Sunspot Magnetic Type Classification Based on Dual-Model Ensemble (Postprint)

Authors: Chen Qingyuan, Jin Fan, Feng Dehua, Wang Yunlong, Liang Yijun

Date: 2022-04-14T00:00:00+00:00

Abstract

Sunspots are phenomena occurring in the solar photosphere and hold significant importance for solar flare prediction. To address the three-class sunspot dataset with imbalanced class sample sizes, a dual-model ensemble classification algorithm is proposed. This method employs one lightweight and one heavyweight model, each responsible for the classification task of two distinct classes, and subsequently integrates their classification results to deduce the classification outcome for the third class. Experimental results indicate that this method can reduce the adverse impacts arising from overfitting and underfitting in individual models on imbalanced datasets, resolving the class imbalance issue of the sunspot dataset from a novel perspective, with an average F1 score reaching 0.931.

Full Text

Classification of Sunspot Magnetic Types Based on Dual Model Integration

Chen Qingyuan[1], Jin Fan[2], Feng Dehua[1], Wang Yunlong[1], Liang Yijun[1],*

[1]Xi' an Jiaotong University, Xi' an 710049, China
[2]Zhejiang University of Science and Technology, Zhejiang 310023, China

Abstract

Sunspots are phenomena occurring in the solar photosphere that hold significant importance for solar flare prediction. To address the class imbalance problem in a three-category sunspot dataset, this paper proposes a dual-model ensemble classification algorithm. The method employs two models—one lightweight and one heavyweight—each specialized for classifying two specific categories, then integrates their results to "squeeze out" the classification for the third category. Experiments demonstrate that this approach mitigates the adverse effects of overfitting and underfitting that occur when single models are trained on imbalanced datasets, offering a novel solution to the class imbalance problem in sunspot datasets and achieving an average F1-score of 0.931.

Keywords: sunspot classification; dual model integration; class imbalance; overfitting; underfitting

Introduction

Sunspots are closely related to solar activity, with solar flares representing the most intense manifestation of such activity. These flares primarily erupt in the atmosphere above sunspots, disrupting Earth' s ionosphere and affecting short-wave radio communications while producing hazardous phenomena such as geomagnetic storms. Observatories worldwide continuously track and classify all visible sunspot groups to enable early flare detection. Accurate automated classification of sunspot groups would enhance our ability to monitor the formation of certain categories, improve flare warning capabilities, and deepen our understanding of solar cycles, space climate, and their impacts on Earth' s climate system.

Traditional classification methods fall into two main categories: conventional digital image processing algorithms, mathematical morphology, and wavelet analysis; and the currently prevalent data-driven machine learning approaches. With advances in observational capabilities, solar activity data has grown rapidly, making deep learning methods increasingly advantageous.

Colak et al. [3] proposed a hybrid system using the McIntosh classification scheme, employing active region data extracted from SOHO/MDI magnetogram images to automatically detect sunspot groups in SOHO/MDI white-light images. After detection, magnetogram images were used for grouping/clustering, with image processing and neural networks integrated for automated classification. However, this system suffered from grouping errors and missed detection of small sunspots. In subsequent work [4], machine learning combined with traditional image processing improved feature extraction but demonstrated poor generalization across different datasets.

Abd et al. [5] adopted a modified seven-class Zurich classification scheme, using Support Vector Machines (SVM) for automated classification of sunspot groups in full-disk white-light images. Their preprocessing involved edge detection, noise removal, and binarization to segment sunspot groups from the solar disk, followed by unsupervised segmentation to merge spots belonging to the same group, feature extraction, and SVM classification. This method' s accuracy depended heavily on image quality and distortion levels, with the segmentation process significantly impacting inference time.

With deep learning development and increasing training data availability, convolutional neural networks have been widely applied to image classification [6,7], segmentation [8,9], and detection [10,11]. Fang et al. [12] used CNNs to classify sunspot group magnetic types, preprocessing images into three categories: white-light, magnetogram, and composite images. Using each as input sources, they found white-light images alone yielded the best results. The authors attributed poor magnetogram performance to their complex structure relative to white-light images and inadequate feature extraction by CNNs. Additionally, class imbalance remained unresolved, with Beta-x category overfitting.

