Article:
Liu, X., Liu, J., Song, S. et al. (2026) Automated detection of coronal mass ejections with few-shot learning. The Astrophysical Journal Supplement Series, 283 (1). 29. ISSN: 0067-0049

https://doi.org/10.3847/1538-4365/ae3d07

The Astrophysical Journal Supplement Series, 283:29 (9pp), 2026 March https://doi.org/10.3847/1538-4365/ae3d07

Automated Detection of Coronal Mass Ejections with Few-shot Learning

Xin Liu[1], Jiajia Liu[2,3], Shuhua Song[4,5], Ye Jiang[6], Xin Huang[7], Zhanpeng Xu[1], Robertus Erdélyi[8,9,10], and Yimin Wang[1]
[1] School of Data Science, Qingdao University of Science and Technology, Qingdao 266061, People's Republic of China; yimin.wang@qust.edu.cn
[2] National Key Laboratory of Deep Space Exploration, School of Earth and Space Sciences, University of Science and Technology of China, Hefei 230026, People's Republic of China
[3] CAS Center for Excellence in Comparative Planetology/CAS Key Laboratory of Geospace Environment/Mengcheng National Geophysical Observatory, University of Science and Technology of China, Hefei, 230026, People's Republic of China
[4] GEOVIS Wisdom Technology Co., Ltd., Qingdao 266100, People's Republic of China
[5] GEOVIS Low-altitude Cloud Technology Co., Ltd., Qingdao 266100, People's Republic of China
[6] School of Information Science and Technology, Qingdao University of Science and Technology, People's Republic of China
[7] Faculty of Electrical Engineering and Computer Science, Ningbo University, People's Republic of China
[8] Solar Physics and Space Plasma Research Centre (SP2RC), School of Mathematical and Physical Sciences, The University of Sheffield, Hicks Building, Hounsfield Road, Sheffield, S3 7RH, UK
[9] Department of Astronomy, Eötvös Loránd University, Budapest, Pázmány P. sétány 1/A, H-1117, Hungary
[10] Gyula Bay Zoltán Solar Observatory (GSO), Hungarian Solar Physics Foundation (HSPF), Petőfi tér 3., Gyula, H-5700, Hungary
Received 2025 November 17; revised 2026 January 6; accepted 2026 January 21; published 2026 February 27

Abstract

Coronal mass ejections (CMEs) are a primary driver of severe space weather impacts, yet accurate delineation of CME structure in coronagraph images remains challenging. We present Segment Anything Model (SAM)–low-rank adaptation (LoRA), an efficient CME segmentation framework that employs the SAM to solar data via LoRA. The approach employs a multichannel preprocessing scheme that combines base- and running-difference coronagraph images to enhance CME morphology and localization. We compile a new, manually annotated Large Angle and Spectrometric Coronagraph dataset with CME-only and mixed CME/non-CME configurations. On CME-only data, SAM-LoRA delivers segmentation quality comparable to state-of-the-art methods while requiring substantially fewer labeled examples. On mixed data, image-level labels derived from the predicted masks (empty versus nonempty) enable high-quality CME presence detection. The parameter-efficient design reduces computational and annotation costs while retaining SAM's strong priors. These results indicate that adapting foundation models is a promising path toward reliable CME detection and segmentation and provides a basis for extensible, multitask pipelines for tracking, parameter extraction, and forecasting in heliophysics.

Unified Astronomy Thesaurus concepts: Solar coronal mass ejections (310)

1. Introduction

Coronal mass ejections (CMEs) are among the most energetic manifestations of solar activity, expelling vast quantities of magnetized plasma into interplanetary space. Lasting from minutes to hours, they emerge as discrete, bright, white-light structures in the corona (A. J. Hundhausen et al. 1984) and propagate radially outward (R. Schwenn 1996). When directed toward Earth, CMEs can trigger severe geomagnetic storms (D. Webb et al. 2000), disrupting communication, navigation, power grids, and satellite operations (J. Liu et al. 2018; Y. Wang et al. 2019; M. B. Korsós et al. 2020; S. Liu et al. 2022; M. K. Georgoulis et al. 2024; J. Chen et al. 2025). Accurate detection of CMEs is required for space-weather forecasting thereby safeguarding technological infrastructure.

Since 1996, the Solar and Heliospheric Observatory (SOHO) and its Large Angle Spectrometric Coronagraph (LASCO) have transformed CME studies, amassing an unprecedented observational record (G. E. Brueckner et al. 1995). Complementing this, the twin Solar-Terrestrial Relations Observatory (STEREO; M. L. Kaiser et al. 2008), carrying the Sun–Earth Connection Coronal and Heliospheric Investigation (SECCHI; R. A. Howard et al. 2008), has provided dual-viewpoint imaging from off the Sun–Earth line, enabling three-dimensional analyses of CMEs. More recently, ESA's Solar Orbiter carries the Metis coronagraph, which images the off-limb corona simultaneously in visible light and UV H I Lyα, expanding diagnostics of CME structure and kinematics (E. Antonucci et al. 2020). In parallel, China's Advanced Space-based Solar Observatory includes the Lyα Solar Telescope, whose Solar Corona Imager is a dual-wave-band coronagraph covering 1.1–2.5 $R_{\odot}$, thereby linking low-corona dynamics to CME onset (B. Chen et al. 2019). Most recently, several new coronagraph missions further broaden coronal-imaging coverage and operational capability. ESA's PROBA-3 employs formation flying to realize the externally occulted ASPIICS coronagraph, enabling high-contrast observations of the inner corona (A. N. Zhukov et al. 2025). NASA's PUNCH mission uses a constellation of four small satellites to provide wide-field polarized imaging that connects coronal structures to the heliosphere (C. DeForest et al. 2026). In the operational space-weather context, NOAA's GOES-19 (GOES-U) carries the Compact Coronagraph to support near-real-time CME monitoring (A. F. Thernisien et al. 2025). In addition, ISRO's Aditya-L1 includes the Visible Emission Line Coronagraph, providing complementary diagnostics of coronal structure and CME evolution (M. Priyal et al. 2024). With the rapid growth of coronal-imaging archives, efficient CME detection and

cataloging are becoming essential—not only to streamline data access for research, but also to support real-time space-weather forecasting from CME-annotated images.

Traditionally, CME detection has largely relied on manual annotation, as in the catalogs maintained by the Coordinated Data Analysis Workshops (CDAW) data center (S. Yashiro et al. 2004). While widely used, such methods are time-consuming and demand substantial human effort, making it difficult to sustain during periods of high solar activity. These constraints have motivated the development of automated detection systems capable of processing large data volumes more efficiently and consistently.

