











Machine Learning–based Identification of the Solar Disk and Plages in Kodaikanal Solar Observatory Historical Suncharts

Dibya Kirti Mishra^{1,2} , Subhamoy Chatterjee³ , Bibhuti Kumar Jha³ , Hemapriya Raju⁴ , Aditya Priyadarshi⁵ ,
Manjunath Hegde⁶ , Srinjana Routh^{1,2} , Dipankar Banerjee^{7,8} , and M. Saleem Khan²

¹ Aryabhata Research Institute of Observational Sciences, Nainital-263002, Uttarakhand, India; dibyakitimishra@aries.res.in

² Department of Applied Physics, Mahatma Jyotiba Phule Rohilkhand University, Bareilly-243006, Uttar Pradesh, India

³ Southwest Research Institute, Boulder, CO 80302, USA

⁴ Tata Institute of Fundamental Research, Colaba, Mumbai 400005, Maharashtra, India

⁵ Curtin University, Karawara, Western Australia, Australia

⁶ Indian Institute of Astrophysics, Koramangala, Bangalore 560034, India

⁷ Indian Institute of Space Science and Technology, Thiruvananthapuram, 695547 Kerala, India; dipu@iiap.res.in

⁸ Center of Excellence in Space Sciences India, IISER Kolkata, Mohanpur 741246, West Bengal, India

Received 2025 August 27; revised 2025 November 7; accepted 2025 November 24; published 2026 February 20

Abstract

Kodaikanal Solar Observatory (KoSO) is one of the oldest solar observatories, possessing an archive of multiwavelength solar observations, including white light, Ca II K, and H α images spanning over a century. In addition to these observations, KoSO has preserved hand-drawn suncharts (1904–2022), on which various solar features such as sunspots, plages, filaments, and prominences are marked on the Stonyhurst grid with distinct color coding. In this study, we present the first comprehensive result that includes the entire dataset from these suncharts using a supervised machine learning (ML) model called “convolutional neural networks” (CNNs), first to identify the solar disks from the charts (1909–2007) and second to identify the plages, spanning nine solar cycles (1916–2007). We train the CNN with the manually identified solar disk and plages. We first detect the solar limb and the north–south line in the suncharts, which enables the extraction of disk center coordinates, radius, and P angle. Following that, we use a CNN similar architecture to achieve accurate image segmentation for the identification of plages. We compare plage areas derived from the suncharts with those obtained from Ca II K full-disk observations, and find good agreement that demonstrates the potential application of such an ML technique for historical data. The results of this study further demonstrate the potential application of sunchart data to fill the existing data gaps in the KoSO multiwavelength observations and contribute toward constructing a composite series over the last century.

Unified Astronomy Thesaurus concepts: [Solar cycle \(1487\)](#); [Solar chromosphere \(1479\)](#); [Plages \(1240\)](#); [Convolutional neural networks \(1938\)](#); [The Sun \(1693\)](#)

Materials only available in the [online version of record](#): data behind figure

1. Introduction

The study of solar magnetism has attracted significant interest over the past century, expanded by advancements in telescopic observations of prominent solar magnetic features such as sunspots (white light), plages (Ca II K), and filaments (H α). Although naked-eye sunspot observations date back millennia, particularly within far eastern civilizations (A. D. Wittmann & Z. T. Xu 1987; K. K. C. Yau & F. R. Stephenson 1988; H. Hayakawa et al. 2017), systematic telescopic observations only began in the early seventeenth century through the works of Harriot (T. Harriot 1613), Galileo (G. Galileo 1613), and Scheiner (C. Scheiner 1615), among others (R. B. Herr 1978; J. M. Vaquero & M. Vázquez 2009; R. Arlt & J. M. Vaquero 2020; V. M. S. Carrasco et al. 2024). Before photographic plates and digital imaging became available, detailed solar drawings provided crucial historical records of magnetic features, supporting reconstructions of solar activity extending back to the Maunder Minimum (H. Hayakawa et al. 2024). Numerous archival datasets globally document solar features, notably from

the Royal Observatory of Belgium (1940–2011), Mount Wilson Observatory (MWO, 1913–2017), Specola Solare Ticinese (1981–2017), Meudon (1919–present), McIntosh archive (1967–present), and Kislovodsk solar station (1979–2021), all providing extensive coverage of solar magnetic activity (A. Priyadarshi et al. 2023). Among these, the Kodaikanal Solar Observatory (KoSO) stands out due to its uniquely continuous record of annotated solar drawings (suncharts hereafter) spanning 1904–2022 (119 yr; B. Ravindra et al. 2020). These KoSO suncharts are concurrently produced with multiwavelength observations in white light (S. Mandal et al. 2017; B. K. Jha et al. 2022), Ca II K (393.37 nm; B. K. Jha et al. 2024), and H α (656.3 nm; S. Chatterjee et al. 2017), and contain detailed annotations of sunspots, plages, and filaments/prominences, offering an invaluable, comprehensive, and consistent dataset for long-term multiwavelength studies of solar variability with its continuity and extensive temporal coverage (see Figure 1).

Plages in Ca II K images are valuable proxies for reconstructing pre-1970 solar magnetic fields due to the correlation between Ca II K intensity and magnetic field strength (A. A. Pevtsov et al. 2016; T. Chatzistergos et al. 2019b; A. V. Mordvinov et al. 2020; G. Shin et al. 2020; A. R. Yeates et al. 2025). They have been used to generate pseudomagnetograms and investigate



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

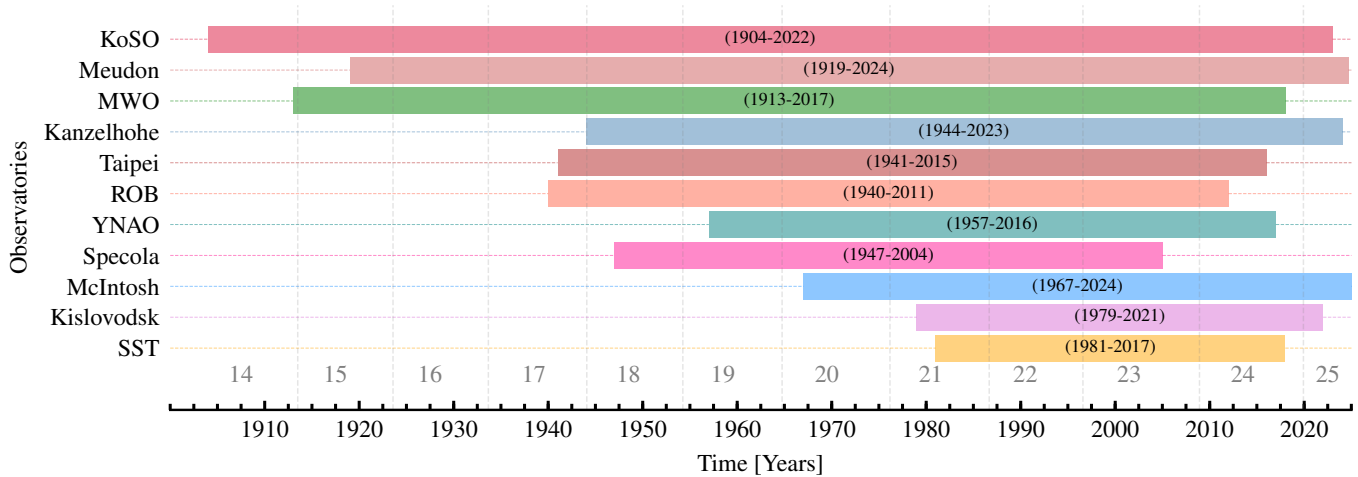


Figure 1. Temporal coverage of solar drawings from various observatories, showing the duration of their observations in years. The KoSO provides the longest continuous dataset (1904–2022), followed by Meudon, MWO, Kanzelhöhe, and others.

long-term solar variability (I. O. I. Virtanen et al. 2019, 2022), benefiting from their long lifetimes and structural stability (H. Zirin 1974). Automatic and semiautomatic detection of plages (M. Priyal et al. 2014; S. Chatterjee et al. 2016; T. Barata et al. 2018; T. Chatzistergos et al. 2019b, 2022; S. Gharat et al. 2023) often suffer from false detections in archival data, as observed for KoSO Ca II K images (post-1980; B. K. Jha et al. 2024; D. K. Mishra et al. 2025). These limitations motivate the use of machine learning (ML) methods, especially for complex datasets such as hand-drawn suncharts containing several kinds of markings apart from solar features. The generalization capability of ML architectures has greatly increased their reliance on the multiple problems of solar physics (see G. Shin et al. 2020; V. Upendran et al. 2020; Q. Li et al. 2021; H. Raju & S. Das 2021; S. Chatterjee et al. 2022, 2023; V. Deshmukh et al. 2023, and references therein), a comprehensive review has been provided by A. Asensio Ramos et al. (2023).

Recently, convolutional neural networks (CNNs) and related ML algorithms have successfully been applied to detect filaments in $H\alpha$ full-disk images from Big Bear Solar Observatory (G. Zhu et al. 2019) and Global Oscillation Network Group and Kanzelhöhe Observatory (A. Diercke et al. 2024), as well as to identify sunspots in Michelson Doppler Imager (C. Chola & J. V. B. Benifa 2022), Helioseismic and Magnetic Imager (C. Chola & J. V. B. Benifa 2022; L. Palladino et al. 2022), and Uccle Solar Equatorial Table (N. Sayez et al. 2023).

Identification of solar features in suncharts is even more challenging due to the coexistence of diverse feature annotations and nonsolar artifacts. Consequently, ML offers an optimal solution for accurate segmentation and classification in these complex datasets. Recently, A. Priyadarshi et al. (2023) specifically applied an unsupervised ML technique called “K-means clustering” for filament detection using color space in digitized KoSO suncharts (1954–1976), digitized using a Canon EOS 800D camera. This work provided a proof-of-concept for the application of such techniques, which actually led to the complete digitization of the suncharts at KoSO. However, neither this nor the prior studies have attempted to detect plages. Identifying plages using unsupervised ML was challenging in the past due to the indistinct color of plages, unlike filaments in KoSO suncharts, irregular morphology, and variability in sunchart quality and annotation

styles. Additionally, the lack of well-labeled datasets has limited the use of supervised ML approaches. Overcoming these obstacles requires the creation of comprehensive, high-quality training sets and the design of transparent, reproducible CNN architectures that can accommodate the heterogeneous nature of archival solar images.

Despite the advances in detection algorithms, the automatic identification of plages employing ML approaches has not yet been attempted. In this paper, we present a deep learning-based supervised ML approach, specifically a CNN model, for the first automated identification of plages from the newly digitized KoSO suncharts covering the period from 1916 to 2007 (sunchart grid style is different for the pre-1916 period and plage marking is unavailable after 2007). This automated plage detection will effectively address the data gaps in historical Ca II K full-disk images from KoSO, as highlighted in recent studies (B. K. Jha et al. 2024; D. K. Mishra et al. 2024, 2025), and will further aid the generation of pseudomagnetograms through corroboration with available $H\alpha$ dataset from KoSO.

This paper is structured as follows. In Section 2, we describe the observational data utilized in this work. Section 3 outlines the methodology employed for plage detection. In Section 4, we present our findings along with comparative analyses. Finally, we conclude in Section 5.

2. Data

KoSO has conducted solar observations across white light, Ca II K, and $H\alpha$ wavelengths since 1904 using photographic plates and films. These have been digitized and calibrated for public access (B. Ravindra et al. 2013; M. Priyal et al. 2014; S. Chatterjee et al. 2016, 2017; S. Mandal et al. 2017; T. Chatzistergos et al. 2018; B. K. Jha et al. 2022; T. Chatzistergos et al. 2023).⁹ The strength of this archive lies in its consistent optical setup and processing, though post-1980 it started to use films instead of photographic plates, and the image contrast also started to reduce.

