Article:

Chen, J., Gyenge, N.G., Jiang, Y. et al. (2025) A bias-free deep learning approach for automated sunspot segmentation. The Astrophysical Journal, 980 (2). 261. ISSN: 0004-637X

https://doi.org/10.3847/1538-4357/adac5e

The Astrophysical Journal, 980:261 (11pp), 2025 February 20 https://doi.org/10.3847/1538-4357/adac5e

A Bias-free Deep Learning Approach for Automated Sunspot Segmentation

Jing Chen[1], Norbert G. Gyenge[2,4], Ye Jiang[3], Robertus Erdélyi[4,5,2], Jiajia Liu[6,7], 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] Gyula Bay Zoltán Solar Observatory (GSO), Hungarian Solar Physics Foundation (HSPF), Petőfi tér 3., Gyula, H-5700, Hungary
[3] School of Information Science and Technology, Qingdao University of Science and Technology, 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, Hungary
[6] National Key Laboratory of Deep Space Exploration, Deep Space Exploration Laboratory/School of Earth and Space Sciences, University of Science and Technology of China, Hefei 230026, People's Republic of China
[7] 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

Received 2024 October 28; revised 2025 January 2; accepted 2025 January 19; published 2025 February 20

Abstract

Solar activities significantly influence space weather and the Earth's environment, necessitating accurate and efficient sunspot detection. This study explores deep learning methods to automate sunspot identification in solar satellite images, keeping personal bias to a minimum. Utilizing observations of the Solar Dynamics Observatory, we leverage active-region data from the Helioseismic Magnetic Imager active-region patches to locate sunspot groups detected between 2011 and 2023. The Morphological Active Contour Without Edges technique is applied to produce pseudo-labels, which are utilized to train the U-Net deep learning architecture, combining their strengths for robust segmentation. Evaluation metrics—including precision, recall, $F1$-score, intersection over union, and Dice coefficient—demonstrate the superior performance of U-Net. Our approach achieves a high Pearson correlation coefficient of 0.97 when compared with the sunspot area estimation of the Space Weather Prediction Center and 0.96 in comparison with the Debrecen Photoheliographic Data. This hybrid methodology provides a powerful tool for sunspot identification, offering the improved accuracy and efficiency crucial for space-weather prediction.

Unified Astronomy Thesaurus concepts: sunspot number (1652)

1. Introduction

Solar activity encompasses various phenomena on the Sun's surface, including sunspots, solar flares, and coronal mass ejections (CMEs; D. H. Hathaway 2010; P. Papathanasopoulos et al. 2016; J. Zhang et al. 2021; J. Liu et al. 2023; L. Miroshnichenko 2023). Sunspots are dark, cooler areas on the Sun's surface associated with strong magnetic activity. They exhibit an approximately 11 yr solar cycle, with increased sunspot numbers during the solar maximum, which correlates with elevated solar activities, including the occurrence of solar flares and CMEs. These highly explosive phenomena release large amounts of energy and bursts of charged particles, influencing space weather and potentially disrupting our high-tech facilities and technology installations, such as, e.g., satellites, telecommunications systems, and power grids. Sunspot monitoring is therefore crucial for predicting these eruptive solar events and understanding the Sun's magnetic dynamics, not to mention the preventive protection of our technosphere (M. K. Georgoulis et al. 2024).

Traditionally, sunspots have been observed and identified using white-light images acquired from ground-based observatories. However, with advancements in observational equipment, the amount of data generated from sunspot observations has increased. The Solar Dynamics Observatory (SDO; W. D. Pesnell et al. 2012), launched by NASA in 2010, continuously monitors the Sun, capturing high-resolution images of and data relating to solar activity, magnetic fields, and the Sun's atmosphere, to enhance our understanding of solar phenomena and their impacts on Earth. Early manual methods for sunspot identification are now inadequate for the growing data volume, leading to the development of semi-automated techniques like the one yielding the Debrecen Photoheliographic Data (DPD; T. Baranyi et al. 2016). Nevertheless, these methods are hindered by factors such as penumbra diffusion, noise, limb darkening, and pixel variations. Moreover, human bias in identification is unavoidable, highlighting the urgent need for a robust and efficient automatic method in solar physics—see, e.g., M. B. Korsós et al. (2021) and H. Morgan & M. B. Korsós (2022).

Many, if not most, sunspot detection methodologies have primarily employed manual or semi-automated threshold techniques, applied to white-light image intensity data. S. Zharkov et al. (2005), J. Curto et al. (2008), and F. Watson et al. (2009) have proposed the use of mathematical morphology prior to applying a threshold for sunspot segmentation. S. Bourgeois et al. (2024) adopted mathematical morphology for sunspot segmentation together with manual feature selection. T. Colak & R. Qahwaji (2008) introduced a method that integrates thresholds with a fully connected neural network to detect and classify sunspot groups. D. Djafer et al. (2012) developed a sunspot recognition technique based on wavelet transformations. S. Goel & S. Mathew (2014) utilized the Selective Binary and Gaussian Filtering Regularized Level Set method for sunspot identification. C. Zhao et al. (2016) proposed a method that begins with Gaussian preprocessing, followed by sunspot identification using a morphological bot-hat operation and Otsu thresholding.

These techniques have enabled sunspot identification and feature extraction but often require threshold selection and struggle with complex images, due to limited feature-learning capabilities. In contrast, convolutional neural network (CNN) models offer a robust solution to this challenge of feature learning (A. Krizhevsky et al. 2017; K. Cao et al. 2023; T. Talaei Khoei et al. 2023). These features, which range from simple edges to complex textures and high-level concepts, are acquired during the training process. Let us briefly recall some of the key steps relevant in the present context.

C. Chola & J. V. B. Benifa (2022) integrated AlexNet with a support vector machine to classify satellite continuum images as either containing sunspots or depicting a quiet Sun. Similarly, J. Santos et al. (2023) applied YOLOv5 for sunspot detection using data from the Geophysical and Astronomical Observatory of the University of Coimbra data set. While these approaches detect sunspots using bounding boxes, they do not perform segmentation.

Segmentation is essential for estimating sunspot properties such as area and count. Traditional intensity-threshold-based techniques can be employed to estimate initial sunspot boundaries, with deep learning models subsequently refining the delineations. For instance, N. Sayez et al. (2023) employed the black-top-hat transform, a mathematical morphology operation, to generate pseudo-labels for training a U-Net model. However, when applied to full-disk continuum images, this transform erroneously classifies black dots distant from active regions as sunspots, resulting in numerous false positives. As an alternative, A. Mourato et al. (2024) utilized a flood fill algorithm to generate pseudo-labels, requiring sunspot coordinates and area data, sourced from the DPD (T. Baranyi et al. 2016). The DPD data set, finalized in 2015 and derived from ground-based observations, may exhibit discrepancies when compared to future data sets collected by advanced instruments or space-borne telescopes with higher resolutions. Furthermore, given the continuous evolution of solar activity, sunspot characteristics are subject to temporal variation. Consequently, using the DPD data set for pseudo-labeling may introduce covariate shift—a phenomenon in machine learning where the distribution of input features diverges between the training and testing data, while the conditional distribution of the target variable, given the input, remains constant.

To address the aforementioned limitations, this work introduces an automated sunspot detection method that integrates physics-based knowledge. By employing deep learning techniques, the method confines the search for sunspots exclusively to active regions. Initially, active-region patches are extracted from full-disk continuum observations based on the Helioseismic Magnetic Imager (HMI) Active Area Patch (HARP) data set. The Morphological Active Contour Without Edges (MorphACWE) algorithm is then applied to these active regions to delineate the initial boundaries of sunspots. This approach ensures that noise far from active regions is not mistakenly identified as sunspots. The resulting sunspot masks are subsequently utilized as pseudo-labels for training a U-Net model, refining the final sunspot outlines. By doing so, we avoid introducing subjective sunspot segmentation standards, thereby achieving more consistent results.

Additionally, our method is also adaptable to observations from other instruments, including ground-based observations, where fewer sunspots may be detected due to atmospheric and weather conditions. It is also well suited for data from new space-borne telescopes with higher resolutions, as the method is robust to variations in resolution. The accurate and consistent sunspot segmentation achieved through this approach will significantly benefit subsequent research, such as automated sunspot number estimation and sunspot group classification.

