A Two-Stream CNN Method for Sunspot Group Magnetic Type Classification: Postprint

Authors: Yang Wugu, Tian Weixin, Shuguang Zeng, Huang Yao

Date: 2021-08-20T00:00:00+00:00

Abstract

Solar flares predominantly occur above magnetically complex sunspot groups; therefore, the magnetic classification of sunspot groups can serve as an important basis for solar flare prediction. Aiming at sunspot classification utilizing both white-light images and magnetogram data, this paper proposes a two-stream CNN method for magnetic classification of sunspot groups. This method employs two network branches to simultaneously extract features from white-light images and magnetograms, fuses the two types of features in the fully connected layer, and finally employs them for training and classification. Experimental results demonstrate that this method can effectively avoid the information loss inherent in single-channel network models as well as the mutual interference between white-light images and magnetograms that occurs when using two-channel image composition, achieving a weighted F1 score of up to 84.87% and an average accuracy of up to 84.93%.

Full Text

A Magnetic Classification Method of Sunspot Groups Based on Dual-Stream CNN

Yang Wugu, Tian Weixin*, Zeng Shuguang, Huang Yao
School of Computer and Information, Three Gorges University, Yichang, Hubei 443000

Abstract: Most solar flares occur above sunspot groups with complex magnetic polarities, making the magnetic type of sunspot groups an important basis for predicting solar flares. This paper proposes a dual-stream CNN method for classifying sunspot group magnetic types when both continuum and magnetogram data are available. The method simultaneously extracts features from continuum and magnetogram images through two network branches, fuses these two types of features in the fully connected layer, and finally uses them for training and classification. Experiments show that this method effectively avoids the information loss of single-channel network models and the mutual interference between continuum and magnetogram images that occurs when using dual-channel image composition, achieving a weighted F1 score of 84.87% and an average accuracy of 84.93%.

Keywords: sunspot; Mount Wilson; dual-stream CNN; ResNet

0 Introduction

Solar flares are caused by the sudden release of magnetic energy on the Sun. Long-term observations reveal that most flares occur above sunspot groups, and the more complex the magnetic polarity configuration of a sunspot group, the higher the probability of a major flare. Consequently, the magnetic type of sunspot groups can serve as an important basis for solar flare prediction. Traditional classification methods for sunspot groups include image-based approaches such as edge detection, mathematical morphology, and wavelet analysis, as well as data-driven machine learning methods. With advances in observational capabilities in recent years, solar activity-related data has grown rapidly, making machine learning approaches increasingly advantageous.

In 2008, Colak et al. proposed a hybrid system using the McIntosh classification scheme that employed active region data extracted from SOHO/MDI magnetogram images to automatically detect sunspot groups on SOHO/MDI continuum images. After detecting sunspots from MDI continuum images, they used MDI magnetogram images for grouping/clustering, and integrated image processing with neural networks to automatically classify the detected sunspot groups. However, their system suffered from grouping errors and missed detection of small sunspots. In 2010, Mehmmood A. Abd et al. adopted a modified seven-class Zurich classification scheme and used Support Vector Machines (SVM) to automatically classify sunspot groups on full-disk solar continuum images. During data preprocessing, they employed edge detection, noise removal, and binarization to segment sunspot groups from the solar disk, followed by unsupervised segmentation to merge sunspots belonging to the same group, extracted attributes for each sunspot group, and finally used SVM for classification. This method’ s accuracy depended on image quality and distortion levels, and the segmentation process significantly impacted output speed.

In 2019, Yuanhui Fang et al. used Convolutional Neural Networks (CNN) to classify sunspot group magnetic types. During data preprocessing, images were divided into three categories: continuum images, magnetogram images, and composite images of continuum and magnetogram. They used CNN to classify these three image types, with results showing that using continuum images alone yielded the best classification performance. The authors attributed the poor performance of magnetogram-based classification to the relatively complex structure of magnetograms, which CNNs could not effectively extract features from.