Alongside photographic observations, systematic hand-drawn suncharts (Figure 2(a)) have also been produced at KoSO. These suncharts document solar features such as sunspots, plages, filaments, and prominences by projecting the

⁹ <https://kso.iiap.res.in/data>

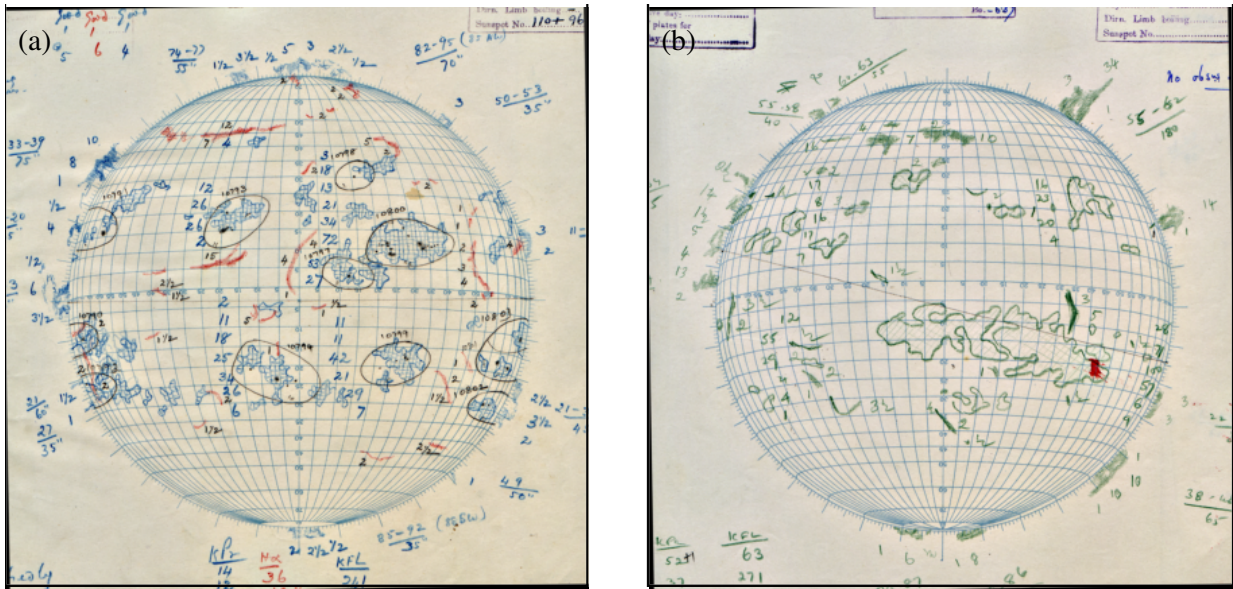


Figure 2. Representative suncharts produced at KoSO. (a) A sunchart from the KoSO dated 1958 January 5, showing detailed hand-drawn annotations of sunspots, plages, filaments, and prominences observed by the KoSO. (b) Another sunchart produced at KoSO dated 1958 February 6, showing solar features with different color because the original observation was made by a different observatory. Please note that suncharts are contrast-enhanced for better visualization.

daily observations (white light for sunspots, Ca II K for plages and prominences, and H α for filaments) onto a Stonyhurst latitude and longitude grid, which is divided into increments of 5 $^\circ$ in both latitude and longitude directions. These drawings were typically created shortly after the plates or films were developed, minimizing the effects of scratches, fungal degradation, and other artifacts that can affect the long-term preservation of the photographic plates/films. The combined use of suncharts along with the digitized photographic plates and films offers a means to generate a consistent and reliable dataset of solar features spanning the period from 1904 to 2007.

Each solar feature is distinctly marked using specific colors: sunspots in black pencil, plages in blue ballpoint pen, filaments in red pencil, and prominences in blue pencil, as seen in Figure 2(a). Additional observational metadata, including the position angle of the solar north (P angle), heliographic latitude of the solar disk center (B angle), and the observation time (white light data timestamp), is provided at the top of each chart. Typically, observers create one sunchart per day, based on the best available photographic observation. In instances where the daily observations from KoSO were unavailable, spectroheliograms from MWO or Meudon Observatory were used to supplement the records and prepare the suncharts. Particularly during the period of 1957–1958, an active exchange program of spectroheliograms was conducted between KoSO and a few other observatories under the scheme of the International Astronomical Union (G. E. Hale et al. 1931; E. Mausam 1951; A. K. Das 1959; S. S. Hasan et al. 2010). To clearly distinguish the observation taken from various observatories, different colors (e.g., green for plages and filaments) are used, as shown in Figure 2(b). Further comprehensive details regarding these data are available in B. Ravindra et al. (2020) and A. Priyadarshi et al. (2023).

2.1. Sunchart Digitization

The initial digitization of KoSO suncharts covered two solar cycles (cycles 19 and 20; 1954–1976) and was conducted

using a Canon EOS 800D camera, as described in detail by A. Priyadarshi et al. (2023). More recently, however, the entire collection of KoSO suncharts spanning from 1904 to 2022 has been systematically digitized utilizing a high-resolution scanner capable of generating 6k \times 6k images. For this purpose, at KoSO the Zeutschel OS12002 scanner is used, which is specifically designed for high-resolution digitization of large-format materials such as books, drawings, and maps. It supports a maximum scanning resolution of 600 pixels per inch, enabling the accurate preservation of fine structural details. Two versions of these digitized suncharts have been produced: high-resolution (6k \times 6k) and lower-resolution (1k \times 1k) images, both stored in “.tif” format, offering distinct fields of view (cropped version). However, we used only high-resolution (6k \times 6k) suncharts in this work. The file names explicitly encode the presence or absence of specific solar features, enabling the quick identification of the sunchart contents without requiring direct inspection of the images. In this convention, the letter “k” denotes features annotated from KoSO and t or m from other observatories (MWO or Meudon), while “x” indicates the absence of a feature. For example, in “sch_19580105_0754_kkkk.tif” all four features (sunspots, filaments, plages, and prominences) are drawn from KoSO observations, whereas in “sch_19580206_0000_xttt.tif” the absence of sunspots is indicated by x, while filaments, plages, and prominences are included based on data from external observatories (MWO and Meudon) as shown in Figure 2.

The temporal distribution of these suncharts, measured in terms of observing days per year, is shown in Figure 3 as open red bars for the period 1904–2022. Furthermore, this histogram also compares the total number of suncharts with those specifically featuring plages (filled red bars), alongside the number of Ca II K observing days (open blue bars). The comparison highlights certain intervals (notably from 1955 to 1986) where the number of available suncharts exceeds the number of corresponding Ca II K images, emphasizing the utility of suncharts in filling existing observational gaps in the KoSO Ca II K dataset. Moreover, the suncharts are particularly

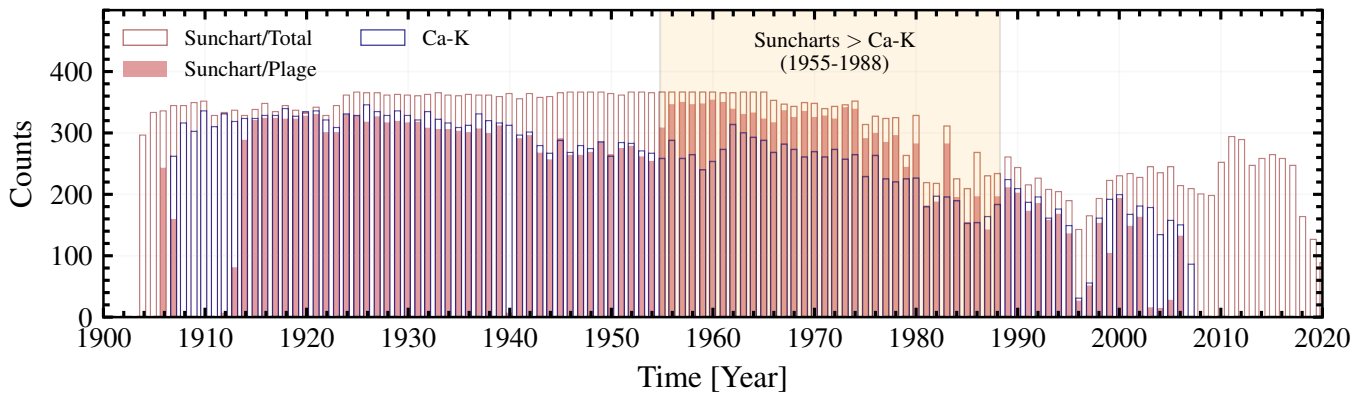


Figure 3. Comparison of the annual number of Ca II K images with the total number of suncharts and those containing plage markings, represented by the number of observing days. This highlights periods where suncharts, particularly those with plages (1955–1988), can supplement or extend the Ca II K observational record.

valuable in addressing the challenge of accurately identifying plage regions after 1980, a period during which data degradation in photographic plates/films significantly affected the quality of Ca II K images in the digitized data.

In the present study, we utilize high-resolution ($6k \times 6k$) digitized suncharts covering the period from 1909 to 2007 (31,438 images) and 1916–2007 (29,070 images) for the automated identification of disk and plages, respectively. Years prior to 1909 are excluded because many of the earlier suncharts do not include solar features overlaid on the Stonyhurst grid, as shown in the Appendix (Figure 14(a)). We also did not utilize the data of the pre-1916 period due to the different pattern of the sunchart grid (blue-filled) for this period (Figure 14(b)).

3. Methodology

In this study, we employ a supervised ML model called U-Net to automate the identification of plages within KoSO hand-drawn suncharts. U-Nets are based on CNNs and are widely used as a highly effective tool, particularly in biomedical image segmentation tasks (O. Ronneberger et al. 2015; Ö. Çiçek et al. 2016). In the domain of solar physics, recent studies have employed U-Net models successfully for the automatic identification of various solar phenomena, such as coronal jets (J. Liu et al. 2024) and solar filaments (G. Zhu et al. 2019). Motivated by these promising results, we adopt the U-Net CNN architecture in our analysis to identify solar plages automatically on digitized suncharts from KoSO.

The U-net architecture comprises fully convolutional layers structured in an encoder–decoder configuration, creating a distinctive U-shaped framework that demonstrates excellent performance even with limited training datasets. The encoder portion uses 3×3 convolutions with ReLU nonlinearity and systematically reduces the spatial dimensions of the input images through max pooling. This results in hierarchical feature learning, which simultaneously extracts and preserves critical image features. After flattening the extracted features from the encoder into vectors (H. Raju & S. Das 2021), the decoder section progressively restores the image to its original dimensions. It combines the encoded high-resolution features with those from up-convolutional operations, ultimately producing the segmented image as shown in Figure 4. The mathematical description of the U-Net model is presented in the Appendix.

We train U-nets to perform two segmentation tasks as described in the following two subsections. The training and inference tasks of our U-Net model are executed on an in-house GPU equipped with a Quadro RTX 8000 graphics card and 50 GB of VRAM.