2. Data

The data employed in this study comprise observations from SDO and HARP (M. Bobra et al. 2014), spanning from 2011 January 1 to 2023 July 1. A subset of the data was deliberately excluded, to prevent data leakage. Specifically, data from January 1 to September 15 of each year were applied as the training set, while the test set comprised data from October 1 to December 15. Given that the Sun has a rotation of approximately 27 days, this approach ensures that the sunspots in the training and test sets do not resemble each other, thereby reducing the risk of overfitting and improving the generalizability. The final training set contains 2464 white-light HMI observations, and the test set has 775 observations. The image dimensions are 4096 by 4096 pixels.

The full-disk continuum observations are downloaded from the JSOC server.[8] Since sunspots usually appear within active regions, additional information is required for identifying the active regions in the raw full-disk images and we utilize the HARP service[9] (J. Hoeksema et al. 2014). The obtained data contain partial-disk, automatically identified active-region patches. However, the HARP data do not provide detailed contours for individual sunspots within the active regions. Instead, they offer an approximate boundary for the whole sunspot group, which is an appropriate tool for choosing the regions of interest in the full-disk images. The full-disk continuum observations and the HARP active-region data are downloaded once a day at 12 PM International Atomic Time.

Before conducting the analysis, the full-disk raw observations must be preprocessed. To address potential misalignment of the Sun's north–south axis in the observational data, a rotation function is employed to correct the image orientation. Limb darkening is corrected using a function that applies a fifth-order polynomial fit to the limb-darkening constants u and v provided by A. Cox (2000). The corrected photon counts p_c in an arbitrary position r of the solar disk can be estimated based on the raw photon count p_r by employing:

$$p_c = \frac{p_r}{1 - u - v + u \cdot \cos(\arcsin(r)) + v \cdot \cos(\arcsin(r))^2}. \quad (1)$$

After applying the limb-darkening correction, 1% of the solar disk is trimmed to maintain a homogeneous background near the solar limb.

3. Methodology

3.1. Pseudo-labeling

Manual annotation of solar sunspots is a labor-intensive process susceptible to human error. To address these limitations and enhance objectivity, we implemented MorphACWE (P. Márquez-Neila et al. 2014), which works well without well-defined boundaries. However, it requires that the inner region of the feature is different from the outer regions on average (T. Chan & L. Vese 1999). In case of a larger sunspot group, the sunspots are usually separated by well-defined

---

[8] http://jsoc.stanford.edu/

[9] http://jsoc.stanford.edu/doc/data/hmi/sharp/sharp.htm

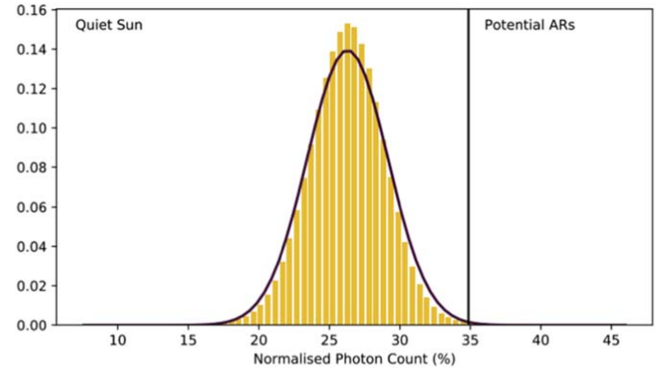


Figure 1. Sample photon-intensity distribution based on an observation without active regions. The proposed threshold separates the background from the sunspots.

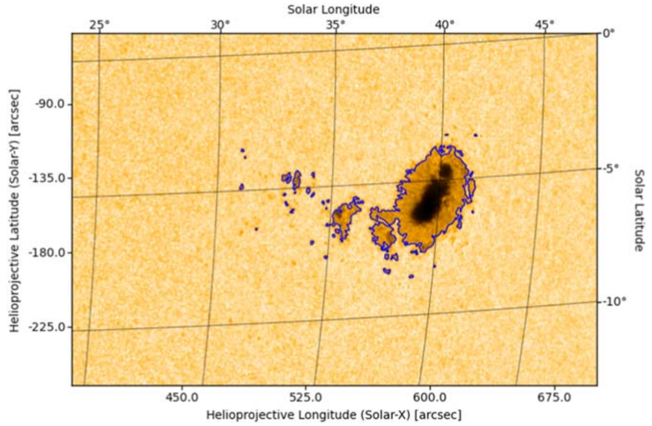


Figure 2. Segmentation results of MorphACWE in Active Region No. 12665 on 2017 July 14.

boundaries, but this may not be true if the investigated region contains smaller sunspots, especially solar pores. It can always be safely assumed that a sunspot is darker than the quiet Sun, hence the MorphACWE technique is perfectly suitable for this application. MorphACWE operates as follows:

Assume C is the evolving boundary of a feature and c_1 and c_2 represent the pixel intensity averages inside and outside the same object. Furthermore, u_0 is the intensity array of the whole region of interest—for instance, a circle-shaped sunspot in the center of the image. Let us introduce the "fitting energy," composed by

$$F_1(C) + F_2(C) = \int_{\mathrm{inside}(C)} |u_0(x, y) - c_1|^2\, dx\, dy + \int_{\mathrm{outside}(C)} |u_0(x, y) - c_2|^2\, dx\, dy. \quad (2)$$

The terms $F_1(C)$ and $F_2(C)$ can be interpreted as two separate actions, i.e., shrinking and expanding. If the first term is larger (smaller), the contour shrinks (expands) in the next iteration. The evolution of the contour stops if

$$\inf\{F_1(C) + F_2(C)\} \approx 0. \quad (3)$$

The algorithm can be significantly speeded up, when the initial contour is relatively close to the final contour. Therefore, we define respective threshold intensities for every active region. Let us consider the quiet-Sun regions as random noise, which can be fitted by a Gaussian distribution X, $X \sim \mathcal{N}(\mu, \sigma^2)$, where μ and σ are the average and the standard deviation of the sample. The initial threshold τ is now described by

$$\tau = \mu + (5\sigma). \quad (4)$$

Every pixel exceeding the threshold τ is omitted from the statistics. The remaining pixels compose the region of interest without noise, which now can be used as an initial condition for the MorphACWE algorithm. Figure 1 shows the distribution of the photon-intensity count with the proposed threshold.

HARP metadata provide the coordinates of the patch boundaries for the active regions derived from SDO/HMI continuum and magnetogram observations, which are used to crop these active-region patches. An initial threshold intensity τ is calculated based on both the continuum and magnetogram of an active-region patch. Using this threshold, MorphACWE is performed on the active-region patch to iteratively refine the boundaries of the penumbra region and produces a binary sunspot segmentation mask. This algorithm identifies the penumbra by detecting contiguous regions of pixels that meet the intensity criteria, effectively distinguishing the penumbra from the surrounding quiet Sun. For visualization purposes, the resulting binary mask is then processed to extract the contours of the penumbra, as demonstrated in Figure 2. This approach avoids hard-coded thresholds, making the method flexible across different active regions and solar activity levels. Each active-region patch with the form of a binary mask is mapped back to its original position on the solar disk. The values outside active regions are set to be zero. In the MorphACWE method, the intensity threshold for detecting sunspots is 5σ, the number of iterations for morphological snakes is 100, and the minimum number of the pixel threshold for filtering out fake sunspots is set to 1.

Solar full-disk continuum satellite images typically exhibit high-contrast features, but edges may be blurred or discontinuous, partly due to solar atmospheric disturbances. MorphACWE is well suited to segmenting sunspots, whose edges are unclear, while the inner and outer aspects of the objects of interest have different averages. The pseudo-label generation is shown in Figure 3.

3.2. Sunspot Segmentation with U-Net

The traditional image processing techniques, such as MorphACWE, are sensitive to parameter selection, requiring tuning to achieve optimal performance. In contrast, U-Net, a deep learning image-segmentation architecture, can automatically learn complex features and patterns from a large amount of labeled training data (O. Ronneberger et al. 2015). In this work, the output of MorphACWE is used as the pseudo-labels to train U-Net.

