The Astrophysical Journal, 997:148 (11pp), 2026 February 1 https://doi.org/10.3847/1538-4357/ae22d6

A Deep Learning Approach for Automated Total Sunspot Number Estimation

Yimin Wang (王宜敏)[1], Jiajia Liu (刘佳佳)[2,3], Jing Chen (陈静)[1], Robertus Erdélyi[4,5,6], Norbert G. Gyenge[4,6], and Ye Jiang (姜也)[7]

[1] School of Data Science, Qingdao University of Science and Technology, Qingdao 266061, People's Republic of China
[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] Solar Physics and Space Plasma Research Centre (SP2RC), School of Mathematical and Physical Sciences, The University of Sheffield, UK
[5] Department of Astronomy, Eötvös Loránd University, Budapest, Pázmány P. sétány 1/A, H-1117, Hungary
[6] Gyula Bay Zoltán Solar Observatory (GSO), Hungarian Solar Physics Foundation (HSPF), Petöfi tér 3., Gyula, H-5700, Hungary
[7] School of Information Science and Technology, Qingdao University of Science and Technology, People's Republic of China; ye.jiang@qust.edu.cn
Received 2025 May 22; revised 2025 November 12; accepted 2025 November 19; published 2026 January 20

Abstract

Accurate sunspot number estimation is essential for understanding the long-term evolution of solar activity and its impact on space weather. Sunspot numbers have been manually determined, leading to inconsistencies and observer-dependent biases. To address this, the World Data Center Sunspot Index and Long-term Solar Observations (WDC-SILSO) aggregates data from a global network of observatories to estimate the daily total sunspot number, enabling cross-validation and calibration across simultaneous observations. This study proposes a novel deep learning framework for automated total sunspot number calculation using solar full-disk continuum images from the Solar Dynamics Observatory. The method integrates U-Net for sunspot segmentation, K-means clustering for distinguishing umbrae from penumbrae, and You Only Look Once model for sunspot group detection. The selection of image-processing thresholds and neural network hyperparameters is optimized with respect to WDC-SILSO reference values during training. The results demonstrate a high correlation of 0.97 between the estimated and the WDC-SILSO daily total sunspot numbers. Furthermore, the framework offers a scalable approach suitable for future high-resolution solar observations.

Unified Astronomy Thesaurus concepts: Sunspot number (1652)

1. Introduction

Sunspots are dark, planet-sized regions on the solar photosphere marked by strong magnetic fields. They emerge when the Sun's internal magnetic field penetrates throughout the photosphere and extends outward into the lower solar atmosphere and the corona. Sunspots exhibit highly intense magnetic fields, typically thousands of times stronger than those in the surrounding regions, which suppress convective heat transfer from the Sun's interior to its surface, resulting in reduced temperatures and their characteristic darkened appearance (J. H. Thomas & N. O. Weiss 2008; T. J. Okamoto & T. Sakurai 2018; A. Siu-Tapia et al. 2019). The magnetic flux emerging at the Sun's surface varies systematically with the progression of the solar cycle, and visible sunspots are one of the manifestations of magnetic field perturbations. The sunspot count is closely correlated with solar activities and serves as a key indicator of eruptive phenomena such as solar flares and coronal mass ejections (CMEs), which have substantial impacts on space weather and the Earth's magnetosphere (J. Zhang et al. 2007; Y. Chi et al. 2016). The observation and feature extraction of sunspots play a significant role in solar activity forecasting. Reliable predictions of solar activity are vital for mitigating potential disruptions to modern communication infrastructures, navigation systems, and high-voltage power grids (X. Huang et al. 2018; 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).

Sunspot features are typically identified through manual inspection, often supplemented by basic image-processing techniques. Many automated sunspot detection methods have predominantly utilized morphological operations (S. Zharkov et al. 2005a, 2005b; J. Curto et al. 2008; U. Dasgupta et al. 2011; T. Baranyi et al. 2016; L. Györi et al. 2016; C. Zhao et al. 2016; S. Carvalho et al. 2020; S. Bourgeois et al. 2024; J. Chen et al. 2025). Beyond morphological approaches, other methodologies have emerged, including intensity filtering and region growing (T. Colak & R. Qahwaji 2008), wavelet transformations (D. Djafer et al. 2012), level set techniques (S. Goel & S. K. Mathew 2014), genetic algorithms (Y. Yang et al. 2018), and Gaussian mixture models (X. Gong et al. 2023), each contributing to the broader effort of improving automated sunspot detection.

These techniques have facilitated sunspot identification and feature extraction. However, they often rely on threshold selection and encounter difficulties when dealing with complex solar images due to limited feature learning capabilities. In contrast, convolutional neural network (CNN) models provide a powerful solution to this feature learning challenge (O. Ronneberger et al. 2015; K. Cao et al. 2023; Y. Zhang et al. 2024). CNNs can automatically extract features—ranging from intricate boundary and texture variations to global contextual relationships—directly from large volumes of labeled training data (X. Huang et al. 2024). For instance, CNN models have been employed by C. Chola & J. V. B. Benifa (2022) to classify continuum images based

on the presence or absence of sunspots and by J. Santos et al. (2023) to detect sunspots using bounding boxes. Additionally, N. Sayez et al. (2023), A. Mourato et al. (2024) and J. Chen et al. (2025) applied CNNs for precise sunspot segmentation.

A notable research gap remains, as few studies have focused on automating the estimation of total sunspot numbers, a process that has traditionally depended on human expert annotations for the past two centuries. C. Zhao et al. (2024) proposed an automated approach for calculating total sunspot numbers. However, it relied on a basic segmentation method based on mathematical morphology. The daily total sunspot number, R, is defined as

$$R = Ns + 10 \cdot Ng, \quad (1)$$

where Ns denotes the number of single spots, and Ng represents the number of sunspot groups observed across the entire solar disk.

Sunspot groups are strongly correlated with perturbations in the Sun's magnetic field. Sunspots within the same group not only are spatially close but often represent manifestations of the same active region. These regions of intense magnetic activity can give rise to solar flares and CMEs. The total sunspot number reflects this relationship by combining the contributions of both single spots and sunspot groups, capturing the fact that groups signify larger, more complex, and magnetically active regions of the Sun.

Moreover, long-term studies of sunspot evolution are fundamental to several branches of solar physics, including space weather (M. J. Owens et al. 2021; N. Buzulukova & B. Tsurutani 2022), helioseismology (L. Gizon 2004; R. Howe 2008), and solar irradiance modeling (T. Chatzistergos et al. 2020; T. N. Woods et al. 2022). Understanding sunspot behavior over extended periods is critical, as it also directly connects to key processes governing the evolution of Earth's climate (B. Owens 2013).

Commonly used sunspot catalogs include (i) daily total sunspot numbers from the Space Weather Prediction Center (SWPC)[8] covering the period 2011–2025; (ii) both daily single-spot counts and group counts from the Debrecen Photoheliographic Data (DPD)[9] spanning 2011–2015; and (iii) daily total sunspot numbers from the World Data Center Sunspot Index and Long-term Solar Observations (WDC-SILSO),[10] which provides records from 1818 to 2025 (see, e.g., F. Clette et al. 2007; S. Mathieu et al. 2019). In addition, WDC-SILSO offers a monthly mean total sunspot number since 1749 and yearly mean total sunspot number since 1700. Serving as the international reference standard, WDC-SILSO aggregates sunspot counts from approximately 80 observing stations worldwide.

This work proposes an automated framework for total sunspot number estimation by integrating deep learning techniques with physics-based knowledge. In addition to its role as the international reference, the WDC-SILSO dataset is adopted as the ground truth in this study due to its inclusion of both single-spot and group counts, as well as its uniquely long temporal coverage.