Yang et al. [13] proposed a dual-stream CNN approach for sunspot magnetic type classification, achieving high accuracy on the Alpha category but suffering from large model size and persistent Beta-x overfitting. Even state-of-the-art detection models like Fu et al.' s SunspotsNet [14] face overfitting on minority classes due to class imbalance.

Our proposed dual-model ensemble algorithm addresses this by assigning separate models to Alpha and Beta-x categories—designing optimal architectures and training strategies for each—complemented by extensive tuning techniques for imbalanced datasets. The final integration of both models' results maximally resolves class imbalance. Unlike mainstream ensemble methods [15,16,17], our two models have explicit task divisions, fewer ensemble components, and superior performance. Notably, this approach secured second place in the Alibaba Tianchi Solar Storm Recognition and AI Challenge (https://tianchi.aliyun.com/competition/entrance/531803/rankingList), with code available at https://github.com/qingyuanchen1997/Dual-Model-Integration.

1.1 Dataset Introduction and Analysis

Deep learning is data-driven, necessitating thorough dataset analysis. This study uses the SOLAR-STORM1 dataset, provided by the Space Environment AI Early Warning Innovation Workshop and publicly available on Tianchi Lab (https://tianchi.aliyun.com/dataset/). Based on the Mount Wilson magnetic classification scheme, which categorizes sunspot groups by magnetic polarity, the dataset comprises three categories: Alpha (unipolar groups), Beta (bipolar groups with simple, clear polarity division), and Beta-x (complex bipolar groups without clear polarity division), totaling 14,469 samples. Each sample contains two co-registered images: a magnetogram and a white-light image.

[Figure 1: see original paper] shows magnetic field (top row) and white-light (bottom row) observations for Alpha (leftmost), Beta (second left), and more complex Beta-x (third and fourth left) magnetic types.

The dataset contains 4,709 magnetogram and white-light image pairs for Alpha (32.54%), 7,353 for Beta (50.82%), and 2,407 for Beta-x (16.64%). While the overall volume is substantial, inter-class distribution is highly imbalanced, with Beta-x significantly underrepresented compared to Alpha and Beta. Moreover, real-world observations confirm that Alpha and Beta categories occur far more frequently than Beta-x.

Proportion of Alpha, Beta and Beta-x sunspot data samples in the training set

1.2 Dual-Model Integration Algorithm

Given the dataset's scale (14,469 samples), we divided it into training, validation, and test sets at a 3:1:1 ratio to ensure reliable evaluation. Due to temporal continuity (96-minute observation intervals) and the slow evolution of sunspot groups, the same group often maintains its magnetic type over extended periods with high similarity. To prevent information leakage between sets, we performed temporal-based partitioning, placing all samples from the same time period into a single set. All experiments employ five-fold cross-validation: the dataset is split into five parts, with one part each for validation and testing and the remaining three for training. Each experiment runs five times with different fold combinations, with final results averaged. For dual-model integration feasibility, both models use identical datasets in each training round.

Initial experiments using a ResNet50 backbone revealed significant overfitting on Beta-x. shows F1-scores on training and validation sets: Beta-x' s training F1-score far exceeds its validation score, while Alpha and Beta show only slight differences, confirming overfitting on the minority class.

Using the lighter AlexNet backbone mitigated Beta-x overfitting and improved its validation F1-score substantially . However, the reduced parameters weakened fitting capacity, causing underfitting on Alpha and Beta categories with higher sample counts, yielding lower F1-scores than ResNet50.

Thus, with extreme class imbalance, no single network can simultaneously accommodate both high-sample and low-sample categories. Heavy-parameter networks require more training data and excel on majority classes, while lightweight networks need less data and perform better on minority classes. This establishes our dual-model integration rationale.

