Compression Method for Solar Polarization Spectra Collected from Hinode SOT/SP Observations

Jargalmaa Batmunkh[a], Yusuke Iida[a], Takayoshi Oba[b], Haruhisa Iijima[c]

[a]*Niigata University, 8050 Ikarashi 2-no-cho, Nishi-ku, Niigata, 950-2181, Niigata, Japan*
[b]*Max Planck Institute for Solar System Research, Justus-von-Liebig-Weg 3, Göttingen, 37077, Germany*
[c]*Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8601, Aichi, Japan*

Abstract

The complex structure and extensive details of solar spectral data, combined with a recent surge in volume, present significant processing challenges. To address this, we propose a deep learning-based compression technique using deep autoencoder (DAE) and 1D-convolutional autoencoder (CAE) models developed with Hinode SOT/SP data. We focused on compressing Stokes I and V polarization spectra from the quiet Sun, as well as from active regions, providing a novel insight into comprehensive spectral analysis by incorporating spectra from extreme magnetic fields. The results indicate that the CAE model outperforms the DAE model in reconstructing Stokes profiles, demonstrating greater robustness and achieving reconstruction errors around the observational noise level. The proposed method has proven effective in compressing Stokes I and V spectra from both the quiet Sun and active regions, highlighting its potential for impactful applications in solar spectral analysis, such as detection of unusual spectral signals.

Keywords: Solar physics, Solar surface, Spectropolarimetry, Astroinformatics, Neural networks, Dimensionality reduction

Email addresses: `f22l006h@mail.cc.niigata-u.ac.jp` (Jargalmaa Batmunkh), `iida@ie.niigata-u.ac.jp` (Yusuke Iida), `oba@mps.mpg.de` (Takayoshi Oba), `h.iijima@isee.nagoya-u.ac.jp` (Haruhisa Iijima)

1. Introduction

Observational spectral data encapsulates important and varied physical information with a multi-dimensional structure about astronomical bodies, necessitating thorough investigation and analysis for a comprehensive understanding of space. The increase in the number of observatory instruments in recent years has led to a substantial growth in the volume of astronomical data. This surge not only emphasizes the significance of studying such data but also opens up promising opportunities for leveraging deep learning techniques in the processing and analysis of these vast datasets in the big data era. One approach to handling such intricate data is the feature extraction technique, which takes the high-dimensional raw data as input, compresses it, and reconstructs it to the original size. The most important features of the original high-dimensional data are extracted in the compressed part, enabling them to serve as representatives of the original complex dataset in subsequent studies. Autoencoders (LeCun, 1987; Kramer, 1991; Goodfellow et al., 2016) play a powerful role in deep learning-based dimensionality reduction. Through this compression approach, further studies such as anomaly detection (Chen et al., 2018; Ryu et al., 2023) and classification (Gogoi and Begum, 2017; Yeom et al., 2021) of data can also be accomplished.

In the latter part of the 2010s, several studies aimed to develop and apply compression methods for spectral data, particularly in the context of galaxy observations. Portillo et al. (2020) utilized variational autoencoder models to compress galaxy spectra by reducing it to six parameters, offering more accurate reconstructions than principal component analysis (PCA). Melchior et al. (2023) introduced an architecture to represent and generate restframe galaxy spectra from 6 to 10 latent parameters, resulting in accurate reconstructions with superresolution and reduced noise.

When compared to data in other fields of space science, solar spectral data stand out in terms of their increased precision and complexity in higher dimensionality, encompassing details about light polarization, temperature, and the magnetic field on the solar surface. Therefore, processing this type of observational data poses a significant challenge. Previous works referring to the representational dimension of solar polarimetric spectra include, Asensio Ramos (2006)'s two-part minimum description length principle for approximation model selection, which suggests the optimal eigenvector dimension for denoising PCA, and Asensio Ramos et al. (2007)'s intrinsic dimensionality estimation method for spectropolarimetry data. López Ariste and Casini

(2002) implemented a PCA inversion technique using 10 eigenprofiles for a single Stokes profile. A feature extraction technique by Socas-Navarro (2005) for simulated solar profiles, based on a multi-layer perceptron, represents one of the first uses of a neural network for solar spectra, achieving higher accuracy than previous methods such as PCA but requiring significant computational expense. Studies conducted on inversion techniques using deep learning, include Gafeira et al. (2021)'s convolutional neural network-based inversion method for Stokes profiles using Hinode (Kosugi et al., 2007) data. Additionally, Asensio Ramos and Díaz Baso (2019) introduced convolutional neural networks that output thermodynamic and magnetic properties from synthetic Stokes profiles, and achieved a precision comparable to the standard technique. Regarding deep learning-based solar spectral compression, Sadykov et al. (2021) used a fully connected autoencoder to reduce one-dimensional quiet Sun spectra, collected by NASA's IRIS (De Pontieu et al., 2014) satellite, from 110 to 4 in size, achieving an average reconstruction error comparable to the variations in the line continuum.

Upon reviewing previous works, it becomes apparent that compression techniques for observational solar spectra have primarily been developed for one-dimensional spectra related to spatial positions in the quiet Sun. However, active regions cannot be disregarded, as they are associated with a variety of significant solar phenomena—such as solar flares, solar jets, and coronal mass ejections—necessitating thorough study as important regions of interest. This motivates our proposal to develop an efficient compression method for solar polarization spectra, applicable to both the quiet Sun and active regions, by utilizing two-dimensional key polarimetric parameters.

We conduct our study using observational solar spectra from Hinode SOT/SP (Tsuneta et al., 2008; Suematsu et al., 2008; Lites et al., 2013), a collaborative mission of JAXA, NASA, and ESA. This mission has been collecting solar spectro-polarimetric data since 2006, constituting an extensive solar spectral database suitable for our work. Our study introduces the compression of solar spectra through the development of two distinct models: a deep autoencoder (DAE) and a 1D-convolutional autoencoder (CAE). Considering the intricate nature of Stokes profiles characterized by high noise levels, we exclusively focus on Stokes I (total intensity) and Stokes V (circular polarization) selected from the set of four parameters.

In Section 2, we provide a description of the Hinode data, followed in Section 3 by a comprehensive explanation of the methods applied in our study. Sections 4 and 5 present the results and discussion, respectively. In

Section 6, we conclude the paper with a brief summary.

2. Hinode SOT/SP Data

The Solar Optical Telescope (SOT) on Hinode is equipped with a spectropolarimeter (SP) that observes the sun, capturing high-resolution spatial images and spectro-polarimetric data. Its long-term collection of solar spectral data over more than 15 years has enabled us to utilize it in the development of our deep learning-based approach.

