 represents the atomic indices.

| No. | Name | Label |
|-----|------|-------|
| 1   | CN   | 1     |
| 2   | CN₂  | 1     |
| 3   | Ca4227 | 0   |
| 4   | G300 | 0     |
| 5   | Fe4383 | 0   |
| 6   | Ca4255 | 0   |
| 7   | Fe4531 | 0   |
| 8   | Fe4668 | 0   |
| 9   | Hβ   | 0     |
| 10  | Fe5015 | 0   |
| 11  | Mg₁  | 1     |
| 12  | Mg₂  | 1     |
| 13  | Mg₃  | 0     |
| 14  | Fe5270 | 0   |
| 15  | Fe5335 | 0   |
| 16  | Fe5406 | 0   |
| 17  | Fe5709 | 0   |
| 18  | Fe5782 | 0   |
| 19  | NaD  | 0     |
| 20  | TiO₁ | 1     |
| 21  | TiO₂ | 1     |
| 22  | Hα   | 0     |
| 23  | Hβ   | 0     |
| 24  | Hγ   | 0     |
| 25  | Hγ P | 0     |

TABLE 4. The denoising time (in seconds) of the five methods for spectra with different S/Ns.

| Data Set Name   | Methods     | Time(s) |
|-----------------|-------------|---------|
| Spectra-GANs    | PCA         | 0.21    |
| Test set of D₁  | DnCNN       | 11.28   |
|                 | RBM         | 0.091   |
|                 | Wavelet     | 0.59    |
| Spectra-GANs    | PCA         | 0.22    |
| Test set of D₂  | DnCNN       | 11.52   |
|                 | RBM         | 0.095   |
|                 | Wavelet     | 0.63    |
| Spectra-GANs    | PCA         | 0.37    |
| Test set of D₃  | DnCNN       | 11.87   |
|                 | RBM         | 0.098   |
|                 | Wavelet     | 0.65    |

* Time refers to the total denoising time of 1000 spectra.

** Each test set contains 1000 spectra.

VI. FAILURE ANALYSIS

Although in most cases Spectra-GANs performed better than other methods in denoising the spectra, it failed in some spectra. In this section, we show some examples where Spectra-GANs faced difficulties. There were three major failure cases:

(1) As shown in the top panels of Fig. 11, there were too many skylights in some low-S/N spectra. The randomness of these skylights was beyond the scope of the effective input, which led to reconstruction failure;

(2) As shown in the bottom panels of Fig. 11, there were some mismatches; that is, the high-S/N spectra and the corresponding low-S/N spectra (as shown in Table 2) were not from the same star. This case could have led to certain training and test errors;

(3) In addition to these two cases, Spectra-GANs may fail on rare objects that were absent or low in number in the training set, because the performance of our method depended on the training set. Spectra-GANs was unable to reconstruct features that were absent in the training set. It is worth noting that although Spectra-GANs did not perform well on some spectra, such cases were very rare. More examples of the results are included in Appendix C, which demonstrate that in most of cases, our proposed method works well in denoising the spectra.

VII. DISCUSSION

The proposed method learns how to derive characteristics from a collection of spectra (such as those with low-S/Ns) and determines how these characteristics can be translated into other types of spectra (such as those with high-S/Ns).

The results demonstrate that Spectra-GANs can effectively denoise the stellar spectra with low-S/Ns from LAMOST, especially stellar spectra with extremely low-S/Ns (2 ≤ S/N ≤ 5). Our method can also be applied to process other types of low-quality spectra, such as defective spectra, incomplete spectra, and so on. We also expect our method to be applied to spectra from other spectroscopic surveys such as the SDSS and 2dF. Although our method works very well when dealing with low-S/N spectra, there are still some shortcomings. The main limitation of our method is that the training set ultimately limits the ability of Spectra-GANs to recover features. We trained spectra in pairs and applied the trained model to spectra in a similar S/N range. If we intend to apply our method to spectra in another S/N range, we should train the model using spectra in similar S/N ranges with an appropriate epoch. Another limitation of our method is that sophisticated details such as weak spectral lines are impossible to recover perfectly, as the probability model is complex. Furthermore, if rare objects are absent from the training set, the method may fail, as illustrated in Section VI.

