Advances in Coronal Mass Ejection Detection (Postprint)

Authors: Guo Min, Li Jingtao, Shang Zhenhong, Liu Hui, XIAN Xianggui, YANG Zhipeng

Date: 2019-09-18T00:00:00+00:00

Abstract

Coronal Mass Ejection (CME) is a large-scale, highly intense eruptive phenomenon and the primary solar eruptive activity affecting Earth. Since this eruptive phenomenon can cause severe disturbances to Earth' s environment, the detection of CMEs is of great significance for forecasting hazardous space weather. To more clearly organize the existing CME detection methods, this article will analyze and summarize typical approaches. First, it introduces coronal mass ejections and their characteristics; then, it provides an overview and analysis of CME detection from both manual methods and automatic detection methods; finally, it discusses some existing problems with current algorithms and proposes future research directions.

Full Text

Research Progress on Coronal Mass Ejection Detection

Guo Min, Li Jingtao, Shang Zhenhong*, Liu Hui, Xian Xianggui, Yang Zhipeng
(School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, Yunnan 650500, China)

Abstract

A Coronal Mass Ejection (CME) is a large-scale, violently explosive phenomenon that represents the primary solar outburst activity affecting Earth. Because such explosive events cause severe interference with the terrestrial environment, CME detection is of great significance for forecasting hazardous space weather. To more clearly organize the existing CME detection methods, this paper analyzes and summarizes typical approaches. First, we introduce coronal mass ejections and their characteristics. Then, we provide an overview and analysis of CME detection from two perspectives: manual methods and automatic detection methods. Finally, we discuss some problems with current algorithms and propose future research directions.

Keywords: solar eruption activity; coronal material; CME detection

---

CME is a dynamic event that ejects plasma from the Sun's corona into interplanetary space. First observed during the Skylab mission in the early 1970s, it is the largest and most energetic eruptive phenomenon originating from the Sun and can be observed in the extended corona through white-light coronagraphs [1]. As CMEs represent enormous eruptions of magnetized plasma, they can propagate at speeds of up to several thousand kilometers per second [2]. Given that such eruptive activities trigger solar energetic particle events and geomagnetic storms that affect aviation safety, satellite operations, communication systems, and power facilities, it is necessary to detect CMEs to provide early warning signals for space weather forecasters, enabling them to take appropriate measures to avoid unnecessary losses [3].

Furthermore, statistical information about CMEs is crucial for better understanding their nature. The concept of CMEs has remained somewhat ambiguous, and this definitional uncertainty leads to variations in CME characteristics and quantities obtained by different detection methods. Hundhausen et al. defined CMEs as new, discrete, bright white-light features appearing in the coronal field of view on timescales ranging from minutes to hours [4]. Since CME phenomena share similar characteristics with other solar activities, Schwenn et al. refined the definition to: a new, discrete, bright white-light feature appearing in the coronal field of view with radially outward velocity. Figure 1 shows a coronal mass ejection eruption process observed by the SOHO/LASCO satellite from the National Space Science Center (NSSC).

Funding: National Natural Science Foundation of China (61462052, 11873027)
Received date: Year-Month-Day; Revised date: Year-Month-Day
Author introductions: Guo Min, female, Master' s degree. Research direction: computer vision and image processing. Email: gmin172@163.com
Corresponding author: Shang Zhenhong, male, Associate Professor. Research direction: computer vision and image processing. Email: shangzhenhong@126.com

With the launch of satellites such as SOHO, Wind, and STEREO, CME detection and identification have become possible. In particular, the LASCO coronagraph aboard the SOHO satellite, launched in 1995, can observe the CME occurrence process and obtain CME image data. Through these data, researchers can detect the motion state of CMEs and derive descriptive parameters (position angle, angular width, velocity, etc.). Currently, significant progress has been made in CME detection research both domestically and internationally, with many CME catalogs compiled to provide benchmarks and references for further study.

To facilitate readers' detailed understanding of current CME detection research progress and provide reference for related research, this paper reviews some domestic and international methods for CME detection, presenting the specific identification processes of representative approaches. For clarity, we categorize current CME detection methods into two major classes: manual identification methods and automatic detection methods. Additionally, based on selected features and employed techniques, we further divide automatic detection methods into four categories for detailed overview: grayscale feature-based, texture feature-based, optical flow-based, and learning-based methods.

This paper is organized into four sections: Section 1 introduces the definition of CMEs and the significance of CME detection; Section 2 describes CME characteristics, including velocity, angular width, and occurrence patterns; Section 3 categorizes and summarizes existing CME detection methods; Section 4 provides summary and outlook, discussing challenges in current CME detection and exploring future research trends. The contributions of this paper are: (1) No similar literature currently exists, providing researchers with the state-of-the-art in this field; (2) It enables researchers to understand in detail the processes of CME detection using different methods; (3) It proposes future research trends addressing current challenges in CME detection, serving as a reference for researchers' future directions.