Hinode SOT/SP data consists of spatio-temporal spectro-polarimetric information covering Fe I line pair profiles at 630.15 and 630.25 nm, along with their nearby continuums. A sampling slit with a width of 0.15" was used to construct these line pair profiles. The data dimension is 2D-space×1D-wavelength×1D-polarimetry. The spatial, wavelength, and polarimetry dimensions correspond to different fields of views (FoV), 112 wavelength points, and four Stokes parameters (I, Q, U, and V). The profiles of the Stokes I and V exhibit distinguishable noise levels at various spatial positions, as shown in Fig. 1. Notably, in various spatial positions with varying magnetic fields, Stokes I exhibits a smoother profile than Stokes V.

We selected Hinode/SP Level 1 data (Lites and Ichimoto, 2013) observed on 2011-09-25 at the timestamp 20:01:04, downloaded from the Community Spectropolarimetric Analysis Center (CSAC, 2006) website, based on its capture near the center of the solar disk containing both sunspots and quiet Sun regions. The data consists of FITS files of individual scans, each with a FoV in the y and x directions of 162.304" and 0.295", respectively. After combining the FITS file scans along the x direction, we reconstructed a 2D spectro-polarimetric (SP) image with a size of 162.304" in the y direction and 151.142" in the x direction. To isolate individual spectral data (each pixel of the FoV) while disregarding spatial information, we transformed the 2D-space into a 1D-pixel dimension, resulting in a new structure of 1D-pixel×1D-wavelength×1D-polarimetry.

3. Compression Model

3.1. Autoencoder

The autoencoder is recognized as one of the most notable representatives of neural network-based feature extraction approaches. Its architecture contains encoding and decoding components, each comprised of neural network
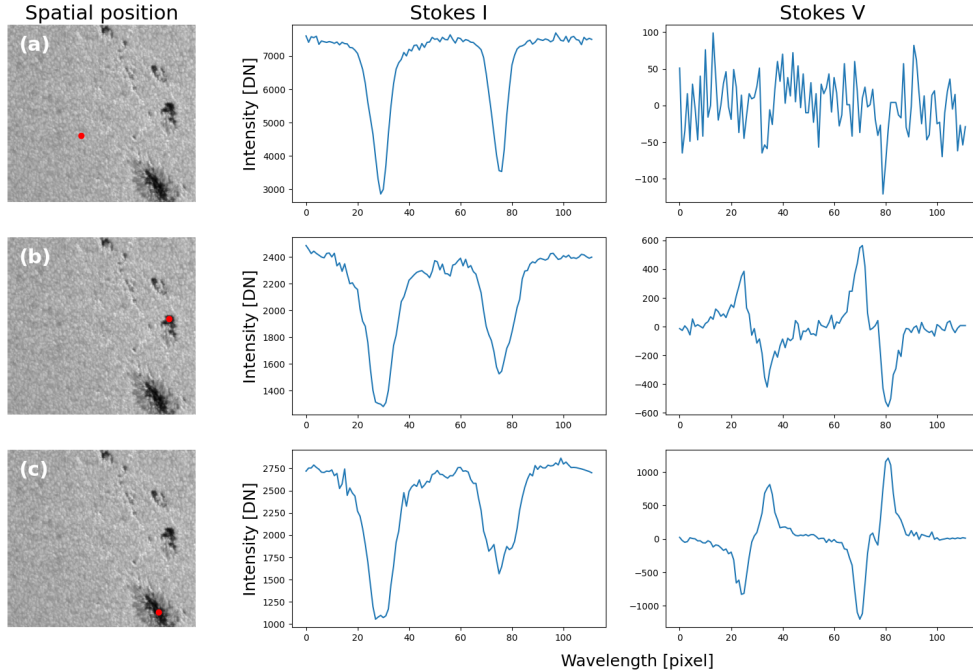
Figure 1: Sample profiles of Stokes parameters corresponding to spatial positions marked in red are provided for (a) quiet Sun, (b) pore, and (c) sunspot core in the FoV image.

layers working together to efficiently reduce the size of the input through reconstruction. This encoder-decoder structured dimensionality reduction technique works effectively on non-linearly connected data. Furthermore, it contributes to reducing noise within the data (Saura et al., 2023), potentially leading to reconstructed spectra with decreased observational and instrumental noise. This characteristic positions it as a strong candidate for model selection. The primary goal of the autoencoder is to reconstruct input data into an output that closely resembles the input. The encoder decreases the dimension of the input, while the decoder performs the reverse operation, increasing the lower dimensional input back to the size of the original input. This lower-dimensional bottleneck, known as the feature vector, serves as the compressed representation of the original input.

Both our DAE and CAE models maintain a simple architecture. Overviews of the models are provided in Fig. 2. The DAE comprises a sequence of fully connected dense layers that decrease in size in the encoder part and increase in size in the decoder part. The CAE includes 1D-convolutional layers and max/average pooling layers in the encoder part, with their opposites, trans-
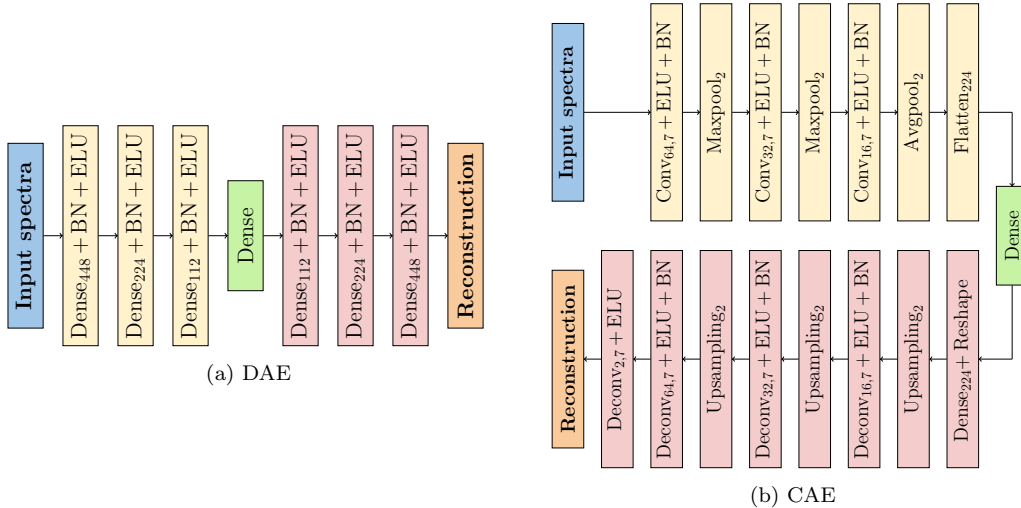
Figure 2: Model architectures for (a) DAE and (b) CAE. Blue and orange blocks represent the input and output (true and reconstructed spectra) of the models, while the encoder, decoder, and bottleneck are respectively depicted in yellow, pink, and green blocks. At each layer name, one index signifies the shape of the layer, while two indices denote the number of filters and the kernel size.