For sunspot segmentation, we follow the methodology of J. Chen et al. (2025), with a key modification: an additional correction layer is applied to remove the smallest and short-lived sunspots, thereby ensuring alignment with WDC-SILSO standard. The masks used for this approximation are generated based on the Uccle Solar Equatorial Table (USET)[11] station at the Royal Observatory of Belgium. Although the Sun can now be observed using advanced techniques such as high-resolution imaging, spectroscopy across wavelengths ranging from X-ray and extreme ultraviolet to optical and radio, and both ground- and space-based instruments capable of detecting even the smallest sunspots. These methods have only been developed in recent decades. Therefore, their records cover just one or a few solar cycles, which is insufficient for studying the long-term variability of the Sun. In contrast, telescopic visual observations have been conducted for over 400 yr. Sunspot counts provide the only direct and continuous record of solar activity spanning tens of solar cycles. Accordingly, incorporating USET-based corrections into the segmentation process improves consistency with the rules adopted by WDC-SILSO.

Within each segmented sunspot region, umbrae that are fully enclosed by penumbrae are distinguished as separate entities. The total number of individual sunspots is determined by counting only the umbrae. Sunspot groups are identified using a You Only Look Once (YOLO) model[12] applied to solar full-disk continuum images containing only the segmented sunspots. The final daily total sunspot number is then computed using Equation (1). Beyond its application to ground-based observations, this framework is designed to be adaptable for data from next-generation space-borne telescopes with higher resolutions, as the method is robust to variations in image resolution. As demonstrated by S. Liu et al. (2022), deep learning networks are insensitive to resolution, provided that the smallest sunspots are visible. The rest of this paper is organized as follows: Section 2 provides a detailed description of the dataset, followed by an explanation of the methodologies in Section 3. The results and corresponding discussions are presented in Section 4, and the paper concludes with a summary of findings in Section 5.

2. Data

The full-disk continuum observations from the Helioseismic and Magnetic Imager (HMI) on board the Solar Dynamics Observatory (SDO) are retrieved from the Joint Science Operations Center (JSOC) database[13] (J. Schou et al. 2012). Since SDO/HMI began routine observations in 2010, this study utilizes data covering the period from 2010 July to 2023 June. The data preprocessing pipeline includes derotating the tilt of the Sun's north–south axes and applying limb darkening correction. The original image resolution of 4096 × 4096 pixels is preserved to retain the smallest sunspots, which are typically represented by only a few pixels.

The training set was constructed using observations collected from January 1 to September 15 of each year, while the test set was composed of data from October 1 to December 15. To mitigate potential data leakage, observations from September 16 to September 30 were excluded from both sets.

To further optimize the dataset for the tiling procedure employed in the segmentation process, only solar disks

[8] https://www.swpc.noaa.gov/products/solar-region-summary
[9] http://fenyi.solarobs.epss.hun-ren.hu/DPD/
[10] https://www.sidc.be/SILSO/datafiles
[11] https://vo-tap.oma.be/
[12] https://github.com/ultralytics/yolov5
[13] http://jsoc.stanford.edu/ajax/lookdata.html

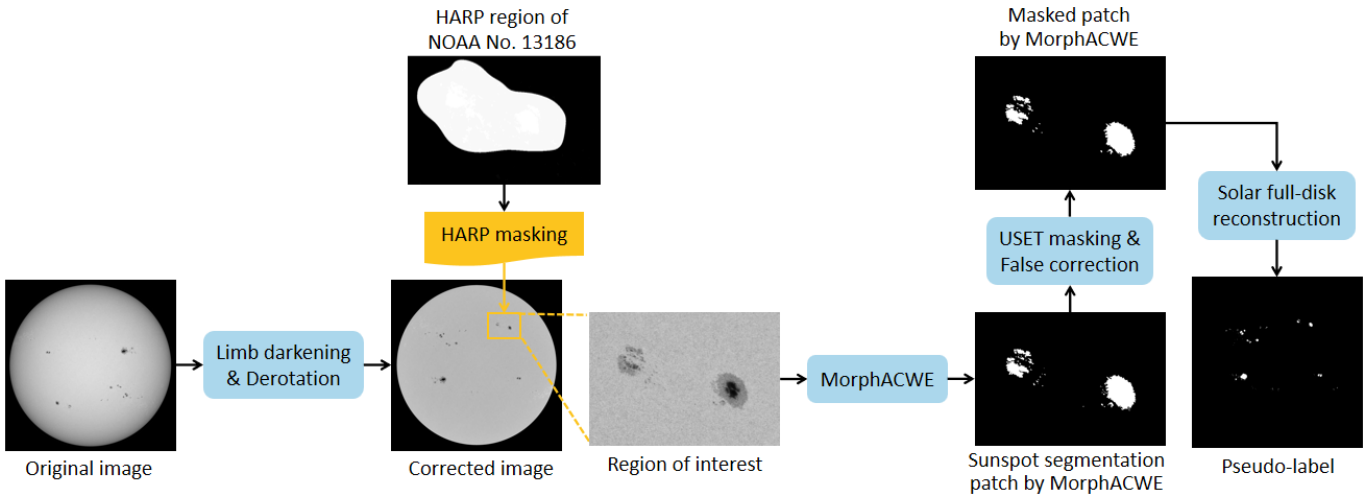


Figure 1. Flowchart for generating pseudolabels, using one HARP region of NOAA No. 13,186 as an example.

containing sunspots were included. The final training set comprises 2464 images, while the test set includes 775 images.

The HMI Active Region Patch (HARP) data are also obtained through the JSOC database to identify active regions within the continuum images (M. G. Bobra et al. 2014). This dataset provides approximate boundaries for entire sunspot groups, making it an effective tool for selecting regions of interest from full-disk observations. Both the full-disk continuum observations and the HARP active region data are downloaded once daily at 12 PM International Atomic Time, ensuring temporal consistency between the two datasets.

Since 1941, USET has compiled an uninterrupted series of sunspot drawings (S. Bechet et al. 1940), with human operators manually annotating these drawings to record sunspot groups and total sunspot counts. In this study, sunspot groups are manually annotated based on the drawings from the USET dataset. For each SDO/HMI continuum image, the closest corresponding USET drawing is paired to provide precise annotations of sunspot groups. These annotations are represented as the smallest-possible bounding boxes, denoted as USET masks, which are used to refine the ground truth for both sunspot segmentation and sunspot group detection, thereby improving the accuracy of the automated framework. The daily total sunspot number is obtained from the WDC-SILSO catalog, as mentioned earlier.

3. Methodology

To determine the daily total sunspot number, as defined in Equation (1), the process begins by identifying sunspots in a solar full-disk continuum image. The detected sunspots are then classified into umbrae and penumbrae, enabling the counting of individual spots. Sunspot groups are identified, yielding the number of groups. The image-processing thresholds and neural network hyperparameters are initially selected to ensure that the model outputs are qualitatively consistent with the manual annotations and/or human visual assessment. Subsequently, these parameters are further optimized by minimizing the discrepancy between the model's outputs and the total sunspot number published by WDC-SILSO. The sunspot segmentation method employed in this study is a refined version of the approach described in our previous work (J. Chen et al. 2025), with a brief summary provided for completeness within the overall framework.

3.1. Pseudolabeling

In the context of full-disk continuum observations, the semantic segmentation method U-Net is employed to segment sunspots. U-Net, a supervised deep learning model, features an encoder–decoder architecture based on CNNs (O. Ronneberger et al. 2015). Rather than relying on manual annotation, a pseudolabeling approach is adopted, in which a model identifies sunspots that serve as labels for training U-Net. This approach is preferred due to the substantial time demands required for manual annotation and the potential introduction of personal bias.

Figure 1 illustrates the workflow for generating pseudolabels. The process begins with limb darkening correction and derotation of the raw continuum images. For pseudolabeling, this study employs the morphological active contours without edges (MorphACWE) method (P. Márquez-Neila et al. 2014). MorphACWE is an unsupervised image segmentation technique designed to handle scenarios where object boundaries may be partially obscured. The method requires that the inner and outer regions of the target feature exhibit distinct average pixel intensities. Given the assumption that sunspots are darker than the surrounding quiet Sun, MorphACWE is considered suitable for this application.