3.1. Disk Identification

Prior to the identification of plages, it is essential to accurately detect the solar limb and the north–south line in the suncharts to extract parameters such as disk center coordinates, radius, and P angle, and assign physical coordinates to the plage locations. The extracted disk parameters are not only essential for plage identification but also serve as crucial information for the future detection of other solar features in suncharts, such as sunspots, filaments, and prominences. In previous work, A. Priyadarshi et al. (2023) utilized the linear Hough transform technique for disk detection in old digitized suncharts covering the period 1954–1976 (solar cycles 19 and 20). However, employing the Hough transform across the entire dataset (1904–2022) proved inefficient due to significant variability in image quality over the extended observational timeframe. To address this challenge, we implemented a U-Net CNN model for robust disk detection across the complete data of digitized suncharts (1909–2007). To train the CNN model, we used previously digitized sunchart images from A. Priyadarshi et al. (1954–1976; 2023), along with corresponding binary masks representing the solar disk and central meridian line derived from known center coordinates, radii, and P -angle values extracted in the same study using Hough transform. From this subset, we systematically selected 7310 high-quality sunchart images, which were subsequently resized to dimensions of 256×256 pixels to standardize the input data for training the neural network. A representative example from the training dataset is presented in Figures 5(a) and (b). During the training procedure, pixel-wise differences between the predicted segmentation masks and the ground truth masks were quantified by the binary cross-entropy loss function defined as

$$\text{loss}(Y, \hat{Y}) = -\frac{1}{n^2} \sum_{i,j=0}^n [y_{ij} \log \hat{y}_{ij} + (1 - y_{ij}) \log(1 - \hat{y}_{ij})], \quad (1)$$

where $Y = y_{ij}$ and $\hat{Y} = \hat{y}_{ij}$ are target and predicted segmentation maps correspondingly. We used the Adam optimizer (D. P. Kingma & J. Ba 2017), a first-order gradient-based

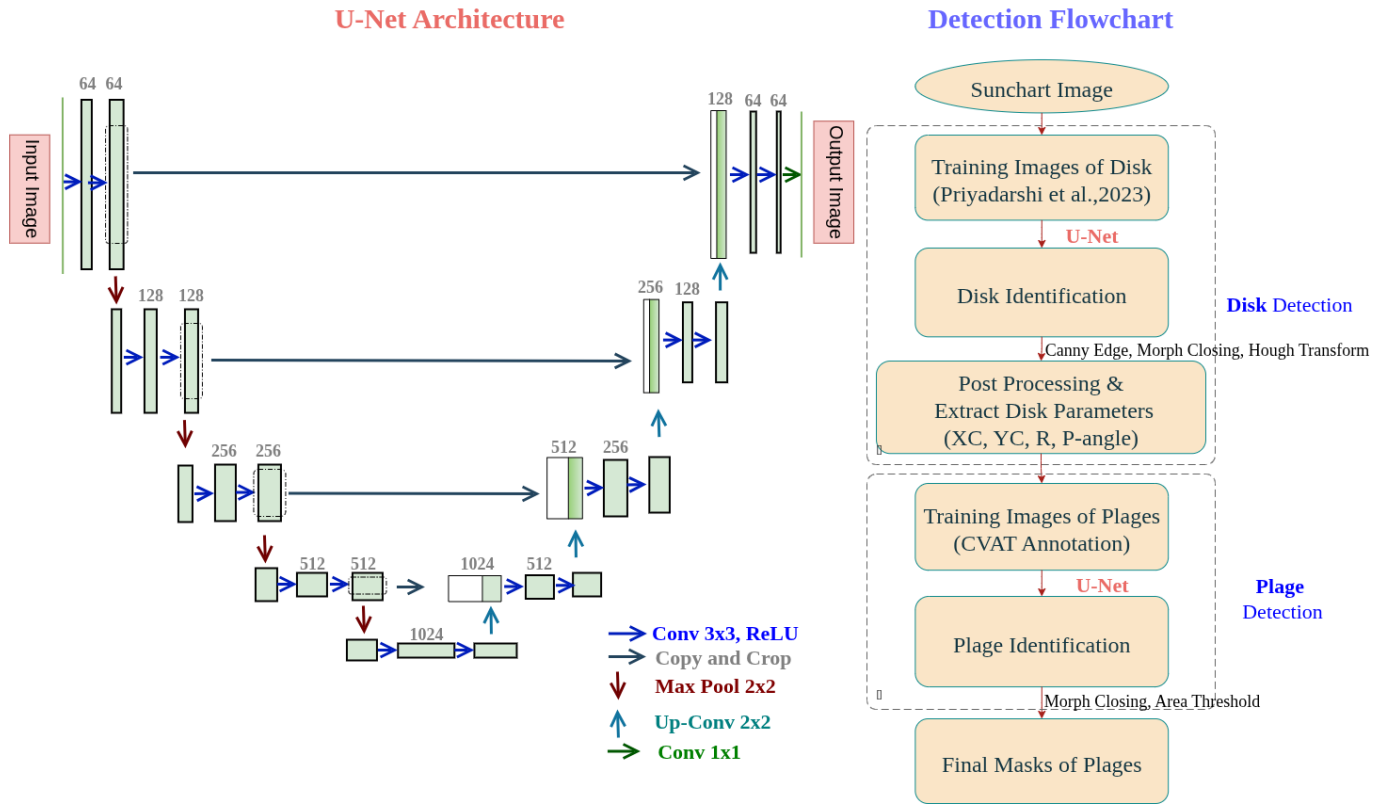


Figure 4. (Left) U-Net architecture (adapted from O. Ronneberger et al. 2015) used for segmenting plages. The architecture consists of a contracting path (left) for feature extraction and an expansive path (right) for precise localization. Key components include 3 × 3 convolutions with ReLU nonlinearity (blue arrows), 2 × 2 max pooling (dark red arrows), 2 × 2 up-convolutions (light blue arrows), and skip connections for combining high-resolution features from the contracting path with the upsampled output. (Right) Flowchart showing the steps for disk and plage detection in suncharts.

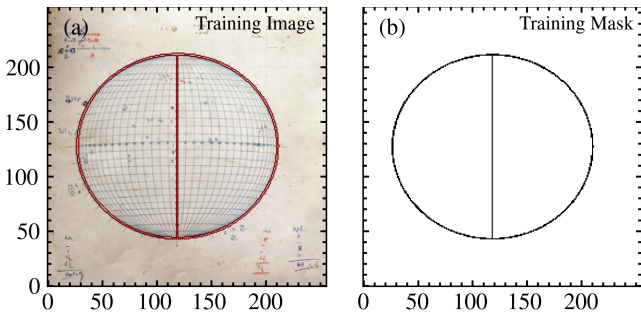


Figure 5. (a) A representative training image with an annotated training mask (red). (b) The corresponding binary mask used for training the disk detection model.

adaptive optimization algorithm, to train the neural network (D. P. Kingma & J. Ba 2017). Adam dynamically adjusts learning rates during each training step, assigning individual adaptive rates for different parameters, thus improving convergence efficiency. The training process involved iterative passes over subsets (batches) of the dataset, where network parameters were updated after processing each batch according to the optimization strategy. An epoch is defined as one complete iteration through the entire dataset. Generally, larger batch sizes can enhance the performance of neural network classification or segmentation, although this choice must be balanced against batch size, model complexity, data dimensionality, and available computational resources (CPU/GPU memory). In this study, the batch size was set to 64, and the

network was trained for a maximum of 25 epochs. The training was concluded at 25 epochs, as this point corresponded to the minimum training and validation loss, along with the intersection-over-union (IoU) score, indicating optimal model performance. With these settings, the training process took approximately 30 minutes to complete.

To evaluate the model’s performance during training, we computed both the training and validation losses. The training loss quantifies the error between the predicted outputs and the true labels within the training dataset, while the validation loss measures the error on a separate validation dataset that is not included by the model during training. It is important to note that 24% of the dataset is allocated for validation purposes in this work. The loss curves for both datasets are shown in Figure 6(a). The convergence of these curves toward the end of training indicates that the model is neither underfitting nor overfitting, suggesting effective learning and generalization. In addition to the loss, we also evaluated the model using the IoU metric for both training and validation datasets. IoU is widely used to assess the accuracy of object detection and image segmentation models. For a predicted region P and the ground truth region G , IoU is defined as

$$\text{IoU} = \frac{\text{Area of Overlap } (P \cap G)}{\text{Area of Union } (P \cup G)}.$$

The IoU score ranges from 0 to 1, where 0 indicates no overlap and 1 indicates a perfect match between the predicted and ground truth regions. The IoU curves for the training and validation datasets are presented in Figure 6(b). The consistent

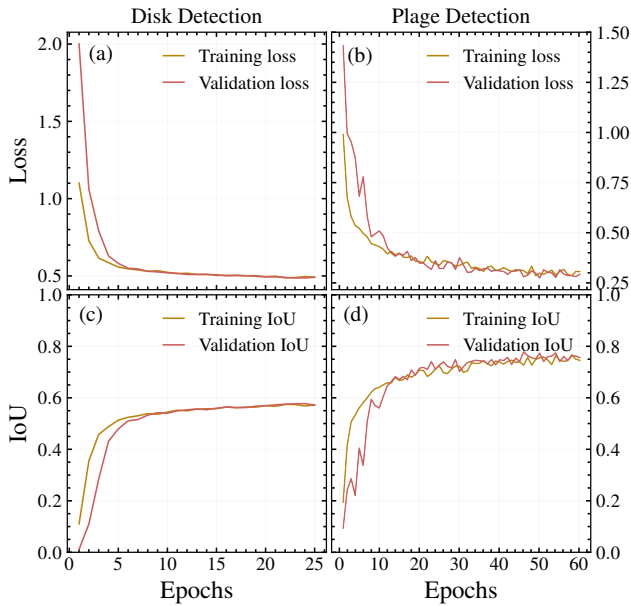


Figure 6. Evolution of model loss and performance with epoch. Panels (a) and (b) show the loss and IoU curves, respectively, for the solar disk detection model. Panels (c) and (d) show the corresponding loss and IoU curves for the plage detection model. Both models exhibit convergence with closely matching training and validation curves, indicating robust performance of the model.

convergence of both curves further confirms that the model is learning effectively to detect the solar disk with reasonable accuracy.

After training, we use the model to infer the solar disk and north–south line from the digitized sunchart images resized to 256×256 pixels, covering the period 1909–2007. An example of a predicted solar disk using the trained model is presented in Figure 7(b) for the sunchart shown in Figure 7(a). To refine the disk segmentation, we apply the Canny edge detector (J. Canny 1986) and morphological operations such as closing, erosion, and dilation to the output of the model, as shown in Figure 7(c). The processed disk contour is then overlaid onto the original sunchart for the observation date of 1917 January 1 and is shown in Figure 7(c) using a red outline. This overlay highlights the accuracy and robustness of the model’s prediction. Subsequently, we apply the linear and circular Hough transform methods (P. V. Hough 1962) to extract geometric parameters of the solar disk. The linear Hough transform is used to determine the solar P angle from the orientation of the detected central meridian line, while the circular Hough transform provides the center coordinates and radius of the disk. Specifically, the Hough transform identifies Hough circle peaks, in terms of the number of intersections, which correspond to the most probable circular features in the parameter space, representing the best-matching circles detected in the image. The resulting fitted circle and line are shown as red dashed contours over the rotation-corrected sunchart image in Figure 7(d), further validating the model’s effectiveness. Since the model was trained and applied on downsampled 256×256 images for computational efficiency, the extracted disk parameters (center, radius, and P angle) were appropriately scaled to correspond to the original $6k \times 6k$ resolution images for the full period from 1909 to 2007. We also manually corrected the P angle for certain images (5%) due to the poor quality of the corresponding sunchart data,

which hindered accurate automatic estimation. This derived disk and P -angle information is subsequently utilized for automated plage detection, which is described in the following section. We also constructed a FITS standard metadata file for the sunchart by extracting the solar disk region, allowing the data to be easily integrated with SunPy (The SunPy Community et al. 2020).¹⁰ An example image generated using SunPy’s Map object overlaid with the Stonyhurst heliographic grid (blue) is shown in the Appendix (Figure 15). The sunchart data, along with adequate metadata information, will be made publicly available in the future. This information will be necessary for any other studies that use these suncharts, e.g., for identifying other solar features, such as sunspots, filaments, and prominences, in suncharts.

3.2. Identification of Plages

The suncharts from KoSO have detailed records of various solar features, including sunspots, plages, filaments, and prominences, which are valuable for filling gaps in historical datasets. In the present study, we primarily focus on the detection of plages, along with accurate solar disk identification, to assess the potential applicability of ML techniques for the analysis of historical observations. For the identification of plages, we utilize the same U-net architecture previously used for solar disk detection, as described in Section 3.1. To prepare the training dataset, we first apply P -angle correction to the digitized sunchart images. Due to computational constraints, the original high-resolution $6k \times 6k$ images are resized to a 2048×2048 format. These resized images are then divided into 512×512 patches using the “patchify” module to meet the input size requirements of the U-Net model during training. The corresponding ground truth masks for these image patches are generated through semiautomated annotation using the open-source Computer Vision Annotation Tool (CVAT.ai Corporation 2024),¹¹ which is widely used in the computer vision community. We utilize the AI-assisted annotation feature of CVAT, which leverages pretrained ML models to support object detection, classification, and segmentation. This feature significantly improves annotation speed and accuracy by allowing users to quickly identify and mark regions of interest, thereby enhancing the overall efficiency of the training dataset preparation. For model training, we select sunchart images from the period 1958 to 1968, since this interval captures many plage regions encompassing the maximum activity phases of two strong solar cycles (SC19 and SC20), for robust training. Additionally, we include images from 1917 to mitigate detection bias and to ensure the model learns to detect plages with diverse morphological characteristics because plage structures from the earlier period differ notably from those in later drawings in terms of hatched lines. Examples of the training patches and their corresponding annotated masks are presented in Figure 8, where the ground truth masks are also overlaid on the original images using red contours. A total of 7875 training patches were utilized to train the model, with a random validation-to-training split ratio of 0.2.