pose, and upsampling layers in the decoder part. The encoder and decoder depths are set to three for each model, as a deeper structure did not yield advantages. Both dense and convolutional layers were augmented with batch normalization (BN; Ioffe and Szegedy, 2015) as the accelerator, and an exponential linear unit (ELU; Clevert et al., 2016) was applied as the activation function. Other parameters were set by default. Considering the 112 wavelength points and 2 polarimetry parameters, the input and output sizes for our models are both 224 for DAE, with a shape of (112, 2) for CAE. The determination and analysis of the bottleneck (feature vector) size are discussed in Section 4.2. The models were implemented using Keras (Chollet, 2015) with TensorFlow (Abadi et al., 2016) in the Python (Van Rossum and Drake, 2009), and the development took place on Google Colaboratory (Bisong, 2019).

3.2. Data preparation

The 2D spatial dimension of the selected data has a shape of (512, 722) resulting in a total of 369,664 spectral pixels. The dataset was partitioned into training, validation, and test sets through manual area selection from the spatial image, ensuring the inclusion of both quiet Sun and active regions in all three sets. This resulted in a ratio of approximately 76% (270,664 px)

for training, 12% (45,000 px) for validation, and 12% (45,000 px) for test sets. Fig. 3 displays the dataset split on the continuum image.
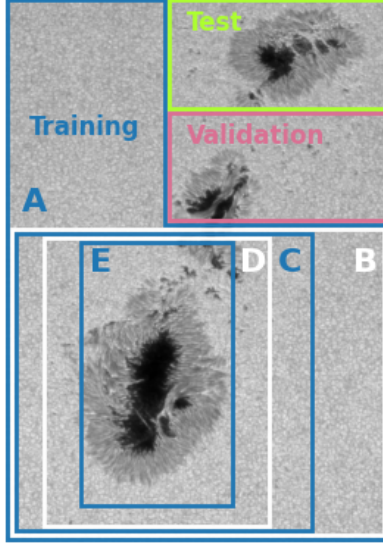

Figure 3: Snapshot of dataset and its partitioning for model training (Versions A to E), validation, and testing.

In the training set, the pixel count of the quiet Sun is distinctly higher than that of the sunspot area. Sunspots are uncommon occurrences on the solar surface, leading to their infrequent capture in observational data. Furthermore, the presence of extremely high magnetic fields around and/or within sunspots results in distinct spectral profiles. In machine learning, the diversity of data in the training set is important, and balancing the incorporation of different data types during training is essential, as insufficient representation can inhibit the model's ability to effectively learn. This limitation may lead to poor predictions for those specific data types in new datasets. Similarly, if the training set contains a significantly smaller number of sunspot pixels compared to the quiet Sun regions within the spatial image, this could potentially give rise to a data imbalance issue in the training of the deep learning model. Since the data is rarely encountered during training, it is likely to be poorly predicted when it appears in entirely new spectra. To address this issue, we manually prepared five versions of our training set (A (270,664 px), B (216,064 px), C (168,800 px), D (120,000 px), and E (70,000 px)) by considering the ratio of sunspot pixels to quiet Sun pixels. This allows us to

7

explore the impact of data balance on model performances. Consequently, we trained each model five times using these five different training sets, while keeping the validation and test sets the same. Fig. 3 depicts dataset versions A to E derived from the initial training set. To quantitatively assess the degree of balance (DoB) for each version of the training set, we calculated the DoB based on the pixel values of the continuum image using Shannon's entropy (Shannon, 1948) method:

$$
\text{DoB} = \frac{-\sum_{i=1}^{k} \frac{c_i}{n} \log \frac{c_i}{n}}{\log k},
\tag{1}
$$

where n is the total number of pixels, k is the total number of bins along the pixel value, and c_i is the number of pixels in the i-th bin. We set k to 100, aiming for a representation of the balance in the training sets that is neither too coarse nor too detailed. The DoB spans from 0 to 1, with a DoB of 0 suggesting unbalanced data and a DoB of 1 indicating balanced data. Fig. 4 shows histograms of the pixel values and their corresponding DoBs.
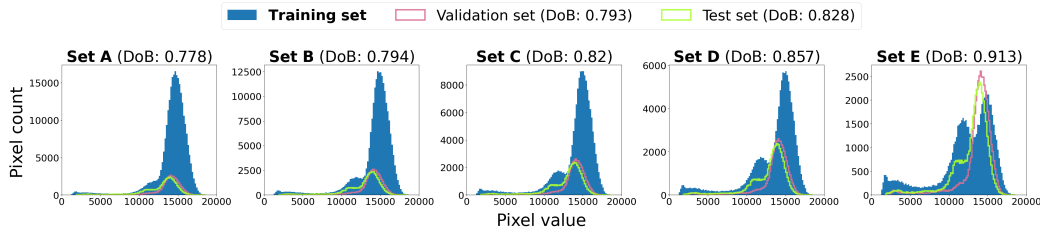

Figure 4: Degree of balance (DoB) in histograms for the five different training sets.

Utilizing normalized data during the model training enhances both performance and training speed. We applied min-max scaling to normalize the input profile of Stokes I, as

$$
x' = (b - a)\frac{x - \min x}{\max x - \min x} + a,
\tag{2}
$$

where x and x' denote the original and normalized data, respectively, and $[a, b]$ signifies the range for scaling, which for Stokes I is $[0, 1]$. For Stokes V, we used zero-mean scaling, considering its property of zero centered values, as

$$
x' = \begin{cases} \dfrac{x}{|\max x|} & \text{if } |\max x| \geq |\min x| \\ \dfrac{x}{|\min x|} & \text{otherwise} \end{cases}.
\tag{3}
$$

8

3.3. Training setups

The models were trained for 1,000 epochs with batch sizes of 512, using the Adam (Kingma and Ba, 2015) optimizer. Attempting smaller batch sizes extended the training process excessively, often leading to a halt, with no improvement in performance. We implemented early stopping and learning rate reduction on plateau optimization techniques to enhance the training effectiveness and conserve computational time. The patience parameter for early stopping was set to 100, while for learning rate reduction on plateau, it was set to 50.