Several automated techniques have been developed for LASCO and SECCHI coronagraph images. The Computer Aided CME Tracking system detects CMEs in SOHO/LASCO C2 and C3 images (E. Robbrecht & D. Berghmans 2004), was adapted for STEREO/SECCHI COR2 (E. Robbrecht et al. 2009), and was later extended to heliospheric imaging with STEREO/HI-1 (V. Pant et al. 2016). The Automatic Recognition of Transient Events and Marseille Inventory from Synoptic Maps applies image filtering and segmentation to identify CMEs in LASCO data (Y. Boursier et al. 2009), while the Solar Eruptive Event Detection System isolates CME leading edges using intensity thresholds and derives kinematic parameters (O. Olmedo et al. 2009). Additional efforts include the Coronal Image Processing catalog (J. P. Byrne et al. 2012) and the Dual-Viewpoint CME catalog (A. Vourlidas et al. 2017). More recent developments have explored alternative strategies for event detection, localization, and characterization, including adaptive background learning in LASCO imagery (Z. Qiang et al. 2019), parabolic Hough-transform-based detection of accelerating eruptions in the inner corona (R. Patel et al. 2021), ALMANAC for estimating eruption coordinates from EUV observations (T. Williams & H. Morgan 2022), and a machine learning–based framework for dual-viewpoint 3D CME reconstruction (R. Lin et al. 2025). Despite these advances, many operational pipelines still rely on thresholding or handcrafted heuristics at key stages, motivating data-driven approaches with stronger generalization and reduced subjectivity.

Machine learning methods have been applied to space-weather problems, notably support vector machines (SVMs; C. Cortes & V. Vapnik 1995) and convolutional neural networks (CNNs; Y. Lecun et al. 1998). For instance, M. Qu et al. (2006) used an SVM classifier to distinguish strong from weak CMEs, while J. Liu et al. (2018) trained an SVM on 182 geoeffective (partial) halo CMEs to predict arrival times. Y. Wang et al. (2019) employed a CNN-based approach for CME propagation time prediction, comparing it with the Inception-ResNet-v2 model. Deep learning, with its superior feature extraction capabilities, has shown advantages over traditional techniques. P. Wang et al. (2019) introduced CME Automatic detection and tracking with MachinE Learning (CAMEL) techniques, which trains CNNs on annotations from the LASCO CME Catalog curated by the CDAW Data Center.[11] K. Simonyan & A. Zisserman (2015) applied the VGG-16 visual model to classify CME-containing images, and R. Alshehhi & P. R. Marpu (2021) developed an unsupervised method that used pretrained VGG-16 features, principal component analysis for dimensionality reduction, and K-means clustering for pixel-level classification. In the same year, X. Xianggui et al. (2022) adopted an improved faster R-CNN for CME detection. However, most of these approaches produced relatively coarse results, outlining CME regions with simple boundaries. Accurate segmentation of CME structures remains essential for precise identification and tracking.

Recent advances in transformer architectures (A. Vaswani et al. 2017) have reshaped computer vision. Originally developed for natural language processing (J. Devlin et al. 2018), transformers were later adapted for visual tasks following their success in language modeling. In 2020, A. Dosovitskiy et al. (2021) introduced the Vision Transformer (ViT), the first pure transformer model to replace CNNs, achieving state-of-the-art performance in image classification. Since then, successive ViT variants have extended its capabilities to object detection, segmentation, and other vision tasks (K. Han et al. 2023). In solar physics, J. Shan et al. (2024) employed the transformer-based SegFormer architecture (E. Xie et al. 2021) for CME image segmentation, reporting higher accuracy than CAMEL and outperforming the CNN-based Pyramid Scene Parsing Network (H. Zhao et al. 2017). More recently, P. Yang et al. (2025) introduced TransCME, a CNN–transformer hybrid encoder, which achieved performance comparable to that of SegFormer.

Following the rise of ViT, deep learning underwent another major shift with the emergence of large-scale models such as DALL-E (A. Ramesh et al. 2021), GPT-4 (OpenAI et al. 2023), the Segment Anything Model (SAM; A. Kirillov et al. 2023), and SegGPT (X. Wang et al. 2023). Trained on massive datasets, these models capture rich representations and exhibit strong generalization, enabling accurate predictions in new domains with minimal fine-tuning—a capability known as few-shot learning. This approach greatly reduces the need for labor-intensive manual annotation.

Among these, SAM (A. Kirillov et al. 2023) was developed as a general-purpose framework for image segmentation that accepts flexible user prompts and is adaptable to a wide variety of tasks. Trained on a dataset of 11 million image-mask pairs, SAM achieves zero-shot performance on many segmentation benchmarks that is comparable to, or even exceeds, that of fully supervised methods. Its versatility has facilitated applications spanning medical imaging (Y. Zhang et al. 2024), anomaly detection (Y. Peng et al. 2025), remote sensing (L. P. Osco et al. 2023), and beyond.

Trained primarily on natural images, SAM performs well in zero-shot segmentation of natural scenes. In solar physics, however, our experiments show that applying SAM directly to CME segmentation yields unsatisfactory results. The gap stems from SAM's lack of exposure to solar observations and the absence of solar-specific inductive biases (e.g., coronal morphology, and the spatial priors needed to localize CME structures within full-disk or coronagraph fields). In addition, full end-to-end fine-tuning is computationally expensive and data intensive, limiting practical deployment.

To make adaptation tractable, parameter-efficient fine-tuning methods have been proposed. Two representative families are adapter modules (N. Houlsby et al. 2019) and prompt/prefix tuning (X. L. Li & P. Liang 2021). Adapters insert small bottleneck layers, increasing parameters and adding some latency at training and inference, whereas

[11] https://cdaw.gsfc.nasa.gov/CME_list/

prompt/prefix tuning can be harder to optimize and its added tokens consume part of the context window. Motivated by these limitations, we employ low-rank adaptation (LoRA; E. J. Hu et al. 2022), which parameterizes weight updates with low-rank matrices applied alongside frozen backbone weights, updating only a small subset of parameters while reducing memory and compute.

In this work, we fine-tune SAM via LoRA on LASCO C2 CME observations, leveraging SAM's strong priors while requiring only a small number of annotated examples. This adaptation markedly improves CME segmentation, as verified on an independent test set spanning diverse viewing angles and background conditions. Our results indicate that large-scale models such as SAM hold promise for CME detection and segmentation, offering effective tools for event identification and a path toward automated analysis in solar physics. Given their capacity to share representations across related subtasks, such models can learn richer, domain-specific features than smaller task-specific networks. This study constitutes an initial exploration of general-purpose segmentation within solar physics.

2. Data

SOHO/LASCO C2 white-light coronal FITS images (Level 0.5) were retrieved using SunPy's Fido interface[12] and converted into base- and running-difference products for segmentation. Because the inputs are difference images, many absolute-calibration effects largely cancel in the subtraction, and the resulting composites retain the morphology and boundaries required for CME segmentation. LASCO C2 observes the solar corona over a heliocentric radial range of 1.5–6 $R_{\odot}$, enabling the detection of CMEs at early stages prior to their propagation into interplanetary space. Since no public dataset exists for CME segmentation, we compiled and manually annotated a new dataset from LASCO C2 observations. To capture solar-cycle variability, we selected LASCO/C2 images from 2012 January–2013 December (near solar maximum) and 2019 January–2021 December (near solar minimum), with the dataset comprising 50% solar-maximum and 50% solar-minimum samples in both the training and test sets. Event annotation is guided by the LASCO CME Catalog maintained by the CDAW Data Center, and each CME instance is classified and archived in an individual event directory.