Sunspot groups are strongly linked to active regions. To improve the accuracy of the pseudolabels, the coordinates of active regions provided by the HARP data are utilized to define regions of interest within a full-disk continuum image. MorphACWE is then applied to these targeted regions to obtain sunspot segmentation patches.

This is the point where the segmentation approach deviates from J. Chen et al. (2025). The WDC-SILSO dataset excludes very small or short-lived sunspots. To ensure consistency with this convention, the outputs of MorphACWE are masked using manual annotations of sunspot groups provided by USET. This masking process excludes certain small sunspots that, although present within the active regions, lie outside the sunspot groups defined by USET. The smallest sunspots typically have a diameter of approximately 1500 km. Given that the pixel size of SDO/HMI is $0\farcs505\,\mathrm{pixel}^{-1}$ (J. Schou et al. 2012), each pixel in a 4096×4096 pixel image corresponds to roughly 362.6 km at the disk center. Consequently, the smallest sunspots generally span about 4.14 pixels at the disk center. However, this apparent size decreases toward the edge of the solar disk due to foreshortening, which follows a cosine

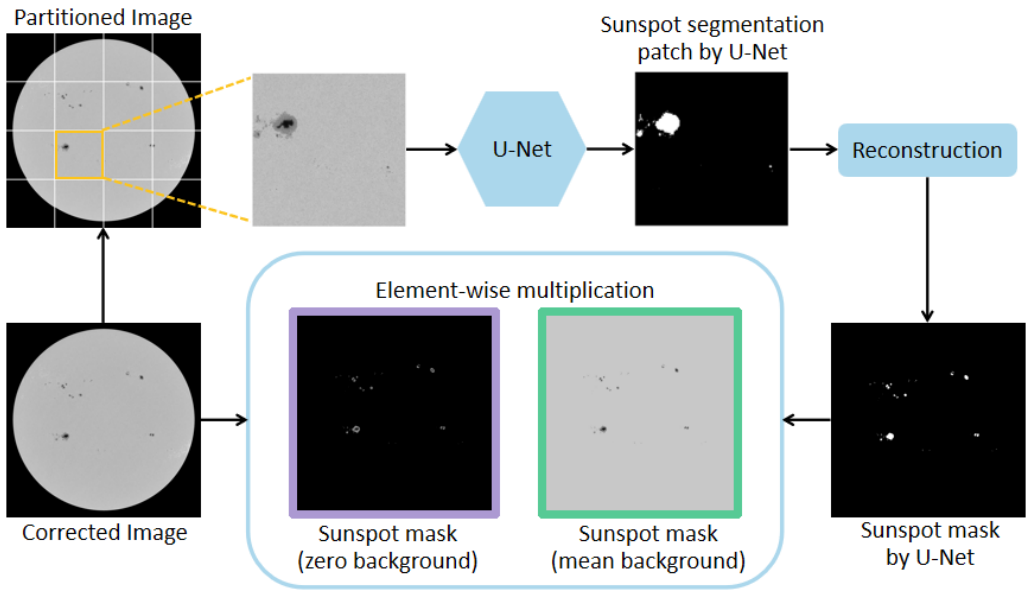


Figure 2. Sunspot segmentation process using U-Net. The purple-highlighted image displays the predicted results generated by U-Net, with nonsunspot areas assigned a value of 0, while the green-highlighted image illustrates the same sunspot segmentation, but with nonsunspot areas represented by their mean value.

dependence on heliographic latitude and longitude. Specifically, the apparent pixel size n of the smallest sunspot at a given location is calculated as

$$n = 4.14 \cdot \cos\phi \cdot \cos\lambda, \tag{2}$$

where ϕ and λ denote the heliographic latitude and longitude, respectively. As such, any dark features identified by the MorphACWE algorithm that occupy fewer than the smallest sunspots are considered false detections and are excluded from the final segmentation mask.

Following the segmentation and correction process, the patches are merged back into a final full-disk sunspot segmentation mask, which is referred to as the pseudolabels.

3.2. Semantic Segmentation

U-Net has emerged as a widely utilized deep learning architecture for semantic segmentation, owing to its accuracy and efficacy, particularly in domains requiring detailed segmentation, such as medical imaging and satellite image analysis (T. P. T. Armand et al. 2024; A. Mourato et al. 2024). However, sunspots are considerably smaller in scale compared to the solar disk and the background of continuum images. Our experiments indicate that, when a full-disk observation is input into U-Net, the model tends to focus primarily on large-scale features, such as the solar disk and the background. This is due to the potential inadequacy of learned features in capturing subtle and intricate details, particularly as a result of downsampling and convolutional operations. This phenomenon is known as small object detection in image processing.

To address this challenge, continuum observations with a resolution of 4096 × 4096 pixels were partitioned into 16 evenly sized patches, each measuring 1024 × 1024 pixels, prior to being fed into U-Net, as illustrated in Figure 2. The loss is computed for each patch to facilitate backpropagation. Our results indicate that reducing the patch size to 512 × 512 pixels degraded U-Net's performance. While smaller patches allow the model to focus more effectively on small sunspots, they also increase the likelihood of splitting a single sunspot across multiple patches, which complicates feature extraction and limits the model's ability to accurately capture sunspot contours.

The output patches from U-Net are reconstructed into 4096 × 4096 pixel images. The output pixels represent probabilities indicating the likelihood of being sunspots, yielding continuous values between 0 and 1. These pixel values are thresholded at 0.5 to generate binary masks, where sunspot areas are assigned a value of 1, and nonsunspot areas are represented by 0. The employed loss function is binary cross-entropy (BCELoss), and the model is optimized using the Adam algorithm. During training, the learning rate is set to 0.001, with a batch size of 2, and the model is trained for nine epochs. Group normalization is utilized in place of batch normalization to ensure consistent performance with the small batch size.

Elementwise multiplication is performed between these binary masks and their corresponding continuum observations to generate masked continuum images, preserving sunspots while setting nonsunspot areas to 0. Furthermore, the mean intensity value of the nonsunspot areas is calculated during sunspot segmentation to facilitate alignment with the annotated sunspot group data.

3.3. Sunspot Number Estimation

Following the identification of sunspots from the full-disk observations, the next step diverges into two branches. One branch involves computing the number of single spots, while the other focuses on determining the number of sunspot groups, as illustrated in Figure 3.

To determine the number of spots, we first identify connected components within the sunspot mask and count the nonzero components. A component is deemed to contain a penumbra if its intensity histogram is bimodal with two prominent peaks (Figure 4). In the example shown, peaks near pixel values ∼25 and ∼145 correspond to the umbra and penumbra, respectively. Operationally, we apply SciPy's peak detector (`scipy.signal.find_peaks`[14]) to each component's histogram

