

# X-ray, Optical, and Radio Solar Event Identification using Convolutional Neural Networks: Data Pipeline Overview

Ted Grosson, Cody Meng, Preston Tracy, Jackson White, Yiwen Zhu

Advisor: Chris Tunnell — Rice University

Submitted March 10, 2021

## Introduction

Solar events, such as flares and prominences, can lead to detrimental effects on Earth systems by disrupting electronics and communication systems, particularly in satellites with limited protection from Earth’s magnetic field. In addition, large flares can produce streams of ionized particles intense enough to kill unprotected astronauts in orbiting spacecraft. Due to the potential harm of these events, it would be remarkably useful to identify these events as soon as they occur to allow rapid preparatory actions to mitigate potential damages as much as possible. We apply machine learning techniques to historical high-resolution multi-waveband solar images from NASA’s Solar Dynamics Observatory (SDO) in order to train a set of convolutional neural networks to identify and predict solar events from live imaging of the Sun. These neural networks will intake SDO images collected in a variety of wavelengths and output particular classifications corresponding to the type and presence of x-ray, optical, and radio solar events. The Solar Dynamics Observatory satellite was launched in 2010 and continuously observes the Sun on roughly 10-minute intervals with multiple imaging instruments. In particular, the Atmospheric Imaging Assembly (AIA) aboard SDO images the sun in wavelengths ranging from the UV to optical, providing a constant stream of live, high-resolution solar images[1]. Using the images obtained from AIA, we are attempting to identify x-ray, optical, and radio events and prominences emanating from the Sun as soon as their signals reach the Earth. By analyzing historical AIA data during documented solar events, we may be able to construct a tool that identifies and categorizes solar events as rapidly as possible, with the potential to further develop into a tool that actively predicts these events before they occur and to answer key outstanding scientific questions regarding the nature of solar events and their origin.

The core scientific questions we are attempting to answer encompass a broad area of yet-unclear information of how solar features arise and present themselves. The cause of solar flares and prominences is most often attributed to magnetic fields, but the precise formation mechanism is unknown [2].

Our primary focus is automated identification and classification of x-ray and optical solar events. These types of events, alongside radio bursts, constitute the most common types of solar events, providing us the largest possible dataset to train, validate, and test our convolutional neural network (CNN) classifier. Such a tool may aid in answering scientific questions on the appearance or shared characteristics of these events. For example, should our tool fail to effectively distinguish between types of solar events in their first stages of formation but succeed when the events have fully matured, this would suggest important commonalities between the visible signatures of formation mechanisms of different types of events. This could serve as an indicator that distinct classes of flares and prominences have similar origins with subtle differences between them invisible in the x-ray, ultraviolet, or optical wavebands, or it could suggest that the signatures of early formative stages of events within a particular classification vary so greatly that automated

classifiers are incapable of grouping them together. Both discoveries would be a notable contribution to the field of solar dynamics. We may additionally provide information regarding the distinctions between flares and prominences both over time and independently of time. Should the duration of a solar event be a necessary factor in distinguishing between types of solar events, or should individual snapshots of solar events be sufficient to classify them at certain stages of event evolution, we can study which features are shared, unique, or similar but offset in time between different types of solar events.

We also hope to determine whether we can precisely identify solar radio events from optical or ultraviolet images of the Sun, as it is unclear as to the degree of precision by which we can identify solar events outside of their principal wavelengths. Should some solar events be recognizable in shorter wavelength images, we may be able to use this information to constrain or classify radio events. For example, if one type of radio event produces an identifiable signature in the optical while another does not, we may be able to use our tools to aid in classifying these events, providing additional clues as to the mechanisms by which either type of event can take place.

## Background

Solar flares are highly energetic events of localized increased brightness on the sun over a time period ranging from milliseconds to over an hour. This brightness can be seen across many wavelengths, including x-ray, optical, and radio, and can be seen in Figure 4. For reference, Figure 5 shows the Sun without an active event. Flares tend to be associated with groups of sunspots — localized regions of cooler material and strong magnetic fields — and are often accompanied by the ejection of charged particles. These particles, as well as high-energy electromagnetic radiation, can affect electrical systems and the Earth’s ionosphere, and have the potential to cause major disruptions [2]. The main goal of the Solar Dynamics Observatory is to understand the mechanisms of these events which have the potential of affecting life on Earth [1]. The precise cause of solar flares remains uncertain, but a leading theory is that energy is suddenly released from the strong magnetic field in sunspots through a process called “reconnection”. This occurs when a magnetic field loop “breaks” and “reconnects” in a more stable path, releasing most of the energy stored in the loop, up to around  $10^{25}$  Joules per event [2]. An additional process which appears in our images is granulation. The surface of the sun appears to be dotted with “granules” or cell-like spots which are each separate convection cells gradually rising up and then falling back down within the Sun’s photosphere. The granules tend to span around 700 km and have lifetimes of five to ten minutes [2]. Since we are analyzing how the Sun is changing over time, the granulation pattern could affect our identifications, as seen in Figure 6.