2 CME Features

CME detection essentially involves detecting CME events through manual labeling or computer automatic identification to obtain CME characteristics. Commonly used CME features include velocity, angular width, occurrence patterns, etc.

2.1 CME Velocity

CME is a dynamic process, and its velocity is an important physical characteristic. Since observed CMEs are typically their projections on the sky plane, CME velocity generally refers to the motion velocity of the CME' s fastest leading edge [5], which is obtained through linear fitting of the "time-height" data of the CME front. Figure 2 [Figure 2: see original paper] shows a time-height plot of a CME occurring at 13:25:42 on July 2, 2017, from the CDAW catalog. CME velocities range from 10-3500 km/s, with those >1000 km/s generally considered fast CMEs and those <100 km/s considered slow CMEs [6].

2.2 CME Angular Width

CME size is generally measured by angular width (angular breadth), which is the angle formed by the two edges of the CME region, numerically equal to the difference in position angles between the two edges. Figure 3 [Figure 3: see original paper] shows a CME detected by CACTus software, where a and b represent the position angles of the CME region at a certain moment, and the angular width is the angle between the white solid lines. CME angular widths range from a few degrees to 360 degrees; a 360-degree CME is called a Halo CME, with widths >120 degrees generally considered wide CMEs and widths <30 degrees considered narrow CMEs [6].

2.3 Relationship Between CMEs and Sunspot Numbers

CME occurrence patterns are related to the solar sunspot cycle, with a rate of 0.5 per day during solar minimum and 2-6 per day during solar maximum. Figure 4 [Figure 4: see original paper] shows the number of CMEs recorded in the CDAW catalog between 2008 and 2017, while Figure 5 [Figure 5: see original paper] presents sunspot numbers over the past 13 years from the Sunspot Index and Long-term Solar Observations (SILSO). We can see that during the overlapping period (2008-2017), the occurrence frequency of CMEs is basically consistent with that of sunspots, demonstrating that CME occurrence patterns are related to the solar sunspot activity cycle.

3 CME Detection Methods

For over a decade, researchers have used various methods to detect CME events and their characteristics (e.g., velocity, angular width, occurrence patterns), compiling CME catalogs for scientific study. Current CME detection methods are mainly divided into two categories: (1) Manual CME identification: Currently, there are two well-known catalogs—the CDAW catalog and the Naval Research Laboratory (NRL) catalog. This method relies on observers manually recording data daily for CME cataloging, which is time-consuming. (2) Automatic CME detection: Since CMEs are dynamic processes with bright intensity features and complex texture structures, we can categorize existing automatic detection methods into four classes based on selected features and techniques, as summarized in Table 1 .

3.1 Comparison Between Manual and Automatic Detection Catalogs

CMEs are related to many phenomena such as flares, solar energetic particles, and geomagnetic storms. Therefore, detecting CME phenomena and compiling event catalogs is very important for researchers worldwide to understand these phenomena and conduct related work. CME catalogs record detected CME event characteristics in data and graphical forms, with main parameters including velocity, angular width, and position angle. Based on the detection methods described above, we can obtain two types of CME catalogs: manual identification catalogs and automatic detection catalogs. By comparing characteristics (velocity and angular width) using the CDAW catalog and CACTus catalog as representatives [6], we identify the features of these two catalog types: (1) Identification of narrow CMEs (W<30°): Studies show that CACTus detects far more narrow CMEs than CDAW during the same period, mainly because the CDAW catalog misses many narrow CMEs during solar maximum years. (2)

Identification of fast CMEs (V>1000 km/s): Although CACTus detects more fast CMEs than CDAW during the same period, only 6% of the detected CMEs are fast among the 32% true CMEs detected.

Therefore, we can conclude that manual and automatic detection catalogs each have advantages and disadvantages. Since manual catalogs are based on human labeling, they inevitably miss many narrow CMEs. While automatic detection catalogs identify more fast CMEs than manual catalogs, their accuracy is not optimal.

3.2 Manual CME Identification

Since the discovery of CME phenomena in 1971, CMEs have been observed by various space-borne instruments. Since the launch of the SOHO spacecraft in 1995, CME observations have primarily used the LASCO coronagraph. LASCO CME identification and cataloging is an important task that provides fundamental knowledge for further scientific research.

The NRL catalog is compiled by LASCO observers who examine sequences of LASCO coronal images and record daily events. This is a preliminary catalog providing information about CME timing and approximate location.

The CDAW catalog is generated and maintained by NASA, the Catholic University of America, and the Naval Research Laboratory at the CDAW Data Center. Each labeled CME is identified by its occurrence date and time. The catalog provides measurements of CME properties, including velocity and angular width, as well as "time-height" plots for determining CME velocity. In addition to data and graphs, the catalog includes CME image sequences from coronagraphs available for download and viewing.