[14] https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html

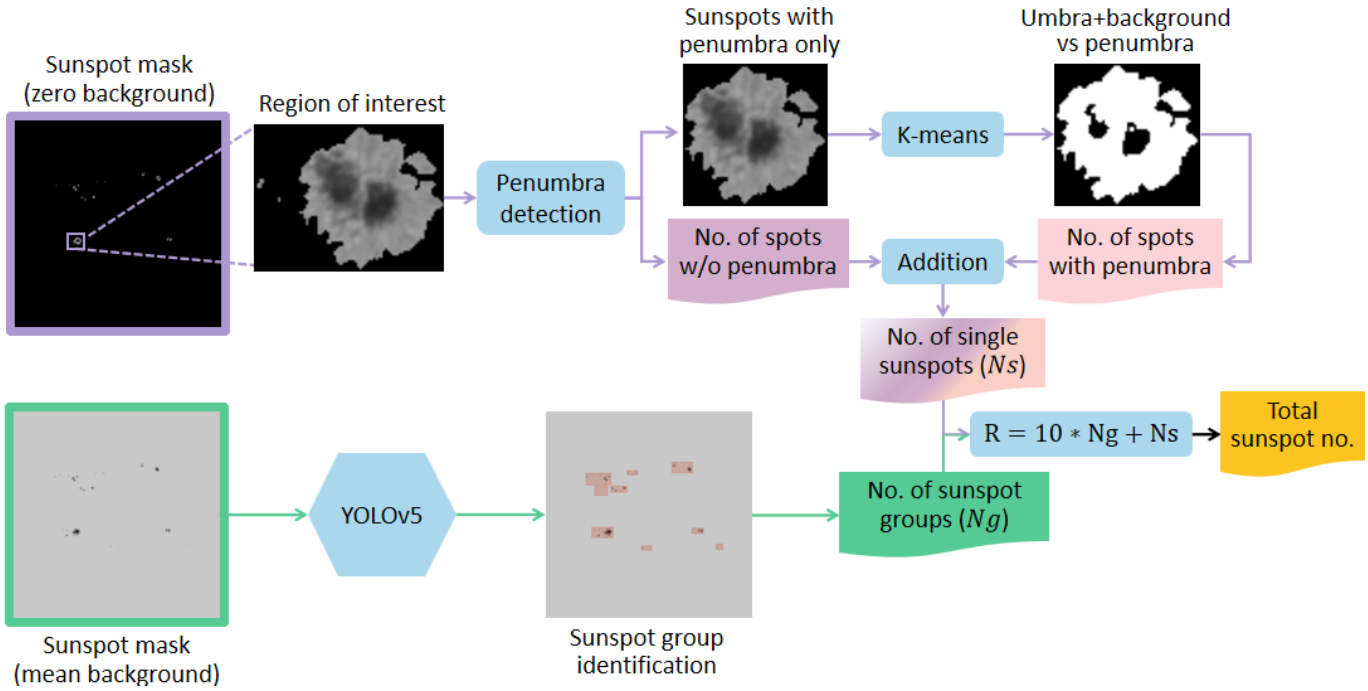


Figure 3. The pipeline for sunspot number estimation.

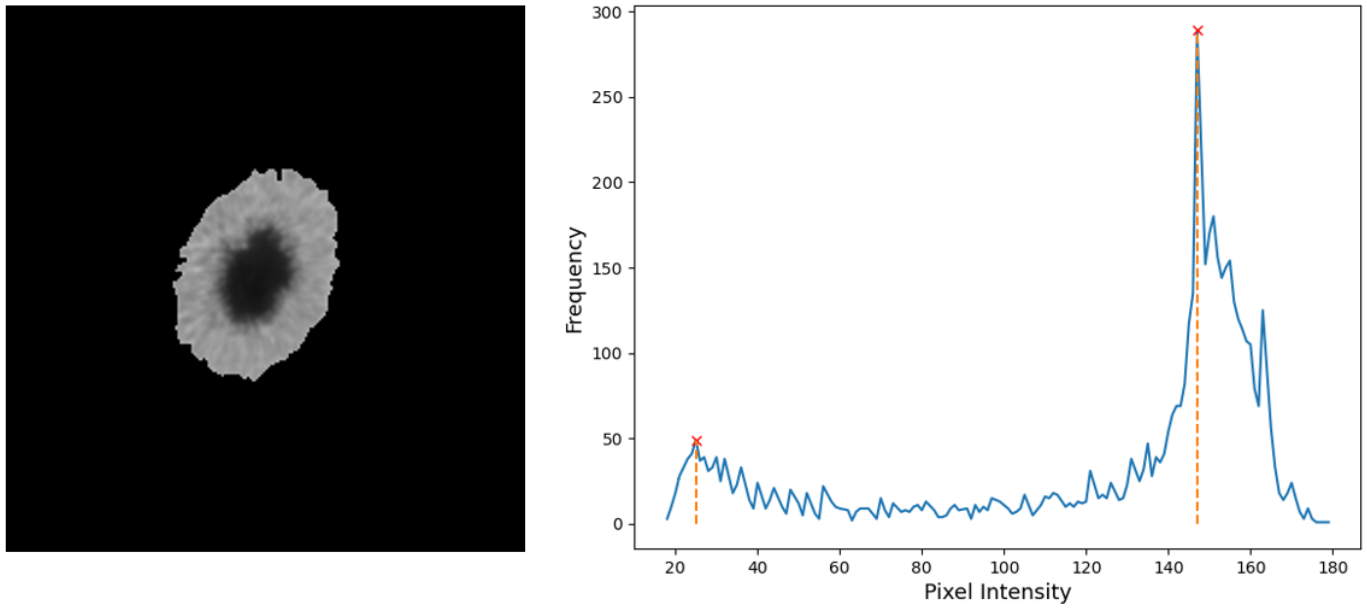
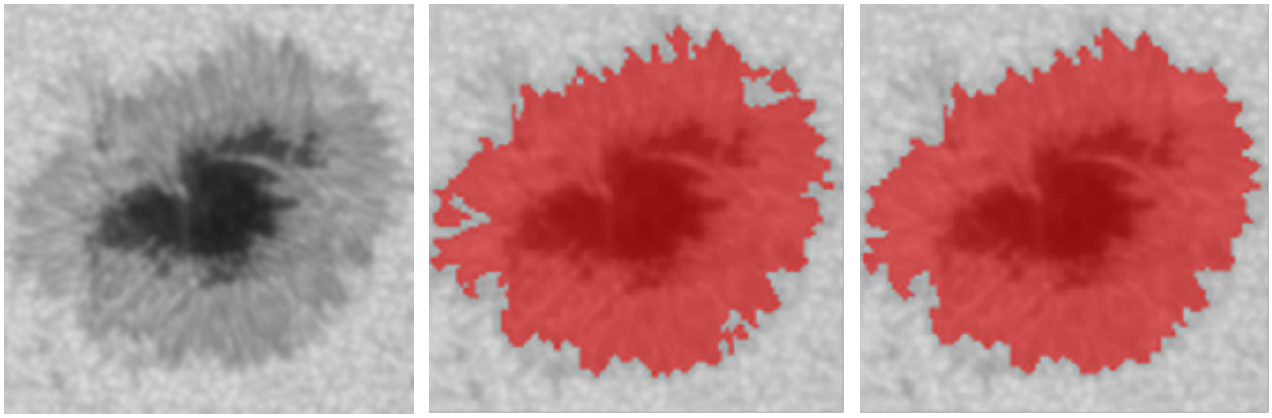


Figure 4. The sunspot region (left) and the corresponding histogram of pixel intensity (right).

with `height` $\geqslant$10, `prominence` $\geqslant$2, and `distance` $\geqslant$90, after inspecting representative components to set these thresholds. Here, `height` is the peak amplitude, which suppresses spurious peaks from noise or isolated pixels; `prominence` is the elevation of a peak above the higher of its two bounding minima, and a threshold of 2 means the peak is meaningfully higher than local fluctuations; `distance` enforces a minimum separation between accepted peaks, preventing closely spaced features from being counted as distinct modes. Components with fewer than two qualifying peaks are classified as sunspots without penumbrae and are subsequently removed from the continuum mask, leaving a final mask that retains only sunspots with penumbrae.