Previous studies which use data from AIA tend to focus on specific details of particular events, with an emphasis on characterizing their dynamics [3][4] or their structure [5]. This is due to the large quantity of high quality images being provided by AIA, which allow detailed studies of solar features. In contrast, our project will focus on the identification of events without worrying about their fine details. Most studies also combine the AIA data with data from the Helioseismic and Magnetic Imager (HMI) instrument on the SDO, which maps out the magnetic field on the solar surface [3][4]. Since solar flares are directly related to changes in the solar magnetic field, HMI data provides a simple method for searching for events. However, we would like to discover whether identification can be achieved through purely visual means, so we are not using data from

HMI.

There are some successful prior efforts to predict the solar flares using machine learning based on changes in the Sun’s magnetic field [6]; however, because solar flare events are associated with changes in the Sun’s magnetic field, it may not be possible to directly predict them through purely visual means. Should this project lead to satisfactory prediction of solar events, it would provide an unprecedented view about the nature of solar flares that the solar flares are correlated to the dynamics of the sun’s surface in a way that we can use only the light from the sun’s surface to detect and predict a solar flare.

## Data Description

### AIA Observations

In this project we will be using two primary datasets: Atmospheric Imaging Assembly (AIA) instrument observations of the sun, from the Solar Dynamics Observatory (SDO), and space weather reports from the National Oceanic and Atmospheric Administration (NOAA). The AIA takes observations at ten different wavelengths, which are summarized in Table 1, and each one probes different layers of the sun. The AIA has been observing the sun since 2010, and observes at 4500 Å once every hour, at 1600 and 1700 Å once every 24 seconds, and at the remaining wavelengths once 12 seconds. The images are stored as 4096x4096 pixel .fits files, where each pixel represents the flux captured by the corresponding pixel on the AIA CCD. The headers of the .fits files also include relevant information such as the exposure time and time of the observation. These files can be queried and downloaded from within python, using the sunpy package. Using sunpy we can select images based on their observation time and filter, which will allow us to easily select and download observations which correspond to events in the space weather reports.

The different wavelengths observed by AIA were chosen to examine specific portions of the Sun’s surface or atmosphere. Each of these wavelengths are centered on a different emission line to examine the features visible at the temperature of the emission line. The EUV wavelengths observe extremely hot temperatures, and are centered on different iron ions, ranging from Fe IX (171 Å) to Fe XXIII (131 Å), and observing temperatures from 60,000 K (304 Å) to 20,000,000 K (193 Å) [7]. The 1700 Å and 4500 Å wavelengths show continuum images of the Sun in the UV and optical, respectively. The 1600 Å wavelength examines the transition region between the chromosphere and the corona. The 304 Å wavelength also examines the transition region, as well as the chromosphere. The 171 Å wavelength shows the “quiet” (low magnetic activity) corona and coronal loops, and the 193 Å examines a hotter region of the corona, as well as the hotter material in solar flares. The 211 Å and 335 Å wavelengths both examine the hot, magnetically active regions of the corona. Finally, the 94 Å and 131 Å wavelengths both examine flaring regions, with wavelengths centered at different temperatures [8]. The temperature coverage provided by these wavelengths allows for a more complete reconstruction of thermal structure than previous missions [9].