VIII. CONCLUSION

In this study, we propose a method called Spectra-GANs to automatically improve the spectral quality of LAMOST spectra. Our method does not assume any prior knowledge and neither does it require any manual intervention. Hence, it is not task-specific but is rather a general-purpose method that can be used for different spectral processing tasks. As verified by the experimental results, compared to
the state is determined according to statistical rules. RBM is of a neuron has only two states (inactive and activated), which (Fig. 12). Its neuron is a kind of random neuron. The output tions within the layers and is fully connected between layers and a hidden layer. This neural network has no connec-

In this section, we provide a brief introduction to RBM. RBM takes a longer time for training. For example, PCA takes less time to extract principal components as well as solve many other scientific challenges. Thus, Spectra-GANs also has limits. For example, PCA takes less time to extract principal components as well as solve many other scientific challenges. Thus, Spectra-GANs also has limits. For

APPENDIX A

RBM

In this section, we provide a brief introduction to RBM. RBM is a stochastic neural network rooted in statistical mechanics, which is a two-layered structures composed of a visible layer and a hidden layer. This neural network has no connections within the layers and is fully connected between layers (Fig. 12). Its neuron is a kind of random neuron. The output of a neuron has only two states (inactive and activated), which are generally represented through binary 0 and 1. The value of the state is determined according to statistical rules. RBM is an energy-based model. The energy of the joint configuration of the visible variable $v$ and the hidden variable $h$ is:

$$E(v, h; \theta) = - \sum_{ij} W_{ij} v_i h_j - \sum_i b_i v_i - \sum_j a_j h_j$$

(14)

where $\theta$ represents the parameter \{W, a, b\}, $W$ is the weight of the edge between the visible and hidden units, and $b$ and $a$ are the biases of the visible and hidden units, respectively. With the energy of the joint configuration of $v$ and $h$, we can get the joint probability of $v$ and $h$:

$$P_\theta(v, h) = \frac{1}{Z(\theta)} \exp(-E(v, h; \theta))$$

(15)

where $Z(\theta)$ is the normalization factor, also known as the partition function. According to Eq. (14), Eq. (15) can be written as:

$$P_\theta(v, h) = \frac{1}{Z(\theta)} \exp(\sum_{i=1}^{D} \sum_{j=1}^{F} W_{ij} v_i h_j + \sum_{i=1}^{D} v_i b_i + \sum_{j=1}^{F} a_j h_j)$$

(16)

The likelihood function $P(v)$ of the observed data is max-
imized. $P(v)$ can be obtained by calculating the marginal distribution of $P(v, h)$ against $h$ in Eq. (16):

$$P_\theta(v) = \frac{1}{Z(\theta)} \sum_h \exp[v^T W h + a^T h + b^T h]$$

(17)

The parameters of the RBM are obtained by maximizing $P(v)$, which is equivalent to maximizing $\log(P(v)) = L(\theta)$:

$$L(\theta) = \frac{1}{N} \sum_{n=1}^{N} \log P_\theta(v^{(n)})$$

(18)

APPENDIX B

DnCNN

In this section, we give a brief introduction to DnCNN and its training sets.

DnCNN is a deep convolutional neural network for Gaus-
sian denoising, which uses residual learning as well as batch normalization to utilize fast convergence and good perfor-

dance for deep network in image restoration. This neural network has 17 layers (Fig. 13), and the structure is explained as follows:

Conv + ReLU: In the first layer, the input is an image of size $35 \times 35 \times c$. Here, $c$ is the number of image channels. After 64 filters of size $3 \times 3 \times c$ and rectified linear units(ReLU) are used, the output is 64 feature maps of size $35 \times 35$. (Because padding is added, the size of the input does not change.)

Conv + BN + ReLU: From 2 to $(d - 1)$ layers, $d$ represents the network depth. There are 64 filters of size $3 \times 3 \times 64$; therefore, the input and output of these layers are feature maps of size $35 \times 35 \times 64$. The batch normalization is used between convolution and ReLU.

Conv: In the last layer, the $c$-dimensional image is recon-
structed as output through $c$ filters of size $3 \times 3 \times 64$.