The U-Net architecture consists of an encoder and a decoder, forming a U shape, as shown in Figure 4. During the encoding phase, U-Net gradually extracts multilevel features, from low to high, of the image through a series of convolutional blocks and pooling layers. The sizes of the feature maps decrease, while the number of feature channels increases. During the decoding phase, deconvolutional layers and upsampling operations are used to restore the image resolution. Skip connections between the encoder and decoder layers ensure that high-resolution features of the encoding phase are preserved, which can prevent small sunspots from being ignored during downsampling of the convolutional operations. Finally, the feature

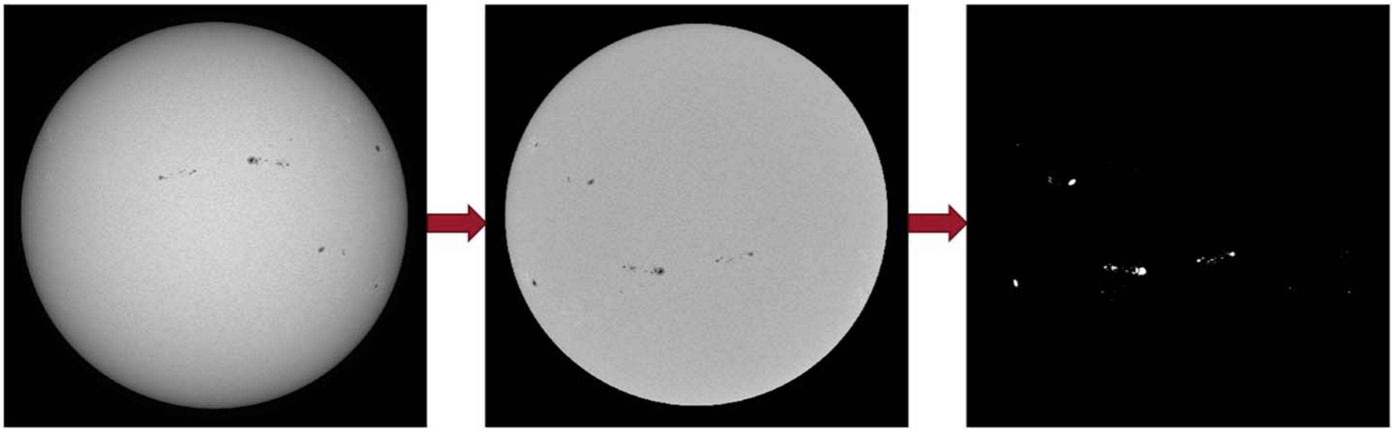


Figure 3. Pseudo-label generation with MorphACWE. The original continuum image (left panel), the preprocessed image (middle panel), and the binary MorphACWE segmentation mask (right panel) are shown.

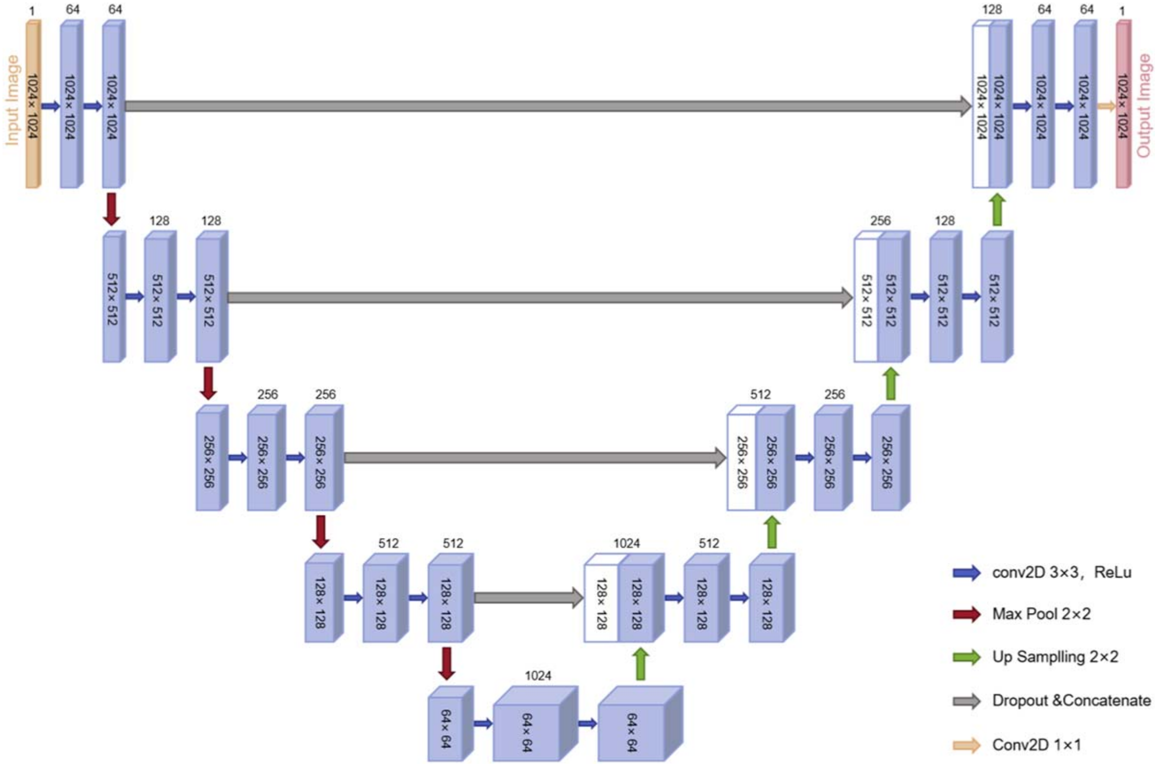
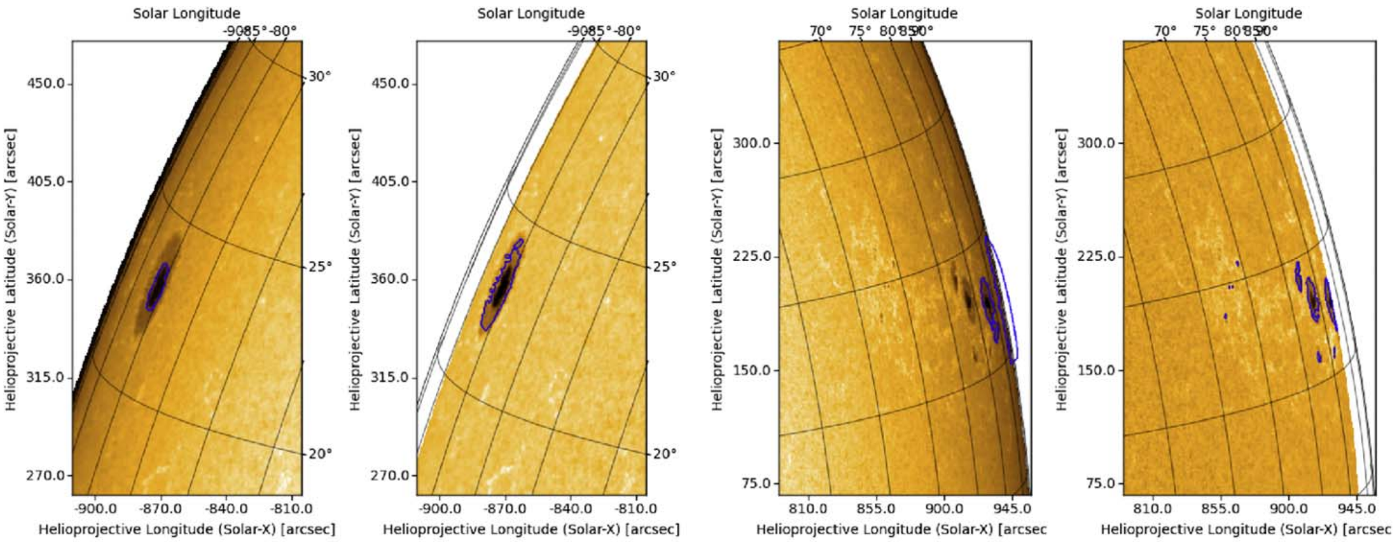


Figure 4. Architecture of U-Net. Adopted from O. Ronneberger et al. (2015).

map is converted to a single-channel image by a layer of convolution. The pixels of the output probability maps are normalized to be between zero and one, using the sigmoid activation function. The value of the output represents the probability that each pixel belongs to a sunspot.