Once the training patches are prepared, we prepare the U-net model where a ResNet-34 backbone (R. A. Nsiah et al. 2023) serves as the encoder, extracting low-dimensional features

¹⁰ <https://docs.sunpy.org>

¹¹ <https://app.cvat.ai/>

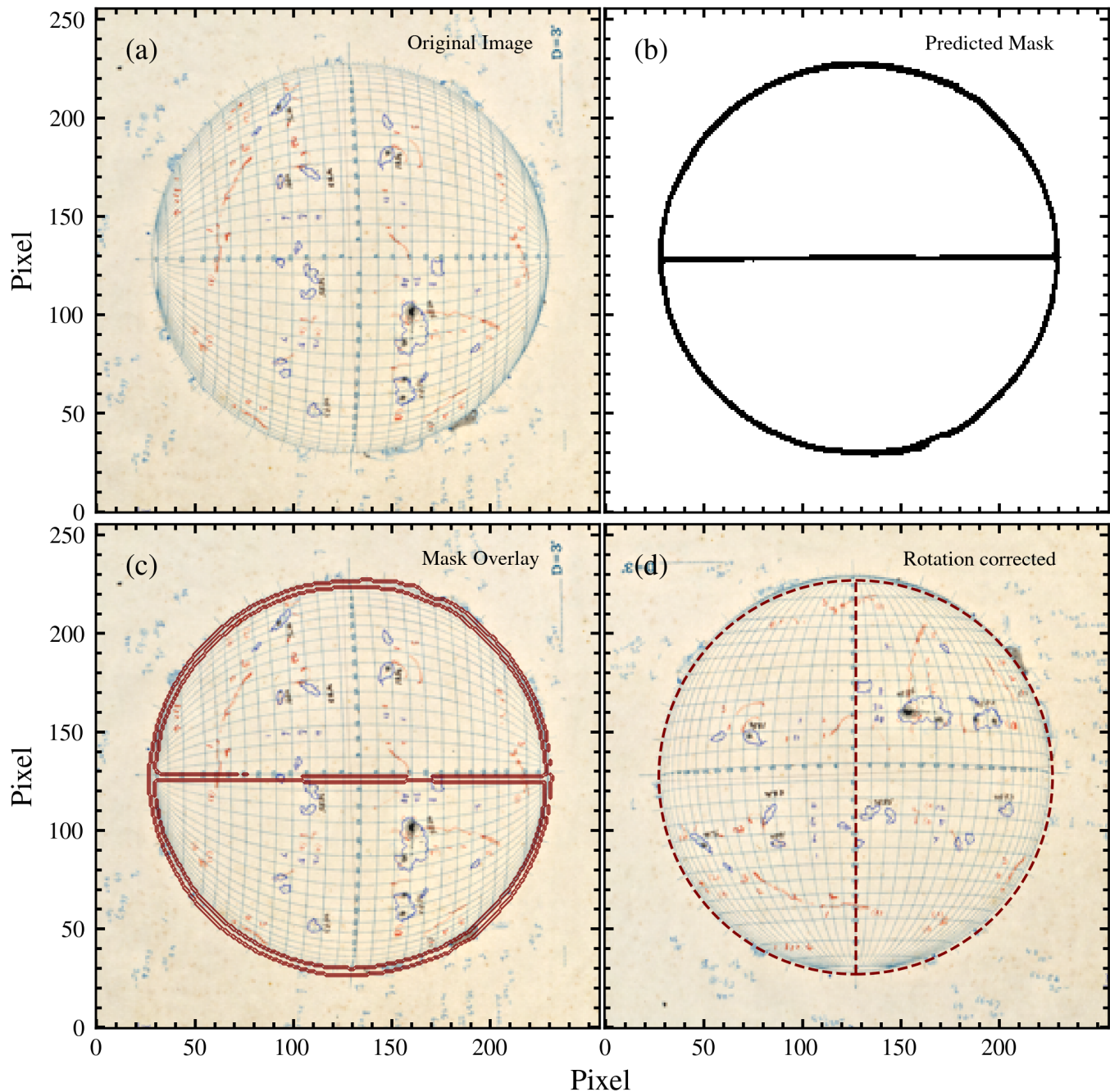


Figure 7. (a) A representative image for a sunchart dated 1917 January 2. (b) Predicted disk mask generated by the trained model. (c) Overlay of the processed mask after postprocessing using Canny edge detection and morphological operations (closing, erosion, dilation), with a contour on the original image. (d) Final rotation-corrected image with a circular fit (dashed red) applied using the Hough circle transform to align the solar disk with the vertical axis. Please note that suncharts are contrast-enhanced for better visualization.

from the images. The encoder is pretrained on large-scale datasets, enabling transfer learning, which significantly reduces training time and enhances model performance, particularly in cases with limited training data. Following preprocessing, we apply data augmentation techniques to reduce the risk of overfitting and to help the model learn invariant features. This augmentation process introduces new transformations of the training images, including rotation, flipping, zooming, and brightness adjustments. For training, we utilize the Adam optimizer and the binary cross-entropy loss function, which is consistent with the configuration used

in the disk identification model. However, we modified the training parameters, using a batch size of 32 and 60 epochs based on the optimal performance of the model with these hyperparameters. Training the model for plage identification on a GPU required approximately 1.5 hr. The performance of the model is shown in Figures 6(c) and (d), where both the training and validation loss curves show convergence after a certain number of epochs, indicating model stability and effective learning. The observed fluctuations in the loss and IoU curves, along with relatively lower IoU scores, are attributed to the morphological complexity of features in the

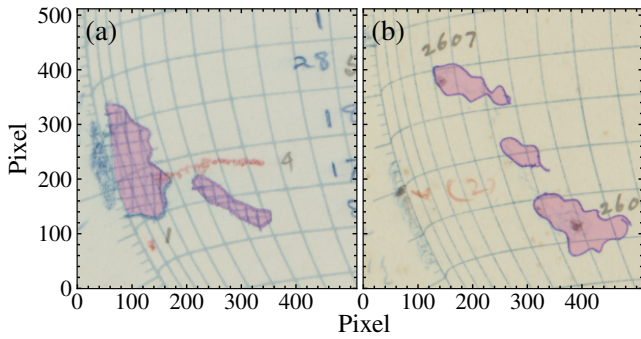


Figure 8. Plage annotation using CVAT. Panels (a) and (b) show input image patches from 1958 March 14 and 1917 January 28, respectively, with pink-filled contours indicating manually annotated plage region masks.

suncharts, which poses a limitation for the model. Nonetheless, the model performs reasonably well, with the validation loss reaching 0.25 and IoU reaching a value of 0.8 at the end of 60 epochs, given the inherent variability and intricacy of the input data.

After completing the training phase, the U-net model is applied to the full dataset (1909–2007) of digitized KoSO sunchart images to automatically identify plages. As for the training dataset, each sunchart is first corrected for its P angle and then divided into 512×512 pixel patches, which are used as input to the model for prediction. An example of the predicted patches for the sunchart observed on 1937 January 2 is shown in Figure 9(a). The predicted mask patches are subsequently recombined to reconstruct the full-disk binary mask of detected plages. To ensure spatial consistency, regions outside the identified solar disk were masked out, eliminating false detections beyond the disk boundary. The resultant full-disk plage masks are shown in Figures 9(b), (c), with the detected plage locations overlaid in pink on the original sunchart image for visual validation. It is important to note that for the period spanning 1909–1915, the model was unable to detect plages due to the suncharts being blue-filled (Figure 14(b)) as mentioned in Section 2, which differs from the annotation style used in the post-1915 charts. Consequently, the analysis of plages based on suncharts is initiated from the year 1916 onward.

These results demonstrate that the model generalizes well to data outside the training period, as shown in Figure 9, which represents data from years not included in the training set (i.e., 1958–1968 and 1917). However, we face some challenges. The model occasionally misclassifies the sunchart grid lines as plages, and some plages are fragmented rather than detected as cohesive structures. To address these limitations, we apply morphological closing operations and area-based thresholding ($A > 600 \text{ pixel}^2$) to merge fragmented plage regions and suppress small false detections, such as grid lines or numbers written in suncharts. This postprocessing step is shown in Figures 9(d)–(f), using the prediction results from the sunchart dated 1975 February 23. The trained model and the processed binary plage masks for the entire period (1916–2007) will be made publicly available in the future.

To evaluate model performance across varying solar activity conditions, we tested the predictions on suncharts from different phases of Solar Cycle 19, specifically the maximum (1958 January 9), intermediate (1956 January 1), and minimum (1954 January 1) periods. The predicted plage masks are overlaid on the original sunchart images as red

contours (Figures 10(a)–(c)), and compared with plage locations in cotemporal Ca II K full-disk images from KoSO (Figures 10(d)–(f)).

These comparisons indicate that the model performs robustly across different phases of the solar cycle. We note that some plages remain undetected during solar maximum, likely due to the increased morphological complexity and density of features. A schematic workflow summarizing the sequential steps involved in identifying the solar disk and plages from the sunCharts is presented in Figure 4.

4. Results

4.1. Plage Time–Latitude Diagram

Following the detection of plages on the suncharts spanning 1916–2007, binary masks of the identified regions are saved in a standardized 2048×2048 format. To evaluate the model’s performance in terms of extracting the spatiotemporal evolution of the Sun, we calculated the rms error in terms of latitude position between the ground truth and predicted plages. This error distribution is shown in Figure 11(a) for the year 1959, where the spatial overlap between predicted and ground truth centroids is visualized in terms of error distribution. The correlation between the predicted and ground truth centroid latitudes is shown in Figure 11(b), yielding a Pearson correlation coefficient (r) of 0.80. These results indicate a generally good agreement between the predicted and actual plage positions. We have also calculated the average balanced accuracy (BA) score to be 0.94, defined as the mean of the true positive rate and the true negative rate. BA addresses the significant disparity in the number of pixels between the background (quiet Sun) and the foreground (plages), driven by the pixel-per-pixel success rate.