Furthermore, testing revealed mutual exclusivity between Alpha and Beta-x. shows that across all models, genuine Alpha samples are rarely misclassified as Beta-x and vice versa; errors primarily involve confusion with Beta. Visualizing extracted three-dimensional features [Figure 2: see original paper] confirms that Alpha and Beta-x sample points are widely separated in feature space with minimal overlap. This validates dual-model integration feasibility—results from different models for Alpha and Beta-x do not interfere, preserving each model' s accuracy on its specialized category.

Based on these findings, we designed separate heavy and light models: the heavy model leverages strong fitting capacity for high-sample Alpha classification, while the light model uses anti-overfitting properties for low-sample Beta-x classification. Leveraging their mutual exclusivity, we integrate results without precision loss, using the heavy model' s Alpha predictions and light model' s

Beta-x predictions to "squeeze out" Beta—assigning unlabeled samples to Beta category [Figure 3: see original paper].

We designed paraResNet (Parallel-ResNet18-D) based on ResNet18 as the heavy model for Alpha, and miniAlexNet based on AlexNet as the lightweight model for Beta-x.

1.3.1 Network Structure

For the Alpha classification task, each sunspot group category includes both magnetogram and white-light images [Figure 4: see original paper]. Initial attempts to concatenate 8-bit single-channel images into 16-bit depth for ResNet18 proved ineffective due to texture differences and unnatural image structure. Therefore, inspired by Yang et al. [13], we designed a dual-channel parallel ResNet18 network to extract features from both modalities. The two parallel channels process white-light and magnetogram images separately, concatenating extracted features before feeding them to fully connected layers [Figure 5: see original paper].

Referencing He et al. [18], we improved ResNet18' s downsampling by replacing the original convolution layer (which simultaneously reduced spatial dimensions and changed tensor depth) with an average pooling layer for spatial reduction and a stride-1 convolution for depth transformation. This task separation reduces the original convolution kernel' s burden and improves downsampling precision. We term this architecture Parallel-ResNet18-D (paraResNet) [Figure 6: see original paper].

1.3.2 Training Strategy

Training employed two single-channel ResNet18 models pretrained separately on white-light and magnetogram images (without ImageNet pretraining), then transferred their convolutional layers to the dual-channel model. Weighted cross-entropy loss assigned higher weights to minority classes to mitigate imbalance effects.

For data augmentation, only horizontal and vertical mirroring were applied to Alpha and Beta categories, as these preserve image information without compromising accuracy. Additional augmentation methods would increase training time without benefit.

1.4.1 Network Structure

For the Beta-x category, we used AlexNet' s convolutional layers as the backbone to reduce parameters. Input consisted of mixed images: Channel 1 for white-light, Channel 2 for magnetograms, and Channel 3 as the weighted sum (0.5 each) of both modalities to guide the network in learning texture correlations while filling the channel requirement for ImageNet-pretrained models. To further reduce fully connected layer parameters, a

3$×3adaptivepoolinglayerwasaddedafterthefinalconvolution,followedbyasinglefullyconnectedlayeroutputt classconfidencescores.Higherresolutionimages(500×$375) compensated for the network' s reduced feature extraction capacity [Figure 7: see original paper].

1.4.2 Training Strategy

Data cleaning was performed first. While Alpha and Beta data with poor temporal continuity were retained entirely, Beta-x data was temporally sorted and downsampled at 3/4 rate based on strong temporal continuity, significantly reducing redundancy and preventing useless learning from repeated information.

Given the subtle differences between Beta-x and Beta, temperature scaling was applied—reducing original logit values before softmax—to increase inter-class differences during learning, reducing training pressure and accelerating convergence.

Data augmentation employed random rotation within 90 degrees, which, while more aggressive, provided superior anti-overfitting benefits for the small dataset.

Analysis of validation set logit distributions revealed that when the difference between Beta-x and Beta logit values fell within [0, 0.5], 7% of Beta samples were misclassified as Beta-x. Therefore, probabilistic correction was implemented: when this logit difference was in [0, 0.5], the Beta logit was increased.