This approach leverages the strengths of both techniques, utilizing MorphACWE to handle images with indistinct edges and noise, while exploiting U-Net for the much finer segmentation. Our experiments show that U-Net struggles with small objects, because it combines features from different levels of the network. If small objects are not well represented in the higher resolution of the earlier layers, they might be overlooked when features are combined. Furthermore, significant computational challenges are raised due to the use of the large images. Therefore, the original images of dimension 4096 by 4096 pixels need to be tiled into 16 patches, with each patch of dimension 1024 by 1024 pixels. These small patches reduce the amount of background information, allowing the model to focus more on the objects within each patch. An evaluation of the overlapping patch technique when slicing the images revealed negligible differences compared to nonoverlapping slicing. This may be because U-Net learns the boundaries of the split sunspots as part of the sunspot features. Consequently, nonoverlapping patch slicing was chosen to optimize computational efficiency.

A binary cross-entropy loss function (BCELoss) and the Adam optimizer are utilized. During training, the learning rate is set to 0.001. Due to computational limitations, the batch size



(a) Active Region No. 11312 on October 4, 2011 **(b) Active Region No. 11302 on October 4, 2011**

Figure 5. Comparison of segmentation results using MorphACWE. The left (right) side of each subplot shows the results without (with) limb-darkening correction. (a) Active Region No. 11312 on 2011 October 4; and (b) Active Region No. 11302 on 2011 October 4.

is set to one. The model is trained for nine epochs, at which point it converges. Notably, since each batch contains only one patch of an image in this application, batch normalization (BatchNorm) is replaced with group normalization (GroupNorm). The normalization method GroupNorm, introduced by Y. Wu & K. He (2018), overcomes the shortcomings of BatchNorm, particularly in situations where the batch size is small or highly variable. GroupNorm normalizes the activation of a neural network across groups of channels, rather than across the entire batch or individual layers, providing consistent performance regardless of batch size.

4. Results and Discussions

4.1. Limb-darkening Correction

The sunspot segmentation results produced by MorphACWE, with and without limb-darkening correction, have been compared. As illustrated on the left sides of Figures 5(a) and (b), the absence of limb-darkening correction prevents MorphACWE from recognizing the penumbra, due to the inhomogeneous intensity in the solar disk's background. Additionally, our experiments indicate that even with limb-darkening correction, it is crucial to remove the dark edge (1%) of the solar disk. Neglecting this step produces results with two types of segmentation errors: first, the outcomes resemble those shown on the left side of Figure 5(a); and second, as depicted on the left side of Figure 5(b), MorphACWE identifies the edge of the solar disk as a penumbra by mistake. These observations reveal that applying limb-darkening correction and trimming the limb by 1% could significantly enhance the performance of MorphACWE.

The solar disk gradually darkens from the center to the edge, with the gradient becoming more pronounced closer to the edge. The U-Net model learns this position-dependent gradient as a feature across all sunspots, which is not essential and may reduce its effectiveness.

4.2. Sunspot Segmentation

In this work, the segmentation results of MorphACWE are utilized as pseudo-labels to train U-Net. The U-Net model learns the mapping between the continuum observations and the binary segmented output results, generating a mask where each pixel value indicates whether it is part of the sunspots. For building the ground-truth data set, the sunspot drawings provided by the Uccle Solar Equatorial Table station of the Royal Observatory of Belgium are referenced, to correct U-Net outcomes manually.[10]

In this section, we present the evaluation results of our sunspot segmentation methods: MorphACWE and U-Net. To assess the performance of our semantic segmentation methods, various standard image-segmentation and image-classification metrics are utilized, including precision, recall, $F1$-score, intersection over union (IoU), and Dice coefficient.

Precision represents the proportion of correctly predicted positive samples among all predicted positives. Recall measures the proportion of actual positive samples that are correctly identified by the model. The $F1$-score is the harmonic mean of precision and recall, providing a balance between precision and recall. IoU quantifies the overlap between the predicted and actual regions relative to their combined area. The Dice coefficient, similar to IoU, emphasizes the agreement between the predicted and ground-truth segments. That is, IoU tends to penalize single instances of bad classification more than the Dice coefficient. The formulas for these evaluation metrics are shown below:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{5}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{6}$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2\text{TP}}{2\text{TP} + \text{FN} + \text{FP}}, \tag{7}$$

---

[10] https://www.sidc.be/uset/

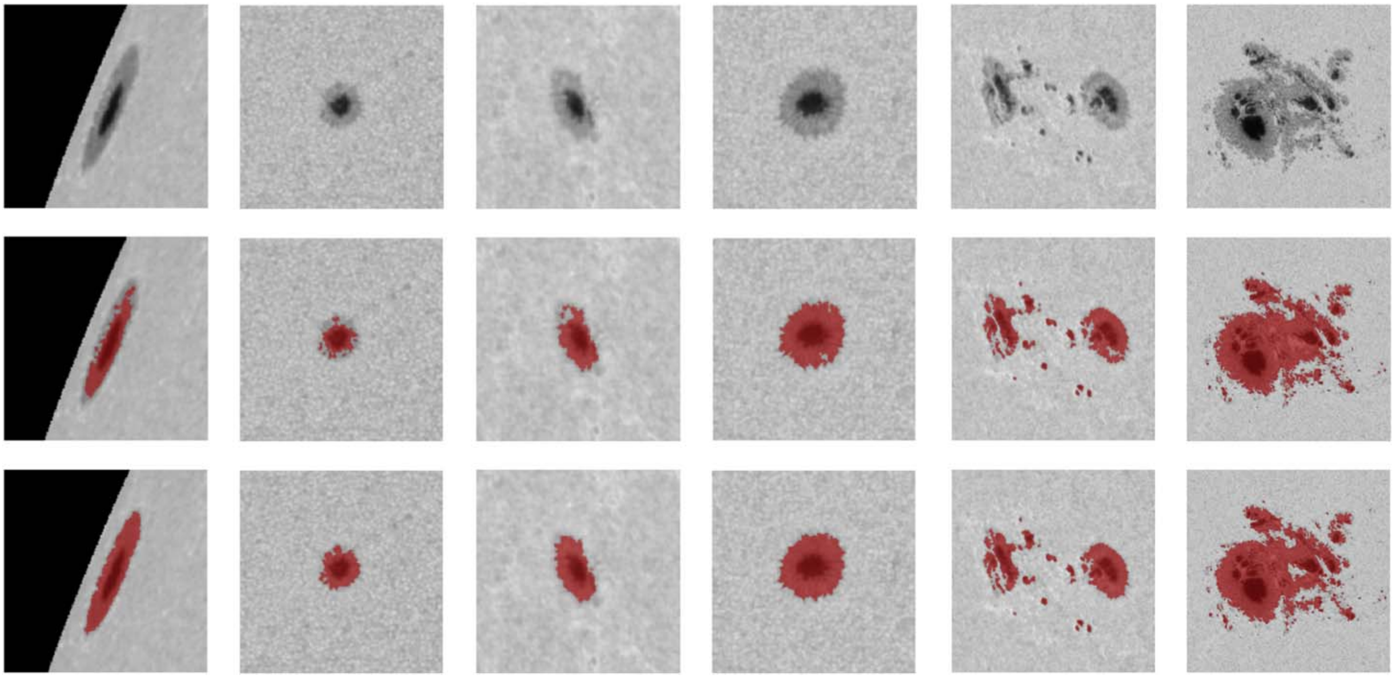


Figure 6. Comparison of the segmentation results: continuum images after preprocessing (top), MorphACWE segmentation results (middle), and U-Net segmentation results (bottom).

Table 1
Evaluation Results of Our Sunspot Segmentation Methods: MorphACWE and U-Net

Method	Precision	Recall	$F1$	IoU	Dice
Morph	0.9212	0.9250	0.9167	0.8608	0.9167
U-Net	0.9705	1.0000	0.9839	0.9705	0.9839

$$\text{IoU} = \frac{\text{Overlap Area}}{\text{Union Area}} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}}, \quad (8)$$

$$\begin{aligned}\text{Dice} &= \frac{2\times \text{Overlap Area}}{\text{Prediction Area} + \text{Ground Truth Area}}\\ &= \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}, \quad (9)\end{aligned}$$

where TP (True Positive) denotes the instances where both the predictions and the actual labels are positive. TN (True Negative) indicates the instances where both the predictions and the actual labels are negative. FP (False Positive) occurs when the actual labels are negative but the predictions are positive. FN (False Negative) arises when the actual labels are positive but the predictions are negative.