CMEs are categorized following the CDAW catalog as narrow (angular width $<30°$), limb (30°–120°), partial halo (120°–360°), and halo (360°). The annotated images comprise cases spanning the full field of view, from near the solar disk to the outer edge, with varying quality (i.e., variations in signal-to-noise ratio (SNR) and CME-background contrast) and scene complexity (single to multiple coexisting CMEs). We design two experimental settings: (i) CME-containing cases (32 training images, 100 test images) and (ii) mixed CME/non-CME cases (40 training images, 125 test images). The 32-image training set is class balanced across the four CDAW angular-width categories, with a 1:1:1:1 proportion: eight narrow ($<30°$), eight limb (30°–120°), eight partial halo (120°–360°), and eight halo (360°). For the mixed CME/non-CME setting, we additionally include eight non-CME images, yielding a 1:1:1:1:1 ratio (eight images per CME width category plus eight non-CME images; 40 training images in total). The same class-balancing strategy is applied to both the training and test sets. To ensure fair sampling and to test robustness and generalization, images in all subsets are drawn across nearly every month in the covered interval rather than clustered in a few periods, and each image corresponds to a distinct CME event. All images were used as full 1024×1024 frames without tiling, consistent with SAM's default 1024×1024 input resolution and positional encoding.

To justify the few-shot training size, we performed a data-scaling study by gradually increasing the number of annotated training images from 20 to 24, 28, 32, 40, and 100. We found that performance saturated quickly: increasing the training set from 32 to 40 images did not yield measurable gains, and expanding to 100 images improved the segmentation metrics by only ~3% compared to the 32-image setting. We therefore adopt 32 training images as a representative few-shot configuration that provides near-saturated performance while keeping manual annotation costs modest.

The raw LASCO C2 FITS images typically exhibit a high dynamic range, with pixel intensities on the order of 10^4–10^5 data numbers. To improve the visibility of faint CME structures against the bright coronal background, the display range was clipped to $[-100,\ 100]$ in data units prior to rendering, which enhances local contrast and suppresses background noise. For each CME, two derived products were generated: (i) base difference image (BDI), obtained by subtracting the last preeruption frame (the one just before the CDAW start time) from the eruption frame, and (2) the running-difference image (RDI), obtained by subtracting the frame immediately preceding the eruption from the current frame.

The processed images were rendered in gray scale and then converted from single-channel to three-channel pseudo-red–green–blue (RGB) representation to satisfy the input requirements of the segmentation model. For single-modality inputs (BDI only or RDI only), pseudo-RGB was formed by channel replication, i.e., copying the same 1024×1024 array into all three channels. For mixed inputs (Concat), pseudo-RGB tensors were formed by channel stacking, e.g., [BDI, BDI, RDI], where each channel corresponds to a difference product.

Manual annotation was conducted with LabelMe, yielding binary PNG masks with values 0 and 1. Annotators consulted both the original white-light frames and the difference products, because various CME morphologies present sharper leading edges in different representations. Angular widths and position angles reported in the LASCO CME Catalog (CDAW) were used to guide precise delineation of CME boundaries.

3. Methodology

3.1. SAM and LoRA

In this work, we propose SAM-LoRA, which automatically generates masks of CME regions under few-shot supervision. SAM comprises an image encoder, a prompt encoder, and a lightweight mask decoder, as illustrated in Figure 1. The image encoder computes a dense representation of each input frame, while the prompt encoder maps user cues, including points, boxes, existing masks, or text, into a compatible embedding

[12] https://docs.sunpy.org/en/latest/generated/gallery/map/lasco_mask.html

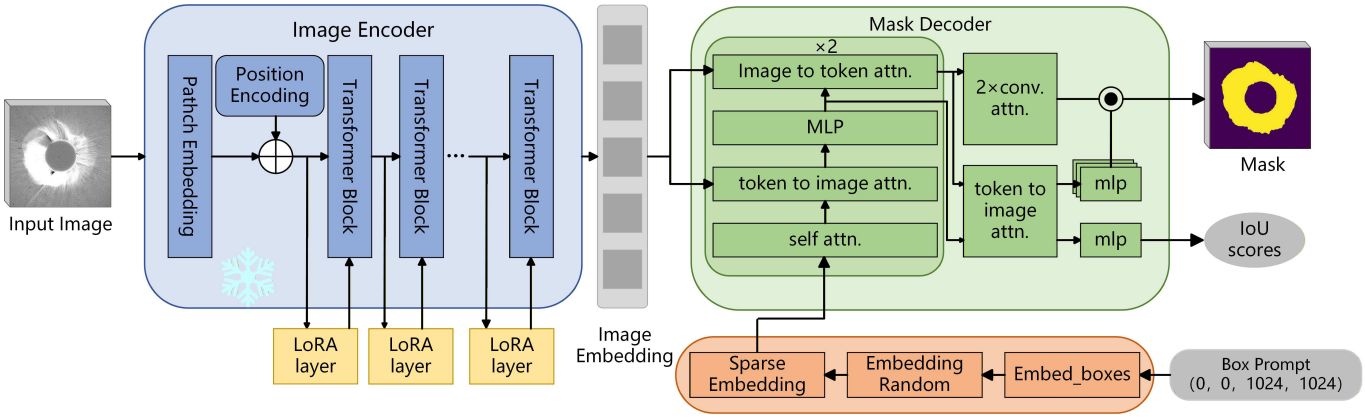

Figure 1. Overview of the SAM-LoRA architecture. The figure also indicates our LoRA integration: the original image-encoder weights are frozen, and trainable low-rank adaptation (LoRA) adapters are inserted into the query (Q) and value (V) projection layers of each transformer block to enable efficient fine-tuning for CME feature extraction.

space. The mask decoder then produces segmentation masks conditioned jointly on the image and prompt embeddings. To automate CME segmentation without manual interaction, we supply a single full-frame bounding box as the prompt (i.e., a box coincident with the entire field of view).

Since SAM was trained primarily on natural images, zero-shot transfer to solar observations is limited. We therefore adapt the model using LoRA. Standard full-parameter fine-tuning, as illustrated in Figure 2, learns a dense update ΔW with the same shape as W:

$$W' = W + \Delta W, \qquad \Delta W \in \mathbb{R}^{d_{\text{out}} \times d_{\text{in}}}. \tag{1}$$

Under LoRA, the pretrained weights $W \in \mathbb{R}^{d_{\text{out}} \times d_{\text{in}}}$ are frozen and each transformer weight matrix is augmented with a trainable low-rank update of rank r:

$$W' = W + BA, \quad A \in \mathbb{R}^{r \times d_{\text{in}}}, \; B \in \mathbb{R}^{d_{\text{out}} \times r}, \\ r \ll \min(d_{\text{in}}, d_{\text{out}}), \tag{2}$$

where r denotes the rank of the low-rank matrices and can be adjusted to set the capacity used during task adaptation. A smaller r yields simpler low-rank updates with fewer learnable parameters, which can accelerate training and reduce computational requirements. However, too small a value may limit the ability to capture task-specific information and degrade performance. Conversely, a larger r increases capacity but also model complexity and the risk of overfitting. Choosing r therefore requires balancing model complexity, adaptability, and the risks of underfitting and overfitting. In practice, we test several values to identify the best trade-off for the target task.