The time–latitude distribution of plage areas, commonly referred to as the “butterfly diagram,” is presented in Figure 12. The diagram shown in Figure 12(a) is constructed using the plage masks derived from suncharts. The plage area was computed as the fractional disk area within 1° latitudinal bins across the range -90° to $+90^\circ$, similarly done in B. K. Jha et al. (2024). As with sunspots, plages tend to emerge at higher latitudes during the early phase of a solar cycle and progressively migrate toward the equator as the cycle advances, consistent with previous findings by S. Chatterjee et al. (2016), M. Priyal et al. (2017), and B. K. Jha et al. (2024). For comparison, the butterfly diagram derived from the plage areas in B. K. Jha et al. (2024) is shown in Figure 12(b). A qualitative comparison between Figures 12(a) and (b) reveals that the sunchart-derived butterfly diagram exhibits reduced scatter and fewer spurious detections, particularly at higher latitudes. Notably, the artifacts present in the Ca II K plage masks, appearing as artificial latitudinal stripes around the years 1961 and 1990 due to artifacts resulting from degradation (or other reasons, such as scratches or fungi) in digitized Ca II K images, are absent in the sunchart-based masks. Although the sunchart record exhibits data gaps, particularly around 1940 and post-2003, most of the observational coverage remains intact. To address the remaining gaps, we constructed a composite time–latitude diagram by supplementing the missing sunchart-derived plage area data with corresponding values from the Ca II K plage masks. The resulting composite diagram, shown in Figure 12(c), provides a more continuous and complete temporal coverage. Minor

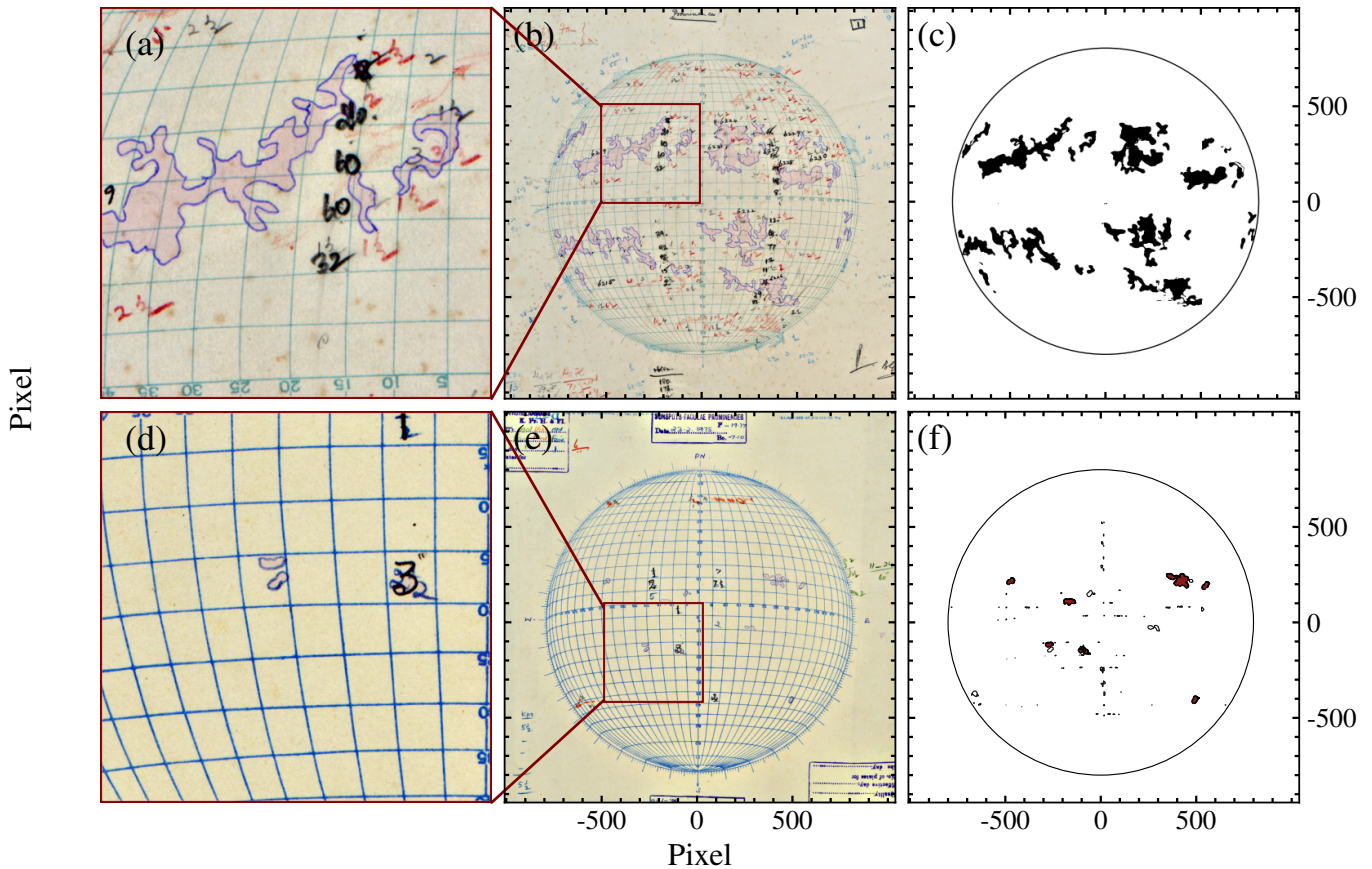


Figure 9. Plage detection from KoSO suncharts. (a) A zoomed-in region of the KoSO sunchart observed on 1937 January 2, with the predicted plage contours shown in pink. (b) The full sunchart image from the same date, with predicted plages overlaid, and (c) the corresponding binary plage mask. (d) A zoomed-in patch from a sunchart with poor-quality annotations observed on 1975 February 23, and (e) the corresponding full-disk image with predicted plage regions. (f) Final processed binary mask from (e) after applying morphological closing and area-based filtering. Red-filled contours in (f) highlight the remaining plage regions after postprocessing. Please note that suncharts are contrast-enhanced for better visualization.

missing data near the year 2007 remain and could be addressed in future work by incorporating additional Ca II K plage datasets from other archives.

In addition to the plage area-based butterfly diagram shown in Figure 12, we also compute the time–latitude distribution of plages using the centroid latitude of each plage region, following the approach adopted by S. Chatterjee et al. (2016) and M. Priyal et al. (2017). The resulting distribution, derived from the sunchart-based plage masks, is shown in Figure 16(a). For comparison, the corresponding diagrams from S. Chatterjee et al. (2016) and M. Priyal et al. (2017) are presented in panels (b) and (c) of Figure 16, respectively. A clear improvement in temporal coverage is evident in the sunchart-derived dataset in comparison to Ca II K image-based plage masks (S. Chatterjee et al. 2016; M. Priyal et al. 2017). The gaps present in the S. Chatterjee et al. (2016) and M. Priyal et al. (2017) Ca II K datasets are largely attributable to the exclusion of lower-quality images during their processing, as discussed in B. K. Jha et al. (2024). Despite minor data gaps around 1940 and after 2003, the sunchart-based centroid latitude distribution offers a more exhaustive record of the long-term spatiotemporal evolution of plages.

4.2. Plage Area Series

In recent times, there have been several studies that have investigated the long-term variation of plage area using Ca II K observations from various historical observatories.

For instance, S. Chatterjee et al. (2016) and M. Priyal et al. (2017) analyzed KoSO Ca II K data, while T. Chatzistergos et al. (2018) utilized data from multiple archives, including Arcetri, Kodaikanal, McMath–Hulbert, Meudon, Mitaka, Mount Wilson, Schauinsland, Wendelstein, and the modern Rome/PSPT series to reconstruct the long-term evolution of plage area. In this study, we present the comparison of the plage area derived from the KoSO sunchart-based plage masks spanning 1916–2007 with previous Ca II K plage area results, provided by (private communication with Theodosios Chatzistergos) T. Chatzistergos et al. (2019a, 2020, 2022) and B. K. Jha et al. (2024) to verify our model output. A quantitative comparison between the monthly KoSO sunchart and Ca II K plage areas yields a Pearson correlation coefficient ($r = 0.80$) and Spearman rank correlation coefficient ($\rho = 0.85$), indicating a strong correspondence between the two independent datasets (see Figure 13(a)). This shows the reliability of plage detection in suncharts from our model in a quantitative way.

The time series of monthly averaged plage area is shown in Figure 13(b) as a red continuous line. For comparison purposes and further verification of the CNN model, we overlay the corresponding plage area obtained from KoSO Ca II K masks (blue dashed line). It is worth noting that in this Ca II K dataset, extremely poor-quality images that could not be processed were excluded. The two time series show a strong agreement overall, with noticeable deviations in solar cycles 16, 20, and

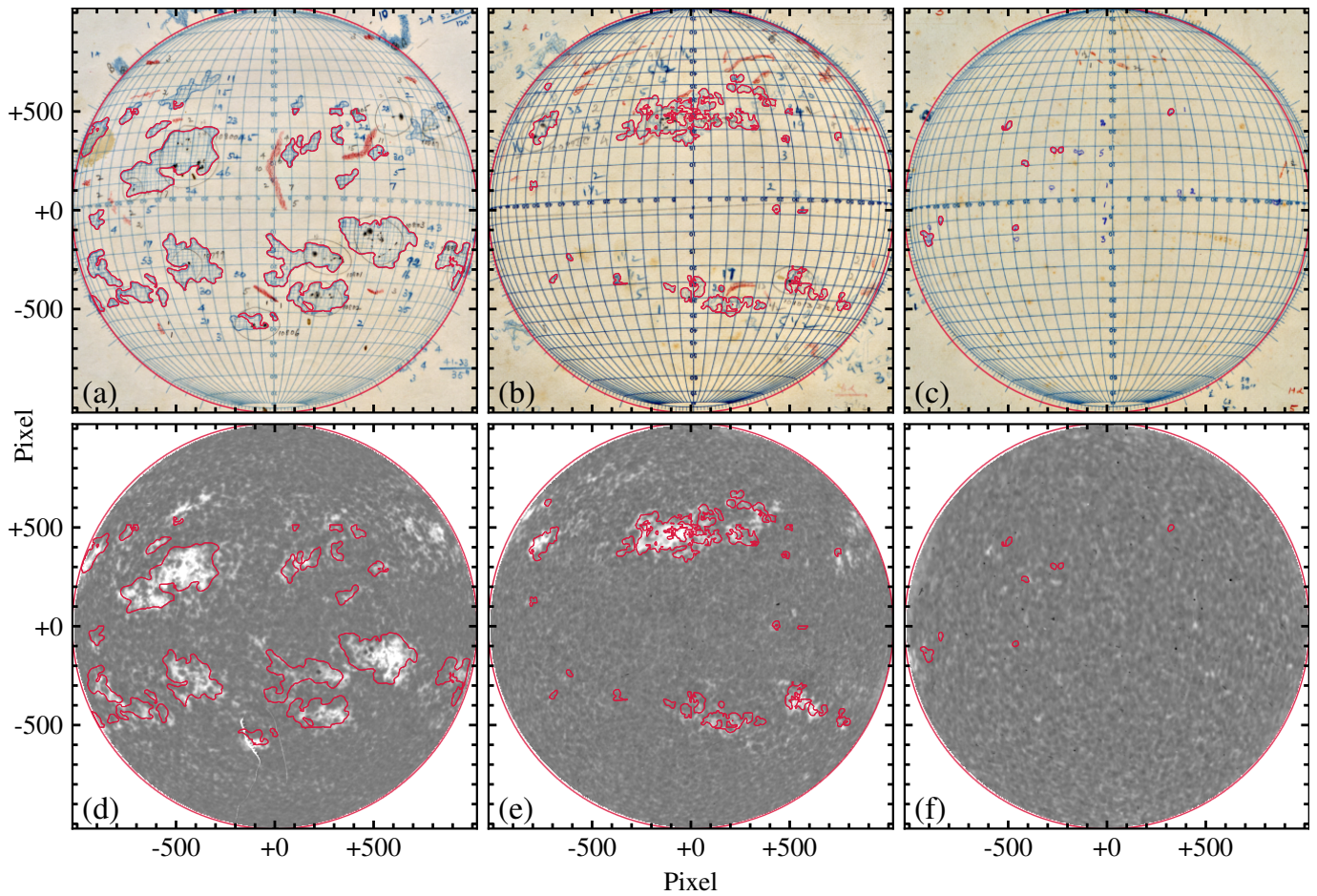


Figure 10. Plage detection in different phases of the solar cycle. (a)–(c) Suncharts representing different phases of solar activity: solar maximum (1958 January 9), intermediate activity (1956 January 1), and solar minimum (1954 January 1). The predicted plage regions are overlaid as blue contours. (d)–(f) Corresponding Ca II K images from the same dates, aligned with the suncharts, and overlaid with the predicted plage masks from the respective suncharts (shown in blue). These comparisons demonstrate the model’s ability to identify plage structures across varying levels of solar activity. Please note that suncharts are contrast-enhanced for better visualization.