A comprehensive comparison between MorphACWE and U-Net across five evaluation metrics is presented in Table 1. The segmentation results of both methods are visualized in Figure 6. U-Net demonstrates superior performance in both quantitative metrics and qualitative visual results, as explained below.

First, MorphACWE assumes distinct average pixel intensities within and outside the target object. Consequently, in the presence of noise or substantial intensity inhomogeneities within sunspot regions, MorphACWE may struggle to delineate boundaries accurately. For instance, as illustrated in the first three columns of Figure 6, the method's failure can be attributed to noise within the penumbra, where the pixel intensities closely resemble those of the background.

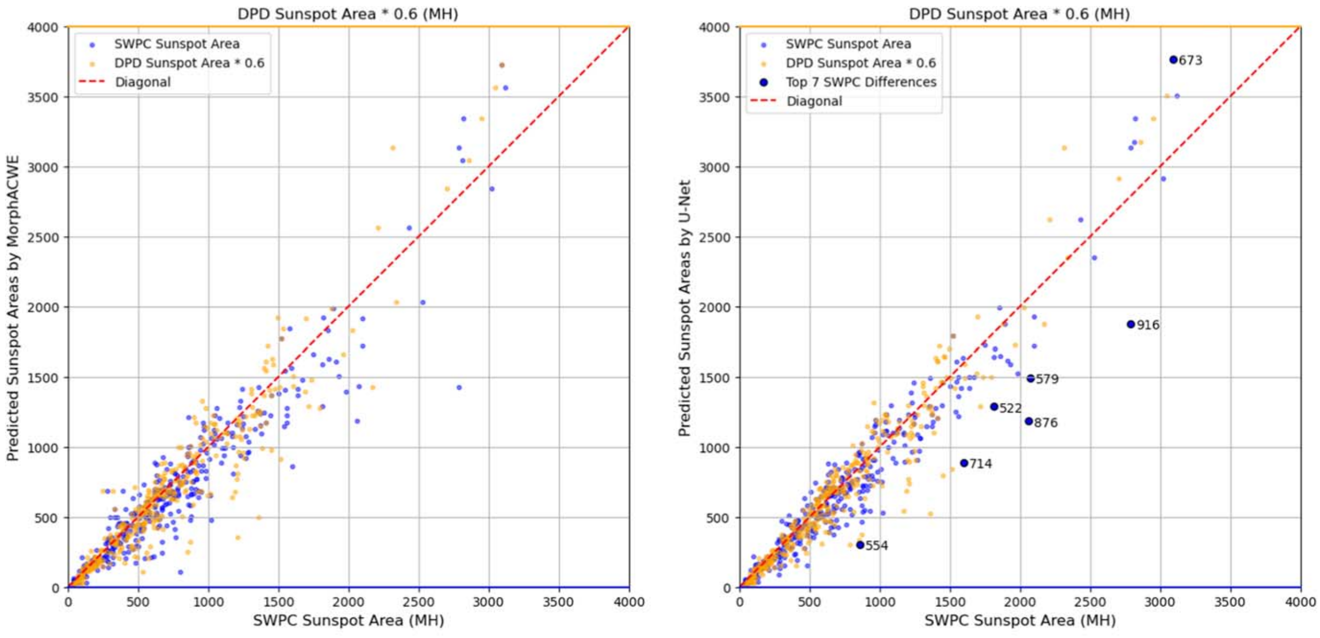
Second, MorphACWE's reliance on local intensity information and its inability to integrate broader contextual information about sunspots and their surroundings further limits its effectiveness. As demonstrated in the fourth and fifth columns of Figure 6, this limitation often leads to boundary breaches of the penumbra.

Third, MorphACWE's predefined energy-minimization criteria are tailored for smooth and homogeneous regions, limiting its accuracy in capturing boundaries with significant boundary variations or spikes. While this simplification facilitates the energy-minimization process, it restricts the method's adaptability to highly textured or heterogeneous regions.

Fourth, MorphACWE is constrained by its initialization and energy-minimization framework, which limits its ability to detect internal holes or discontinuities within sunspot regions. As shown in the last column of Figure 6, MorphACWE fails to segment the complex internal structure.

In contrast, U-Net leverages a deep CNN to learn hierarchical and complex features from the entire sunspot training set. This enables U-Net to capture intricate patterns, including edges, textures, and global contextual relationships. The incorporation of skip connections allows U-Net to integrate fine-grained local details with global contextual information. This architecture helps U-Net overcome the limitations inherent in MorphACWE. U-Net's robust feature extraction capabilities enable it to effectively handle regions with substantial texture or intensity variation and identify internal structures within sunspots. Moreover, this architecture enhances its robustness to noise and artifacts commonly present in solar imagery, ensuring more reliable segmentation performance.

The metrics presented in Table 1 are evaluated at the pixel level, assessing the accuracy of each pixel classification. The significantly higher recall achieved by U-Net is directly



(a) MorphACWE vs SWPC/DPD

(b) U-Net vs SWPC/DPD

Figure 7. Comparison of sunspot areas for each image in the test set between 2011 and 2015. The small blue dots show the sunspot areas predicted by the proposed method compared to SWPC, while the large blue dots highlight the days with the most significant discrepancies, with the corresponding difference values labeled. The small yellow dots represent the comparison between DPD sunspot areas (scaled by 0.6) and the proposed method. The red dashed line indicates the point where the predicted areas match the actual values.

Table 2
Comparison Between the Seven Largest Discrepancies of Sunspot Areas by SWPC and U-Net

	Date	SWPC	U-Net	Difference
1	2014.10.28	2790	1874	916
2	2014.11.25	2060	1184	876
3	2012.11.12	1600	886	714
4	2014.10.23	3090	3763	673
5	2014.11.24	2070	1491	579
6	2014.12.08	860	306	554
7	2013.11.12	1810	1288	522

attributed to its ability to address the limitations of MorphACWE.

It is worth noting that U-Net presents a high IoU of 97%, surpassing MorphACWE by 11%. However, the superiority of U-Net over MorphACWE for the Dice coefficient is only 6%. The results indicate that the vast majority of the inferences are moderately better with U-Net than with MorphACWE, but some of them are significantly worse using the latter. The evaluation results highlight the effectiveness of combining MorphACWE with U-Net for the sunspot segmentation, resulting in a robust and precise segmentation solution.

4.3. Sunspot Areas

To further validate the accuracy of sunspot recognition in this work, the sunspot areas are calculated for the test set. The results are then compared with those provided by the Space Weather Prediction Center (SWPC)[11] for the period 2011–2022 and the DPD[12] for the period 2011–2015, to conduct a comprehensive comparative analysis. SWPC uses a combination of ground-based and space-based observational data to detect and track sunspots, mainly focusing on space-weather monitoring. DPD primarily relies on full-disk white-light observations from the Debrecen and Gyula observatories, though gaps in their data are supplemented with solar images from other ground-based and space-based observatories. Sunspot detection in both SWPC and DPD is carried out using a combination of manual methods and thresholding-based automated techniques.