Since both the NRL and CDAW catalogs are manually identified by observers, even for the same observer, identification capability is not constant. Therefore, the provided events and measured parameters are significantly influenced by human subjective factors, and the cataloging process is also very time-consuming.

3.3 Automatic CME Detection Methods

Traditional CME detection methods are based on human observation, which is inefficient and susceptible to individual subjective factors. With the rapid development of automatic recognition technology in recent years, many algorithms for automatic CME detection have emerged, providing new methods for CME detection research. Since CMEs have different morphologies, scales, and features, typically appearing as bright, texture-complex enhanced structures trailing a dimmer dark region [3], we can categorize existing automatic CME detection methods into the following four classes.

3.3.1 Grayscale Feature-Based Methods Grayscale features represent image brightness based on pixel values. Unlike background grayscale features, CMEs have brighter structures, typically exhibiting bright white-light features. Therefore, detecting CMEs by extracting grayscale features is a straightforward approach. Currently, research on CME detection based on grayscale features is relatively mature, with many researchers achieving good results using this method. The Computer Aided CME Tracking Software Package (CACTus) and the Solar Eruptive Event Detection System (SEEDS) are two representative detection methods.

The CACTus method was proposed by Berghmans et al. [7, 8]. This method completes preprocessing through filtering and denoising, polar coordinate transformation, inter-frame differencing, and integrating LASCO C2 and C3 images. By integrating difference images [, r] at different times t, a [t, , r] data cube is obtained. A ridge line in the slice [t, r] along a specified angle in this cube represents a CME' s ejection performance along that angle. To clearly see the ridge line, the slice image color is inverted to obtain inclined dark ridge lines, as shown in Figure 6 [Figure 6: see original paper]. Since the Hough transform [26] is an image processing technique that can detect line segments from images, it is used to detect ridge lines in [t, r] slices. Meanwhile, the ridge lines detected by Hough transform are CMEs along a single angle slice. To obtain complete CMEs, images are projected into a [v, , t] data cube, where velocity v can be calculated from ridge segment length. At this point, CME identification is converted into a point clustering problem in the [v, , t] cube. Since CACTus assumes that radial velocities along different directions are similar for the same CME, and v-direction information contributes little to clustering, to reduce computational complexity, [v, , t] is integrated along the v-direction, converting clustering to a two-dimensional space [, t] to obtain cluster positions in the CME overview map. These positions represent the start time and angle of CME occurrence, with the vertical length of clusters indicating CME eruption duration.

CACTus software was the first to implement CME detection using automatic methods. Compared with manual identification, this method is fast and has good detection accuracy, detecting about 75% of CMEs in the CDAW catalog and also detecting weak CMEs not found in the CDAW catalog. Currently, CACTus software has compiled an online catalog from 1996 to the present.

The SEEDS method was proposed by Olmedo et al. [9, 10], using image segmentation technology based on a region-growing algorithm. Similar to CACTus, this method completes preprocessing through normalized input images, filtering and denoising, and polar coordinate transformation. Then, the [, r] image sequence obtained from polar coordinate transformation is projected along the direction to obtain signal intensity curves, as shown in Figure 7 Figure 7: see original paper. Next, threshold processing is applied to obtain the center and brightest part of the CME—the core angle—while region-growing algorithm is used on the core angle to obtain the entire CME angular width. CME leading edge determination is achieved by projecting images along the r direction to obtain signal intensity curves, as shown in Figure 7(b), where curve peaks correspond to Max-Height, and half of the curve peak corresponds to Half-Max-Lead (leading edge) and Half-Max-Follow (trailing). By comparing two consecutive frames, if the leading edge, trailing edge, and their difference in the later image are higher than in the previous image, it indicates CME expansion. Tracking stops when the leading edge exits the field of view or the CME becomes too dim to identify, at which point the next CME is sought.

In terms of detection accuracy, similar to CACTus, SEEDS can detect about 75% of CMEs in the CDAW catalog. However, in terms of total detected CMEs, the number detected by CACTus and SEEDS is more than double that of the CDAW catalog. Additionally, the SEEDS method can obtain CME leading edge contours, which can be viewed in the SEEDS catalog.

Besides the two methods described in detail above, there are many other CME detection methods based on grayscale features. For example, Qu et al. [11] proposed a CME detection method based on Support Vector Machine (SVM). To distinguish CMEs from other structures, this algorithm generates difference images and segmented images through preprocessing, then uses threshold methods again to segment these preprocessed images, with the final segmentation result being the sum of the two segmentation results. All segmented regions are then treated as CME candidate regions, and thresholds are set to classify candidate regions into CME regions and background. Finally, an SVM classifier is used in CME regions to distinguish strong CMEs from weak CMEs. Compared with other methods, this method' s greatest advantage is its ability to accurately and quickly detect both strong CMEs and most weak CMEs.