BN represents the Batch Normalization, which normal-
izes the activations of the previous layer at each batch, i.e. applies a transformation that maintains the mean activation close to 0 and the activation standard deviation close to 1. Therefore, the goal of BN is to achieve a stable distribution of activation values throughout training, which can significantly accelerate the convergence and avoid gradient explosion. ReLU is on behalf of the Rectified Linear Unit, which is a activation function and shown in Eq. (19). Rectifier units help to find better minima than other activation functions during training [44].

$$f(x) = \max(0, x)$$

(19)

In addition, the receptive field size of the network is $(2d + 1) \times (2d + 1)$, that is, $35 \times 35$. Therefore, the corresponding network depth $d$ is 17. The input is $y = x + v$, where $y$ is the noise observation, $x$ is the clean image, and $v$ is the residual image (noisy image). The network expects to learn $R(y) = v$, that is, noisy images. The loss function is the average mean square error between the predicted residual image and the true
residual image:

\[ L(\theta) = \frac{1}{2N} \sum_{i=1}^{N} \| R(y_i; \theta) - (y_i - x_i) \|_F^2 \]  

(20)

Here, \( \theta \) represents the trainable parameters to be learned, and \( N \) represents the number of train images.

DnCNN needs pure data and pure noise to train itself. The pure data used in this paper are synthetic spectra from the KURUCZ model \( [33] \), which are calculated using the SPECTRUM program, and the resolution of the spectra is convolved to 2000, which is basically consistent with the resolution of the LAMOST spectra. The wavelength range is 3000Å~10000Å. In order to facilitate the subsequent processing of the Spectra-GANs, we have selected the wavelength ranging from 4000Å to 8095Å.

The stellar spectra contain many sources of noise, such as shooting noise, skylight background, cosmic noise, readout noise, dark current noise, and so on, so its noise model is very complicated. In order to simplify the calculation, WGN is commonly used as the noise model in the stellar spectra, so the pure noise used in this paper is also WGN. Therefore, in the DnCNN, the pure data set and pure noise set consist of synthetic spectrum and WGN, respectively, and the synthetic low-S/N spectrum is the superposition of synthetic spectrum and WGN (as shown in Fig. 14). The S/N of the synthetic spectrum is calculated by the following formula:

\[ \frac{S}{N} = \frac{Flux}{WGN} \]  

(21)

so the calculation of the WGN under a certain S/N is as follows:

\[ WGN = \frac{Flux}{S/N} \]  

(22)

where \( Flux \) represents a one-dimensional random array with Gaussian distribution. In order to keep the data distribution consistent with the training sets of Spectra-GANs (as shown in Table 1), we build three data sets for training DnCNN (as shown in Table 5).

### APPENDIX C

**LICK/IDS INDEX**

In this section, we briefly introduce the Lick index. The Lick/IDS index system measures 25 optical absorption features, 19 of which are atomic indices and 6 of which are molecular bands (as shown in Table 3). Each index was measured according to the following scheme: Two pseudo-continuum bandpasses were defined on each side of a central bandpass. A line representing the continuum between the midpoints of the two flanking bandpasses was drawn, and the flux difference between this line and the central bandpass flux determined the index.

There are two kinds of indices, i.e., molecular bands and atomic features. Molecular bands were expressed in magnitudes, while atomic features were expressed in angstroms of equivalent width. In the pseudo-continuum \( (P) \) bandpass, we calculated the average bandpass flux from the spectrum:

\[ F_P = \int_{\lambda_1}^{\lambda_2} F_{\lambda} \, d\lambda / (\lambda_2 - \lambda_1) \]  

(23)

A molecular band measured in magnitudes was then

\[ EW = \int_{\lambda_1}^{\lambda_2} (1 - \frac{F_{\lambda}}{F_{C\lambda}}) \, d\lambda \]  

(24)
and an equivalent width as the atomic feature was

\[ \text{Mag} = -2.5 \log \left( \frac{1}{\lambda_2 - \lambda_1} \int_{\lambda_1}^{\lambda_2} \frac{F_{I\lambda}}{F_{C\lambda}} d\lambda \right) \]  

(25)

where \( F_{I\lambda} \) and \( F_{C\lambda} \) were the fluxes per unit wavelength in the index passband and the straight-line continuum flux in the index passband, respectively.