This parameterization substantially reduces the number of trainable parameters and memory footprint, preserves the architecture and near-original inference speed after merging the adapters, and enables efficient domain specialization with few labeled CME datasets.

3.2. Architecture of SAM-LoRA

This section describes our implementation of SAM-LoRA for training on CME observations. In the original code base, SamPredictor is designed for single-image, interactive inference and thus is not suitable for batched, gradient-based training. We therefore built a dataloader-friendly preprocessing pipeline that normalizes images to the required resolution

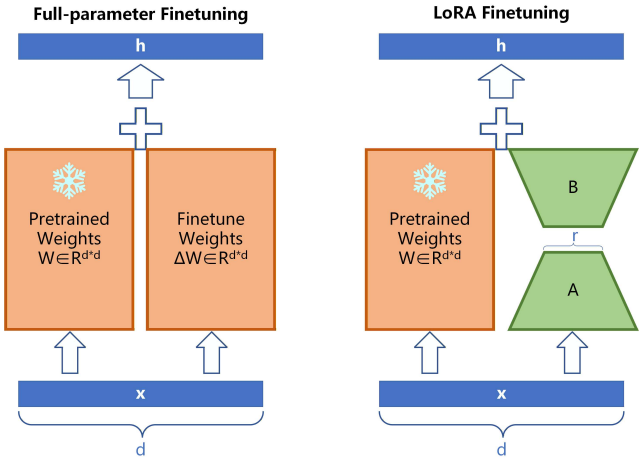

Figure 2. Comparison between full-parameter fine-tuning (left) and LoRA (right). Full fine-tuning updates all parameters in W, whereas LoRA freezes the pretrained weights and trains only a low-rank update $\Delta W = BA$ with $r \ll \min(d_{\text{in}}, d_{\text{out}})$, greatly reducing the number of trainable parameters.

of the encoder and collates them into fixed-size batches for the SAM backbone.

We incorporate LoRA into the attention modules to adjust the ability of the model to capture features in CME images. In particular, we make two modifications: (i) attach low-rank adapters to the query (Q) and value (V) projections; and (ii) apply these adapters within the image encoder. For each projection matrix $W_Q \in \mathbb{R}^{d_{\text{out}} \times d_{\text{in}}}$ and $W_V \in \mathbb{R}^{d_{\text{out}} \times d_{\text{in}}}$, we learn rank-$r$ factors $A_{\{\cdot\}} \in \mathbb{R}^{r \times d_{\text{in}}}$ and $B_{\{\cdot\}} \in \mathbb{R}^{d_{\text{out}} \times r}$ and form

$$W'_Q = W_Q + B_Q A_Q, \tag{3}$$

$$W'_V = W_V + B_V A_V, \tag{4}$$

where only $A_{\{\cdot\}}$ and $B_{\{\cdot\}}$ are trainable. We implement these adapters as two additional linear layers and initialize their weights with Kaiming-uniform initialization to promote stable early training and faster convergence.

Applied off the shelf, the baseline SAM checkpoint tended to output near full-frame masks on CME events, meaning masks covering most of the 1024 × 1024 field of view with near-uniform foreground, which indicates that the pretrained model did not capture the positional structure needed to localize CME regions in astrophysical observations. We

therefore placed the LoRA blocks in the image encoder (Figure 1), which is the most computationally intensive component and aggregates positional information over patch tokens; adapting this stage improved the model's ability to localize CME structures.

To balance performance and GPU memory, we adopt SAM's sam_vit_b checkpoint, the smallest of the official pretrained models. Training is conducted with the Adam optimizer, and an initial learning rate of 1×10^{-4}. The optimization objective is the Dice loss.

4. Results and Discussions

4.1. Evaluation of CME-containing Images

To quantify segmentation performance, we report four pixel-wise metrics: pixel accuracy (PA), mean pixel accuracy (MPA), intersection over union (IoU), and mean intersection over union (MIoU). For binary masks (CME versus background), let $\mathrm{TP}_{\mathrm{seg}}$, $\mathrm{TN}_{\mathrm{seg}}$, $\mathrm{FP}_{\mathrm{seg}}$, and $\mathrm{FN}_{\mathrm{seg}}$ denote true positives (CME pixels correctly predicted as CME), true negatives (background pixels correctly predicted as background), false positives (background pixels predicted as CME), and false negatives (CME pixels predicted as background), respectively. The metrics are defined as

$$\mathrm{PA} = \frac{\mathrm{TP}_{\mathrm{seg}} + \mathrm{TN}_{\mathrm{seg}}}{\mathrm{TP}_{\mathrm{seg}} + \mathrm{TN}_{\mathrm{seg}} + \mathrm{FP}_{\mathrm{seg}} + \mathrm{FN}_{\mathrm{seg}}}, \tag{5}$$

which measures the fraction of correctly classified pixels. Because coronagraph images are typically background dominated, PA can be inflated by the majority class. We therefore also report the class-balanced accuracy

$$\mathrm{MPA} = \frac{1}{2}\left(\frac{\mathrm{TP}_{\mathrm{seg}}}{\mathrm{TP}_{\mathrm{seg}} + \mathrm{FN}_{\mathrm{seg}}} + \frac{\mathrm{TN}_{\mathrm{seg}}}{\mathrm{TN}_{\mathrm{seg}} + \mathrm{FP}_{\mathrm{seg}}}\right), \tag{6}$$

i.e., the mean of per-class accuracies. MPA mitigates class imbalance by weighting classes equally, providing a more faithful assessment than PA when background pixels vastly outnumber CME pixels. For the CME class, the IoU is

$$\mathrm{IoU}_{\mathrm{CME}} = \frac{\mathrm{TP}_{\mathrm{seg}}}{\mathrm{TP}_{\mathrm{seg}} + \mathrm{FP}_{\mathrm{seg}} + \mathrm{FN}_{\mathrm{seg}}}, \tag{7}$$

i.e., the ratio of intersection to union of the predicted and ground-truth CME regions. The MIoU averages the per-class IoUs (CME and background):

$$\mathrm{MIoU} = \frac{1}{2}\left(\frac{\mathrm{TP}_{\mathrm{seg}}}{\mathrm{TP}_{\mathrm{seg}} + \mathrm{FP}_{\mathrm{seg}} + \mathrm{FN}_{\mathrm{seg}}} + \frac{\mathrm{TN}_{\mathrm{seg}}}{\mathrm{TN}_{\mathrm{seg}} + \mathrm{FP}_{\mathrm{seg}} + \mathrm{FN}_{\mathrm{seg}}}\right). \tag{8}$$

SAM is trained and typically applied to three-channel natural images. To use SAM with coronagraph data, each observation is represented as a pseudo-RGB tensor of size $1024 \times 1024 \times 3$. Because CMEs are often low-contrast and spatially diffuse in raw white-light frames, contrast-enhanced difference products commonly used in CME analysis are adopted. BDI emphasizes the overall CME envelope and faint extensions by removing the preevent background, whereas RDI highlights sharp leading edges and short-timescale motion while suppressing slowly varying coronal structures.