1.5 Model Integration

Based on the mutual exclusivity between Alpha and Beta-x discussed in Section 1.3, we fuse paraResNet and miniAlexNet results by adopting the former's Alpha predictions and the latter' s Beta-x predictions, assigning unlabeled samples to Beta category to derive the final three-class classification.

2.1 Parameter Settings

For Alpha model training, we used the Adam optimizer [20] with an initial learning rate of 2e-6. As an SGD extension, Adam adapts gradient coefficients based on parameter history, making it suitable for large datasets. Weighted cross-entropy loss used weights of Alpha: 1.56, Beta: 1.0, and Beta-x: 3.05, set based on training set proportions and fine-tuned using validation performance to reduce class imbalance effects at the loss level.

For Beta-x model training, we used SGD with momentum (learning rate 0.0008, momentum 0.9). Two additional preprocessing steps were applied: (1) input resolution increased to 500$×$375 to preserve information given Beta-x' s small sample size; (2) redundancy filtering via temporal downsampling of Beta-x sequences at stride 4, retaining 891 training samples.

2.2 Evaluation Metrics

We adopt F1-score as the evaluation metric for the three-class sunspot classification task. As the harmonic mean of precision and recall, F1-score ranges from 0 to 1, with values approaching 1 indicating better performance. The calculation is:

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

The dataset contains three categories (Alpha, Beta, Beta-x), each with its own F1-score. Since Beta sunspots occur most frequently in reality, we prioritize Beta' s F1-score, followed by Alpha and Beta-x.

2.3 Ablation Study

This paper employs a specialized dual-model fusion algorithm for training data imbalance: a dual-channel parallel ResNet18 for Alpha classification and miniAlexNet for Beta-x classification, integrated based on their mutual exclusivity to produce final three-class results. Evaluation follows competition requirements using F1-score to measure test set performance.

2.3.1 Alpha-model Ablation studies validate each strategy' s effectiveness for the Alpha model . Every technique significantly improves performance: mirroring-based data augmentation (horizontal + vertical) mitigates limited training data with minimal information loss; the dual-channel architecture better extracts features from both modalities while additional parameters enhance fitting capacity; loss weighting balances Alpha and Beta training volumes; improved downsampling boosts feature extraction effectiveness. Most critically, the specialized paraResNet achieves excellent Alpha classification.

2.3.2 Beta-x-model Ablation studies for the Beta-x model show similar improvements: random rotation augmentation alleviates small-data overfitting risks; higher-resolution inputs compensate for the small network' s limited feature extraction; temperature scaling and probabilistic correction improve complex boundary fitting through manual adjustment; the compact miniAlexNet architecture with fewer parameters outperforms dual-channel networks on the small Beta-x dataset.

2.3.3 Dual Model Integration Final integration of paraResNet and miniAlexNet combines the former' s Alpha results with the latter' s Beta-x results, assigning unlabeled samples to Beta category . The fused model fully preserves paraResNet' s Alpha F1-score and miniAlexNet' s Beta-x F1-score, yielding an even higher Beta F1-score through "squeezing." Final test set F1-scores are 0.970 for Alpha, 0.946 for Beta, and 0.877 for Beta-x.

2.4 Comparison with Other Algorithms

We evaluated Fang et al. [12], Yang et al. [13], and mainstream ResNet architectures [21] on SOLAR-STORM1 with mirroring augmentation and loss weighting .

Results show our paraResNet achieves exceptional Alpha performance (F1-score 0.970), while miniAlexNet excels on Beta-x (F1-score 0.877). Integration yields optimal Beta classification (F1-score 0.946), surpassing all single-network models.

Conclusion

This paper analyzes sunspot datasets and proposes a dual-model ensemble algorithm. By training specialized paraResNet for Alpha and miniAlexNet for Beta-x, then fusing results to "squeeze out" Beta classification, we avoid heavy-model overfitting on minority classes and light-model underfitting on majority classes while preserving each model' s strengths. This novel approach mitigates class imbalance effects, achieving superior average F1-score on SOLAR-STORM1 compared to all single-network models.