In calculating the reconstruction loss function, we computed the mean absolute error (MAE) of intensity values independently for the Stokes I and V parameters at each wavelength point. The total reconstruction loss was then determined by summing the MAE of Stokes I over the MAE of Stokes V, as expressed in

$$L_{\mathrm{recons}} = \mathrm{MAE}_I + \mathrm{MAE}_V. \tag{4}$$

3.4. Evaluation methods

Stokes I features two clearly recognizable absorption lines in the left and right halves of its profile, whereas Stokes V displays four lobes, each pair corresponding to the two absorption line cores. Considering these properties, we defined four target areas for model evaluation based on the root mean square deviation (RMSD) at the respective wavelength ranges—left and right line cores for Stokes I (LLC_I, RLC_I), and similarly, left and right line cores for Stokes V (LLC_V, RLC_V). The ranges for the left and right line cores are the same for both parameters: 10–45 and 60–95.
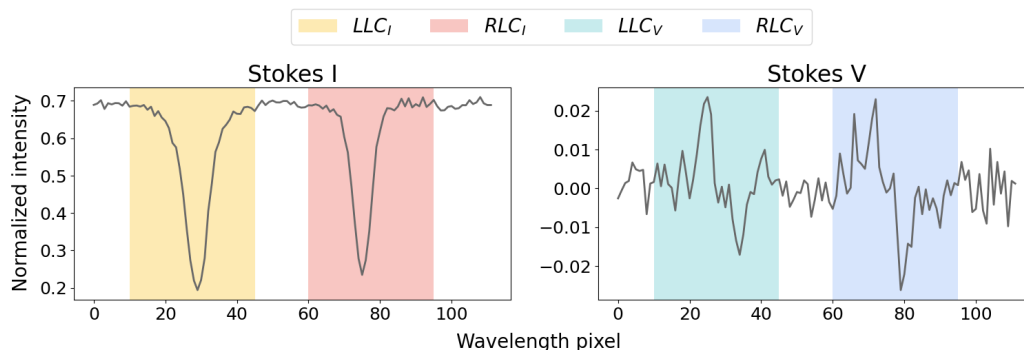

Figure 5: Evaluation areas of Stokes profiles. Colored shaded regions indicate the calculations of RMSD for left and right cores of both Stokes parameters (LLC_I, RLC_I, LLC_V, RLC_V).

9

4. Results

4.1. Model training

We configured the training process with 1000 epochs and implemented early stopping, set to activate if there was no reduction in the reconstruction loss on the validation set for 100 consecutive epochs. The training dynamics are depicted in the loss-epoch dependency graphs for models trained on set B, as shown in Fig. 6. While the DAE model exhibited several sudden jumps in loss during the validation process, the CAE model demonstrates a consistent and smooth decline in validation loss, mirroring the decrease in training loss.
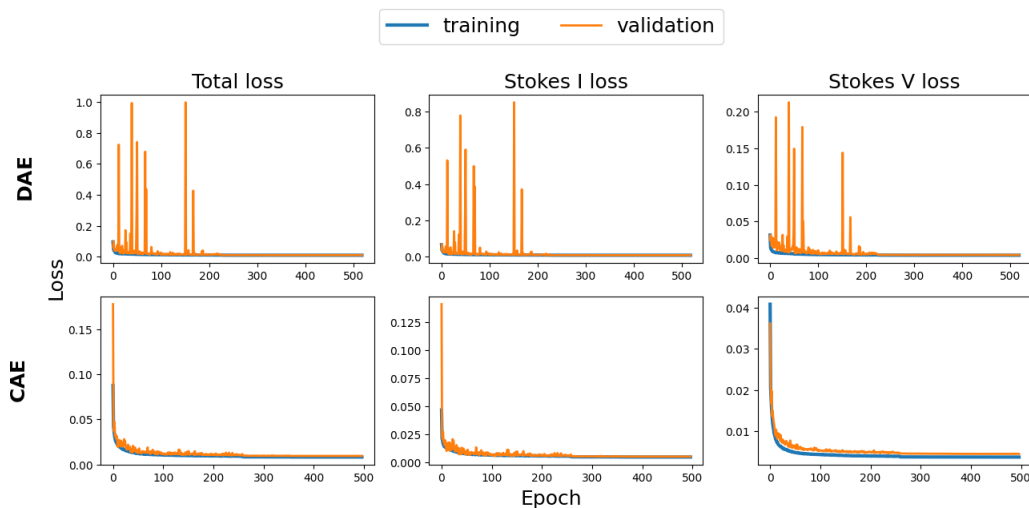

Figure 6: Loss-epoch graphs of models trained on set B.

4.2. Compression rate analysis

To determine the optimal feature vector size for the compression, we conducted experiments by training the models with different bottleneck sizes. The results were analyzed across the four line core areas (LLC_I, RLC_I, LLC_V, RLC_V), where we compared them with the observational noise levels of Stokes I and V ($\sigma_{obs,I}$, $\sigma_{obs,V}$). For each Stokes parameter, σ_{obs} is calculated similarly as

$$\sigma_{obs} = \frac{\sum_{i=1}^{N} \sigma_i^{[0,15]}}{N}, \tag{5}$$

where N represents the number of spectra in the test set, and $\sigma_i^{[0,15]}$ refers to the standard deviation of the continuum within the wavelength range $[0,$

10

15] of each spectrum. The reference continuum level for Stokes V was considered to be 0. Fig. 7 shows the dependency between bottleneck size and RMSD values in the target areas of LLC_I, RLC_I, LLC_V, and RLC_V for the DAE and CAE models. The horizontal axis indicates the number of nodes in the bottleneck and model types, while the vertical axis displays the RMSD values. The DAE model achieved the lowest RMSDs, approaching the observational noise levels at 28 nodes. However, at 56 nodes, the RMSDs increased, suggesting an overfitting issue. The CAE model also showed a decreasing RMSD trend up to 28 nodes. Notably, its performance continued to improve even at 56 and 112 nodes, with RMSDs falling below the observational noise levels, highlighting the model's robust performance. Given
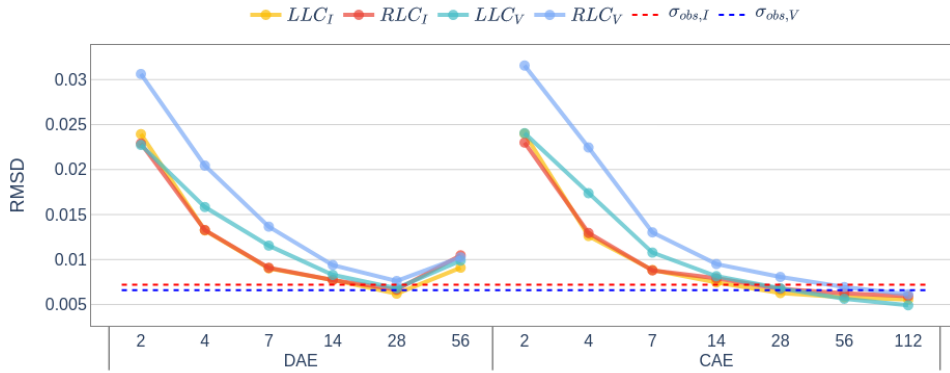

Figure 7: Bottleneck-RMSD dependency graphs of the DAE and CAE models.

that 28 nodes in the bottleneck yielded the best possible results for the DAE model, we proceeded by comparing the performances of the DAE and CAE models, both configured with 28 parameters, for spectral reconstruction in the next phase of our analysis.