Table 1
Segmentation Performance for CME Masks with BDI/RDI Inputs

Method	PA	MPA	IoU	MIoU
BDI only	0.97	0.91	0.74	0.86
RDI only	0.97	0.90	0.75	0.86
Concat (BDI + RDI + original)	**0.98**	**0.93**	0.77	0.87
Concat (BDI + RDI + (BDI + RDI)/2)	**0.98**	**0.93**	**0.80**	0.88
Concat (BDI × 1 + RDI × 2)	**0.98**	**0.93**	**0.80**	0.88
Concat (BDI × 2 + RDI × 1)	**0.98**	**0.93**	**0.80**	**0.89**

Note. BDI = base difference image; RDI = running-difference image. Concat denotes three-channel composites constructed by stacking BDI and RDI maps to satisfy SAM's input format. Boldface indicates the best value(s) in each column.

Table 1 quantitatively confirms the complementarity of these two representations: mixed three-channel composites that combine BDI and RDI (Concat) consistently outperform single-modality inputs (BDI only or RDI only) across PA, MPA, IoU, and MIoU. In contrast, including the original intensity channel degrades segmentation performance, likely because the quiescent coronal background and instrumental structure dominate the raw signal. Among the tested inputs, the best results are obtained with the [BDI, BDI, RDI] composite (Table 1). Figure 3 illustrates this configuration, where two BDI channels and one RDI channel are concatenated to form a pseudo-RGB image for SAM.

Let us compare now SAM-LoRA against CAMEL (P. Wang et al. 2019), SegFormer (J. Shan et al. 2024), and TransCME (P. Yang et al. 2025) in terms of pixel-wise performance and data efficiency (training/test set sizes), as shown in Table 2. Using substantially fewer training samples, SAM-LoRA attains performance comparable to SegFormer and competitive with TransCME. When sufficient annotated data are available, fully fine-tuning SAM (i.e., updating all parameters) could yield substantial additional gains.

In CME segmentation, PA can appear inflated (SegFormer: 0.997, TransCME: 0.997, SAM-LoRA: 0.98) because most pixels are background, so vast numbers of background pixels (TN) dominate even when the CME mask is imperfect. MPA, which is the average of the per-class accuracies for CME and background, reduces background dominance (SegFormer: 0.98, SAM-LoRA: 0.93). However, MPA can still appear optimistic under, e.g., thin oversegmentation (a narrow FP band around the CME front). It barely lowers background class accuracy because FPs are tiny compared with TNs. It may even raise CME class accuracy by capturing faint rim pixels. Consequently, MPA may remain high (or increase) despite boundary overreach. By contrast, IoU/MIoU more directly track scientific usefulness: despite being trained on only 32 images, SAM-LoRA attains competitive IoU (0.80) and the best MIoU (0.89), exceeding SegFormer (0.83) and the value reported for CAMEL (0.49). We therefore emphasize IoU/MIoU alongside PA/MPA.

Note that the metric values in Table 2 are not directly comparable across studies because the test sets differ. To ensure fair sampling, as described in Section 2, our test set comprises images drawn from nearly every month in the covered interval (rather than clustered in a few periods), with each image corresponding to a distinct CME event. Class

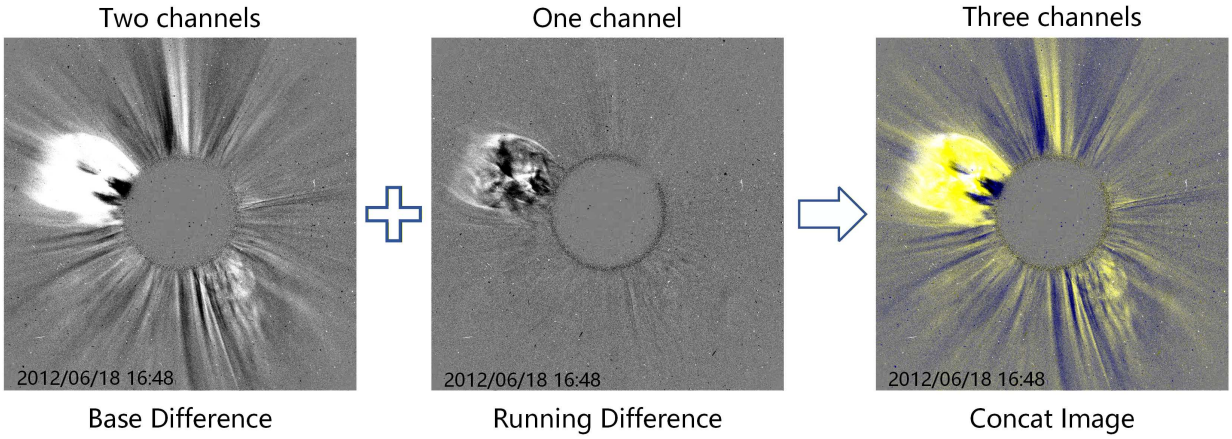

Figure 3. Three-channel composite used as the SAM input. Channels 1–2 are BDI images, and channel 3 is an RDI image, yielding a pseudo-RGB composite.

Table 2
Comparison of Different Models for CME Segmentation

Model	PA	MPA	IoU	MIoU	Training Set Size	Test Set Size
CAMEL	⋯	0.95	⋯	0.49	⋯	⋯
SegFormer	**0.997**	**0.98**	0.79	0.83	300	75
TransCME	**0.997**	⋯	**0.84**	⋯	100	252
SAM-LoRA	0.98	0.93	0.80	**0.89**	**32**	100

Note. Best available value(s) per column are typeset in bold. ⋯ indicates the metric was not reported in the cited work.

balance is enforced at approximately 1:1:1:1 across the four CME types. To enable a direct, side-by-side comparison, we adopt the test images presented in P. Yang et al. (2025) and compare those two state-of-the-art methods with our approach. Our method achieves markedly higher completeness in all cases, as shown in Figure 4. The CME in column (a) is classified as a halo event in the CDAW catalog. Inspection of the LASCO/C2 time sequence indicates that the sharp upper extension and the narrow lower feature belong to the same expanding CME envelope. SegFormer and TransCME tend to fragment these extensions into disconnected components, whereas SAM-LoRA recovers a single contiguous mask that better follows the overall structure. In addition, the faint region bridging the upper extension and the main body is likely part of the CME envelope, but it is close to the noise level (low SNR and weak CME-background contrast) and therefore is not reliably recovered by any of the models. In columns (d)–(f), the faint white CME envelope is consistently identified only by our method. In this study, our annotations follow the CDAW catalog conventions for central position angle and angular width to support subsequent physical-parameter extraction. Consequently, a faint shock-like or wavelike outer front may be included in the predicted CME region when it appears visually contiguous with the CME envelope (e.g., Figures 4(d)–(f) and Figure 5(d)). In the current version, SAM-LoRA is trained with binary masks (CME versus background) and therefore does not explicitly distinguish the CME body from shock-like outer fronts. Separating CME and shock would require dedicated shock-specific annotations and extending the current formulation to a multiclass objective or a multimask output head (instance-level masks).