The dual-model ensemble algorithm is applicable beyond sunspot classification to any imbalanced three-class task, offering 启发性意义 for machine learning classification problems with uneven class distributions.

Future Work

To reduce model complexity, parameter count, inference time, and deployment difficulty, future research will explore knowledge distillation and few-shot learning to enable single-network models that simultaneously perform well on both majority and minority classes, potentially replacing the dual-model architecture.

References

[1] Liang Bo, Lin Yuqi, Dai Wei, et al. 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, et al. A Detection and Tracking Method for Active Regions Based on Deep Learning[J]. Astronomical Research & Technology, 2020, 17(02):191-200.

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

[4] Colak T and Qahwaji R. Automated solar activity prediction: a hybrid computer platform using machine learning and solar imaging for automated prediction of solarflares[J]. Space Weather, 2009, 7(6).

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

[6] Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, and Rabinovich A. Going deeper with convolutions[C]// Computer Vision and Pattern Recognition. 2015, Boston, United States: IEEE: 1-9.

[7] Zhang Ke, Feng Xiaohan, Guo Yurong, Su Yukai, Zhao Kai, Zhao Zhenbing, Ma Zhanyu and Ding Qiaolin. Overview of deep convolutional neural networks for image classification[J]. Journal of Image and Graphics, 2021, 26(10):2305-2325.

[8] Ronneberger O. U-Net Convolutional Networks for Biomedical Image Segmentation[C]// Bildverarbeitung für die Medizin 2017. 2017, Berlin, Germany: Springer Berlin Heidelberg: 3-3.

[9] Zhou Tao, Dong Yali, Huo Bingqiang, Liu Shan and Ma Zongjun. U-Net and its applications in medical image segmentation: a review[J]. Journal of Image and Graphics, 2021, 26(09):2058-2077.

[10] Cai Z W and Vasconcelos N. Cascade R-CNN: Delving Into High Quality Object Detection[C]// Computer Vision and Pattern Recognition. 2018, Salt Lake City, United States: IEEE: 1483-1498.

[11] Zhao Yongqiang, Rao Yuan, Dong Shipeng P and Zhang Junyi. Survey on deep learning object detection[J]. Journal of Image and Graphics, 2020, 25(04): 0629-0654.

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

[13] Yang Wugu, Tian Weixin, Zeng Shuguang and Huang Yao. A magnetic classification method of sunspot group based dual-stream CNN[J]. Astronomical Research Technology, 2021, https://doi.org/10.14005/j.cnki.issn1672-7673.20210823.003.

[14] Fu Xiaona, Liao Chengwu, Bai Xianyong, Liang Bo, Feng Song, Yang Hongjuan and Yang Yunfei. Sunspot detection method based on lenet-5 convolutional neural network[J]. Astronomical Research & Technology, 2018, $1672-7673(2018)\ 03-0340-07$.

[15] Chen T Q and Carlos G. XGBoost: A Scalable Tree Boosting System[C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016, San Francisco, United States: ACM: 785-794.

[16] Breiman L. Bagging predictors[J]. Machine Learning. 1996, 24: 123-140.

[17] Breiman L. Random Forest[J]. Machine Learning. 2001, 45(1): 5-32.

[18] He T, Zhang Z, Zhang H, Zhang Z Y, Xie J Y and Li M. Bag of Tricks for Image Classification with Convolutional Neural Networks[C]// Computer Vision and Pattern Recognition. 2019, Long Beach, United States: IEEE: 558-567.

[19] Krizhevsky A, Sutskever I, and Hinton G. ImageNet Classification with Deep Convolutional Neural Networks// Neural Information Processing Systems. 2017, New York, United States: ACM: 84-90.

[20] Kingma D P and Ba J. Adam: a method for stochastic optimization[EB/OL][R]. 2017, [2020-07-11]. https://arxiv.org/pdf/1412.6980.pdf.

[21] Zhang X Y, Ren S Q and Sun J. Deep Residual Learning for Image Recognition[C]// Computer Vision and Pattern Recognition. 2016, Las Vegas, United States: IEEE: 770-778.

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