4.3. Data imbalance analysis

The DAE and CAE models, each having 28 nodes in the bottleneck, were trained using five versions (A to E) of the training sets for evaluations on a common test dataset. Fig. 8 illustrates the dependencies between DoB and RMSD values in the target areas. The horizontal axis aligns the training set names and model types, while the vertical axis represents the RMSD values.

The DAE exhibited noticeable fluctuations in results, whereas we observed minimal differences in the performance of the CAE. Models trained on sets A, B, and C emerged as top performers, indicating the potential of both models to mitigate training set imbalances.
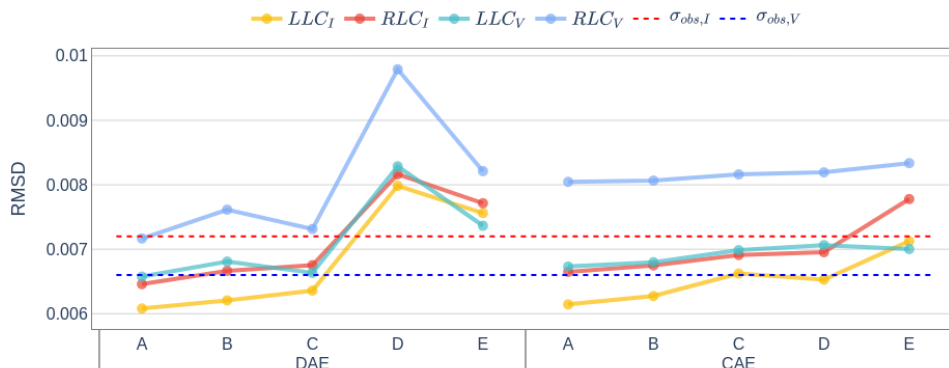

Figure 8: Training set-RMSD dependency graphs of the DAE and CAE models.

4.4. Comparison of observed and reconstructed spectra

We compared the original and reconstructed profiles from the models with the best performances based on chi-squared and RMSD metrics. The reconstructed profiles are unscaled to original intensity range from the scaled model output. The observational test profiles with original intensity and the unscaled reconstructions are then normalized to the quiet Sun continuum (I_c). Quiet Sun regions selected from the test set, as shown in Fig. 9, are used to define I_c. Chi-square values (χ^2) are calculated wavelength-wise with respect to σ_{obs} for each profile as

$$\chi^2 = \frac{1}{d} \sum_{i=1}^{d} \frac{(S'(\lambda_i) - S(\lambda_i))^2}{\sigma_{obs}^2}, \tag{6}$$

where d represents the number of data points in the continuum within the wavelength range [0, 15], and $S(\lambda_i)$ and $S'(\lambda_i)$ refer to the spectral intensity at i-th wavelength point of observational and reconstructed spectra, respectively. Fig. 10 shows the chi-square histograms of the distribution across all test data for Stokes I and V, comparing the performances of the DAE
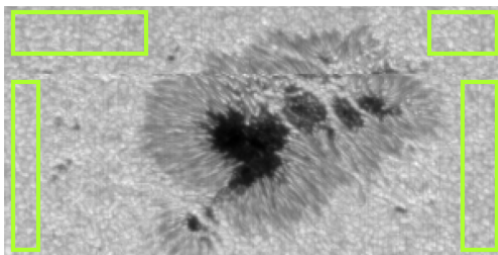
12

Figure 9: Quiet Sun regions, depicted by green rectangles in the test dataset, are selected to define the continuum intensity of the quiet Sun.

and CAE models. Stokes I histograms resulted in mean values less than 1, suggesting that the average χ^2 values are within the observational noise. For Stokes V, both models yielded mean values slightly above 1, particularly in the case of CAE. Similarly, RMSD histograms of Stokes I and V reconstructions obtained from the DAE and CAE models are depicted in Fig. 11, using the calculation for each profile as

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^{d}(S'(\lambda_i) - S(\lambda_i))^2}{d}}, \tag{7}$$

suggesting that all histograms resulted in mean RMSD values comparable to the observational noise.

To display reconstruction samples, we chose eight different spatial positions, including the quiet Sun, pores, and both the penumbra and umbra of a sunspot in the continuum image of the test set. Figure 12 illustrates comparisons between the true observational profiles of selected pixel positions and their respective reconstructions from the DAE and CAE models. Overall, both models produced smooth and comparable reconstructions that accurately fit the entire profiles, including fluctuations such as Stokes I continuums. Importantly, the reconstructions effectively captured Stokes V shapes in the quiet Sun despite the initial high noise levels, achieving a good balance of noise removal without overfitting. In high magnetic field regions like the sunspot center in (c3), where the Stokes profiles are rarer and more complex, both models faced challenges in accurate reconstruction, with the DAE model showing particular difficulty.
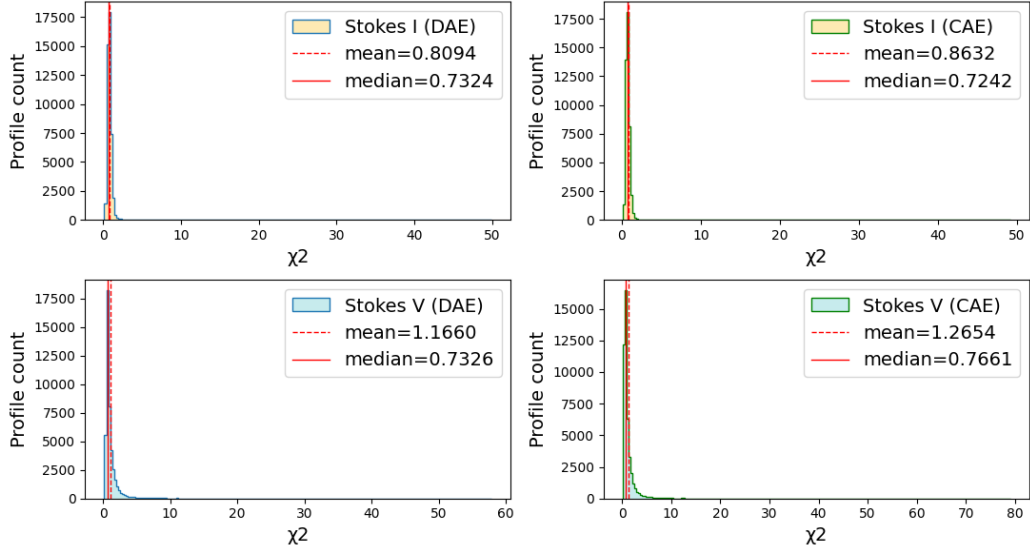
13