Unlike the three methods above that convert LASCO images to differentially processed polar coordinate maps, the ARTEMIS method proposed by Boursier et al. [12, 13] converts LASCO C2 images to synoptic maps. As shown in Figure 8 [Figure 8: see original paper], the synoptic map' s horizontal axis represents time, indicating CME outflow time, while the vertical axis represents latitude, indicating CME angular width. The algorithm uses a median filter to remove noise, then thresholds the synoptic map to return a two-dimensional mask defining the region of interest. Finally, prior knowledge of CME characteristics is introduced to correctly identify CMEs. The algorithm ultimately produces an ARTEMIS catalog listing each detected CME event and its main parameters. Comparison with the CDAW catalog reveals that this method' s advantage is its ability to detect some small, weak CMEs. Like CACTus and SEEDS, it detects more CME events than the CDAW catalog.

The four grayscale feature-based methods first preprocess LASCO images to obtain polar coordinate maps or synoptic maps, then apply thresholding and morphological processing to extract CME features, and finally calculate various CME properties through recognition or tracking of CME features. We can see that these methods share a common point: they all use brightness enhancement to highlight regions of interest in coronal images (i.e., CME regions). Meanwhile, all these methods detect more CMEs than the CDAW catalog when applied to LASCO images from the same period, yet none can detect 100% of CMEs in CDAW. This is related to the image preprocessing techniques, detection rules, threshold selection used in automatic detection methods, and the ambiguous definition of CMEs.

3.3.2 Texture Feature-Based Methods CMEs not only have obvious brightness structures but also complex texture structures. For some relatively dim CME regions with obvious texture structures, using texture features for detection is a good choice. Texture feature extraction generally requires setting a window region of certain size containing multiple pixels, from which texture features are extracted. This approach performs well especially when retrieving images with obvious coarseness, density, and other characteristics.

Gonzalez-Gomez et al. used wavelet analysis methods [14] to classify CMEs. This algorithm is based on frequency domain analysis. By placing a fixed-pixel window on LASCO C2 images (experimental data show the window can be placed anywhere on the image), convolution is performed on the window using a standard Fast Fourier Transform algorithm, while the Mexican Hat function is used as the basis function to obtain the wavelet spectrum of the window region. The resulting wavelet spectrum plot has scale on the horizontal axis and flux on the vertical axis. The curve representing the window region has an inflection point that distinguishes the speed of flux change with scale. The inflection point' s scale (ac) and flux (fc) are obtained by calculating the second derivative of each curve. As shown in Figure 9 [Figure 9: see original paper], in the ac-fc plot, the wavelet spectrum is clearly divided into two groups: low-flux, small spatial scale group (Homogeneous Group) and high-flux, large spatial scale group (Collimated Group), where box1 and box2 represent two windows on the image. Applying this method to a larger sample of CME images and plotting inflection point scale and flux values in the ac-fc plot, linear least squares fitting of these points can classify which group a CME image belongs to. This method is an image filtering analysis approach that uses high-pass and low-pass characteristics in the frequency domain to classify CME images, enabling direct classification of coronal mass ejections obtained by different satellites.

CME image features have multi-scale characteristics, and multi-scale image processing techniques are important for enhancing CME front visibility and suppressing noise. Since line segments better reflect image information, while wavelets are more suitable for identifying point-like features such as noise or background stars and not ideal for detecting CME feature line structures, Gallagher et al. [15] studied higher-order multi-scale techniques such as ridgelets and curvelets. Unlike wavelet transforms, ridgelet transforms first perform Radon transforms, replacing point parameters with line parameters, followed by wavelet transforms. This transform can effectively represent straight singularity features but cannot well represent image edge curves. Curvelet transforms compensate for ridgelet transforms' shortcomings and can optimally represent image curve singularity features. Wavelets, ridgelets, and curvelets are similar in that they all use inner products of basis functions and signals to achieve sparse signal representation. The difference is that ridgelet and curvelet transforms have better denoising effects and can better express image edge information. This paper uses wavelet transforms and curvelet transforms to filter original images separately. Comparative experimental results show that compared with wavelets, curvelet transforms achieve better denoising effects and effectively enhance CME image front structures.