## AIA Filter Summary

Wavelength (Å)	Primary Emission Source	Region of Atmosphere	log(T)
4500	Continuum	Photosphere	3.7
1700	Continuum	Photosphere	3.7
1600	C IV + Continuum	Upper Photosphere / Transition Region	4.7
335	Fe XVI	Flaring Regions	6.8
304	He II	Chromosphere / Transition Region	5.8
211	Fe XIV	Active-Region Corona	6.3
193	Fe XII, XXIV	Corona and Hot Flare Plasma	6.1, 7.3
171	Fe IX	Quiet Corona / Upper Transition Region	5.0
131	Fe VIII, XX, XXIII	Flaring Regions	5.6, 7.0, 7.2
94	Fe XVIII	Flaring Regions	6.8

Table 1: The wavelengths observed by the AIA, along with the primary source of emission at each wavelength, and the approximate location and temperature of the portion of the Sun that each wavelength probes. [9]

## NOAA Space Weather Reports

The second dataset, the space weather reports, are created once per day by the NOAA and can contain 13 different types of space weather events, which are listed in Table 2. These reports exist back until 1997, so we have corresponding reports for all AIA data. A sample report from 2015 is depicted in Figure 3. The entire collection of reports amounts to only 18 MB in storage size, so we are able to store all the reports on google drive.

FLAs are flares on the solar surface which peaks in the optical wavelength, it is observed in H-Alpha emissions; XRAs are flare event which peaks in the X-ray wavelength, detected by an increase in X-ray flux at Earth; RBRs are a bursty radio event of narrow-band bursts, corresponds to Type I solar radio bursts; RSPs are A bursty radio event where the radio burst sweeps through a range of radio frequencies over the duration of the event, which corresponds to Type II through V solar radio bursts. BSLs are bright streams of gas emanating from the chromosphere which move out at least 0.15 solar radii above the limb. DSFs are solar filaments which disappear suddenly on a timescale of minutes to hours. EPLs are solar prominences that become activated and are seen to ascend away from the earth, which can be associated with coronal mass ejections. FILs are masses of gas suspended over the chromosphere by magnetic fields, which are threaded over the solar disk. (As opposed to those which are outside the limb and visible against the sky which are called prominences). FORs are abrupt decreases (at least 10%) in the background cosmic ray intensity, which is associated with plasma and magnetic field enhancements in the solar wind. GLEs are sharp increases (again at least 10%) in the background cosmic ray intensity. LPSs are systems of prominences which take the form of loops, and are associated with major flares that bridge the magnetic inversion line on the sun. PCAs are anomalous conditions where the polar ionosphere reflects certain radiowaves at lower altitudes than normal. This typically occurs within hours or days of major flares. RNSs are transient enhancements of solar radio emission, consisting of elevated background emission. SPYs are defined when significant luminous material is ejected

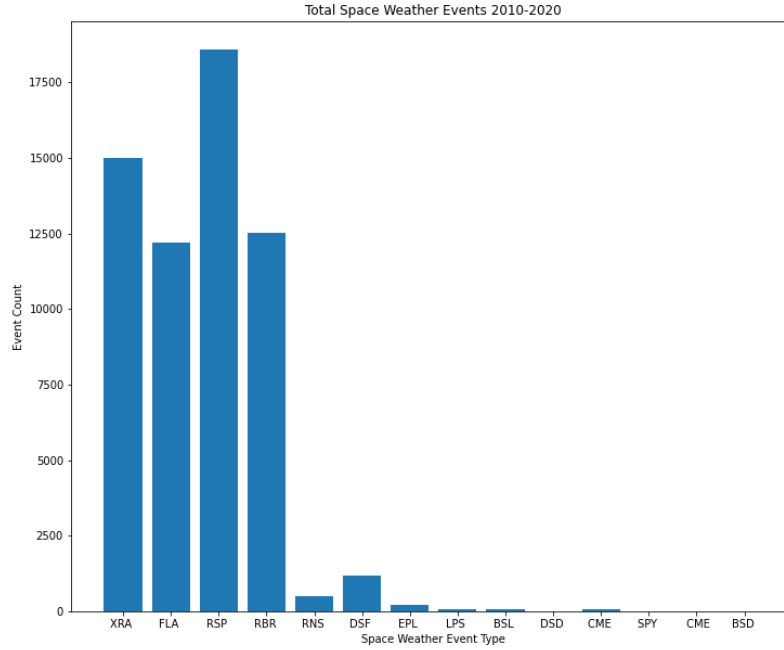


Figure 1: Total number of space weather events observed since 2010, grouped by event type (Event types are listed in table 2)

from a solar flare with sufficient velocity to escape the sun [10].