IoU and MIoU are more sensitive than PA/MPA to boundary fidelity and faint, low-contrast structure because they penalize both missed CME pixels and boundary overreach. The superior performance of our approach in Figure 4 indicates higher IoU/MIoU at CME fronts and improved handling of fine detail in complex backgrounds. Additional examples spanning the four CME types and multi-CME scenes are shown in Figure 5.

Performance varies systematically with CME angular width (Table 3). Narrow events ($<30^\circ$) yield the lowest IoU (0.78) and MIoU (0.89) and a comparatively low MPA (0.93), despite a high PA (1.0). This reflects their limited spatial extent: small boundary errors constitute a large fraction of the CME region, depressing IoU/MIoU, while PA remains high because correctly classified background dominates. Moreover, near the 30° threshold, narrow CMEs are occasionally confused with jets. This ambiguity further lowers IoU/MIoU while leaving PA largely unchanged. IoU rises (0.82–0.84) and MIoU stabilizes near 0.91, indicating better recovery of the CME envelope, as width increases (30°–120° and 120°–360°). Halo events yield the lowest PA (0.94), MPA (0.89), IoU (0.76), and MIoU (0.84). Because halos occupy much more of the field of view, background pixels no longer dominate, so PA becomes more sensitive and decreases. At the same time, halo fronts are often faint and diffuse with complex streamer and background structure, increasing both missed CME pixels and boundary overreach, which lowers IoU/MIoU.

4.2. Evaluation of Mixed CME/Non-CME Images

In a second part of the experiments, we evaluated the model's ability to distinguish CME-containing from non-CME images by augmenting both the training and test sets with additional non-CME samples. Specifically, we added eight non-CME images to the training set and 25 to the test set, resulting in 40 training images and 125 test images in total. For image-level classification, we derive the label from the segmentation output: if the predicted CME mask is empty, the image is treated as non-CME; otherwise it is treated as CME. Performance on this binary task (CME = positive class) was quantified using the following standard metrics:

$$\text{Accuracy} = \frac{\text{TP}_{\text{cls}} + \text{TN}_{\text{cls}}}{\text{TP}_{\text{cls}} + \text{TN}_{\text{cls}} + \text{FP}_{\text{cls}} + \text{FN}_{\text{cls}}}, \quad (9)$$

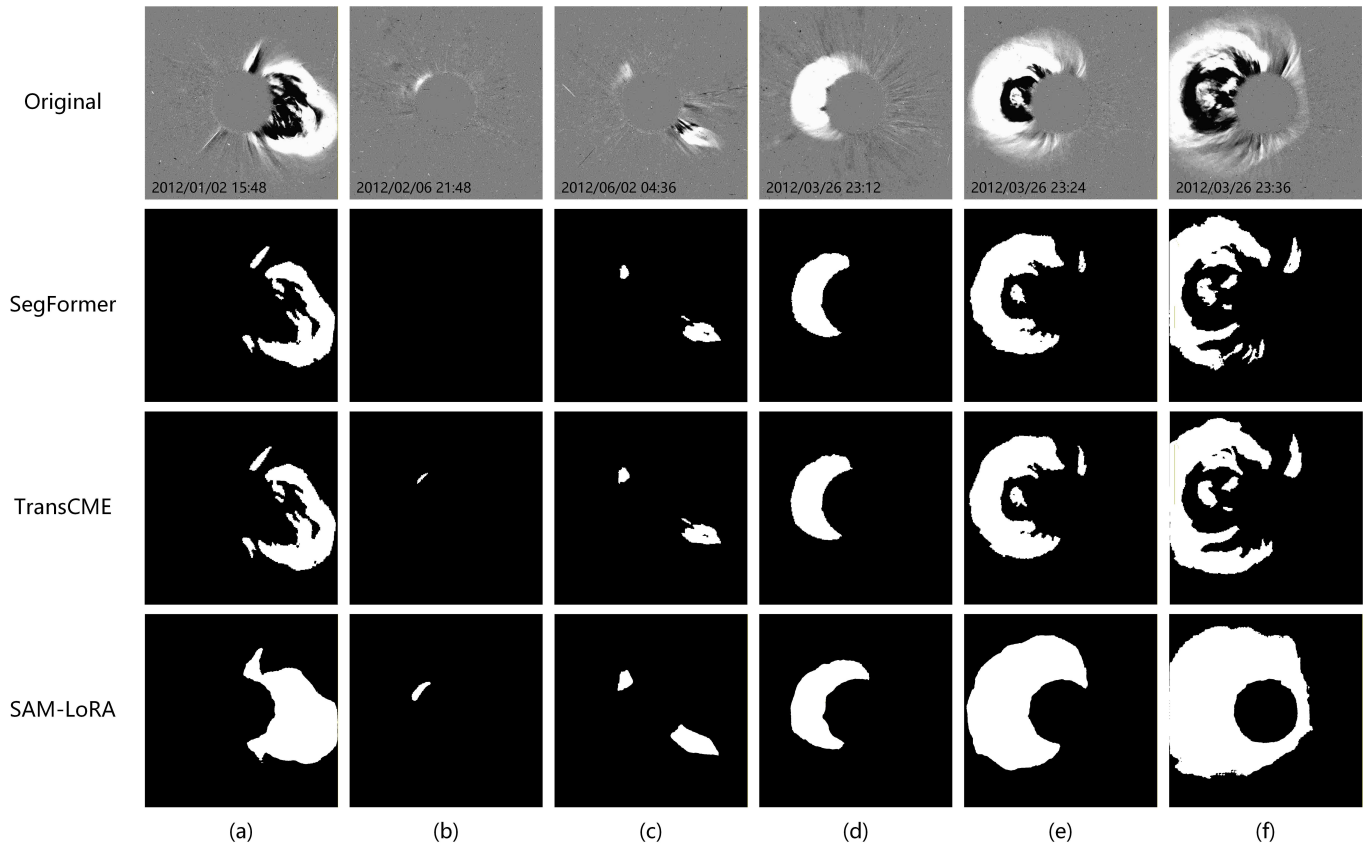

Figure 4. Qualitative comparison of CME segmentation on six representative SOHO/LASCO C2 examples. The top row shows the input difference images (timestamps indicated in each panel). The remaining rows show the predicted binary masks from SegFormer, TransCME, and SAM-LoRA.