Image texture can be expressed as a function of spatial variation of pixel grayscale values. Besides using linear multi-scale transform methods in signal processing such as wavelets, ridgelets, and curvelets to enhance image edge information for CME front detection, another effective method is using Gray-Level Co-occurrence Matrix [27] to capture and characterize texture information in different regions. Goussies et al. proposed a non-parametric supervised CME segmentation method based on gray-level co-occurrence matrix [16, 17], called the CORSET algorithm (COROnal Segmentation Technique), which is an improvement on supervised region competition model segmentation methods. The goal of region competition is to segment images into multiple regions where points in each region have similar image features. However, when using each pixel' s grayscale value as the feature vector in region competition methods, supervised region competition model segmentation of CMEs has two problems: first, grayscale histograms of CME regions and background overlap; second, grayscale histograms of CME events and background do not follow normal distributions (i.e., statistical models are unknown). These two problems lead to CME classification errors. To address the first problem, the new algorithm uses gray-level co-occurrence matrix to describe CME texture information. For the second problem, chi-square statistical test in non-parametric models is used, which can assess whether observed events follow a specific distribution. By modifying the region competition motion equation to introduce chi-square tests and texture information, and using fast level set algorithms to achieve segmentation curve evolution in image segmentation. This segmentation algorithm first forms an annular region around the C2 occulting disk, as shown in Figure 10 Figure 10: see original paper. If a CME exists, the curve evolves into a CME contour, as shown in Figure 10(b-d). If no CME exists, the contour disappears, as shown in Figure 11 [Figure 11: see original paper]. To track CMEs, the segmentation result of the current image is used as the initial contour for the next image, and the contour evolves according to the modified region competition motion equation. If the contour disappears, it means a CME event has ended. This algorithm uses chi-square tests and gray-level co-occurrence matrix to correctly capture CME texture information observed in difference images. It can detect and track CMEs with different shapes and intensities, while obtaining a CME overview map similar to CACTus.

Although the CORSET algorithm provides boundary information for CME events, it does not provide explicit quantitative evaluation of CME parameters. Building on Goussies et al.' s work, Braga et al. [18] enhanced the algorithm' s functionality by adding several new features, namely automatic calculation of different morphological and kinematic parameters, expanding CORSET, and comparing obtained parameters (central position angle, angular width, and velocity) with existing manual and automatic detection catalogs.

Since CMEs are in constant motion, they can be detected by distinguishing CME structures from the background. Currently, most studies use difference-based methods to detect moving regions in images. However, such numerical differencing enhances noise to levels comparable to targets. Although median filters suppress noise, they also smooth small CME features, and differencing methods introduce spatiotemporal crosstalk problems. To address this, Morgan et al. [19] proposed a deconvolution method that uses a Normalized Radial-Graded Filter (NRGF) [28] to separate CME images into static background structures and dynamic CME motion structures. Experimental results show that this method can well detect weak CMEs.

CME regions differ significantly from background and other solar structures such as streamers in morphological and textural features. Visually, streamers have relatively uniform spectral distributions, while CME regions exhibit frequent spectral mutations. Therefore, Zeng Zhaoxian et al. [20] proposed a CME identification method based on spectral mutation analysis. Like the multi-scale transform methods above, this algorithm is also frequency domain-based. It uses Fourier transforms to obtain spectral plots of preprocessed coronal images. After separating non-mutated and mutated information, inverse Fourier transforms are applied to obtain corresponding information of mutated structures in the original image for preliminary CME separation. Finally, local stable extremum region detection methods are used to determine CME region contours. Since CME region contours are irregular, a circumscribed ellipse method based on regional covariance is proposed to obtain parameters such as angular width and velocity. Because this method analyzes image spectra, it performs well in identifying multiple and faint CMEs.

The above CME detection methods are based on time domain, frequency domain, and multi-scale analysis. Through linear multi-scale transform methods such as wavelets, ridgelets, and curvelets in signal processing, studying image spectral information, and using gray-level co-occurrence matrix in statistics to describe texture features, these methods perform well especially in detecting weak and dark CMEs.

3.3.3 Optical Flow-Based Methods Optical flow is a powerful image processing tool for measuring motion in digital images, containing information about target motion in images. Optical flow algorithms can estimate velocity vectors for each pixel from continuous image sequences, forming a motion field of the image that can be used to determine target motion conditions.

Colaninno et al. [21] proposed a CME detection and tracking algorithm based on optical flow, adding a smoothing regularization term Markov Random Field to the Optical Flow Constraint Equation (OFCE). Optical flow estimation is then expressed as a global optimization problem. To find the global minimum, a multigrid relaxation method [29] is used for solving. Applying this optical flow estimation method to LASCO C2 images, approximately 330 CMEs were observed between 1999 and 2004 and cataloged. Finally, 10 CMEs with obvious structures were analyzed to derive their velocity fields. Results are shown in Figure 12 [Figure 12: see original paper], where the upper row shows input images and the lower row shows optical flow results, with shaded colors indicating velocity magnitude and arrow directions indicating velocity directions sampled on an 8×8 grid. The obtained velocity measurements enable visualization of CME plasma evolution. This algorithm is very fast and can be easily applied to all available CME images, while CME expansion can be easily seen in images detected by the optical flow algorithm.