Dual-stream CNN can fuse features from different dimensions of an object, providing more feature information for classification tasks. In action recognition, one branch of dual-stream CNN extracts spatial information from RGB images while the other extracts temporal information from optical flow images; after fusion, action recognition is performed. Using this method to recognize fighting behavior achieved an average recognition rate of 89.3%, higher than single-branch methods. The Mount Wilson scheme for determining sunspot group magnetic types requires feature information from different dimensions: first observing sunspot group structure through continuum images, then examining magnetic polarity distribution in corresponding magnetograms, and finally combining both to determine the magnetic type. Given these characteristics of the Mount Wilson scheme, this paper applies dual-stream CNN to classify sunspot group magnetic types, with one branch extracting continuum features and the other extracting magnetogram features, fusing different feature types before the fully connected layer. Classification results demonstrate that dual-stream CNN achieves higher accuracy for sunspot group magnetic type classification than CNN using single continuum or magnetogram images.

1 Dataset and Preprocessing

This study uses the SOLAR-STORM dataset, compiled and provided by the Space Environment AI Warning Interdisciplinary Innovation Workshop and publicly available on the Tianchi Lab platform (https://tianchi.aliyun.com/dataset/). Based on the Mount Wilson sunspot group magnetic classification scheme, the dataset consists of two parts: continuum images and magnetogram images. Since the same observation target at the same time has both continuum and magnetogram images, they have a one-to-one correspondence. Each continuum or magnetogram portion contains three types: Alpha, Beta, and Betax.

We assigned objects numbered after 6205, 6207, and 6206 in Alpha, Beta, and Betax respectively to the test set, with the remainder used as the training set. The specific distribution is shown in Table .

The original dataset is in FITS format, with continuum image pixel values ranging from [200, 74000] and magnetogram pixel values ranging from [-5000, 5000]. We compressed the continuum image value range to [0, 255] and converted them to JPG format, while compressing the magnetogram range [-800, 800] to [0, 255] to enhance features. Figure [Figure 1: see original paper] shows continuum and magnetogram images for Alpha, Beta, and Betax types. It can be observed that Alpha class represents unipolar sunspot groups, Beta class represents bipolar sunspot groups with opposite polarity, and Betax class represents complex multipolar sunspot groups.

Since sunspot images are captured at 96-minute intervals, sunspot group evolution is not significant in short periods, resulting in high similarity between images of the same observation target within a certain timeframe. As shown in

Figure [Figure 2: see original paper], the continuum and magnetogram images of object 26 in the Beta class taken at different times within a single day show almost no visible changes.

To avoid artificially high accuracy caused by similar images appearing in both training and validation sets during data splitting, we aggregated data by object and then split the training and validation sets by object, using ten-fold cross-validation for training and testing.

2.1 Dual-Stream CNN Classification Method

We first establish two different CNN networks using ResNet50 as the backbone. Continuum and magnetogram images are input separately into each branch. Since the images enter different branches, the CNN weights are not shared. After both image types pass through the networks to extract features, they form feature vectors that are concatenated and then fed into fully connected layers for Softmax classification. The network structure is shown in Figure [Figure 3: see original paper].

2.1.1 ResNet50

ResNet was proposed by Kaiming He et al. to address the degradation phenomenon in deep networks. The basic idea is to set skip connections within ResNet units, allowing original information to be directly transmitted to later layers. Common ResNet networks have depths of 18, 34, 50, and 101 layers. The basic ResNet structure is shown in Figure Figure 4: see original paper, where x is the input, f(x) is the mapping function, and H(x) is the output.

From the ResNet module, we know that f(x) = H(x) - x, meaning the ResNet module learns the difference between input and output. When the network reaches an optimal state at a certain layer, CNN would need to further update weights, whereas the ResNet module only needs to perform identity mapping, thereby avoiding network degradation.