Acronym	Full Name
BSL	Bright Surge on Limb
DSF	Filament Disappearance
EPL	Eruptive Prominence on Limb
FIL	Filament
FLA	Optical Flare in H-Alpha
FOR	Forbush Decrease (CR Decrease)
GLE	Ground-Level Event (CR Increase)
LPS	Loop Prominence System
PCA	Polar Cap Absorption
RBR	Fixed-Frequency Radio Burst
RNS	Radio Noise Storm
RSP	Sweep-Frequency Radio Burst
SPY	Spray
XRA	X-ray Event

Table 2: List of Event Types tracked in space weather reports

## Exploratory Data Analysis

While exploring the space weather reports we have found 79.6% of the reports since 2010 have at least one observed space weather event, and there are on average 15.0 events per day. There are significant frequency discrepancies between the different types of space weather events. As shown in Figure 1, the most common events are X-Ray events (XRA), optical flares in H-Alpha (FLA), and Sweep-Frequency Radio Bursts (RSP), and Fixed-Frequency Radio Bursts (RBR). Each of these events comprise almost a quarter of the dataset. Because each of these four weather types have thousands of observed events to use in our CNNs, and the remaining types have at most hundreds of observed events, we have decided to focus primarily on these four types of events for this project.

Another key feature that we have identified in the data is the typical size of a visible space weather event. Most flares and prominences are on the order of about 100 pixels in each dimension, as shown in Figure 2. This is good because it should allow us to down sample our images considerably from their original size of 4096x4096 pixels, which would be challenging to run through a CNN.

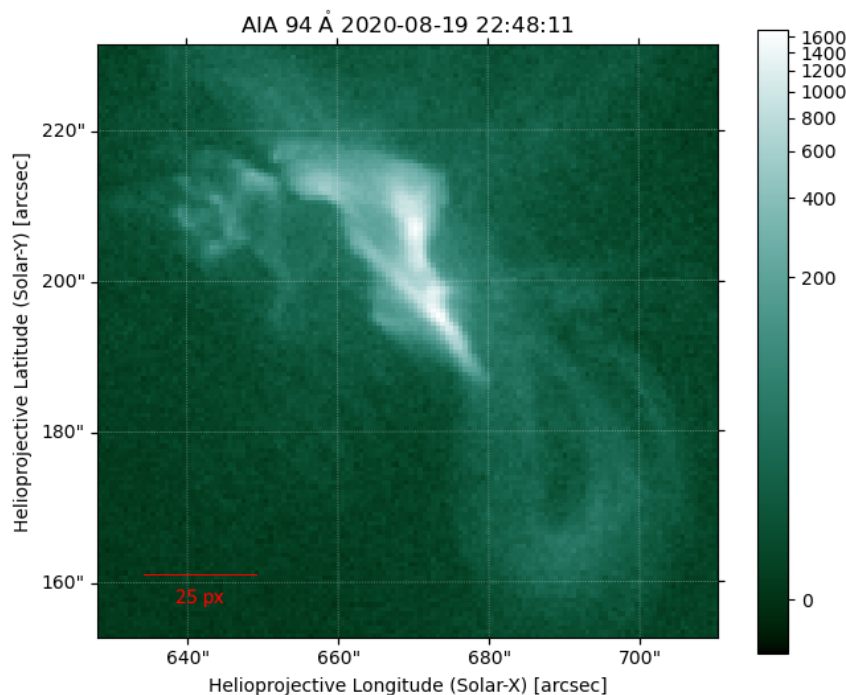


Figure 2: Typical scale for an x-ray flare.

## Project Objectives

Our project objectives fall under one major task with multiple subtasks, and two more possible objectives depending on our success.

First, we will attempt to train a neural network to identify solar events from solar images, using historical SDO AIA solar images in 5-6 wavebands between 90 and 300 Angstroms alongside solar event reports obtained from NOAA's Space Weather Prediction Center. The images and solar event reports will be aligned in time, so that the neural network will learn to associate features within the images with specific types of solar events. We will initially focus on two types of events, x-ray events and optical flares. These events are visible in the wavelengths available to us from AIA, and they are the most frequent events in the dataset, aside from the radio burst events.

After creating a model for identifying x-ray and optical flare events, we will apply this same methodology to radio bursts. Identifying these events with our available wavelengths is a more difficult task, because these events peak at wavelengths far outside our available range. We will attempt to use the longest wavelengths available to us on the AIA, but radio emission is well outside of the low-frequency limit of the accessible bandpasses. However, we will still train a neural network on similar sets of AIA images alongside simultaneous radio burst reports. If radio bursts carry signature features present in higher-frequency wavebands, our neural network may be able to identify them. If we are unable to create a neural network which accurately identifies these radio events, that would also be an indicator that these events do not have significant activity in our range of wavelengths.