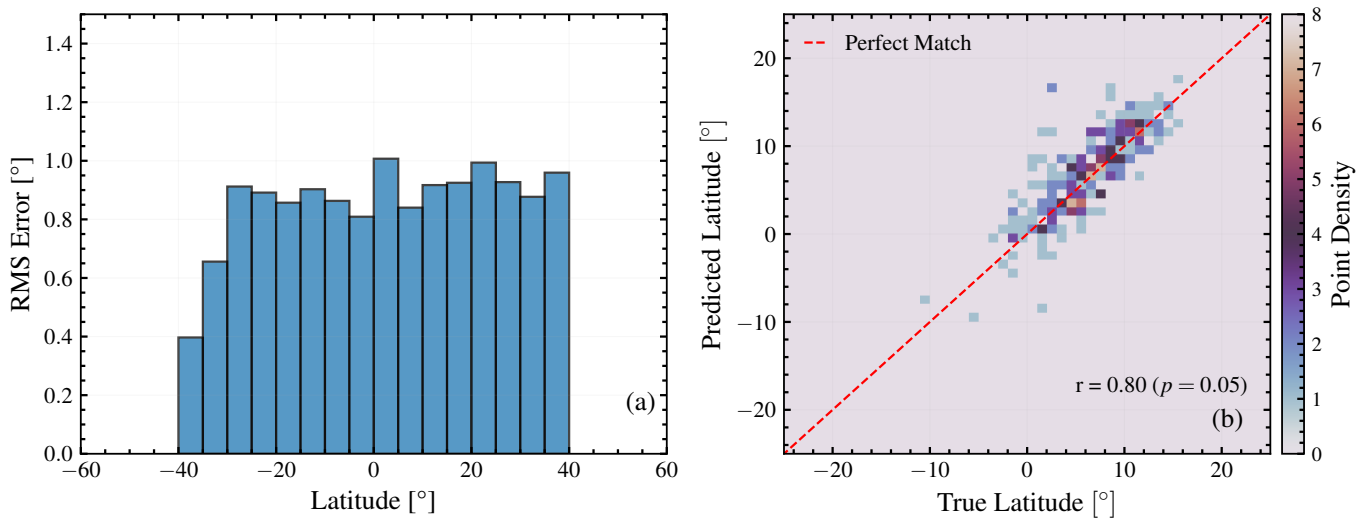


Figure 11. Comparison of physical coordinates of detected plages with ground truth (a) rms latitude prediction error as a function of latitude. The model shows nearly uniform performance across all latitude bins. (b) A 2D histogram presentation comparing predicted versus ground truth latitudes for plage centroids. The color scale indicates the point density of predicted and ground truth latitudes. The dashed line represents the one-to-one relation (perfect match), and the Pearson correlation coefficient ($r = 0.80$, $p = 0.05$) quantifies the level of agreement.

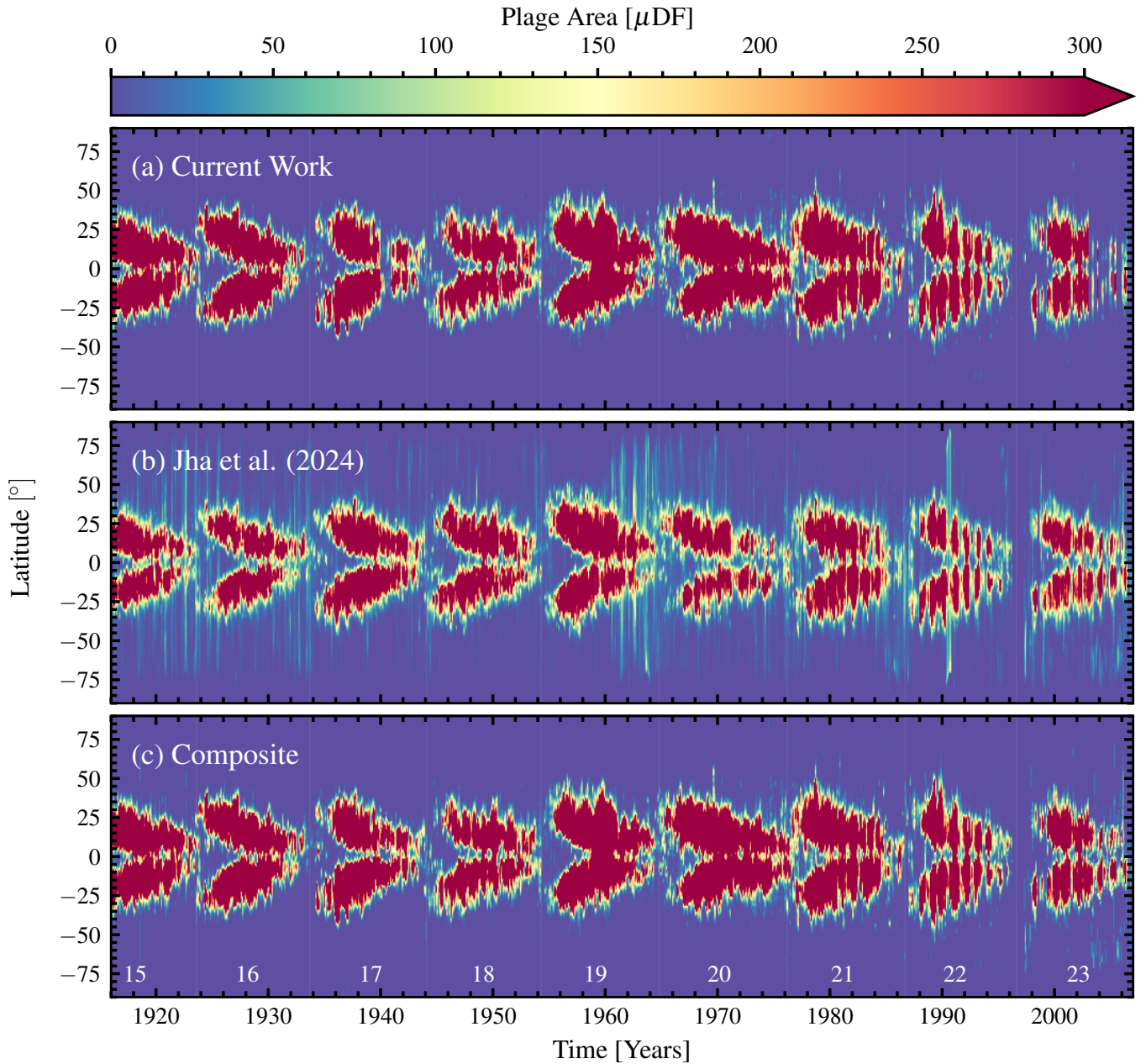


Figure 12. Time–latitude distribution of plage area (in millionths of solar disk fraction, μDF) binned by 1° latitude for each observation day. In cases with multiple observations per day, values are averaged. (a) Results from the current study using the U-net-based detection on suncharts. (b) Corresponding butterfly diagram based on plage detections from Ca II K data (B. K. Jha et al. 2024). (c) Composite series combining (a) and (b), using (b) to fill temporal data gaps in (a). The color bar represents plage area, highlighting latitudinal and solar cycle variation of plage area from 1916 to 2007 (solar cycles 15–23).

(The data used to create this figure are available in the [online article](#).)

especially 22. In cycle 22, a sudden spike is observed around 1989 in the sunchart series, attributable to significantly larger plage markings by the observer during that period, as evident in Figure 17(a). Additionally, a general trend of larger plage annotations is observed in suncharts after 1964, likely reflecting a shift in observer bias or the definition of plage boundaries. A notable drop in the sunchart plage area around 1990 arises from model failure in detecting plages due to their unusual representation in black ink for that year, see Figure 17(b). This mislabeling also resulted in physically implausible high-latitude detections. The data gap in the series near 1940 is due to the absence of plage markings in the suncharts. Notably, after 1978, the Ca II K plage area series

(Figure 13(b)) exhibits a decline in the number of available data points. This gap is supplemented by plage area measurements derived from suncharts. In future work, we plan to combine the sunchart-derived plage area with the Ca II K series to investigate the long-term variability of plage properties.

5. Summary and Conclusion

This study presents the first successful application of an ML approach for the detection of plages in historical suncharts spanning nine solar cycles. Initially, we identify the solar disk and north–south line in suncharts over the period 1909–2007 using a U-Net model, extracting key disk parameters such as

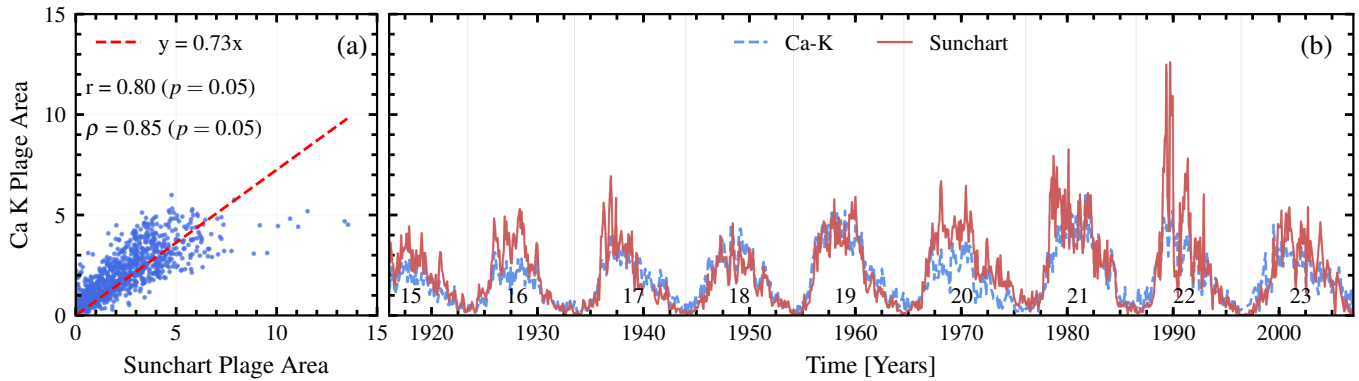


Figure 13. (a) Scatter plot showing the correlation between monthly mean plage areas ($10^4 \mu\text{Hem}$) derived from suncharts and Ca II K observations. The red dashed line represents a linear fit with a slope of 0.73. Pearson correlation coefficient ($r = 0.80$) and Spearman rank correlation coefficient ($\rho = 0.85$) both suggest a strong agreement between the two quantities ($p = 0.05$). (b) Monthly averaged plage area series from suncharts (in $10^4 \mu\text{Hem}$) derived from sunchart-based detections (red) compared with corresponding values from Ca II K full-disk observations (blue) spanning solar cycles 15–23.

center coordinates, radius, and P angle. These disk parameters will be useful for future studies in the analysis of other solar features (sunspots, filaments, prominences) in suncharts and will be available online in “fits” format. Subsequently, plages are segmented using a similar U-net architecture, and the resulting masks are saved in a standardized format for the period 1916–2007. We systematically derive physical coordinates of detected plages and compare them with those obtained from Ca II K full-disk observations to evaluate consistency. We construct time–latitude diagrams from the sunchart-derived plage masks and compare them against similar diagrams produced from Ca II K data, finding good agreement. To enhance temporal coverage, we also develop a composite butterfly diagram by integrating data from both sources. Furthermore, we compare the monthly variation in sunchart-derived plage area from 1916 to 2007 with that from Ca II K observations, resulting in a strong correlation ($r = 0.80$ and $\rho = 0.85$), confirming the effectiveness of the detection method. These results collectively demonstrate the efficacy and robustness of the proposed ML approach for automated plage detection in historical hand-drawn suncharts.

Despite the successful detection of plages in the KoSO suncharts, several limitations remain in the current model. The model fails to identify plages in the pre-1916 period and in 1990 due to inconsistencies in the representation of the Stonyhurst grid and variations in plage annotations in these datasets. This highlights the need to develop a more generalized and robust model capable of handling diverse sunchart formats across different observational periods. Additionally, suncharts from the pre-1909 period exhibit a significantly different grid structure, which has hindered the reliable detection of solar disks using the present method. Future work will focus on enhancing the detection method by incorporating training datasets from multiple solar cycles, thereby accounting for variations in plage appearance, particularly the color inconsistencies observed in years such as 1990. This will be achieved by standardizing training datasets using gray-scale images to improve model generalization. Additionally, implementing a calibration procedure for plage area estimates will help mitigate observer-dependent biases inherent in the manual annotation of suncharts.