2.1.2 Overfitting Prevention Measures

To prevent model overfitting, we implemented data augmentation and added three regularization methods: Dropout, L2 regularization, and Label smoothing.

Data augmentation uses rotation, horizontal mirroring, and scaling to enrich data distribution and improve model generalization capability. Dropout is set to 0.5, randomly setting half of the feature vector values to 0, which essentially introduces noise at that layer to avoid obtaining correct results due to certain insignificant coincidences.

L2 regularization forces weight parameters to take smaller values to prevent overfitting by adding the square of weight coefficients to the loss function. The cross-entropy loss function used is:

$$\text{Loss} = -\frac{1}{n}\sum_{i=0}^{n} q_i \log(p_i)$$

where n is the total number of samples, i is the class index, q is the actual label, and p is the predicted label.

After adding L2 regularization:

$$\text{Loss} = -\frac{1}{n}\sum_{i=0}^{n} q_i \log(p_i) + \lambda \sum_i w_i^2$$

where λ is the regularization parameter set to 0.001, and w represents weights. As shown in formula (2), if weight values become large, the Loss value increases, so to reduce the Loss value, the model gradually decreases w values, ensuring that even with large input differences, model predictions do not change significantly.

Label smoothing adjusts q in formula (1). When sample labels are converted to One-hot form, the true value is 1 and others are 0. However, p calculated through Softmax gradually approaches q during training. The Softmax calculation formula is:

$$S_i = \frac{e^i}{\sum_j e^j}$$

where i is the class index and j is the total number of classes. Through formula (3), model output becomes a probability distribution. Combined with formula (2), since sample labels are represented as non-zero 即-1, the Loss value approaches 0 when classification is correct and approaches -∞ when incorrect. This incentive-penalty mechanism is too extreme, suitable only for cases where datasets cover all possibilities or annotations are completely correct. To mitigate this extreme mechanism, we modify the One-hot label representation as:

$$\begin{cases} 1-\varepsilon & \text{if } y = k \\ \varepsilon/(k-1) & \text{if } y \neq k \end{cases}$$

where   is a hyperparameter set to 0.1, k is the class label, and y is the true label. As shown in formula (4), by reducing inter-class distance and increasing intra-class distance, the model' s generalization capability is improved.

2.2 Experimental Comparison Group Methods

To establish experimental comparisons, we additionally implemented ResNet50 and the CNN model used in literature [6]. All models used identical data splitting, hyperparameters, and overfitting prevention measures except for network structure. The CNN model structure is shown in Figure [Figure 5: see original paper].

This structure has three convolutional layers with kernel size (5,5) and each layer' s number of kernels being twice that of the previous layer to extract image features. After feature extraction, feature vectors are formed and input into fully connected layers, followed by Softmax classification. For the input layer, continuum images, magnetogram images, and dual-channel images composed of both were separately input into the network for training, ultimately obtaining seven groups of data.

2.3 Model Evaluation Metrics

This experiment evaluates model classification performance using accuracy, precision, recall, F1-score, and confusion matrix. Taking Alpha class as an example, the calculation formulas are:

$$\text{Accuracy} = \frac{TA + TB + TX}{TA + TB + TX + AB + AX + BA + BX + XA + XB}$$

$$\text{Precision} = \frac{TA}{TA + BA + XA}$$

$$\text{Recall} = \frac{TA}{TA + AB + AX}$$

$$\text{F1-score} = \frac{2TA}{2TA + BX + BA + AX + AB}$$

The confusion matrix for the three classes is shown in Table , where TA represents correctly identifying Alpha as Alpha, TB correctly identifying Beta as Beta, TX correctly identifying Betax as Betax, AB incorrectly identifying Alpha as Beta, AX incorrectly identifying Alpha as Betax, BA incorrectly identifying Beta as Alpha, BX incorrectly identifying Beta as Betax, XA incorrectly identifying Betax as Alpha, and XB incorrectly identifying Betax as Beta.