## Event Identification Pipeline

### Data Wrangling

In order to generate our testing, training and validation data sets we have written python scripts which utilize the space weather reports and the sunpy python package to download and store AIA observations. These scripts download one observation per event per wavelength, for a specified event type and set of wavelengths, which are chosen at random from all the observations taken over the duration of each event which fit the specified parameters.

Each observation, as a fits file, is 12 MB in size, and we have approximately 60,000 observed events. This means that a full collection of images would require almost a full terabyte of storage space per wavelength. In addition, the large resolution of each image (4096 x 4096 pixels) is unreasonably large to be used in our CNNs. To remedy both of these challenges we have also written a script which iterates over our downloaded images, down samples them, and saves the underlying pixel value array as a compressed .gz file. The down sampling is performed as a meanpool on 8x8 arrays of pixels, reducing the resolution to 512x512 pixels, and reducing the file size to less than 1 MB. The down sampling script is quite fast, but the image downloading script is quite slow and is the rate limiting step in this process. At the moment it takes about 60-70 seconds to download one image through google colab, compared to about 50-60 seconds per image on a local jupyter notebook connected to the Rice wifi, so building up a substantial dataset of thousands of images currently takes considerable time.

## Modeling and Validation

To achieve our goal of identifying solar events from AIA images, we will create a convolutional neural network using Keras, which is available in python and more user friendly than other deep learning libraries such as TensorFlow, which Keras is built on. We will create a CNN within Keras which identifies whether a solar event is present within a given image. After downsampling the observations to 512x512 pixels, the dataset of images can then be input into the neural network as numpy arrays with the number of channels used being equal to the number of wavelengths used. We will test the network subject to different parameters to discover the network design which produces the model best able to accurately identify solar events. We will modify things within the network design such as the learning rate, the batch size, and the number of convolution and pooling layers. We will test the effect of modifying these parameters by comparing the cross entropy between the different network designs.

We will validate the success of our model by setting aside one third of our dataset in each event category for three stages: training, testing, and validation. By partitioning data outside that which is used for forming the CNN, we will be able to avoid validating and testing our models on potentially overfitted data that could misrepresent the actual veracity of our model. For all CNN models, we will examine key parameters such as false positive and false negative rates in order to comprehensively evaluate their successes and shortcomings.

## Communication

Upon completion of our work at the end of the semester, we will present our results in an oral presentation to the course as well as in a comprehensive final report detailing our progress, results, and the potential for future work.

## References

- [1] W. Dean Pesnell, B. J. Thompson, and P. C. Chamberlin. The Solar Dynamics Observatory (SDO). *Solar Physics*, 275(1-2):3–15, January 2012.
- [2] Bradley W. Carroll and Dale A. Ostlie. *An Introduction to Modern Astrophysics*. Cambridge University Press, 2 edition, 2017.
- [3] J. Dai, Q. M. Zhang, Y. N. Su, and H. S. Ji. Transverse oscillation of a coronal loop induced by a flare-related jet. *Astronomy & Astrophysics*, 646:A12, February 2021.
- [4] L. P. Chitta, H. Peter, and S. K. Solanki. Nature of the energy source powering solar coronal loops driven by nanoflares. *Astronomy & Astrophysics*, 615:L9, July 2018.
- [5] Markus J. Aschwanden and Hardi Peter. The Width Distribution of Loops and Strands in the Solar Corona—Are We Hitting Rock Bottom? *The Astrophysical Journal*, 840(1):4, May 2017.
- [6] Abbas Raboonik, Hossein Safari, Nasibe Alipour, and Michael Wheatland. Prediction of solar flares using unique signatures of magnetic field images. *The Astrophysical Journal*, 834, 10 2016.



- [7] J.R. Lemen, A.M. Title, and D.J. et al. Akin. The atmospheric imaging assembly (aia) on the solar dynamics observatory (sdo). *Solar Physics*, 275:17–40, January 2012.
- [8] Holly Zell. How sdo sees the sun. <https://www.nasa.gov/content/goddard/how-sdo-sees-the-sun/>, Feb 2015.
- [9] AIA Science Team. Atmospheric Imaging Assembly for the Solar Dynamics Observatory, Concept Study Report, Appendix A (AIA Science Plan) , Feb. 2004.
- [10] NOAA. Space weather glossary. <https://www.swpc.noaa.gov/content/space-weather-glossary>.