A sunspot area is the total area of the pixels inside the obtained contour lines. The real physical area can be easily determined, if the sunspot is located around the center of the observations. However, the area estimation may become a challenging task if the investigated sunspot is close to the solar limb. The area of the feature must be corrected for foreshortening. The method introduced by H. Çakmak (2014) for estimating the area is implemented. In the first stage, the proposed models predict sunspot masks, which are binary arrays indicating the positions of the sunspots. When a pixel is inside a sunspot, the pixel is assigned a Boolean variable TRUE, otherwise its value is FALSE. Next, the Carrington heliographic (HG) coordinates are estimated for each TRUE pixel. Two additional mask layers on top of the binary layer are implemented. The first layer is the latitude mask, where each pixel represents three properties—the x and y positions of a pixel and the latitudinal position of the same pixel. The top layer shows the same with HG longitude. Assuming the latitudinal and longitudinal widths for the pixel (x, y) are $\Delta B_p(x, y)$ and $\Delta L_p(x, y)$, respectively:

$$\Delta B_p(x, y) = |P_B(x, y + 1) - P_B(x, y)|, \tag{10}$$

$$\Delta L_p(x, y) = |P_L(x + 1, y) - P_L(x, y)|, \tag{11}$$

---

[11] https://www.swpc.noaa.gov/

[12] http://fenyi.solarobs.epss.hun-ren.hu/DPD/

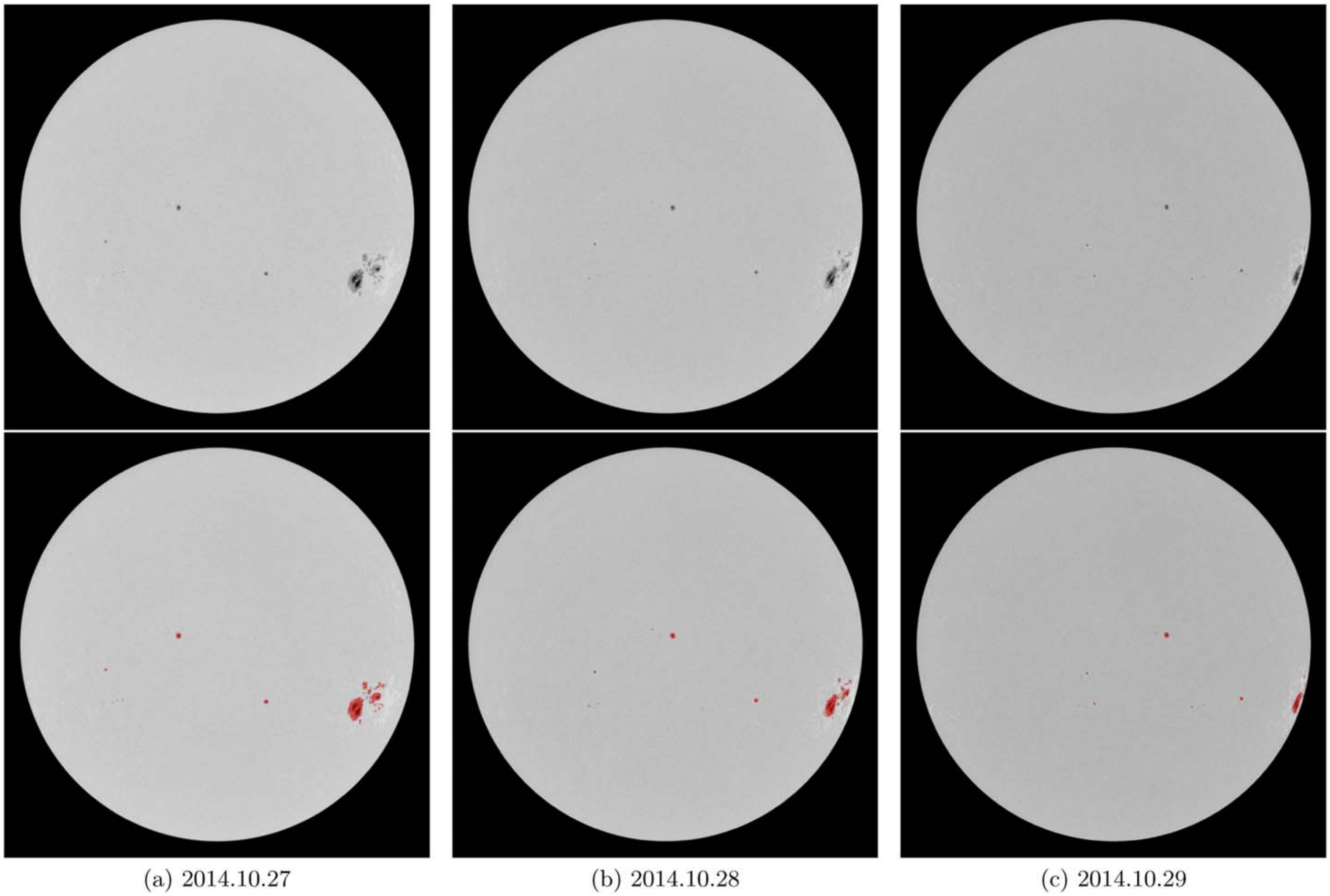


Figure 8. Evolution of the sunspots from 2014 October 27 to 29. The upper row presents the original continuum observations, while the lower row displays the corresponding sunspot masks predicted by U-Net.

Table 3
Comparison of Sunspot Areas Between SWPC and U-Net from 2014 October 27 to 29

Date	SWPC	U-Net
2014.10.27	2530	2347
2014.10.28	2790	1874
2014.10.29	1630	1018

where $P_B(x, y)$ and $P_L(x, y)$ are the HG latitude and longitude of the pixel (x, y), respectively. $P_L(x + 1, y)$ is the HG longitude of the pixel $(x + 1, y)$, and $P_B(x, y + 1)$ is the HG latitude of the pixel $(x, y + 1)$. Since the HG area must be positive, the absolute values of the widths are used.

When the parameters n and m are the total number of pixels of the active regions in the x- and y-dimensions, the total area A_{deg^2} of the sunspot can be estimated by

$$A_{\mathrm{deg}^2} = \sum_{x=1}^{n}\sum_{y=1}^{m}(\Delta B_P(x, y)\Delta L_P(x, y)). \quad (12)$$

The total sunspot area A_{deg^2} is computed within the framework of the Carrington coordinate system, which is defined on a spherical surface. This approach ensures that the sunspot area is measured in angular units (square degrees) rather than in image-based units (square pixels). Consequently, this method eliminates distortions, including those near the solar limb, which are typically introduced by projection effects. The total area can be converted to units of millionths of hemisphere (MH) by applying

$$A_{\mathrm{MH}} = \frac{A_{\mathrm{deg}^2}}{2\pi(180/\pi)^2/10^7}. \quad (13)$$

The transformed sunspot areas, as compared to those provided by the SWPC, are presented in Figure 7. The analysis reveals a strong alignment between the predictions of U-Net (MorphACWE) and the SWPC data, with Pearson correlation coefficients of 0.97 (0.96) over the 2011–2022 period. This high correlation highlights the reliability of both U-Net and MorphACWE in capturing sunspot areas consistent with SWPC measurements. The success of our approach can be attributed to the automated sunspot detection and segmentation process. Pseudo-labels are generated using the MorphACWE thresholding technique, which is applied to the active regions detected by HARP. MorphACWE performs effectively in the absence of well-defined boundaries. Restricting the analysis to active regions ensures the removal of false sunspots. This label generation process is not biased by subjective sunspot segmentation criteria. Therefore, our method is adaptable to data from various observational instruments, ranging from ground-based observatories to space-based telescopes, as the pseudo-label generation is independent of specific human annotations for sunspot segmentation.

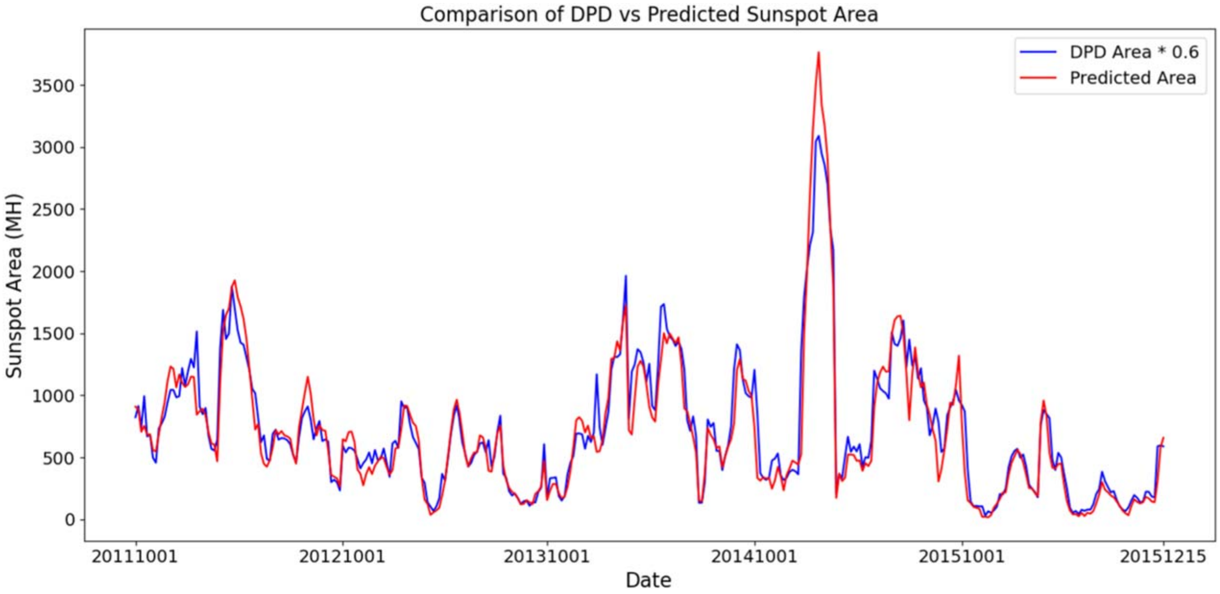


Figure 9. Daily comparison between DPD and the predicted sunspot areas for the test set. The blue line represents DPD sunspot areas scaled by a factor of 0.6, while the red line represents predicted sunspot areas.