Given all the limitations of the current study, we believe the detected plage information from this century-long series of hand-drawn suncharts will make a valuable addition to the historical solar chromospheric observations. The Ca II K sunchart composite

will not only enhance the observation statistics by mitigating data gaps but also help constrain the solar magnetic evolution in the past, complementing sunspot records. While the current study has focused solely on the identification of plages, future extensions of this work will target the automated detection of other annotated solar features in suncharts, including sunspots, filaments, and Ca II K prominences. This would produce a more complete picture of the long-term spatiotemporal evolution of the solar magnetic field.

Acknowledgments

The data and suncharts used in this analysis were developed at the Kodaikanal Solar Observatory. We are grateful to the many observers who consistently carried out observations and prepared the suncharts, and to Mr. P. Manikandan (Sathish) for their patient work in digitizing them. We gratefully acknowledge ARIES for providing the GPU server essential to developing our model. We gratefully acknowledge Theodosios Chatzistergos for providing the Ca II K plage area series and for his valuable suggestions. The funding support for D.K.M.’s research is from the Council of Scientific & Industrial Research (CSIR), India, under file No .09/0948(11923)/2022-EMR-I. This research used version 7.0.1 (S. J. Mumford et al. 2025) of the SunPy open-source software package (The SunPy Community et al. 2020). This study has utilized the SAO/NASA Astrophysics Data System’s bibliographic services.

Software: IDL, sunpy, tensorflow, keras.

Appendix U-net Model Description

Here, we discuss the mathematical operations underlying the U-Net architecture. As mentioned in Section 3, U-Net comprises an encoder–decoder pair. The encoder performs a sequence of convolutions with appropriate padding, followed by downsampling (e.g., max pooling), to extract hierarchical features while progressively reducing the spatial resolution. For encoder layer Le , the feature map (O_{Le}^{en}) is produced from the output (O_{Le-1}^{en}) of the previous layer according to Equation (A1), i.e., the convolution followed by a ReLU nonlinearity. At each level of the encoder and decoder, two consecutive convolutional operations are applied, each

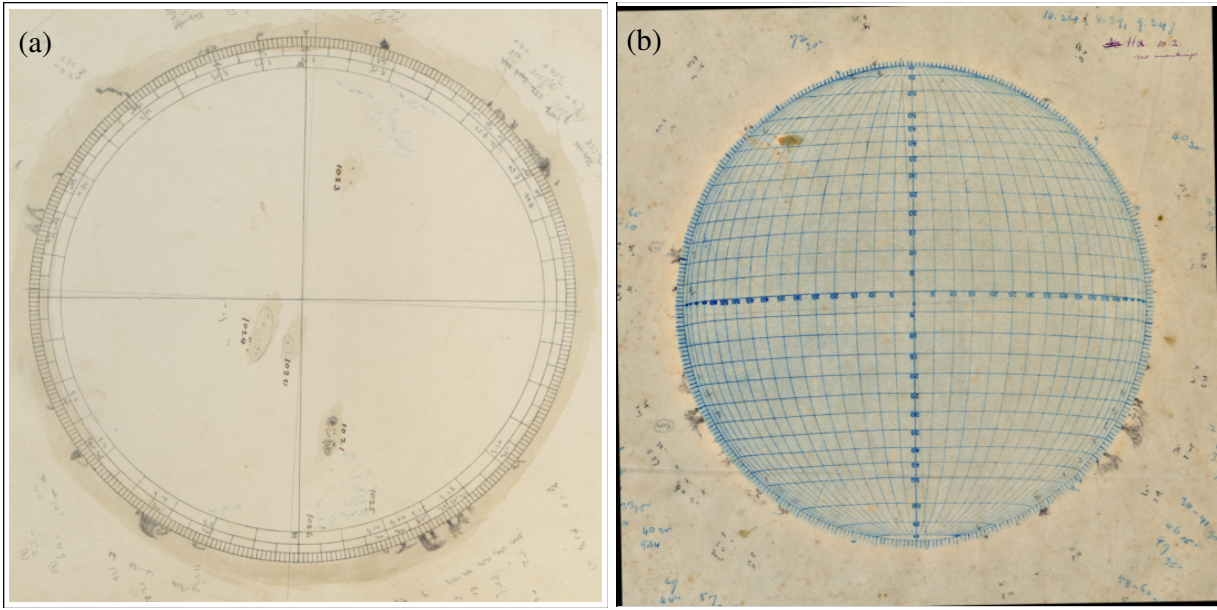


Figure 14. (a) Sunchart observed on 1907 January 1, without the Stonyhurst grid overlay. (b) Sunchart observed on 1912 January 1, featuring a blue-filled Stonyhurst grid for coordinate reference and feature annotation. Please note that suncharts are contrast-enhanced for better visualization.

followed by a nonlinear activation function,

$$O_{Le}^{en} = \text{ReLU}(\text{Conv}(O_{Le-1}^{en})). \quad (\text{A1})$$

The central part of the “U” represents the region (“bottleneck layer”) where the network encodes the most compact and abstract representation of the input features, effectively forming the bridge between the encoder and decoder. Within this bottleneck layer, both the feature dimensionality and spatial resolution are maintained to preserve consistency in feature representation. This compressed latent space is designed to retain only the most relevant information necessary for the accurate reconstruction of the input or the generation of the corresponding segmentation map. The decoder performs learnable upsampling to restore spatial resolution, utilizing transposed convolutions (also known as deconvolutions) with appropriate padding. These operations expand the feature maps back to the input scale. For the decoder layer Ld , the output feature map O_{Ld}^{de} is computed from the preceding representation O_{Ld-1}^{de} via a deconvolution-based upsampling, as specified in Equation (A2)

$$O_{Ld}^{de} = \text{ReLU}(\text{DeConv}(O_{Ld-1}^{de})). \quad (\text{A2})$$

Skip connections constitute the defining mechanism of the U-Net, transmitting encoder feature maps to their decoder counterparts at matching resolutions. This fusion (typically via concatenation) preserves spatial detail and stabilizes gradient flow; the operation is shown in Equation (A3),

$$O_{Ld}^{de} = \text{concat}(O_{Ld}^{de}, O_{nL-Le}^{en}), \quad (\text{A3})$$

where nL is the number of layers in the encoder. If the network has nL levels, the output from the encoder layer $nL - Le$ has the same spatial dimensions as the decoder layer. Concatenating them aligns features pixel-by-pixel. In the U-Net, the output head is a 1×1 convolution that linearly projects the decoder features at each spatial location to the target number

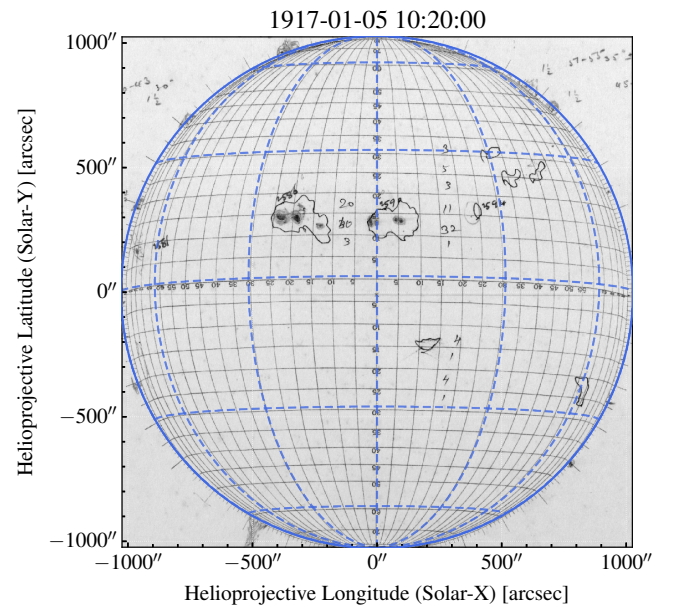


Figure 15. Sunchart observed on 1917 January 5, shown as a SunPy map with the Stonyhurst heliographic grid (blue) overlaid.

of classes, yielding the segmentation output ($O_{\text{final}}(x, y)$) as defined in Equation (A4). This operation preserves spatial resolution while performing channel-wise mixing only; see B. Baştuğ et al. (2024) for further details,

$$O_{\text{final}}(x, y) = \text{sigmoid}(\text{Conv}_{1 \times 1}(O_{Ld}^{de})). \quad (\text{A4})$$

Figure 14 illustrates early twentieth-century KoSO suncharts, showing the solar disk with and without the Stonyhurst grid. Figure 15 shows a SunPy map representation of the 1917 January 5 sunchart with the Stonyhurst heliographic grid overlaid to demonstrate the accuracy of coordinate extraction. Figure 16 presents latitude–time diagrams comparing plage centroid distributions from this work (CNN-based

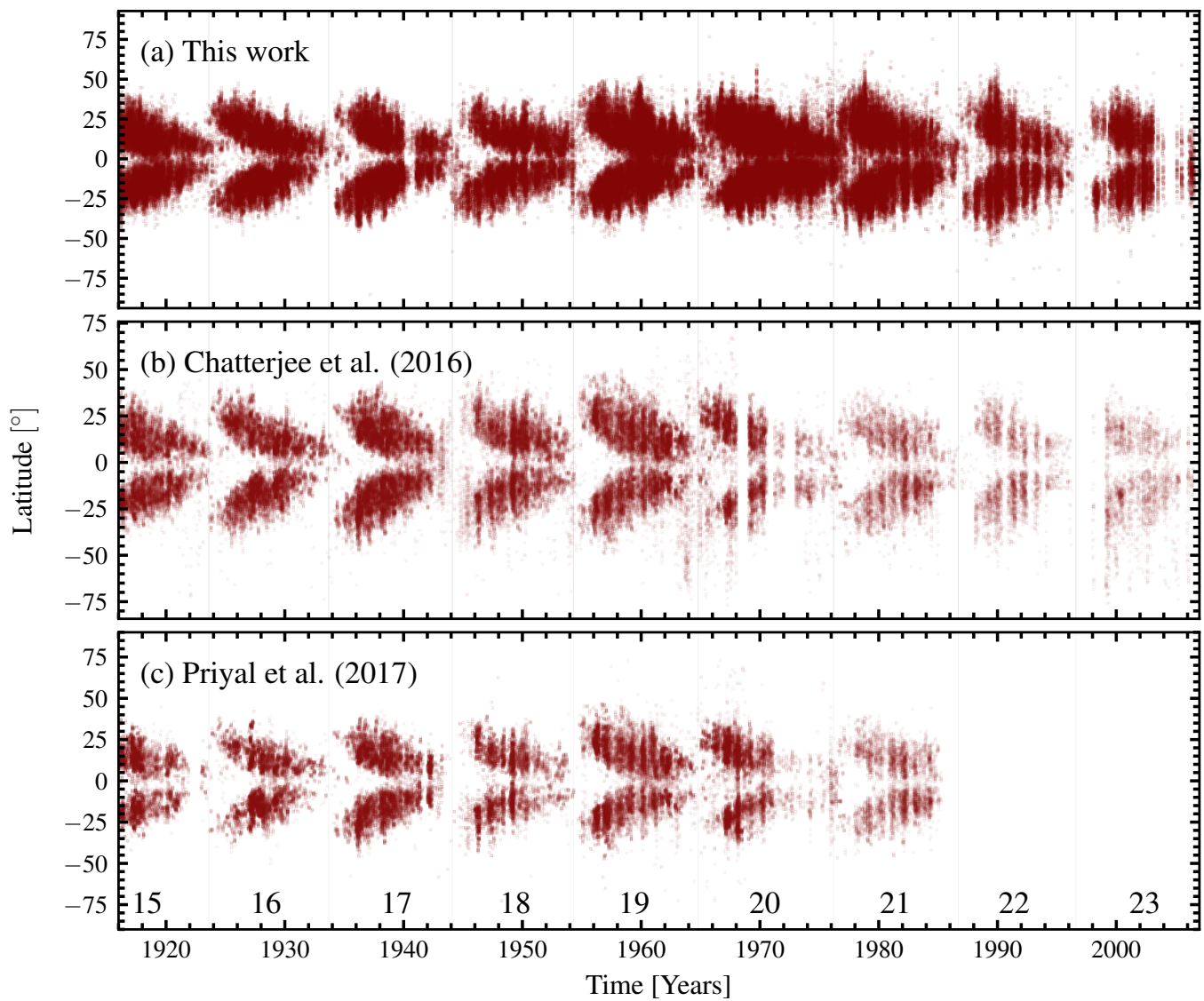


Figure 16. Latitude–time distribution of the centroid positions of individual plage regions derived from (a) this work using CNN-based segmentation on suncharts, (b) generated using plage regions from S. Chatterjee et al. (2016), and (c) M. Priyal et al. (2017) on KoSO Ca II K images.