Unlike Colaninno et al.' s global optimization optical flow algorithm, Gissot et al. [22] proposed a local parametric optical flow method using gradient estimation to detect CMEs. The Lucas-Kanade algorithm calculates the movement (velocity vector) of each pixel position between two adjacent frames during a certain time period. However, this optical flow algorithm has certain constraints and is only suitable for small target motions. When target motion speed is large (i.e., large inter-frame motion), algorithm errors become significant. Therefore, pyramid-based multi-resolution constraints are introduced in motion analysis, i.e., reducing image size through subsampling while applying low-pass filters to images before subsampling to reduce the impact of intensity changes on motion estimation, improving algorithm robustness. This method proposes a reliable motion region extraction strategy to obtain a dense velocity field and tests it on a series of consecutive coronal images to detect and track motion information between adjacent frames.

The basic idea of using optical flow to detect CME moving targets is to assign a velocity vector to each pixel in the image, forming a vector field for dynamic image analysis. If there are no moving targets in the image, optical flow vectors vary continuously throughout the image region. When moving objects exist in the image, relative motion between target and background necessarily creates different velocity vectors, thus enabling calculation of CME positions in the image.

3.3.4 Learning-Based Methods Traditional detection methods mainly process through artificially defined features or simple threshold settings, which cannot accurately detect CME phenomena. Moreover, CMEs typically have bright structures and rich texture features, making single-feature identification of CME regions insufficiently accurate. With the increasingly widespread application of machine learning and deep learning technologies in computer vision and image processing in recent years [30, 31], many effective classifiers in machine learning have been used to solve various classification and detection problems. Since CME detection is a binary classification problem—identifying whether CMEs exist in images—learning-based methods can be applied to CME detection.

Since CMEs vary greatly in shape and spatial scale, using a single classifier is insufficient. In machine learning, the AdaBoost algorithm proposed by Yoav

Freund and Schapire [32] can integrate different weak classifiers trained on the same training set to form a strong classifier whose performance exceeds any of the integrated weak classifiers. Based on AdaBoost, Zhang et al. [23] proposed a CME image classification algorithm. Unlike most methods above that detect CMEs in polar coordinate transformed images, this method directly segments the differential image of fan-shaped regions into blocks and analyzes the grayscale of the brightest blocks in the image. Experiments show that slice size, grayscale threshold, and bright spot score threshold all affect detection results. Therefore, this method obtains learning samples and designs a classification model for them to obtain the optimal parameter group of these three influencing factors to design weak classifiers. Finally, AdaBoost is used to combine these weak classifiers to obtain the final strong classifier for CME classification.

Similar to Zhang et al., Yin et al. [25] also believe that CMEs correspond to the brightest blocks in images, thus modeling CME detection as a classification problem for the brightest blocks in differential images and proposing a CME image classification algorithm based on AdaBoost. The difference is that this algorithm performs polar coordinate transformation on differential images, extracting grayscale, texture, and HOG features of the brightest blocks after transformation, using multi-feature fusion to construct classifiers. Since CMEs have different appearances, grayscale, and texture features, selecting appropriate features and classifiers is very important. Decision trees can automatically achieve feature selection and classification, so they are designed as weak classifiers, while AdaBoost as an ensemble classifier can improve the classification capability of individual classifiers. Due to the collective wisdom of ensemble learning and the feature selection capability of decision trees, the proposed multi-feature-based detection algorithm ensemble can achieve better detection results. However, since this method mainly considers the brightest block region while ignoring dark cavity regions and other solar structures similar to CMEs, it is prone to missing relatively weak CME events.

To address the slow convergence speed of classifiers, Zhang et al. [24] also proposed a new detection algorithm based on Extreme Learning Machine (ELM) [33]. This method completes preprocessing through inter-frame differencing, polar coordinate transformation, and filtering denoising. Similar to their previously proposed method, it also uses block-based image segmentation. The difference is that the previous method segmented the brightest parts into blocks, while this algorithm uses fixed-size blocks to traverse the entire preprocessed image. Block size selection is as follows: if the average grayscale value of all pixels contained in a block is higher than a specified threshold, the block is considered a bright block. The block size with the highest classification accuracy (ratio of correctly detected samples to all samples) is selected for image segmentation. Simultaneously, grayscale and texture features in blocks form feature vectors to construct an ELM-based classifier. This method' s advantages are that the ELM-based detection method can select appropriate features, has fast convergence speed, and by using spatiotemporal continuity decision rules, can remove most solar structures similar to CMEs, improving classification performance.