Figure 10: Chi-square histograms of observed and reconstructed spectra.
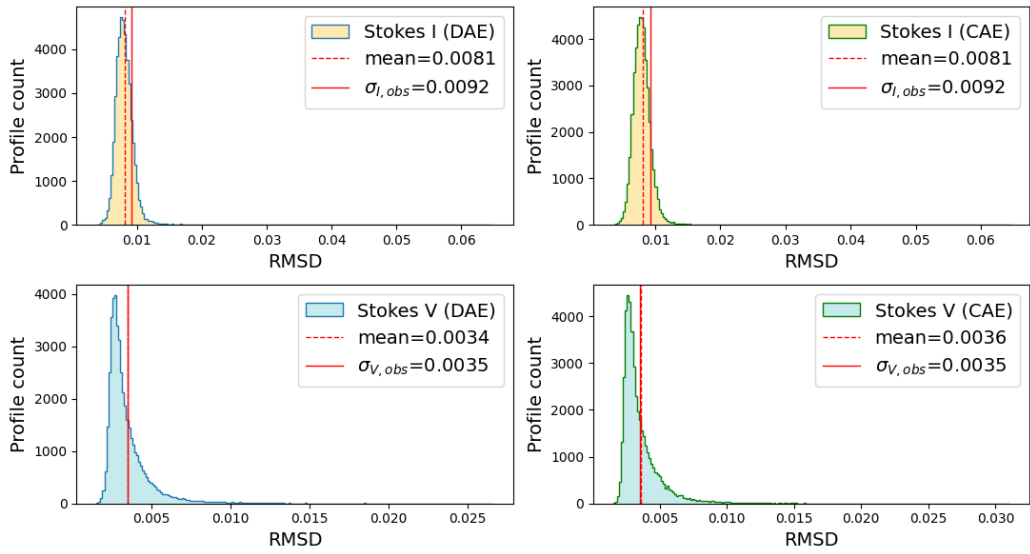

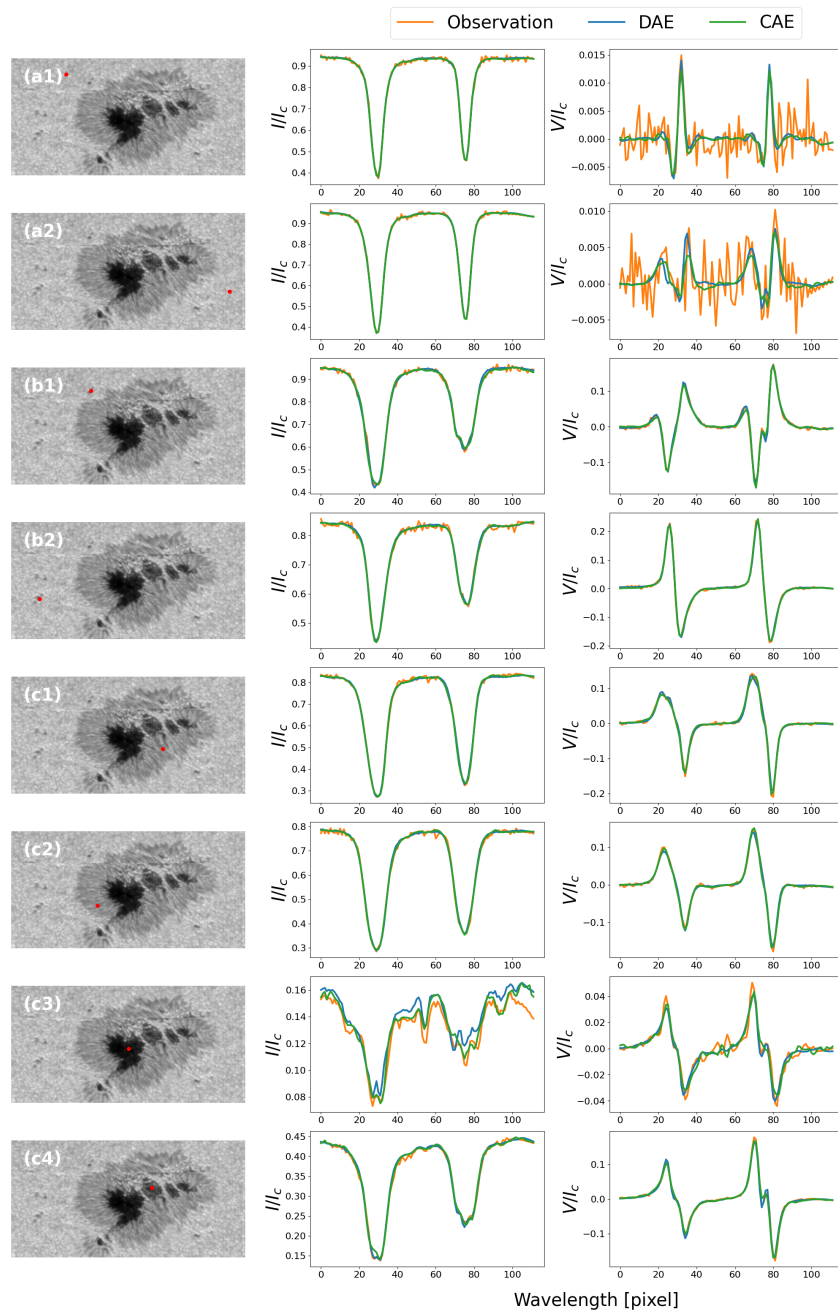Figure 11: RMSD histograms of observed and reconstructed spectra.

14

Figure 12: Samples of observational and reconstructed spectra at various spatial positions, including (a1–2) quiet Sun, (b1–2) pore, and (c1–4) sunspot penumbra and umbra, in the test set.