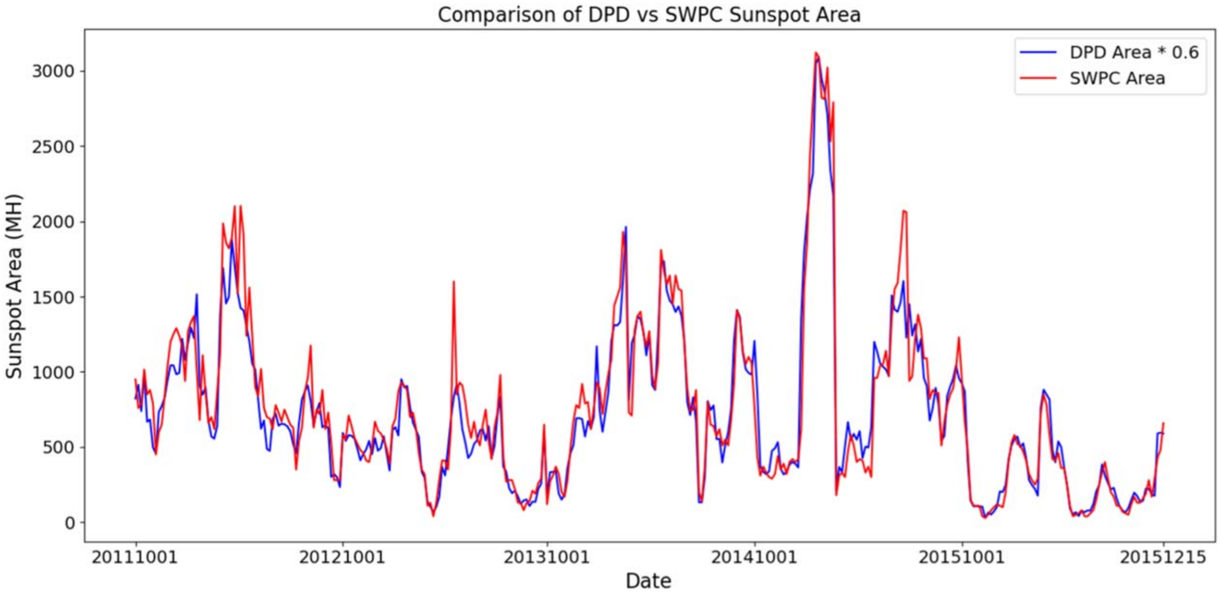


Figure 10. Daily comparison between SWPC and DPD sunspot areas for the test set. The blue line represents the DPD sunspot areas scaled by a factor of 0.6, while the red line represents the SWPC sunspot areas.

In order to understand the reasons for some significant discrepancies between SWPC and the predicted results, we analyzed the top seven cases where the discrepancies are the largest, as shown in Table 2. The first entry in Table 2 corresponds to 2014 October 28, where the discrepancy in the sunspot areas between the SWPC and U-Net models reached 916, marking the largest difference observed in the test set. A reduction in the sunspot areas is expected between 2014 October 27 and 29, as illustrated in Figure 8. The predictions generated by U-Net, as presented in Table 3, align with this expected trend. In contrast, the SWPC data indicate an increase in the sunspot areas from 2014 October 27 to 28, which appears to be inconsistent with the anticipated behavior. The second, third, sixth, and seventh entries all demonstrate similar results, where the SWPC results are inconsistent with the observed evolution of the continuum images, but the U-Net results are consistent. The fourth and fifth entries present cases where it is challenging to determine which method is more accurate. In 71.42% of the seven cases, our method is supported by the data, and in the remaining cases, it is indeterminate which approach is superior. These findings suggest that our proposed method may offer a more reliable sunspot area estimation than the approach employed by SWPC. We analyze the top seven cases with the most significant discrepancies and infer that similar factors likely contribute to the remaining differences.

When comparing the U-Net-predicted sunspot areas to those reported in the DPD data for the period 2011–2015, a Pearson correlation coefficient of 0.96 is achieved, whereas MorphACWE yields a slightly lower correlation coefficient of 0.95, as illustrated in Figure 7. Notably, when a reduction factor of 0.6 is applied to the DPD sunspot areas, they align closely with the predicted areas, as shown in Figure 9. To determine the

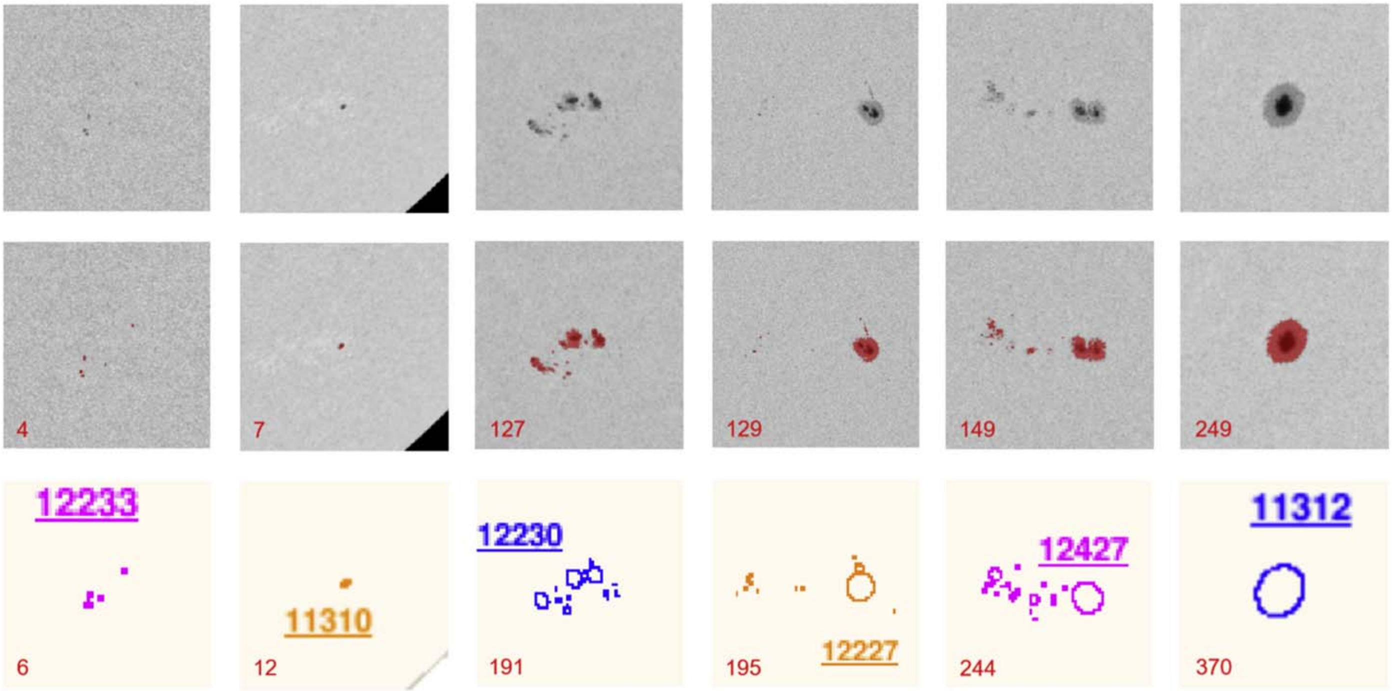


Figure 11. Examples of sunspot identification results. The first row shows the original continuum observations. The second row presents the predicted sunspot masks generated by U-Net. The third row displays the sunspot identification from DPD. The red numbers in the lower left corners of the images in the second and third rows indicate the sunspot areas.