The above several machine learning-based methods all consider the brightest blocks in images as CMEs, thus detecting and classifying the brightest blocks, while selected classifiers also have different characteristics. However, these methods all extract features through artificial definition. Since CMEs have multiple features, manually selected features may not achieve good detection results. Deep learning has powerful learning and feature expression capabilities and can automatically extract features for classification. Yao Hai et al. [3] proposed a CME detection method based on Convolutional Neural Networks (CNN). CNN uses supervised learning with massive labeled training samples, combined with backpropagation algorithms to update weights, training a network with classification capability. This method automatically extracts image features suitable for CME detection and establishes a detection model. The network structure uses a six-layer convolutional neural network containing convolutional layers, pooling layers, and fully connected layers. To enable the network model to learn better parameters for classification tasks during training, obvious CME image datasets are used for training. Meanwhile, using these parameters as a benchmark, weak CME datasets are used for parameter fine-tuning in the fine-tuning stage, enabling the network model to achieve better classification results. Compared with previous detection methods, the classification method based on convolutional neural networks far surpasses other algorithms in image processing efficiency under GPU mode, with good real-time performance.

Currently, research on CME detection based on learning methods is still limited, with few publications. Compared with traditional methods, these methods have good classification effects and high detection accuracy, making them an important future research direction for CME detection.

4 Summary and Outlook

CME is a violent solar eruption activity. The ejected material carries enormous energy, causing strong disturbances in Earth's space environment and potentially catastrophic impacts on high-tech systems such as communications, navigation, and spacecraft. Therefore, CME detection is very important for preventing these hazardous space weather events.

Although many methods have been applied to CME detection, achieving great progress in CME detection research, this remains a challenging problem. First, CMEs have different scales, sizes, and shapes, making it difficult to select appropriate features to detect all CMEs. Second, CMEs are accompanied during ejection by interference sources with very similar structures, making it difficult to distinguish CMEs from these structures. Third, small, weak CMEs often suffer from false detection and missed detection.

With the great success of machine learning technology in image processing and computer vision in recent years, machine learning techniques can be used to replace simple threshold segmentation technology to improve CME detection performance. Meanwhile, deep neural networks can automatically extract appropriate features and have achieved good results in speech recognition, image processing, etc. Additionally, big data and high-performance computing equipment are now fully available, providing conditions for deep learning applications in CME detection. Future main research directions for CME detection can be carried out from the following aspects: (1) Continue using machine learning methods to select appropriate features and classifiers to further improve CME classification; (2) Since CME eruption is a dynamic process, deep learning-based moving target detection methods can be applied; (3) CME detection can be considered as a process of separating foreground moving targets from coronal images, so deep learning-based foreground detection and background subtraction methods can be applied.

A difficulty and focus of using deep learning-based methods lies in dataset creation. Manual labeling requires significant human resources and time, while labeling weak, small CMEs may have errors and ambiguities. Therefore, continuous in-depth research is needed.

In summary, with the development and progress of artificial intelligence technology, various learning-based technologies for big data analysis and mining will have broader application space and prospects in solar physics research and applications. Therefore, machine learning and deep learning technologies can be well applied to CME detection, promising faster and more accurate detection results.

References

[1] WEBB D F, HOWARD T A. Coronal mass ejections: observations [J]. Living Reviews in Solar Physics, 2012, 9(1): 3.
[2] YASHIRO S, GOPALSWAMY N, MICHALEK G, et al. A catalog of white light coronal mass ejections observed by the SOHO spacecraft [J]. Journal of Geophysical Research: Space Physics, 2004, 109(A7):
[3] 姚海, 尹建芹, 林佳本, et al. 一种端到端的日冕物质现象检测新方法 [J]. 科学通报, 2017, 62(23): 2680-90. Yao Hai, Yin Jianqin, Lin Jiaben, et al. A new end-to-end detection method for corona material phenomena[J]. Chinese Science Bulletin, 2017, 62(23): 2680-90.
[4] HUNDHAUSEN A J, SAWYER C B, HOUSE L, et al. Coronal mass ejections observed during the Solar Maximum Mission: Latitude distribution and rate of occurrence [J]. Journal of Geophysical Research: Space Physics, 1984, 89(A5): 2639-46.
[5] 高朋鑫, 李可军. 日冕物质抛射基本物理参数的统计特征 [J]. 天文学进展, 2008, 26(2): 117-28. Gao Pengxin, Li Kejun. Statistical Characteristics of Basic Physical Parameters of Coronal Mass Ejection[J]. Advances in Astronomy, 2008, 26(2): 117-28.
[6] YASHIRO S, MICHALEK G, GOPALSWAMY N. A comparison of coronal mass ejections identified by manual and automatic methods; proceedings of the Annales geophysicae: atmospheres, hydrospheres and space sciences, F, 2008 [C].