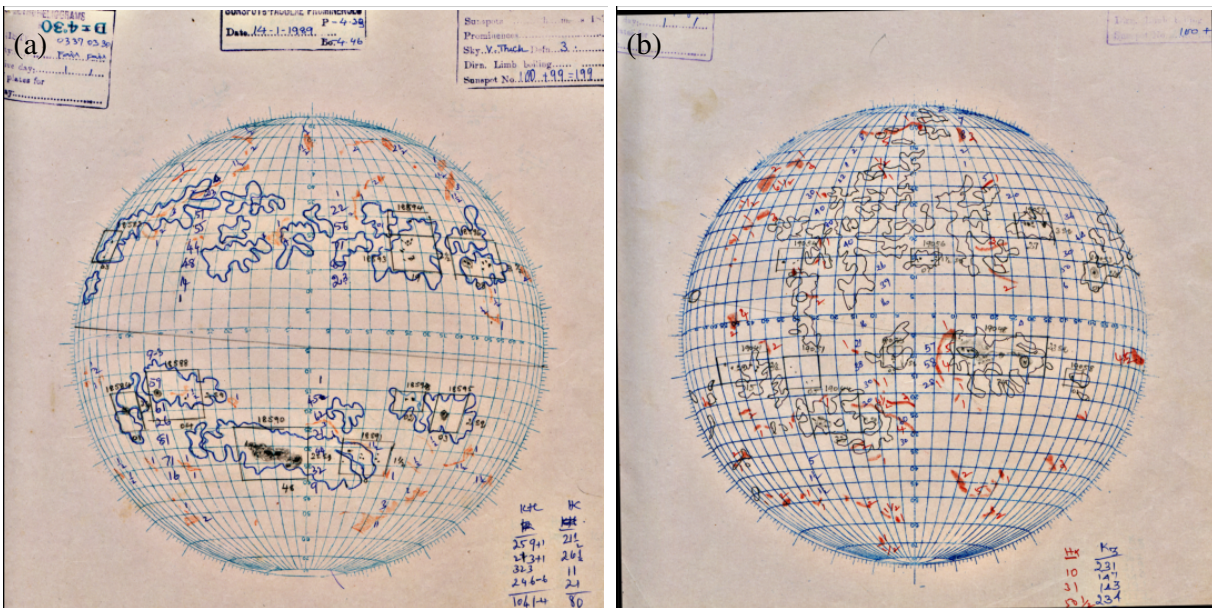




Figure 17. (a) Sunchart observed on 1989 January 14, showing comparatively large-sized plage regions. (b) Sunchart from 1990 January 26, where plages are marked using different colors and are concentrated at relatively higher latitudes. Please note that suncharts are contrast-enhanced for better visualization.

segmentation) with earlier studies, highlighting consistency across datasets and solar cycles. Figure 17 displays suncharts from 1989 and 1990, showing large and high-latitude plages.

ORCID iDs

Dibya Kirti Mishra  <https://orcid.org/0009-0003-1377-0653>
Subhamoy Chatterjee  <https://orcid.org/0000-0002-5014-7022>

Bibhuti Kumar Jha  <https://orcid.org/0000-0003-3191-4625>
Hemapriya Raju  <https://orcid.org/0000-0002-5829-2697>
Aditya Priyadarshi  <https://orcid.org/0000-0003-2476-1536>
Manjunath Hegde  <https://orcid.org/0000-0002-8163-3322>
Srinjana Routh  <https://orcid.org/0009-0008-5834-4590>
Dipankar Banerjee  <https://orcid.org/0000-0003-4653-6823>

References

- Arlt, R., & Vaquero, J. M. 2020, *LRSP*, **17**, 1
Asensio Ramos, A., Cheung, M. C. M., Chifu, I., & Gafeira, R. 2023, *LRSP*, **20**, 4
Barata, T., Carvalho, S., Dorotović, I., et al. 2018, *A&C*, **24**, 70
Baştuğ, B., Güneri, G., Yıldırım, M., Çorbacı, K., & Dandil, K. 2024, *J. Clin. Med.*, **13**, 5893
Canny, J. 1986, *ITPAM*, **PAMI-8**, 679
Carrasco, V. M. S., Aparicio, A. J. P., Chatzistergos, T., et al. 2024, *ApJ*, **968**, 65
Chatterjee, S., Banerjee, D., & Ravindra, B. 2016, *ApJ*, **827**, 87
Chatterjee, S., Hegde, M., Banerjee, D., & Ravindra, B. 2017, *ApJ*, **849**, 44
Chatterjee, S., Muñoz-Jaramillo, A., Dayeh, M. A., Bain, H. M., & Moreland, K. 2023, *ApJS*, **268**, 33
Chatterjee, S., Muñoz-Jaramillo, A., & Lamb, D. A. 2022, *NatAs*, **6**, 796
Chatzistergos, T., Ermolli, I., Banerjee, D., et al. 2023, *A&A*, **680**, A15
Chatzistergos, T., Ermolli, I., Krivova, N. A., et al. 2020, *A&A*, **639**, A88
Chatzistergos, T., Ermolli, I., Solanki, S. K., & Krivova, N. A. 2018, *A&A*, **609**, A92
Chatzistergos, T., Ermolli, I., Solanki, S. K., et al. 2019a, *SoPh*, **294**, 145
Chatzistergos, T., Ermolli, I., Solanki, S. K., et al. 2019b, *A&A*, **626**, A114
Chatzistergos, T., Krivova, N. A., & Ermolli, I. 2022, *FrASS*, **9**, 336
Chola, C., & Benifa, J. V. B. 2022, *GloTP*, **3**, 177
Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. 2016, arXiv:1606.06650
CVAT.ai Corporation 2024, Computer Vision Annotation Tool (CVAT), v2.16.1, Zenodo, doi:10.5281/zenodo.12771595
Das, A. K. 1959, *MNRAS*, **119**, 396
Deshmukh, V., Baskar, S., Berger, T. E., Bradley, E., & Meiss, J. D. 2023, *A&A*, **674**, A159
Diercke, A., Jarolim, R., Kuckein, C., et al. 2024, *A&A*, **686**, A213
Galileo, G. 1613, *Istoria e Dimostrazioni Intorno alle Macchie Solari* (Accademia dei Lincei)
Gharat, S., Bose, B., Borthakur, A., & Mazumder, R. 2023, *RASTI*, **2**, 393
Hale, G. E., Adams, W. S., & Seares, F. H. 1931, *MWOAR*, **3**, 171
Harriot, T. 1613, *Harriot Papers*, Vol. 8: Spots on the Sun, <http://echo.mpiwg-berlin.mpg.de/MPIWG:FAYG83FB>
Hasan, S. S., Mallik, D. C. V., Bagare, S. P., & Rajaguru, S. P. 2010, *ASSP*, **19**, 12
Hayakawa, H., Murata, K., Teague, H., Bechet, S., & Sôma, M. 2024, *ApJL*, **970**, L31
Hayakawa, H., Tamazawa, H., Ebihara, Y., et al. 2017, *PASJ*, **69**, 65
Herr, R. B. 1978, *Sci*, **202**, 1079
Hough, P. V. 1962, Method and Means for Recognizing Complex Patterns, US Patent 3069654, <https://patents.google.com/patent/US3069654A/en>
Jha, B. K., Chatzistergos, T., Banerjee, D., et al. 2024, *SoPh*, **299**, 166
Jha, B. K., Hegde, M., Priyadarshi, A., et al. 2022, *FrASS*, **9**, 1019751
Kingma, D. P., & Ba, J. 2017, arXiv:1412.6980
Li, Q., Wan, M., Zeng, S.-G., Zheng, S., & Deng, L.-H. 2021, *RAA*, **21**, 184
Liu, J., Ji, C., Wang, Y., et al. 2024, *ApJ*, **972**, 187
Mandal, S., Hegde, M., Samanta, T., et al. 2017, *A&A*, **601**, A106
Mausam, E. 1951, *Mausa*, **2**, 85
Mishra, D. K., Jha, B. K., Chatzistergos, T., et al. 2025, *ApJ*, **982**, 78
Mishra, D. K., Routh, S., Jha, B. K., et al. 2024, *ApJ*, **961**, 40
Mordvinov, A. V., Karak, B. B., Banerjee, D., et al. 2020, *ApJL*, **902**, L15
Mumford, S. J., Freij, N., Stansby, D., et al. 2025, sunpy: A Core Package for Solar Physics, v7.0.1, Zenodo, doi:10.5281/zenodo.16638197
Nsiah, R. A., Mantey, S., & Ziggah, Y. Y. 2023, *SCSC*, **1**, 21
Palladino, L., Ntagiou, E., Klug, J., Palacios, J., & Keil, R. 2022, in 2022 IEEE Aerospace Conf. (AERO) (IEEE)
Pevtsov, A. A., Virtanen, I., Mursula, K., Tlatov, A., & Bertello, L. 2016, *A&A*, **585**, A40
Priyadarshi, A., Hegde, M., Jha, B. K., et al. 2023, *ApJ*, **943**, 140
Priyal, M., Singh, J., Belur, R., & Rathina, S. K. 2017, *SoPh*, **292**, 85
Priyal, M., Singh, J., Ravindra, B., Priya, T. G., & Amareswari, K. 2014, *SoPh*, **289**, 137
Raju, H., & Das, S. 2021, *SoPh*, **296**, 134
Ravindra, B., Pichamani, K., Selvendran, R., et al. 2020, *Ap&SS*, **365**, 14
Ravindra, B., Priya, T. G., Amareswari, K., et al. 2013, *A&A*, **550**, A19
Ronneberger, O., Fischer, P., & Brox, T. 2015, arXiv:1505.04597

- Sayez, N., Vleeschouwer, C., Delouille, V., Bechet, S., & Lefèvre, L. 2023, *JGRA*, **128**, e2023JA031548
- Scheiner, C. 1615, Sol Ellipticus: Hoc est Novum et Perpetuum Solis Contrahi Soliti Phaenomenon, Quod Noviter Inventum Strenae Loco (C. Mangius)
- Shin, G., Moon, Y.-J., Park, E., et al. 2020, *ApJL*, **895**, L16
- The SunPy Community, Barnes, W. T., Bobra, M. G., et al. 2020, *ApJ*, **890**, 68
- Upendran, V., Cheung, M. C. M., Hanasoge, S., & Krishnamurthi, G. 2020, *SpWea*, **18**, e02478
- Vaquero, J. M., & Vázquez, M. 2009, The Sun Recorded Through History: Scientific Data Extracted from Historical Documents, 361 (Springer)
- Virtanen, I. O. I., Pevtsov, A. A., Bertello, L., & Mursula, K. 2022, *A&A*, **667**, A168
- Virtanen, I. O. I., Virtanen, I. I., Pevtsov, A. A., et al. 2019, *A&A*, **627**, A11
- Wittmann, A. D., & Xu, Z. T. 1987, *A&AS*, **70**, 83
- Yau, K. K. C., & Stephenson, F. R. 1988, *QJRAS*, **29**, 175
- Yeates, A. R., Bertello, L., Pevtsov, A. A., & Pevtsov, A. A. 2025, *ApJ*, **978**, 147
- Zhu, G., Lin, G., Wang, D., Liu, S., & Yang, X. 2019, *SoPh*, **294**, 117
- Zirin, H. 1974, *IAUS*, **56**, 161