To further extract the umbra from the penumbra, K-means clustering with two clusters is applied, separating the umbra and background (both characterized by lower intensity values) from the penumbra. By default, K-means assigns the label zero to the cluster corresponding to the umbra and background, and the label one to the penumbra cluster. Because the connected component analysis counts the number of components with a pixel value of 1, the cluster labels must be swapped. After this reassignment, pixels belonging to the umbra and background take the value of 1, while those of the penumbra are assigned 0. Applying a connected component analysis then yields the total number of individual umbrae and background components. Since all umbrae are embedded within the penumbrae,



(a) Original image (b) MorphACWE predicted mask (c) U-Net predicted mask

Figure 5. Comparison of the segmentation results produced by MorphACWE and U-Net. (a) The original continuum image. (b) Shows the predicted segmentation mask by MorphACWE. (c) Demonstrates the predicted segmentation mask by U-Net. MorphACWE leaks across penumbral boundaries, whereas U-Net more accurately delineates sunspot edges.

the background itself forms a single connected component. Consequently, the number of umbrae is given by the total number of connected components minus 1, which accounts for the background. Finally, the total number of sunspots is computed by summing the counts of single spots without penumbrae and those with penumbrae.

The deep learning algorithm YOLO is employed to detect sunspot groups due to its efficacy in real-time processing, multiscale detection, and robust performance in accurately localizing objects with varying sizes and overlapping structures. This capability makes it particularly well suited for dynamic solar observation and large-scale solar image analysis. In this study, the performance of the widely recognized YOLOv5[15] and YOLOv9,[16] (C.-Y. Wang et al. 2025) models is examined, with both models demonstrating comparable detection accuracy. Finally, the number of single spots and groups is used as inputs in Equation (1) to compute the total sunspot number. The training process is conducted with a batch size of 9 over 30 epochs, maintaining the original loss functions and optimizers of the YOLO models. A learning rate of 0.01 is applied throughout the training process.

3.4. Inference

Despite the multiple unsupervised and supervised machine learning methods and calibrations involved in the training process, the testing procedure remains relatively straightforward. For estimating the total sunspot number, a preprocessed continuum observation is input into U-Net for sunspot identification, which generates a mask of the continuum image (refer to Figure 2).

The masked continuum image, with the background intensity value set to 0, is processed to separate the umbrae from the penumbrae, followed by K-means clustering to count the number of single spots. Simultaneously, the masked continuum image with the mean background intensity value is input into YOLOv5 to count the number of sunspot groups (refer to Figure 3).

[15] https://github.com/ultralytics/yolov5, released under the AGPL-3.0 license.
[16] https://github.com/WongKinYiu/yolov9 released under the GPL-3.0 license.

4. Results and Discussions

4.1. Sunspot Identification

U-Net is adopted on top of MorphACWE due to its ability to automatically learn and extract adaptive, hierarchical features essential for a complex solar image analysis. Traditional image-processing techniques, such as MorphACWE, rely primarily on pixel-level gradient information, and lack the ability to capture higher-level structural features. U-Net effectively captures the surrounding context of sunspot boundaries, a capability that MorphACWE inherently lacks, as illustrated in Figure 5.

Initial segmentation masks are generated with MorphACWE. These masks are subsequently constrained by USET masks (manually annotated, smallest-possible bounding boxes around sunspot groups), which removes small detections outside the boxes and produces the pseudolabels used to train the U-Net. Small detections within the boxes are retained at this stage. The resulting U-Net segmentations are then inspected and refined against sunspot drawings from the USET database to establish the ground truth, during which remaining small sunspots across the image are removed. This ground truth is used to further train U-Net and to evaluate and compare the performance of MorphACWE and U-Net. The performance of the segmentation models is briefly evaluated for the completeness of the proposed framework. A more detailed evaluation can be found in J. Chen et al. (2025). The following quantitative metrics are utilized:

Precision measures the proportion of true positive predictions among all positive predictions made by a model. It is defined as

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{3}$$

where TP represents true positives (the model successfully detects the sunspot), and FP represents false positives (the model incorrectly identifies sunspots when none is actually present). High precision indicates that the model is effective in reducing the occurrence of false sunspots.

Recall (or sensitivity) assesses the proportion of true positive predictions among all actual positives. It is calculated

Table 1
Comparison between MorphACWE and U-Net Segmentation Results in the Test Set

Model	Precision	Recall	F1-score	IoU
MorphACWE	0.810	0.930	0.847	0.756
U-Net	0.989	0.992	0.986	0.981

as

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{4}$$

where *FN* denotes false negatives (the model fails to detect sunspots). High recall reflects the model's ability to capture as many positive sunspot pixels as possible.

F1-score is the harmonic mean of precision and recall, providing a single metric that balances both aspects. It is defined as

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{5}$$

This metric is particularly useful when dealing with imbalanced datasets, as it considers both *FPs* and *FNs*.

Intersection over union (IoU) quantifies the overlap between the predicted and ground-truth regions in object detection tasks. It is defined as

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}, \tag{6}$$

where the area of overlap is the intersection of the predicted and ground-truth regions, and the area of union is their union. IoU provides a measure of how well the predicted sunspot masks align with the actual ones.

Table 1 presents the precision, recall, F1-score, and IoU for both MorphACWE and U-Net. The metrics presented are evaluated at the pixel level, assessing the accuracy of each pixel classification. The notable differences in precision, and consequently in the F1-score, between the two segmentation models presented in Table 1 mean MorphACWE tends to produce a larger number of false sunspot detections compared to U-Net. This can partially be explained by the removal of small sunspots according to USET dataset when constructing the ground truth.

Table 1 presents the precision, recall, F1-score, and IoU for both MorphACWE and U-Net. The metrics presented are evaluated at the pixel level, assessing the accuracy of each pixel classification. The lower precision of MorphACWE relative to U-Net, and the associated reduction in F1, indicates a higher rate of *FP* sunspot pixels for MorphACWE. This discrepancy is partly attributable to the ground-truth construction, which removes small sunspots based on USET dataset. Detections of such small spots are therefore counted as *FPs* affecting MorphACWE. The high recall achieved by U-Net indicates its ability to capture a larger proportion of actual sunspot pixels, highlighting its capability to learn the hierarchical structure of sunspots, as demonstrated in Figure 5. Overall, the superior metric values of U-Net underscore its effectiveness in segmenting sunspots from full-disk continuum images. Notably, the additional refinement layer based on USET data, applied to the MorphACWE results, is unnecessary when the application is not constrained to align with the WDC-SILSO convention, enabling our method with greater future scalability.

4.2. Detection of Single Spot and Sunspot Group

The K-means clustering algorithm is employed to effectively distinguish umbrae from penumbrae, facilitating the accurate identification and counting of individual sunspots encompassed by penumbrae. The umbrae, characterized by its significantly cooler temperature, exhibit much lower intensity compared to the penumbrae. The reduced intensity of the umbrae causes them to cluster together with the dark background, as illustrated in Figure 6. This classification process isolates the penumbrae into a separate cluster, ensuring a clear and reliable distinction between these two regions of sunspots.

YOLOv5 is employed for detecting sunspot groups on full-disk solar segmentation images, where the background is assigned the mean intensity of nonsunspot regions. The representative examples demonstrate the effectiveness of YOLOv5 in accurately detecting and localizing sunspot groups, as depicted in Figure 7. The performance of YOLOv5 is evaluated using precision, recall, F1-score, and average precision at a confidence level of 0.5 (AP50), as presented in Table 2.

Precision is defined as the proportion of correctly detected sunspot groups out of all groups detected by the model. In this context, *TP* refers to predictions where the detected group correctly matches the ground truth with an IoU $\geqslant 0.5$. Conversely, *TF* represents predictions where the detected group either does not match any ground-truth groups or exhibits an IoU below the specified threshold of 0.5.

Recall quantifies the ability of YOLOv5 to correctly identify all ground-truth groups. In this context, *FN* corresponds to ground-truth groups that were not correctly detected by the model, i.e., the model fails to identify the group.

AP50 calculates the area under the precision–recall curve when the IoU threshold is set to 0.5. It serves as an indicator of how well the model detects objects with moderate localization accuracy.