[7] BERGHMANS D, FOING B, FLECK B. Automated detection of CMEs in LASCO data; proceedings of the From solar min to max: Half a solar cycle with SOHO, F, 2002 [C].
[8] ROBBRECHT E, BERGHMANS D. Automated recognition of coronal mass ejections (CMEs) in near-real-time data [J]. Astronomy & Astrophysics, 2004, 425(3): 1097-106.
[9] OLMEDO O, ZHANG J, WECHSLER H, et al. Automatic detection and tracking of coronal mass ejections in coronagraph time series [M]. Solar Image Analysis and Visualization. Springer. 2008: 275-89.
[10] OLMEDO O, ZHANG J. Partial torus instability [J]. The Astrophysical Journal, 2010, 718(1): 433.
[11] QU M, SHIH F Y, JING J, et al. Automatic detection and classification of coronal mass ejections [J]. Solar Physics, 2006, 237(2):
[12] BOURSIER Y, LLEBARIA A, GOUDAIL F, et al. Automatic detection of coronal mass ejections on LASCO-C2 synoptic maps [J]. Proceedings of SPIE - The International Society for Optical Engineering, 2005,
[13] BOURSIER Y, LAMY P, LLEBARIA A, et al. The ARTEMIS catalog of LASCO coronal mass ejections [J]. Solar Physics, 2009, 257(1): 125-47.
[14] GONZÁLEZ-GÓMEZ D, BLANCO-CANO X, RAGA A. CME classification based on wavelet spectra [J]. Solar Physics, 2010, 266(2): 337-47.
[15] GALLAGHER P T, YOUNG C A, BYRNE J P, et al. Coronal mass ejection detection using wavelets, curvelets and ridgelets: applications for space weather monitoring [J]. Advances in space research, 2011, 47(12): 2118-26.
[16] GOUSSIES N, STENBORG G, VOURLIDAS A, et al. Tracking of coronal white-light events by texture [J]. Solar Physics, 2010, 262(2): 481-94.
[17] GOUSSIES N A, MEJAIL M E, JACOBO J, et al. Detection and tracking of coronal mass ejections based on supervised segmentation and level set [J]. Pattern Recognition Letters, 2010, 31(6): 496-501.
[18] BRAGA C R, DAL LAGO A, STENBORG G. Pseudo-automatic characterization of the morphological and kinematical properties of coronal mass ejections using a texture-based technique [J]. Advances in Space Research, 2013, 51(10): 1949-65.
[19] MORGAN H, BYRNE J P, HABBAL S R. Automatically detecting and tracking coronal mass ejections. I. Separation of dynamic and quiescent components in coronagraph images [J]. The Astrophysical Journal, 2012, 752(2): 144.
[20] 曾昭宪，刘毅，魏雅利．基于频谱突变分析的日冕物质抛射识别方法 [J]. 2012, Zeng Zhaoxian, Liu Yi, Wei Yali. Identification method of corona material projection based on spectrum mutation analysis [J]. 2012
[21] COLANINNO R C, VOURLIDAS A. Analysis of the velocity field of CMEs using optical flow methods [J]. The Astrophysical Journal, 2006, 652(2): 1747.
[22] GISSOT S F, HOCHEDEZ J-F, DIBOS F, et al. Extracting the apparent motion from two successive EIT images; proceedings of the Solar Variability as an Input to the Earth' s Environment, F, 2003 [C].
[23] ZHANG L, YIN J-Q, LIN J-B, et al. Detection of coronal mass ejections using AdaBoost on grayscale statistic features [J]. New Astronomy, 2016, 48(49-57.

[24] ZHANG L, YIN J-Q, LIN J-B, et al. Detection of coronal mass ejections using multiple features and space–time continuity [J]. Solar Physics, 2017, 292(7): 91.
[25] YIN J, YAO H, LIN J, et al. Coronal Mass Ejections detection using multiple features based ensemble learning [J]. Neurocomputing, 2017, 244(123-30.
[26] J HNE B. Digital image processing: concepts, algorithms, and scientific applications [M]. Springer Heidelberg, 1997.
[27] HARALICK R M, SHANMUGAM K. Textural features for image classification [J]. IEEE Transactions on systems, man, and cybernetics, 1973, 6): 610-21.
[28] MORGAN H, HABBAL S R, WOO R. The depiction of coronal structure in white-light images [J]. Solar Physics, 2006, 236(2):
[29] HEITZ F, P REZ P, BOUTHEMY P. Multiscale minimization of global energy functions in some visual recovery problems [J]. CVGIP: image understanding, 1994, 59(1): 125-34.
[30] LI Z, LIU J, TANG J, et al. Robust structured subspace learning for data representation [J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(10): 2085-98.
[31] TANG J, LI Z, WANG M, et al. Neighborhood discriminant hashing for large-scale image retrieval [J]. IEEE Transactions on Image Processing, 2015, 24(9): 2827-40.
[32] SCHAPIRE R E, FREUND Y, BARTLETT P, et al. Boosting the margin: A new explanation for the effectiveness of voting methods [J]. The annals of statistics, 1998, 26(5): 1651-86.
[33] HUANG G-B, ZHU Q-Y, SIEW C-K. Extreme learning machine: theory and applications [J]. Neurocomputing, 2006, 70(1-3):

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