3 Experimental Results and Analysis

Since the experiment uses ten-fold cross-validation, StratifiedKFold was used to split the training set into 10 folds, with one fold used as the validation set and the remaining nine as the training set in each iteration. Finally, model classification performance on the test set was statistically analyzed. Evaluation metrics use weighted averages, and confusion matrices are calculated as the sum of corresponding confusion matrices from each model. Experimental data is shown in Table .

As shown in Table 3, dual-stream ResNet50 achieves the highest scores across all metrics. Even ResNet50 with strong mapping and optimization capabilities cannot outperform dual-stream ResNet50 when missing partial data, indicating that dual-stream ResNet50 has better classification capability when facing the same test set.

Since the test set has class imbalance, the weighted F1-score is used as the primary reference metric, with other metrics as supplementary references. Comparing experimental results from the same data source but different models reveals that ResNet50' s F1-score is significantly higher than CNN' s. For continuum, magnetogram, and dual-channel images, the weighted F1-score increased by 6.41%, 9.41%, and 7.97% respectively. When using continuum, magnetogram, and dual-channel images, improvements were 6.48%, 9.29%, and 7.71% respectively. This demonstrates that the feature extraction capability of deep networks like ResNet50 is stronger than shallow networks like CNN, enabling them to extract more useful features.

Comparing results from the same model but different data sources shows that when using CNN, continuum images yield the highest F1-score, followed by dual-channel images, with magnetograms lowest. When using ResNet50, magnetograms achieve the highest F1-score, followed by dual-channel images, with continuum images lowest. Continuum and magnetogram images show different classification advantages under different models, likely because the CNN architecture is simple and shallow, while ResNet50 is complex and deep. Continuum images have simple structures, while magnetograms have complex structures. When magnetograms enter CNN, CNN cannot extract features beneficial for classification. When continuum images enter CNN, the extracted features are more beneficial for classification than those from magnetograms. ResNet50 has stronger classification capability, so when magnetograms enter ResNet50, their complex structure provides more information beneficial for classification. When continuum images enter ResNet50, their simple structure provides relatively less classification information compared to magnetograms.

Since ResNet50 has strong classification capability, comparing ResNet50 and dual-stream ResNet50 results shows that models using dual-channel images have decreased classification capability compared to those using magnetograms alone. This indicates that when continuum and magnetogram images form a dual-channel input, they influence each other, negatively affecting the formation of feature vectors needed for classification. Finally, using magnetogram-based ResNet50 as the baseline and comparing dual-channel ResNet50 and dual-stream ResNet50 results reveals that both incorporate continuum images. However, the dual-channel model' s F1-score decreases while dual-stream ResNet50' s F1-score increases, demonstrating that dual-stream ResNet50 is superior to other models in terms of total data utilization and feature extraction.

According to literature [6], model F1-scores using continuum, magnetogram, and dual-channel images were 94.20%, 87.22%, and 88.61% respectively, all higher than dual-stream ResNet50. Possible reasons for this discrepancy include: (1) Data quality differences—literature [6] used sunspot group images within $±$75 degrees of solar disk center longitude with low background noise, while SOLAR-STORM1 contains some limb sunspot groups and noisy images; (2) Different data splitting methods—literature [6] used random splitting, which may result in highly similar images appearing in both training and test sets when sunspot groups evolve slowly, whereas this paper splits by object, ensuring images from the same sunspot group do not appear in both sets; (3) Different training methods—literature [6] performed a single data split, while this paper uses ten-fold cross-validation with averaged results.

Table shows the confusion matrix for dual-stream ResNet50. The results indicate that Alpha and Betax are more frequently misclassified as Beta, while Beta is misclassified as Alpha at twice the rate it is misclassified as Betax.

Taking Alpha class object 7169 as an example (Figure [Figure 6: see original paper]), the object appears as a unipolar sunspot group on October 27 and 29, correctly classified by the model. However, on October 28, another sunspot appears at the upper right of the group. After this sunspot' s features are extracted by the model, the entire image is classified as Beta. Therefore, transient sunspots appearing during evolution may be one reason for misclassification.