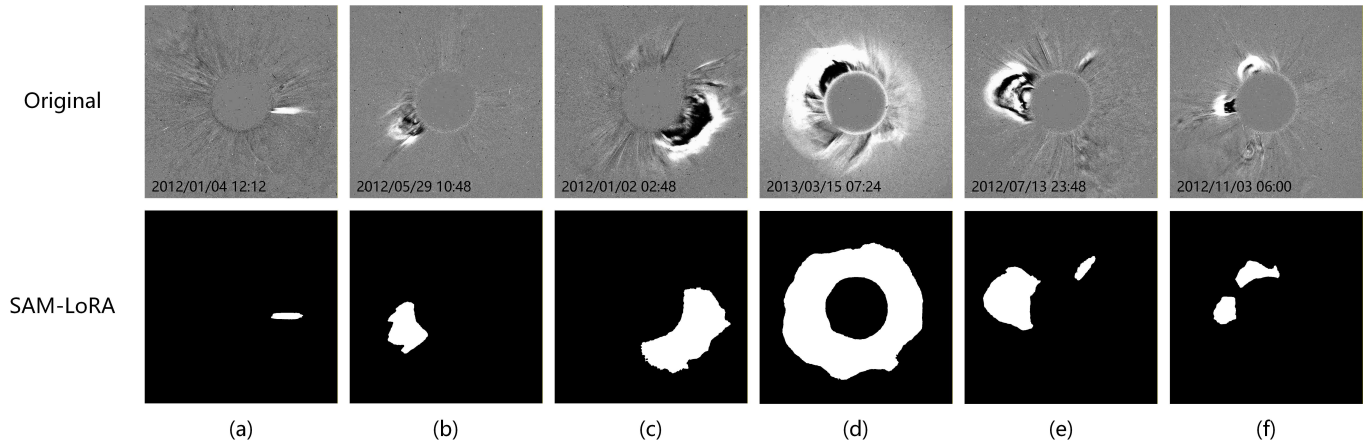

Figure 5. Representative segmentation examples across CME types. From left to right: (a) narrow (0°–30°), (b) limb (30°–120°), (c) partial halo (120°–360°), (d) halo (360°), and (e), (f) scenes with two CMEs occurring simultaneously.

Table 3
Segmentation Performance versus CME Angular Width

Angular Width	PA	MPA	IoU	MIoU
<30°	1.00	0.93	0.78	0.89
30°–120°	0.99	0.95	0.82	0.91
120°–360°	0.98	0.95	0.84	0.91
Halo (360°)	0.94	0.89	0.76	0.84

$$\text{Precision} = \frac{\text{TP}_{\text{cls}}}{\text{TP}_{\text{cls}} + \text{FP}_{\text{cls}}}, \tag{10}$$

$$\text{Recall} = \frac{\text{TP}_{\text{cls}}}{\text{TP}_{\text{cls}} + \text{FN}_{\text{cls}}}, \tag{11}$$

$$\text{F1} = \frac{2\,\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{12}$$

where TP_{cls} denotes the number of ground-truth CME images predicted as CME (nonempty predicted CME mask), TN_{cls} the number of ground-truth non-CME images predicted as non-

Table 4
Binary CME/Non-CME Classification with Different Image Preprocessing Methods

Method	Accuracy	Precision	Recall	F1-score
RDI only	0.93	0.96	0.95	0.95
Concat	0.97	0.99	0.98	0.98

CME (empty mask), FP the number of ground-truth non-CME images predicted as CME (spurious nonempty mask), and FN_{cls} the number of ground-truth CME images predicted as non-CME (missed mask).

These metrics provide complementary views of performance: accuracy reflects overall correctness, precision measures the reliability of CME detections, recall quantifies sensitivity to true CMEs, and F1 summarizes the precision–recall trade-off. Table 4 shows that segmentation-derived image labels enable high-quality CME/non-CME discrimination. Using RDI-only inputs yields strong performance, while the BDI + RDI Concat composite further improves all metrics. The gain is driven primarily by a reduction in both false positives and false negatives, indicating that combining BDI and RDI provides more reliable cues for detecting CME presence than RDI alone. Overall, these results suggest that SAM-LoRA retains robust segmentation behavior on the mixed dataset and that a simple empty-mask versus nonempty-mask rule can support accurate CME detection at the image level, with the remaining errors most likely arising from borderline narrow events whose morphology can resemble jetlike outflows.

5. Conclusion

We presented an efficient and streamlined pipeline for CME image segmentation that adapts the foundation segmentation model SAM with LoRA. Fine-tuning SAM via LoRA yields precise CME masks on preprocessed, manually annotated images despite complex coronal backgrounds and diverse viewing geometries. Our main contributions and findings are that:

1. To the best of our knowledge, this is among the first studies to adapt a large, general-purpose segmentation foundation model for CME detection and segmentation in coronagraph imagery. The pretrained encoder serves as a task-oriented backbone for CME analysis, enabling efficient transfer with minimal task-specific tuning.
2. We introduced a multichannel scheme that combines BDI and RDI images, enhancing the ability of the model to capture CME morphology and positional evolution.
3. On the CME-only test set, while using substantially fewer training samples, the proposed SAM-LoRA achieves performance comparable to that of SegFormer and remains competitive with TransCME.
4. On the auxiliary task of discriminating CMEs from non-CME images using segmentation-derived cues, the model achieves strong binary classification performance.
5. Under a few-shot regime, the pipeline reduces annotation effort while keeping computational demands modest. LoRA fine-tuning achieves high-quality CME segmentation with minimal labeled data; on a single NVIDIA RTX 4090 (24 GB), training completes in ≈8 minutes.

While the proposed approach is robust, future work will refine the methodology by incorporating temporal context. Conditioning on short sequences of preceding and subsequent frames could better distinguish narrow CMEs from jets by leveraging their different temporal evolution and morphology, such as coherent outward expansion, front persistence, and changes in angular extent, and it can also reduce background fluctuations from transient or instrumental variations. When sufficient annotations are available, fully fine-tuning SAM, which is to updating all parameters rather than applying parameter-efficient adapters, could yield additional gains. In large-data regimes, adapter methods are capacity limited and may fail to absorb the necessary new information, whereas fully fine-tuning better captures CME morphology and improves calibration under domain shifts (e.g., exposure and solar-cycle phase).

Future work will also couple segmentation with temporal association and tracking across consecutive frames to enable objective CME kinematics and characterization, and will benchmark both detections and derived parameters against existing manual and automated CME catalogs to quantify agreement and reduce subjectivity. In addition, the framework will be validated and adapted beyond SOHO/LASCO C2 to other coronagraph datasets, particularly more challenging low-SNR observations such as MLSO K-Cor and STEREO/SECCHI COR1.

Acknowledgments

We acknowledge the use of the data from the Solar Dynamics Observatory (SDO). This research is supported by the National Natural Science Foundation of China (NSFC 12303103). J.L. acknowledges support from the National Natural Science Foundation (NSFC 42521007, 42188101, 12373056). Y.J. acknowledges the support from the Natural Science Foundation of Shandong Province (ZR2023QF151). X.H. acknowledges supports from the National Key R&D Program of China under grant 2021YFA1600504, and the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB0560000). R.E. acknowledges the NKFIH (OTKA, grant No. K142987) Hungary for enabling this research. R.E. is also grateful to Science and Technology Facilities Council (STFC, grant No. ST/M000826/1) UK, PIFI (China, grant No. 2024PVA0043), and the NKFIH Excellence Grant TKP2021-NKTA-64 (Hungary).