The sunspot butterfly diagram is presented in Figure 8. The corrected areas are computed for each sunspot group, providing a quantitative measure of the associated magnetic flux. The latitude of each sunspot group is determined based on the central position of a group. This diagram offers valuable insights into the cyclical nature of solar activity and the latitudinal migration of sunspots throughout the solar cycle. Specifically, it illustrates how sunspots initially emerge at higher latitudes ($\sim 30^\circ$–40°) at the onset of a solar cycle and gradually migrate toward the equator ($\sim 5^\circ$) as the cycle progresses. The characteristic pattern of the butterfly diagram confirms that the identified sunspots and groups align with the expected latitude evolution across multiple solar cycles.

4.3. Estimation of Sunspot Numbers

The comparison of daily total sunspot numbers between various human annotations and the estimated values from the proposed framework is presented in Table 3. Both WDC-SILSO and USET are operated by the Royal Observatory of Belgium, with USET serving as one of the observational resources that directly contributes to WDC-SILSO's solar monitoring and sunspot number calculations.

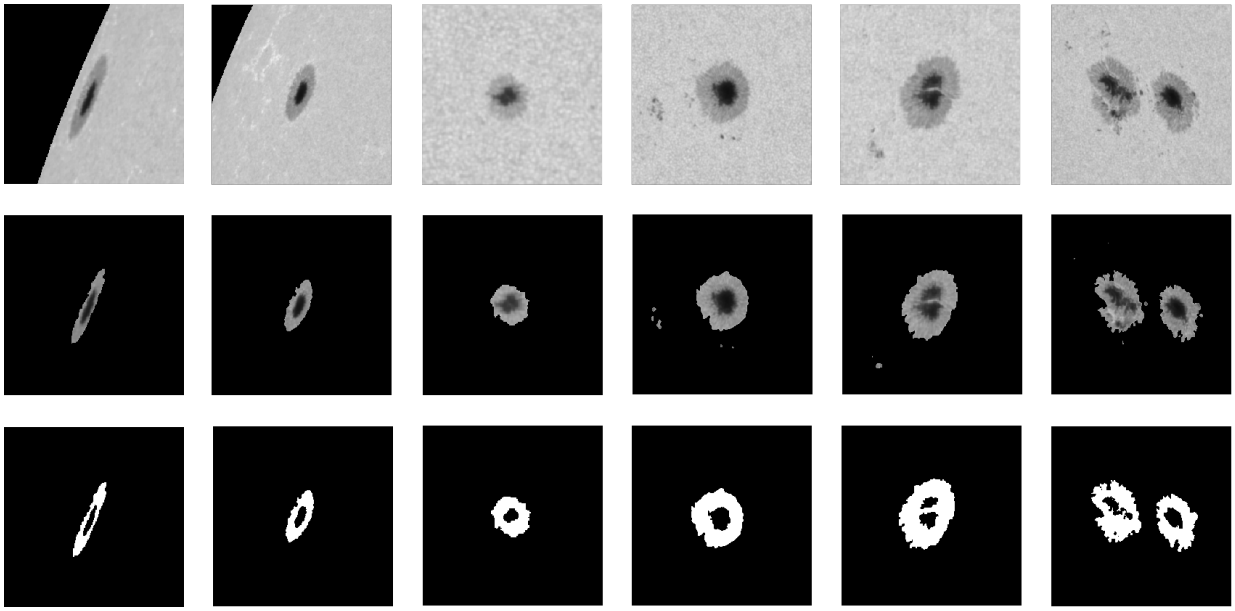


Figure 6. Examples of separation of umbrae and penumbrae with K-means. The first row displays patches from SDO/HMI continuum observations. The second row presents the sunspot segmentation results. The third row shows the clustering outcomes generated using K-means.

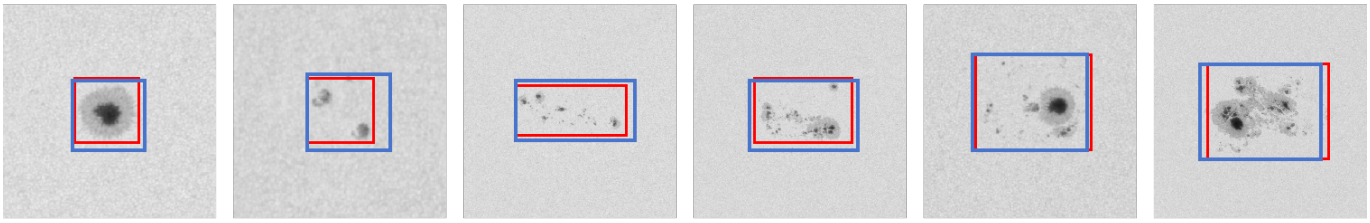


Figure 7. Examples of sunspot group detection. The red boxes highlight the annotated sunspot groups based on SDO/HMI continuum observations, while the blue boxes indicate the detection results of sunspot groups generated using YOLOv5.

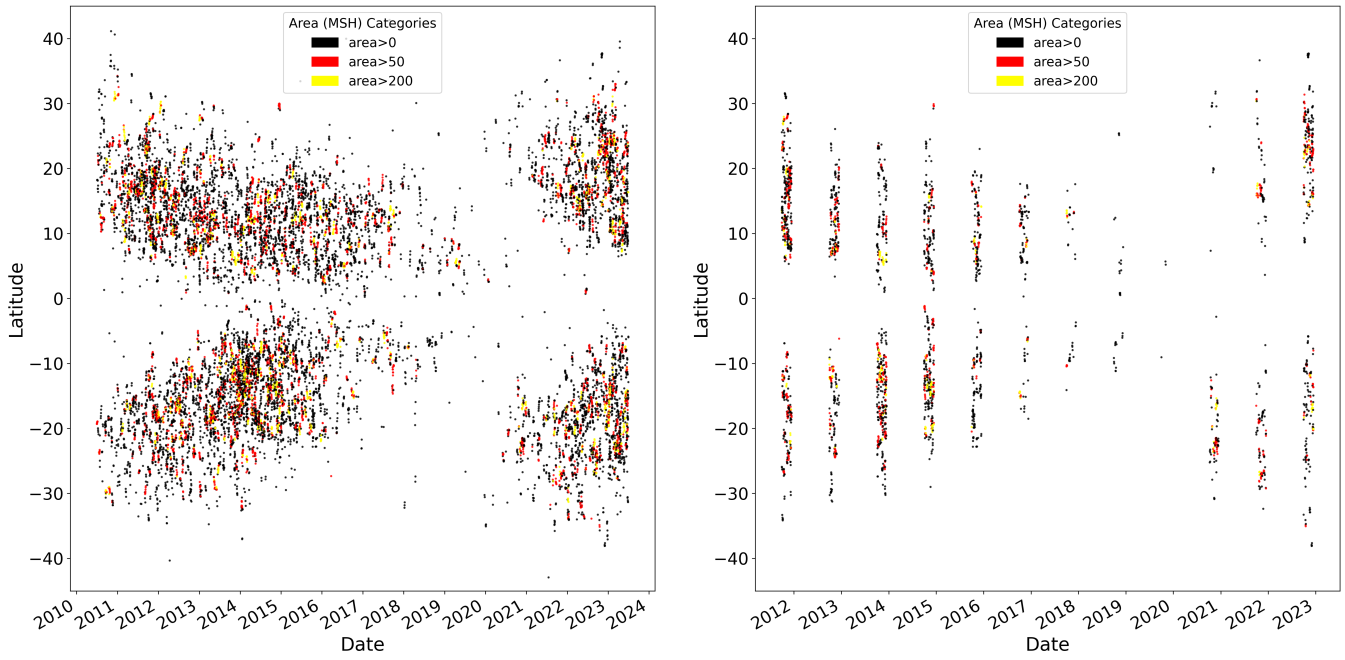
Table 2
Evaluation Results of the Proposed Sunspot Group Detection Method Using YOLOv5

Method	Precision	Recall	F1-score	AP50
YOLOv5	0.757	0.732	0.744	0.732

The WDC-SILSO database does not currently provide daily counts of single spots or sunspot groups. The Uccle station, recognized for its stability (F. Clette et al. 2016), has produced an extensive observational record spanning nearly six decades (F. Clette et al. 2014). Therefore, the estimated sunspot numbers in this study are validated against the corresponding daily reports from USET.