Taking Beta class object 6242 as an example (Figure [Figure 7: see original paper]), the sunspot group shows clear bipolar characteristics in its early stage, but the right-side sunspot gradually dissipates during evolution while the left-side sunspot remains. Consequently, images from the late stage are easily classified as Alpha when input to the model. This formation of ambiguous type features due to evolution may be a primary reason for suboptimal classification accuracy.

Taking Betax class object 6975 as an example (Figure [Figure 8: see original paper]), the object gradually approaches the solar limb in its late stage. Due to projection effects and limb darkening, the continuum image features become blurred and magnetic field measurement accuracy decreases. Therefore, reduced image quality when sunspot groups approach the solar limb may also contribute to suboptimal classification accuracy.

The SOLAR-STORM1 dataset in FITS format is 7GB with slow read speeds. After conversion to JPG format, the file size is 600MB with faster read speeds, while sunspot group features are preserved. Therefore, the JPG-converted dataset was used in experiments.

Advances in observational conditions have provided massive amounts of solar data. This paper addresses the characteristics of the SOLAR-STORM1 dataset by employing a dual-stream CNN method for sunspot group magnetic type classification and conducting comparative experiments with other methods. Experimental results demonstrate that dual-stream CNN achieves better classification performance for sunspot group magnetic types. Compared with single-channel methods, dual-stream CNN avoids data loss from using only one image type and maximizes data utilization. Compared with dual-channel methods, dual-stream CNN avoids mutual interference between continuum and magnetogram images during feature extraction. With the same data source, deep models have stronger feature extraction capability than shallow models. When the algorithm has strong feature extraction capability, using magnetograms alone yields higher classification accuracy than using continuum images alone, while using dual-channel images yields accuracy between the two.

Although dual-stream CNN demonstrates good classification performance, misclassifications occur due to feature dissipation during evolution for some objects and limb darkening effects in some images. Future work will focus on further optimizing dual-stream CNN to address these two issues.

References

[1] Liang Bo, Lin Yuqi, Dai Wei, Feng Song, Yang Yunfei. Prediction and Analysis of Sunspot Activity Based on Multivariable LSTM Network[J]. Astronomical Research & Technology, 2020, 17(03): 322-333.

[2] Zhu Jian, Yang Yunfei, Su Jiangtao, Liu Haiyan, Li Xiaojie, Liang Bo, Feng Song. A Detection and Tracking Method for Active Regions Based on Deep Learning[J]. Astronomical Research & Technology, 2020, 17(02): 191-200.

[3] Liu Hui, Ji Kaifan, Jin Zhenyu. The application of machine learning in solar physics (in Chinese). Sci Sin-Phys Mech Astron, 2019, 49: 105-117.

[4] T. Colak and R. Qahwaji, "Automated mcintosh-based classification of sunspot groups using MDI images," Solar Physics, vol. 248, no. 2, pp. 277–296, 2008.

[5] M. A. Abd, S. F. Majed, and V. Zharkova, "Automated classification of sunspot groups with support vector machines," in Proceedings of the Technological Developments in Networking, Education and Automation, pp. 321–325, 2010.

[6] Fang Y, Cui Y, Ao X. Deep Learning for Automatic Recognition of Magnetic Type in Sunspot Groups[J]. Advances in Astronomy, 2019, 2019(123): 1-10.

[7] Gao Yang. Fighting Behavior Detection In Surveillance Video By Two-Stream Convolutional Networks[D]. Xi' an University of Technology, 2018.

[8] The dataset of sunspot group magnetic type provided by Space Environment Warning and AI Technology Interdisciplinary Innovation Working Group[FANG Y, et al., 2019].

[9] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE

Computer Society, 2016.

Note: Figure translations are in progress. See original paper for figures.