5. Discussion

To achieve the highest compression rate possible while ensuring robust performance relative to observational noises, we opted to use 28 nodes in the bottleneck for both the DAE and CAE models. A larger bottleneck in the CAE model has the potential to further improve its ability to reconstruct spectral shapes.

Stokes V profiles typically exhibit higher and less clearly definable noise levels compared to Stokes I profiles, which prompted concern about their potential impact on the training process and the risk of undertraining Stokes I signals. Importantly, we found that optimizing the balance between the contributions of Stokes I and Stokes V in the reconstruction loss during training—such as by customizing the loss function with a specific weight for the Stokes V component—was unnecessary to achieve satisfactory model performance.

When using different datasets for training, the DAE model exhibited a higher sensitivity to the DoB in the training set. It faced challenges in accurately reconstructing spectral profiles that were less presented during training. In contrast, the CAE models demonstrated greater flexibility in this regard, consistently delivering stable performances across varying DoB in the training dataset.

The novel aspect of this work lies in the integrated compression of two polarimetric parameter profiles, applied not only to the quiet Sun but also to various positions on the solar surface, including active regions. This approach enables the analysis of spectral profiles in regions of interest with strong magnetic fields, which could potentially drive a range of solar behaviors. The study is limited by its reliance on only the Stokes I and V parameters from the four Stokes spectra. To address this, future research should incorporate the Stokes Q and U parameters, which will provide a more comprehensive understanding of the solar atmosphere's structure, physical conditions, and complex magnetic fields. Additionally, while our current analysis is confined to data from the disk center, future studies should extend this focus to encompass other regions across the solar disc.

Our compression method shows potential for a wide range of applications, such as detection of anomalous spectra. In this scenario, the model is training on normal data to ensure a reconstruction error below a specified threshold. Subsequently, the pre-trained model is applied to reconstructing data containing previously unseen anomalous signals, resulting in a recon-

struction error that exceeds the threshold. Furthermore, unusual events, such as solar flares, could possibly be detected based on their distinctive spectral signatures observed in data preceding actual flare occurrences. Additionally, the suggested approach potentially facilitates the comparison and analysis of observational and numerical simulation data by leveraging the compact representations provided by the compression model.

6. Conclusion

In this work, we developed two distinct deep learning model architectures, a deep autoencoder (DAE) and a 1D-convolutional autoencoder (CAE), specifically tailored for compressing Hinode SOT/SP spectral data, with a primary focus on the Stokes I and V polarization parameters. Our experiments aimed to determine the optimal compression rate, evaluate different model architectures, and assess their performance across various balanced training datasets.

The results demonstrate that our compression models effectively reduced the spectral data dimensionality from 224 to 28 parameters, yielding reconstruction residuals comparable to the observational noise while also eliminating high noise levels. Notably, the CAE model outperformed the DAE model, offering greater stability in handling data imbalance and robustly maintaining the reproducibility of complex profile shapes.

The novelty of our study lies in the compression of two-dimensional observational solar polarimetric spectra in both the quiet Sun and active regions. This method provides a more effective analysis technique, significantly broadening its applicability for solar physics studies.

In future work, we aim to develop a universal compression model that improves detailed spectral analysis by incorporating full Stokes parameters and can be applied to a broad range of snapshots, extending beyond the disk center.

Hinode is a Japanese mission developed and launched by ISAS/JAXA, collaborating with NAOJ as a domestic partner and with NASA and STFC (UK) as international partners. Scientific operation of the Hinode mission is conducted by the Hinode science team organized at ISAS/JAXA. This team mainly consists of scientists from institutes in the partner countries. Support for the post-launch operation is provided by JAXA and NAOJ(Japan), STFC (U.K.), NASA, ESA, and NSC (Norway).

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X., 2016. TensorFlow: A System for Large-Scale Machine Learning, in: Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation, pp. 265–283.

Asensio Ramos, A., 2006. The Minimum Description Length Principle and Model Selection in Spectropolarimetry. The Astrophysical Journal 646, 1445–1451. doi:`10.1086/505136`.

Asensio Ramos, A., Díaz Baso, C.J., 2019. Stokes Inversion Based on Convolutional Neural Networks. Astronomy & Astrophysics 626, A102. doi:`10.1051/0004-6361/201935628`.

Asensio Ramos, A., Socas-Navarro, H., López Ariste, A., Martínez González, M.J., 2007. The Intrinsic Dimensionality of Spectropolarimetric Data. The Astrophysical Journal 660, 1690. doi:`10.1086/513069`.

Bisong, E., 2019. Google Colaboratory, in: Building Machine Learning and Deep Learning Models on Google Cloud Platform, pp. 59–64.

Chen, Z., Yeo, C.K., Lee, B.S., Lau, C.T., 2018. Autoencoder-based network anomaly detection, in: 2018 Wireless Telecommunications Symposium (WTS), pp. 1–5. doi:`10.1109/WTS.2018.8363930`.

Chollet, F., 2015. Keras. `https://github.com/fchollet/keras`.

Clevert, D.A., Unterthiner, T., Hochreiter, S., 2016. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs), in: Proceedings of the 4th International Conference on Learning Representations.

Community Spectropolarimetric Analysis Center (CSAC), 2006. Hinode-SpectroPolarimeter (SP) level 1 (calibrated) full stokes data. `https://www2.hao.ucar.edu/csac`. doi:`10.5065/D6T151QF`.

De Pontieu, B., Title, A.M., Lemen, J.R., Kushner, G.D., Akin, D.J., Allard, B., Berger, T., Boerner, P., Cheung, M., Chou, C., Drake, J.F., Duncan, D.W., Freeland, S., Heyman, G.F., Hoffman, C., Hurlburt, N.E., Lindgren, R.W., Mathur, D., Rehse, R., Sabolish, D., Seguin, R., Schrijver, C.J., Tarbell, T.D., Wülser, J.P., Wolfson, C.J., Yanari, C., Mudge, J., Nguyen-Phuc, N., Timmons, R., van Bezooijen, R., Weingrod, I., Brookner, R., Butcher, G., Dougherty, B., Eder, J., Knagenhjelm, V., Larsen, S., Mansir, D., Phan, L., Boyle, P., Cheimets, P.N., DeLuca, E.E., Golub, L., Gates, R., Hertz, E., McKillop, S., Park, S., Perry, T., Podgorski, W.A., Reeves, K., Saar, S., Testa, P., Tian, H., Weber, M., Dunn, C., Eccles, S., Jaeggli, S.A., Kankelborg, C.C., Mashburn, K., Pust, N., Springer, L., Carvalho, R., Kleint, L., Marmie, J., Mazmanian, E., Pereira, T.M.D., Sawyer, S., Strong, J., Worden, S.P., Carlsson, M., Hansteen, V.H., Leenaarts, J., Wiesmann, M., Aloise, J., Chu, K.C., Bush, R.I., Scherrer, P.H., Brekke, P., Martinez-Sykora, J., Lites, B.W., McIntosh, S.W., Uitenbroek, H., Okamoto, T.J., Gummin, M.A., Auker, G., Jerram, P., Pool, P., Waltham, N., 2014. The Interface Region Imaging Spectrograph (IRIS). Solar Physics 289, 2733–2779. doi:`10.1007/s11207-014-0485-y`.