The Pearson correlation coefficients between the USET single-spot counts and sunspot group numbers, in comparison with the estimated values, are 0.88 and 0.93, respectively. As shown in Figure 9, USET's total sunspot numbers are generally lower than the estimated values, with a correlation coefficient of 0.93. Notably, despite these discrepancies, a strong correlation of 0.97 is achieved between the estimated daily total sunspot numbers and those reported by WDC-SILSO, as illustrated in Figure 9(a). Furthermore, the correlation coefficient between USET and WDC-SILSO total sunspot numbers is 0.95. The better correlation between the proposed method and WDC-SILSO, as compared to that between the proposed method and the single USET station, is demonstrated in Figure 9(a), where the latter exhibits more scatter.

The proposed approach mimics the logic of human counting while minimizing subjective bias. This forms the first pillar of our contribution. According to S. Mathieu et al. (2019), manual sunspot counts are prone to short-term inconsistencies caused by variable seeing conditions (e.g., weather, atmospheric turbulence), the involvement of multiple observers at one station, and counting or transcription errors. Long-term errors may arise from gradual instrument degradation and systematic biases in the counting process, with biases being particularly pronounced during solar minima when short-lived sunspots are common. To mitigate these potential issues, we rely on satellite imagery, which eliminates various atmospheric seeing conditions, and ensures more homogeneous observations. Further, an automated system also avoids interobserver variability and transcription errors. Furthermore, because the image-processing thresholds and neural network hyperparameters are optimized with respect to the WDC-SILSO total sunspot number during training, long-term systematic errors can be effectively reduced. Nevertheless, the automated system may still be subject to residual biases introduced by human intervention in parameter settings or by the limited span of solar cycles represented in the training data. The performance and stability of the system are expected to further improve as longer-term satellite observations become available.



(a) Training and test data (b) Test data only

Figure 8. Sunspot butterfly diagram illustrating (a) both the training and test datasets, and (b) the test dataset only. The black, red, and yellow dots represent the corrected areas of sunspot groups within the ranges of 0–50 millionths of solar hemisphere (MSH), 50–200 MSH, and greater than 200 MSH, respectively.

Table 3
Correlation Coefficients between the Proposed Method and the WDC-SILSO, USET, SWPC, and DPD Datasets Evaluated Based on the Number of Single Spots, Sunspot Groups, and Total Sunspot Counts, Where Available

Dataset	No. of Single Spots	No. of Groups	Total Sunspot No.
SILSO	None	None	0.97
USET	0.88	0.93	0.93
SWPC	None	None	0.94
DPD	0.78	0.74	0.81

Note. "None" means no record is available from this dataset.

Our second contribution deals with the fact that traditional sunspot masks are typically available only as hand-drawn sketches, which limits their utility for large-scale analysis. Our method generates accurate by-products including sunspot masks, umbra/penumbra segmentation, and positional information, all of which are calibrated to be consistent with the relevant WDC-SILSO records.

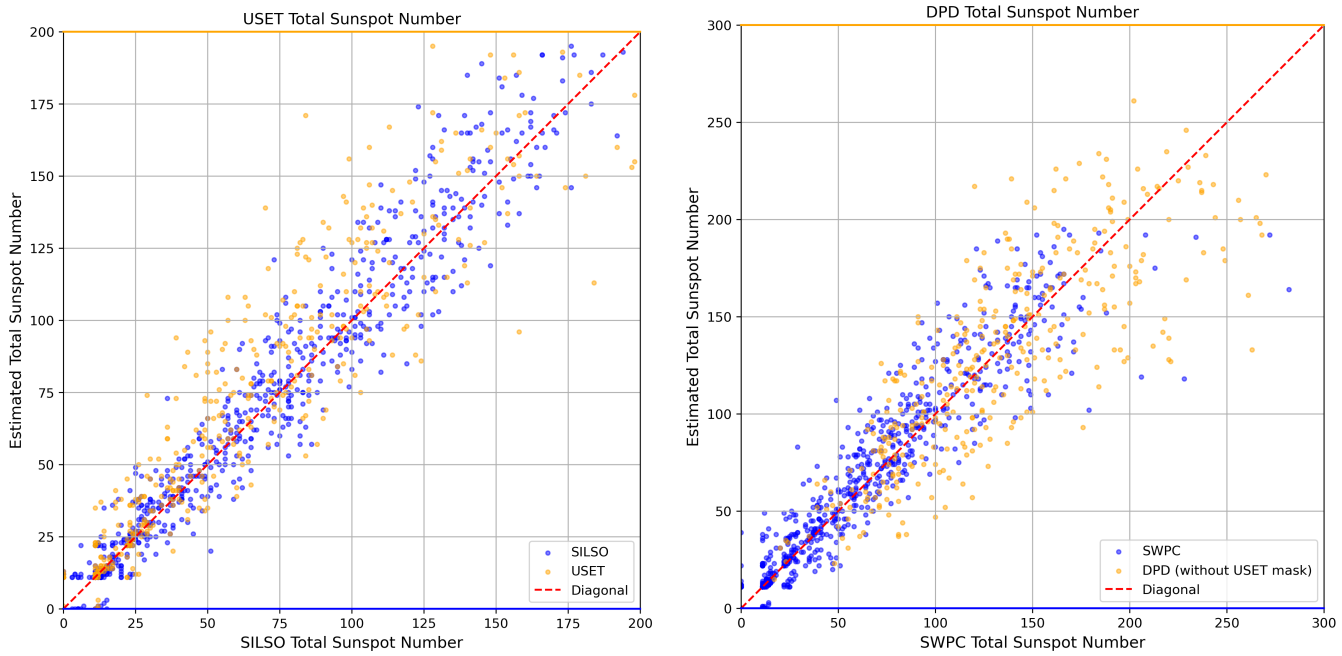
Third, the incorporation of advanced techniques (such as limb darkening correction, feature extraction, and image segmentation) enhances the model's sensitivity to low-contrast features, enabling the detection of faint or small sunspots that may be missed by human observers. Therefore, the proposed method provides an automated and scalable approach capable of replicating the composite outputs of a multistation network such as WDC-SILSO. Notably, the performance of the method should continue to be monitored, as it remains uncertain whether it will perform equally well across different solar cycle regimes (S. Mathieu et al. 2023). This nevertheless highlights its potential as a valuable tool for future high-resolution solar observations and, more broadly, for long-term monitoring of solar activity.

The correlation coefficient between the SWPC total sunspot numbers and the estimated values is 0.94. The discrepancy could be attributed to the higher consistency and reliability of the proposed sunspot segmentation approach, as discussed by J. Chen et al. (2025). As shown in Figure 9(b), without applying the additional USET-based mask, the correlation coefficient between DPD and the estimated total sunspot numbers is 0.81. When the mask is applied, this correlation decreases slightly to 0.79, and DPD continues to report higher total sunspot numbers than the proposed method. This difference likely arises from distinct sunspot identification criteria, as DPD tends to include significantly smaller sunspots (L. Lefèvre & F. Clette 2011).

5. Conclusions

This study presents a deep learning-based approach for automated total sunspot number estimation. The proposed framework employs semantic segmentation methods to identify sunspots, machine learning clustering techniques to distinguish umbrae from penumbrae, and object detection models to detect sunspot groups, enabling a fully automated determination of daily total sunspot numbers. The Pearson correlation coefficients between the estimated total sunspot numbers and the WDC-SILSO dataset reach 0.97, demonstrating the reliability of our approach.