ORCID iDs

Xin Liu https://orcid.org/0009-0004-2687-9866
Jiajia Liu https://orcid.org/0000-0003-2569-1840
Ye Jiang https://orcid.org/0000-0002-6683-0205
Xin Huang https://orcid.org/0000-0003-0569-5932
Zhanpeng Xu https://orcid.org/0009-0000-1836-3414
Robertus Erdélyi https://orcid.org/0000-0003-3439-4127
Yimin Wang https://orcid.org/0000-0002-8835-3825

References

Alshehhi, R., & Marpu, P. R. 2021, SoPh, 296, 104
Antonucci, E., Romoli, M., Andretta, V., et al. 2020, A&A, 642, A10
Boursier, Y., Lamy, P., Llebaria, A., Goudail, F., & Robelus, S. 2009, SoPh, 257, 125
Brueckner, G. E., Howard, R. A., Koomen, M. J., et al. 1995, SoPh, 162, 357
Byrne, J. P., Morgan, H., Habbal, S. R., & Gallagher, P. T. 2012, ApJ, 752, 145

Chen, B., Li, H., Song, K.-F., et al. 2019, RAA, 19, 159
Chen, J., Gyenge, N. G., Jiang, Y., et al. 2025, ApJ, 980, 261
Cortes, C., & Vapnik, V. 1995, MachL, 20, 273
DeForest, C., Gibson, S., Killough, R., et al. 2026, SoPh, 301, 16
Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. 2018, arXiv:1810.04805
Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. 2021, arXiv:2010.11929
Georgoulis, M. K., Yardley, S. L., Guerra, J. A., et al. 2024, AdSpR, in press
Han, K., Wang, Y., Chen, H., et al. 2023, ITPAM, 45, 87
Houlsby, N., Giurgiu, A., Jastrzebski, S., et al. 2019, in Proc. of Machine Learning Research 97, Proc. of the 36th Int. Conf. on Machine Learning, ed. K. Chaudhuri & R. Salakhutdinov (PMLR), 2790, https://proceedings.mlr.press/v97/houlsby19a.html
Howard, R. A., Moses, J. D., Vourlidas, A., et al. 2008, SSRv, 136, 67
Hu, E. J., Shen, Y., Wallis, P., et al. 2022, Int. Conf. on Learning Representations 2022 (OpenReview) https://openreview.net/forum?id=nZeVKeeFYf9
Hundhausen, A. J., Sawyer, C. B., House, L., Illing, R. M. E., & Wagner, W. J. 1984, JGR, 89, 2639
Kaiser, M. L., Kucera, T. A., Davila, J. M., et al. 2008, SSRv, 136, 5
Kirillov, A., Mintun, E., Ravi, N., et al. 2023, arXiv:2304.02643
Korsós, M. B., Georgoulis, M. K., Gyenge, N., et al. 2020, ApJ, 896, 119
Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998, IEEEP, 86, 2278
Li, X. L., & Liang, P. 2021, in Proc. of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int. Joint Conf. on Natural Language Processing, Vol. 1, ed. C. Zong et al. (Association for Computational Linguistics), 4582
Lin, R., Yang, Y., Shen, F., Pi, G., & Li, Y. 2025, ApJS, 280, 44
Liu, J., Ye, Y., Shen, C., Wang, Y., & Erdélyi, R. 2018, ApJ, 855, 109
Liu, S., Xu, L., Zhao, Z., et al. 2022, ApJ, 941, 20
Olmedo, O., Zhang, J., Wechsler, H., Poland, A., & Borne, K. 2009, in Automatic Detection and Tracking of Coronal Mass Ejections in Coronagraph Time Series, ed. J. Ireland & C. A. Young (Springer), 275
OpenAI, Achiam, J., Adler, S., et al. 2023, arXiv:2303.08774
Osco, L. P., Wu, Q., de Lemos, E. L., et al. 2023, IJAEO, 124, 103540
Pant, V., Willems, S., Rodriguez, L., et al. 2016, ApJ, 833, 80
Patel, R., Pant, V., Iyer, P., et al. 2021, SoPh, 296, 31
Peng, Y., Lin, X., Ma, N., et al. 2025, Knowl. Based Syst., 314, 113176
Priyal, M., Singh, J., Prasad, B. R., et al. 2024, AdSpR, 74, 547
Qiang, Z., Bai, X., Zhang, Q., & Lin, H. 2019, AdAst, 2019, 6582104
Qu, M., Shih, F. Y., Jing, J., & Wang, H. 2006, SoPh, 237, 419
Ramesh, A., Pavlov, M., Goh, G., et al. 2021, in Proc. of Machine Learning Research 139, Proc. of the 38th Int. Conf. on Machine Learning, ed. M. Meila & T. Zhang (PMLR), 8821, https://proceedings.mlr.press/v139/ramesh21a.html
Robbrecht, E., & Berghmans, D. 2004, A&A, 425, 1097
Robbrecht, E., Berghmans, D., & Van der Linden, R. A. M. 2009, ApJ, 691, 1222
Schwenn, R. 1996, Ap&SS, 243, 187
Shan, J., Zhang, H., Lu, L., et al. 2024, ApJS, 272, 18
Simonyan, K., & Zisserman, A. 2015, arXiv:1409.1556
Thernisien, A. F., Chua, D. H., Carter, M. T., et al. 2025, arXiv:2508.13467
Vaswani, A., Shazeer, N., Parmar, N., et al. 2017, in Advances in Neural Information Processing Systems 30, ed. I. Guyon et al. (NeurIPS) https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
Vourlidas, A., Balmaceda, L. A., Stenborg, G., & Lago, A. D. 2017, ApJ, 838, 141
Wang, P., Zhang, Y., Feng, L., et al. 2019, ApJS, 244, 9
Wang, X., Zhang, X., Cao, Y., et al. 2023, in 2023 IEEE/CVF Int. Conf. on Computer Vision (ICCV) (IEEE), 1130
Wang, Y., Liu, J., Jiang, Y., & Erdélyi, R. 2019, ApJ, 881, 15
Webb, D., Cliver, E., Crooker, N., Cyr, O., & Thompson, B. 2000, JGRA, 105, 7491
Williams, T., & Morgan, H. 2022, SpWea, 20, e2022SW003253
Xianggui, X., Zhenhong, S., Meiyu, Y., Zhipeng, Y., & Zhenping, Q. 2022, AstTI, 19, 65
Xie, E., Wang, W., Yu, Z., et al. 2021, in Advances in Neural Information Processing Systems 34, ed. M. Ranzato et al. (NeurIPS), 12077, https://proceedings.neurips.cc/paper_files/paper/2021/file/64f1f27bf1b4ec22924fd0acb550c235-Paper.pdf
Yang, P., Fu, H. S., Cao, J. B., & Shen, F. 2025, ApJS, 279, 23
Yashiro, S., Gopalswamy, N., Michalek, G., et al. 2004, JGRA, 109, A07105
Zhang, Y., Shen, Z., & Jiao, R. 2024, Comput. Biol. Med., 171, 108238
Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. 2017, in 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (IEEE), 6230
Zhukov, A. N., Thizy, C., Galano, D., et al. 2025, arXiv:2509.00253