optimal reduction factor, values ranging from 0.5 to 0.7 with a step size of 0.01 were systematically tested. The mean absolute errors (MAEs) between the predicted areas (U-Net or MorphACWE) and the adjusted DPD areas were calculated for each reduction factor. A reduction factor of 0.6 yielded the smallest MAE, confirming its suitability for aligning the predicted sunspot areas with the DPD measurements. Compared to the sunspot areas of SWPC or DPD, U-Net achieves higher Pearson correlation coefficients than MorphACWE. Figure 7 further supports this, showing that the predicted sunspot areas from MorphACWE exhibit more scatter than those from U-Net. Similarly, the Pearson correlation coefficient between the SWPC and DPD sunspot areas reaches 0.96, and applying a reduction factor of 0.6 to the DPD areas aligns them with the SWPC values, as shown in Figure 10. To explore the reasons behind the consistently larger DPD sunspot areas compared to those obtained from the proposed methods, various scenarios involving individual sunspots or sunspot groups are visualized in Figure 11. In each case, the sunspot areas reported by DPD exceed those identified by the proposed method by approximately 40%–50%. A detailed analysis reveals two primary factors contributing to this gap. First, DPD and the proposed method utilize distinct segmentation techniques, resulting in different contours for the same sunspots. Second, the methods used to estimate sunspot areas, based on the pixel counts within the segmented contours, may also differ between DPD and the proposed approach. Despite these differences, the two data sets exhibit consistent trends in sunspot area variation.

Solar cycles are approximately 11 yr long, and each cycle varies in terms of intensity, duration, and characteristics. These differences in sunspot activity across cycles can potentially influence the performance of both MorphACWE and U-Net. MorphACWE is sensitive to local intensity information and typically performs well when sunspots are well defined and distinct. However, during weaker cycles, with fewer or less prominent sunspots, it may struggle to accurately detect or segment smaller, less distinct sunspots. Furthermore, if sunspots in a given cycle are larger, more complex, or have intricate penumbra structures, MorphACWE may encounter difficulties, as its reliance on local intensity is insufficient to capture such complex details. In contrast, U-Net, with its deep learning architecture and use of skip connections, is capable of capturing both local and global context, enabling it to better adapt to varying sunspot distributions and complex structures across different solar cycles.

5. Conclusion

In this study, we address the challenge of segmenting sunspots in high-resolution solar images by applying MorphACWE within active-region boundaries, which are subsequently utilized to train a U-Net model. The U-Net model achieves a mean performance of 0.98 across five evaluation metrics over the entire test data set, whereas MorphACWE alone yields a mean performance of 0.91. Additionally, the Pearson correlation coefficients between the sunspot areas reported by SWPC and DPD and those predicted by our method are 0.97 and 0.96, respectively.

Traditional machine learning methods, relying on pixel intensity thresholds, fall short in handling the intricate and blurred edges presented in high-resolution images. MorphACWE, with its robustness to noise and ability to manage blurred boundaries, provides a foundation for generating initial segmentation labels. These labels are then used to train a U-Net model, which excels in learning complex features from labeled data. The integration of MorphACWE and U-Net leverages the strengths of both methods, resulting in accurate and efficient sunspot segmentation. The use of MorphACWE eliminates the need for manual annotation, which is not only time-consuming but also requires specific expertise and may introduce human bias. While MorphACWE is computationally intensive, the trained U-Net model offers significant improvements in time efficiency, segmenting a full-disk continuum image several minutes faster than MorphACWE.

Both MorphACWE and U-Net may be affected by the differences in sunspot characteristics across solar cycles, but U-Net is likely to be more adaptable. To improve performance across cycles, it would be beneficial to train these models on a diverse data set including sunspot images from different solar cycles, with varied sunspot sizes, shapes, and distributions.

The proposed method segments sunspots based on single continuum images. However, examining multiple preceding and subsequent images for sunspot segmentation could help mitigate noise that might otherwise be mistaken for small sunspots. As such, future work will explore deep learning approaches that incorporate the evolutionary properties of sunspots, which may further enhance sunspot segmentation accuracy.

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). N.G.G. acknowledges the NVIDIA Academic Hardware Grant Program for their support in providing essential computational resources (NVIDIA A5000 GPU). R.E. acknowledges the NKFIH (OTKA; grant No. K142987), Hungary for enabling this research. R.E. is also grateful to the Science and Technology Facilities Council (STFC; grant No. ST/M000826/1), UK, to PIFI (China; grant No. 2024PVA0043), and to the NKFIH Excellence Grant TKP2021-NKTA-64 (Hungary). J.L. acknowledges support from the National Natural Science Foundation (NSFC 42188101 and 12373056).

Software: Github (https://github.com/cjing00923/Automated-Sunspot-Segmentation).

ORCID iDs

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

References

Baranyi, T., Győri, L., & Ludmány, A. 2016, SoPh, 291, 3081
Bobra, M., Sun, X., Hoeksema, J., et al. 2014, SoPh, 289, 3549
Bourgeois, S., Barata, T., Erdélyi, R., Gafeira, R., & Oliveira, O. 2024, SoPh, 299, 10
Çakmak, H. 2014, ExA, 37, 539
Cao, K., Xia, Y., Yao, J., et al. 2023, Nat. Med., 29, 3033
Chan, T., & Vese, L. 1999, Scale-space Theories in Computer Vision, Lecture Notes in Computer Science 1682 (Berlin: Springer), 141
Chola, C., & Benifa, J. V. B. 2022, GloTP, 3, 177
Colak, T., & Qahwaji, R. 2008, SoPh, 248, 277
Cox, A. 2000, Allen's Astrophysical Quantities (New York: Springer)
Curto, J., Blanca, M., & Martínez, E. 2008, SoPh, 250, 411
Djafer, D., Irbah, A., & Meftah, M. 2012, SoPh, 281, 863
Georgoulis, M. K., Yardley, S. L., Guerra, J. A., et al. 2024, AdSpR, in press
Goel, S., & Mathew, S. 2014, SoPh, 289, 1413
Hathaway, D. H. 2010, LRSP, 7, 1
Hoeksema, J., Liu, Y., Hayashi, K., et al. 2014, SoPh, 289, 3483
Korsós, M. B., Erdélyi, R., Liu, J., & Morgan, H. 2021, FrASS, 7, 113
Krizhevsky, A., Sutskever, I., & Hinton, G. 2017, CACM, 60, 84
Liu, J., Song, A., Jess, D., et al. 2023, ApJS, 266, 17
Miroshnichenko, L. 2023, Solar-terrestrial Relations: From Solar Activity to Heliobiology (Cham: Springer)
Morgan, H., & Korsós, M. B. 2022, ApJL, 933, L27
Mourato, A., Faria, J., & Ventura, R. 2024, Eng. Appl. Artif. Intell., 129, 107636
Márquez-Neila, P., Baumela, L., & Alvarez, L. 2014, ITPAM, 36, 2
Papathanasopoulos, P., Preka-Papadema, P., Gkotsinas, A., et al. 2016, Clin. Neurol. Neurosurg., 146, 82
Pesnell, W. D., Thompson, B. J., & Chamberlin, P. C. 2012, SoPh, 275, 3
Ronneberger, O., Fischer, P., & Brox, T. 2015, in Medical Image Computing and Computer-assisted Intervention, MICCAI 2015 (Cham: Springer), 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
Talaei Khoei, T., Ould Slimane, H., & Kaabouch, N. 2023, Neural Comput. Appl., 35, 23103
Watson, F., Fletcher, L., Dalla, S., & Marshall, S. 2009, SoPh, 260, 5
Wu, Y., & He, K. 2018, Group Normalization, in Proc. European Conf. on Computer Vision (ECCV), ed. V. Ferrari (Cham: Springer), 3
Zhang, J., Temmer, M., Gopalswamy, N., et al. 2021, PEPS, 8, 56
Zhao, C., Lin, G., Deng, Y., & Yang, X. 2016, PASA, 33, e18
Zharkov, S., Zharkova, V., Ipson, S., & Benkhalil, A. 2005, EURASIP J. Adv. Signal Processing, 2005, 318462