Gafeira, R., Orozco Suárez, D., Milić, I., Quintero Noda, C., Ruiz Cobo, B., Uitenbroek, H., 2021. Machine learning initialization to accelerate Stokes profile inversions. Astronomy & Astrophysics 651, A31. doi:`10.1051/0004-6361/201936910`.

Gogoi, M., Begum, S.A., 2017. Image Classification Using Deep Autoencoders, in: 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), pp. 1–5. doi:`10.1109/ICCIC.2017.8524276`.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, in: Proceedings of the 32nd International Conference on Machine Learning, pp. 448–456.

Kingma, D.P., Ba, J.L., 2015. Adam: A Method for Stochastic Optimization, in: Proceedings of the 3rd International Conference on Learning Representations.

Kosugi, T., Matsuzaki, K., Sakao, T., Shimizu, T., Sone, Y., Tachikawa, S., Hashimoto, T., Minesugi, K., Ohnishi, A., Yamada, T., Tsuneta, S., Hara, H., Ichimoto, K., Suematsu, Y., Shimojo, M., Watanabe, T., Shimada, S., Davis, J.M., Hill, L.D., Owens, J.K., Title, A.M., Culhane, J.L., Harra, L.K., Doschek, G.A., Golub, L., 2007. The Hinode (Solar-B) mission: An overview. Solar Physics 243, 3–17.

Kramer, M.A., 1991. Nonlinear principal component analysis using autoassociative neural networks. AIChE Journal 37, 233–243.

LeCun, Y., 1987. Phd thesis: Modeles connexionnistes de l'apprentissage (connectionist learning models) .

Lites, B.W., Akin, D.L., Card, G., Cruz, T., Duncan, D.W., Edwards, C.G., Elmore, D.F., Hoffmann, C., Katsukawa, Y., Katz, N., Kubo, M., Ichimoto, K., Shimizu, T., Shine, R.A., Streander, K.V., Suematsu, A., Tarbell, T.D., Title, A.M., Tsuneta, S., 2013. The Hinode Spectro-Polarimeter. Solar Physics 283, 579–599. doi:`10.1007/s11207-012-0206-3`.

Lites, B.W., Ichimoto, K., 2013. The SP_PREP data preparation package for the Hinode spectro-polarimeter. Solar Physics 283, 601–629. doi:`10.1007/s11207-012-0205-4`.

López Ariste, A., Casini, R., 2002. Magnetic Fields in Prominences: Inversion Techniques for Spectropolarimetric Data of the He I D_3 Line. The Astrophysical Journal 575, 529–541. doi:`10.1086/341260`.

Melchior, P., Liang, Y., Hahn, C., Goulding, A., 2023. Autoencoding Galaxy Spectra. I. Architecture. The Astronomical Journal 166, 74. doi:`10.3847/1538-3881/ace0ff`.

Portillo, S.K.N., Parejko, J.K., Vergara, J.R., Connolly, A.J., 2020. Dimensionality Reduction of SDSS Spectra with Variational Autoencoders. The Astronomical Journal 160, 45. doi:`10.3847/1538-3881/ab9644`.

Ryu, S., Jeon, B., Seo, H., Lee, M., Shin, J.W., Yu, Y., 2023. Development of deep autoencoder-based anomaly detection system for HANARO. Nuclear Engineering and Technology 55, 475–483. doi:`10.1016/j.net.2022.10.009`.

Sadykov, V.M., Kitiashvili, I.N., Dalda, A.S., Oria, V., Kosovichev, A.G., Illarionov, E., 2021. Compression of Solar Spectroscopic Observations: a Case Study of Mg II k Spectral Line Profiles Observed by NASA's IRIS Satellite, in: 2021 International Conference on Content-Based Multimedia Indexing (CBMI), pp. 1–6. doi:`10.1109/CBMI50038.2021.9461879`.

Saura, N., Garrido, D., Benkadda, S., Ibano, K., Ueda, Y., Hamaguchi, S., 2023. Spectroscopic analysis improvement using convolutional neural networks. Journal of Physics D: Applied Physics 56, 354001.

Shannon, C.E., 1948. A Mathematical Theory of Communication. The Bell System Technical Journal 27, 379–423. doi:`10.1002/j.1538-7305.1948.tb01338.x`.

Socas-Navarro, H., 2005. Feature Extraction Techniques for the Analysis of Spectral Polarization Profiles. The Astrophysical Journal 620, 517–522. doi:`10.1086/426811`.

Suematsu, Y., Tsuneta, S., Ichimoto, K., Shimizu, T., Otsubo, M., Katsukawa, Y., Nakagiri, M., Noguchi, M., Tamura, T., Kato, Y., Hara, H., Kubo, M., Mikami, I., Saito, H., Matsushita, T., Kawaguchi, N., Nakaoji, T., Nagae, K., Shimada, S., Takeyama, N., Yamamuro, T., 2008. The Solar Optical Telescope of Solar-B (Hinode): The Optical Telescope Assembly. Solar Physics 249, 197–220. doi:`10.1007/s11207-008-9129-4`.

Tsuneta, S., Ichimoto, K., Katsukawa, Y., Nagata, S., Otsubo, M., Shimizu, T., Suematsu, Y., Nakagiri, M., Noguchi, M., Tarbell, T., Title, A., Shine, R., Rosenberg, W., Hoffmann, C., Jurcevich, B., Kushner, G., Levay, M., Lites, B., Elmore, D., Matsushita, T., Kawaguchi, N., Saito, H., Mikami, I., Hill, L.D., Owens, J.K., 2008. The Solar Optical Telescope for the Hinode Mission: An Overview. Solar Physics 249, 167–196. doi:`10.1007/s11207-008-9174-z`.

Van Rossum, G., Drake, F.L., 2009. Python 3 Reference Manual.

Yeom, S., Choi, C., Kim, K., 2021. AutoEncoder Based Feature Extraction for Multi-Malicious Traffic Classification, in: The 9th International Conference on Smart Media and Applications, pp. 285–287. doi:10.1145/3426020.3426093.