**APPENDIX D**

**DENOISING EXAMPLES**

In this section, we randomly present some denoising examples using our method and other methods in Fig. 15, Fig. 16 and Fig. 17.

**ACKNOWLEDGMENT**

This research has made use of data products from LMOST. The Guo Shou Jing Telescope (the Large Sky Area Multi-Object Fiber Spectroscopic Telescope, LAMOST) is a National Major Scientific Project built by the Chinese Academy of Sciences. Funding for the project has been provided by the National Development and Reform Commission. LAMOST is operated and managed by National Astronomical Observatories, Chinese Academy of Sciences.

**REFERENCES**

[1] D. G. York et al., “The Sloan digital sky survey: Technical summary,” Astron J., vol. 120, p. 1579, Mar. 2000.

[2] X. Q. Cui et al., “The large sky area multi-object fiber spectroscopic telescope (LMOST),” Res. Astron. Astrophys., vol. 12, no. 9, pp. 1197–1242, Aug. 2012.

[3] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” 2014, arXiv:1406.2661. [Online]. Available: http://arxiv.org/abs/1406.2661

[4] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” 2014, arXiv:1411.1784. [Online]. Available: http://arxiv.org/abs/1411.1784

[5] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, “Improved training of wasserstein GANs,” 2017, arXiv:1704.00028. [Online]. Available: http://arxiv.org/abs/1704.00028

[6] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. Metaxas, “StackGAN: Text to photo-realistic image synthesis with stacked generative adversarial networks,” 2016, arXiv:1612.03242. [Online]. Available: http://arxiv.org/abs/1612.03242

[7] E. Denton, S. Chintala, A. Szlam, and R. Fergus, “Deep generative image models using a Laplacian pyramid of adversarial networks,” 2015, arXiv:1506.05751. [Online]. Available: http://arxiv.org/abs/1506.05751

[8] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” 2015, arXiv:1511.06434. [Online]. Available: http://arxiv.org/abs/1511.06434

[9] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, “Learning from simulated and unsupervised images through adversarial training,” 2016, arXiv:1612.07828. [Online]. Available: http://arxiv.org/abs/1612.07828

[10] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” 2017, arXiv:1703.10593. [Online]. Available: http://arxiv.org/abs/1703.10593

[11] K. Schawinski, C. Zhang, H. Zhang, L. Fowler, and G. K. Santhanam, “Generative adversarial networks recover features in astrophysical images of galaxies beyond the deconvolution limit,” Monthly Notices Roy. Astronomical Soc., Lett., vol. 467, pp. L110–L114, Jan. 2017.

[12] D. Stark, B. Launet, K. Schawinski, C. Zhang, M. Koss, M. D. Turp, L. F. Sartori, H. Zhang, Y. Chen, and A. K. Weigel, “PSFGAN: A generative adversarial network system for separating quasar point sources and host galaxy light,” Monthly Notices Roy. Astronomical Soc., vol. 477, no. 2, pp. 2513–2527, Jun. 2018.

[13] I. T. Jolliffe, *Principal Component Analysis*. Berlin, Germany: Springer, 1986.

[14] T. J. Deeming, “Stellar spectral classification: I. Application of component analysis,” Monthly Notices Roy. Astronomical Soc., vol. 127, no. 6, pp. 493–516, May 1964.
ZHENPING YI received the B.S. and M.S. degrees from Shandong University, China, in 2002 and 2005, respectively, and the Ph.D. degree from the University of Chinese Academy of Sciences, China, in 2015. She is currently an Associate Professor with the School of Mechanical, Electrical and Information Engineering, Shandong University. Her research interests include big data and stellar spectra data analysis.

XIAOMING KONG received the B.S. degree from the Shandong University of Technology, China, in 2002, the M.S. degree from Shanghai University, China, in 2005, and the Ph.D. degree from the University of Chinese Academy of Sciences, China, in 2018. She is currently an Associate Professor with the School of Mechanical, Electrical and Information Engineering, Shandong University. Her research interests include big data processing, data mining, and stellar spectral analysis.