Traditional sunspot counting methods rely on manual annotations or threshold-based image processing, both of which are susceptible to inconsistencies and biases. Our framework improves upon these approaches by leveraging deep learning to automatically count individual sunspots and detect sunspot groups. The correction mask based on USET data enhances the consistency of automated results with the WDC-SILSO convention regarding which sunspots should be



(a) Estimated vs WDC-SILSO/USET (b) Estimated vs SWPC/DPD

Figure 9. Comparison of daily total sunspot numbers for each test image between (a) the estimated values and WDC-SILSO (blue dots) or USET (yellow dots) observations; and (b) the estimated values and SWPC (blue dots) or DPD without USET mask (yellow dots) observations. The red-dashed line indicates the points where estimated values match the corresponding observations.

counted. In addition, the framework parameters are optimized with respect to WDC-SILSO reference values, ensuring a more objective and reliable estimation process.

While the proposed approach is robust, future work will focus on refining the calculation method by incorporating temporal information. Examining multiple preceding and subsequent images could help differentiate persistent sunspots from transient noise. Additionally, incorporating solar cycle variability into the training dataset may further enhance the model's adaptability across different phases of solar activity. By integrating advanced time-series models, future studies aim to improve the accuracy and reliability of sunspot number estimation, contributing to a deeper understanding of solar activity and its implications for space weather forecasting.

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). Y.J. acknowledges the support from the Natural Science Foundation of Shandong Province (ZR2023QF151). J.L. acknowledges support from the National Natural Science Foundation (NSFC 42521007, 42188101, 12373056). 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). N.G.G. acknowledges NVIDIA Academic Hardware Grant Program for their support in providing essential computational resources (NVIDIA A5000 GPU).

ORCID iDs

Yimin Wang (王宜敏) https://orcid.org/0000-0002-8835-3825
Jiajia Liu (刘佳佳) https://orcid.org/0000-0003-2569-1840
Jing Chen (陈静) https://orcid.org/0009-0005-8109-1497
Robertus Erdélyi https://orcid.org/0000-0003-3439-4127
Norbert G. Gyenge https://orcid.org/0000-0003-0464-1537
Ye Jiang (姜也) https://orcid.org/0000-0002-6683-0205

References

Armand, T. P. T., Bhattacharjee, S., Choi, H.-K., & Kim, H.-C. 2024, in 2024 Int. Conf. on Artificial Intelligence in Information and Communication (ICAIIC) (IEEE), 238
Baranyi, T., Győri, L., & Ludmány, A. 2016, SoPh, 291, 3081
Bechet, S., Clette, F., & Verstringe, F. 1940, KSB-ROB, VO Resource Provided by the KSB-ROB, https://vo-tap.oma.be/__system__/services/root/info
Bobra, M. G., Sun, X., Hoeksema, J. T., et al. 2014, SoPh, 289, 3549
Bourgeois, S., Barata, T., Erdélyi, R., Gafeira, R., & Oliveira, O. 2024, SoPh, 299, 10
Buzulukova, N., & Tsurutani, B. 2022, FrASS, 9, 429
Cao, K., Xia, Y., Yao, J., et al. 2023, NatMe, 29, 3033
Carvalho, S., Gomes, S., Barata, T., Lourenço, A., & Peixinho, N. 2020, A&C, 32, 100385
Chatzistergos, T., Ermolli, I., Giorgi, F., Krivova, N. A., & Puiu, C. C. 2020, JSWSC, 10, 45
Chen, J., Gyenge, N. G., Jiang, Y., et al. 2025, ApJ, 980, 261
Chi, Y., Shen, C., Wang, Y., et al. 2016, SoPh, 291, 2419
Chola, C., & Benifa, J. V. B. 2022, GloTP, 3, 177
Clette, F., Berghmans, D., Vanlommel, P., et al. 2007, AdSpR, 40, 919
Clette, F., Lefèvre, L., Cagnotti, M., Cortesi, S., & Bulling, A. 2016, SoPh, 291, 2733
Clette, F., Svalgaard, L., Vaquero, J. M., & Cliver, E. W. 2014, SSRv, 186, 35
Colak, T., & Qahwaji, R. 2008, SoPh, 248, 277
Curto, J., Blanca, M., & Martínez, E. 2008, SoPh, 250, 411
Dasgupta, U., Singh, S., & Jewalikar, V. 2011, in 2011 Third National Conf. on Computer Vision, Pattern Recognition, Image Processing and Graphics (IEEE), 171
Djafer, D., Irbah, A., & Meftah, M. 2012, SoPh, 281, 863

Georgoulis, M. K., Yardley, S. L., Guerra, J. A., et al. 2024, COSPAR Scientific Assembly, 45, PSW.5–0004-24
Gizon, L. 2004, SoPh, 224, 217
Goel, S., & Mathew, S. K. 2014, SoPh, 289, 1413
Gong, X., Zhong, L., Rao, C., et al. 2023, A&A, 670, A132
Györi, L., Ludmány, A., & Baranyi, T. 2016, MNRAS, 465, 1259
Howe, R. 2008, AdSpR, 41, 846
Huang, X., Wang, H., Xu, L., et al. 2018, ApJ, 856, 7
Huang, X., Zhao, Z., Zhong, Y., et al. 2024, ScChD, 67, 3727
Korsós, M. B., Georgoulis, M. K., Gyenge, N., et al. 2020, ApJ, 896, 119
Lefèvre, L., & Clette, F. 2011, A&A, 536, L11
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
Márquez-Neila, P., Baumela, L., & Alvarez, L. 2014, ITPAM, 36, 2
Mathieu, S., Lefèvre, L., von Sachs, R., et al. 2023, JQT, 55, 104
Mathieu, S., von Sachs, R., Ritter, C., Delouille, V., & Lefèvre, L. 2019, ApJ, 886, 7
Mourato, A., Faria, J., & Ventura, R. 2024, Eng. Appl. Artif. Intell., 129, 107636
Okamoto, T. J., & Sakurai, T. 2018, ApJL, 852, L16
Owens, B. 2013, Natur, 495, 300
Owens, M. J., Lockwood, M., Barnard, L. A., et al. 2021, SoPh, 296, 82
Ronneberger, O., Fischer, P., & Brox, T. 2015, Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015, 9351, 234
Santos, J., Peixinho, N., Barata, T., et al. 2023, ApSci, 13, 5833
Sayez, N., De Vleeschouwer, C., Delouille, V., Bechet, S., & Lefèvre, L. 2023, JGRA, 128, e2023J
Schou, J., Scherrer, P. H., Bush, R. I., et al. 2012, SoPh, 275, 229
Siu-Tapia, A., Lagg, A., van Noort, M., Rempel, M., & Solanki, S. K. 2019, A&A, 631, A99
Thomas, J. H., & Weiss, N. O. 2008, Sunspots and Starspots, Cambridge Astrophysics (Cambridge Univ. Press)
Wang, C.-Y., Yeh, I.-H., & Mark Liao, H.-Y. 2025, in Computer Vision—ECCV 2024, ed. A. Leonardis et al. (Springer), 1
Wang, Y., Liu, J., Jiang, Y., & Erdélyi, R. 2019, ApJ, 881, 15
Woods, T. N., Harder, J. W., Kopp, G., & Snow, M. 2022, SoPh, 297, 43
Yang, Y., Yang, H., Bai, X., et al. 2018, PASP, 130, 104503
Zhang, J., Richardson, I. G., Webb, D. F., et al. 2007, JGRA, 112, A10102
Zhang, Y., Shen, Z., & Jiao, R. 2024, Comput. Biol. Med., 171, 108238
Zhao, C., Lin, G., Deng, Y., & Yang, X. 2016, PASA, 33, e018
Zhao, C., Yang, S., Wang, T., et al. 2024, AJ, 167, 52
Zharkov, S., Zharkova, V., Ipson, S., & Benkhalil, A. 2005b, EURASIP Journal on Advances in Signal Processing, 2005, 318462
Zharkov, S., Zharkova, V. V., & Ipson, S. S. 2005a, SoPh, 228, 377