## Appendix

```

:Product: 20150617events.txt
:Created: 2015 Jun 18 2127 UT
:Date: 2015 06 17
# Prepared by the U.S. Dept. of Commerce, NOAA, Space Weather Prediction Center
# Please send comments and suggestions to SWPC.Webmaster@noaa.gov
#
# Missing data: ////
# Updated every 5 minutes.
#
# Edited Events for 2015 Jun 17
#
#Event      Begin      Max      End Obs  Q  Type  Loc/Frq  Particulars      Reg#
#-----
3910        0040      ////      0041 PAL  C   RSP  025-075  III/1
4060        A0116      ////      A1226 HOL  3   DSF  N39W23   7
3930        0120      0120      0120 LEA  G   RBR  245      110
3940        0205      ////      0205 PAL  C   RSP  075-180  III/1
3950        0207      0207      0207 LEA  U   RBR  245      240
3960        0546      0606      0635 G15  5   XRA  1-8A     C1.6    4.5E-03  2367
3960 +      B0548      U0558      A0653 SVI  2   FLA  S19E04   1F      ERU      2367
4280        0654      ////      0739 LEA  U   RSP  025-037  VI/1
3970        0706      0706      0706 LEA  G   RBR  245      120

```

Figure 3: Sample Space Weather Report text file From June 6, 2015. Along with event types and times, the reports also contain information about the event strengths and locations.

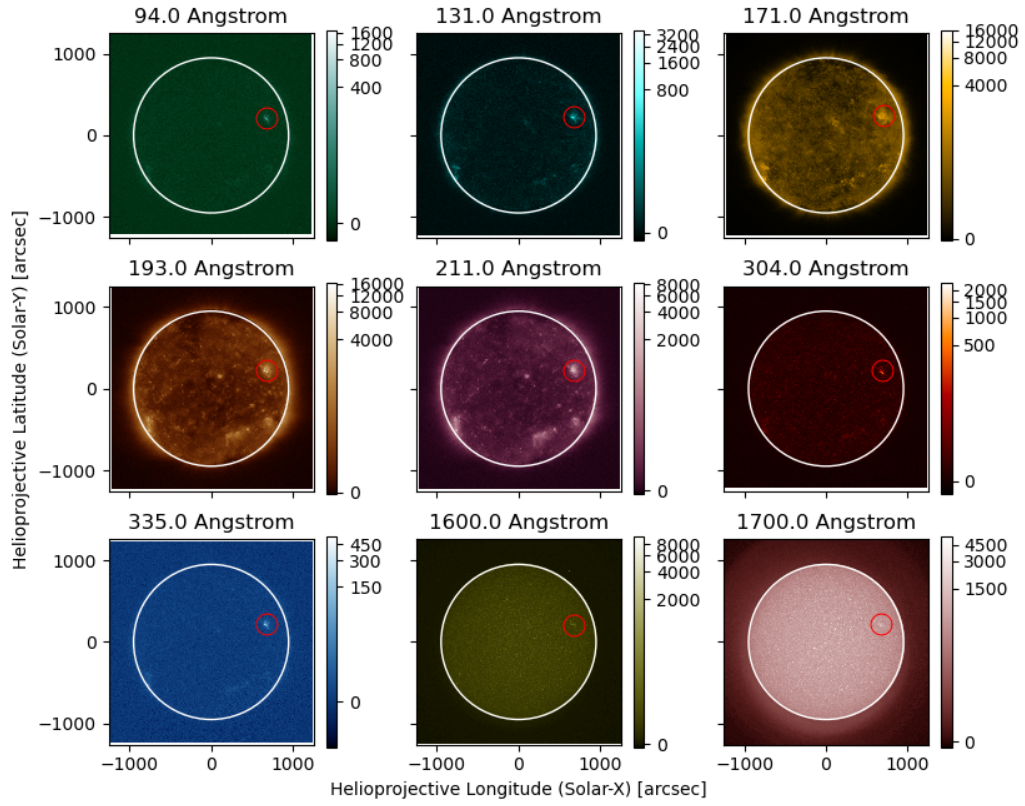


Figure 4: A small x-ray flare which occurred August 19, 2020 in each AIA wavelength apart from 4500 Å. The flare, circled in red on each image, shows up in all wavelengths to varying degrees.

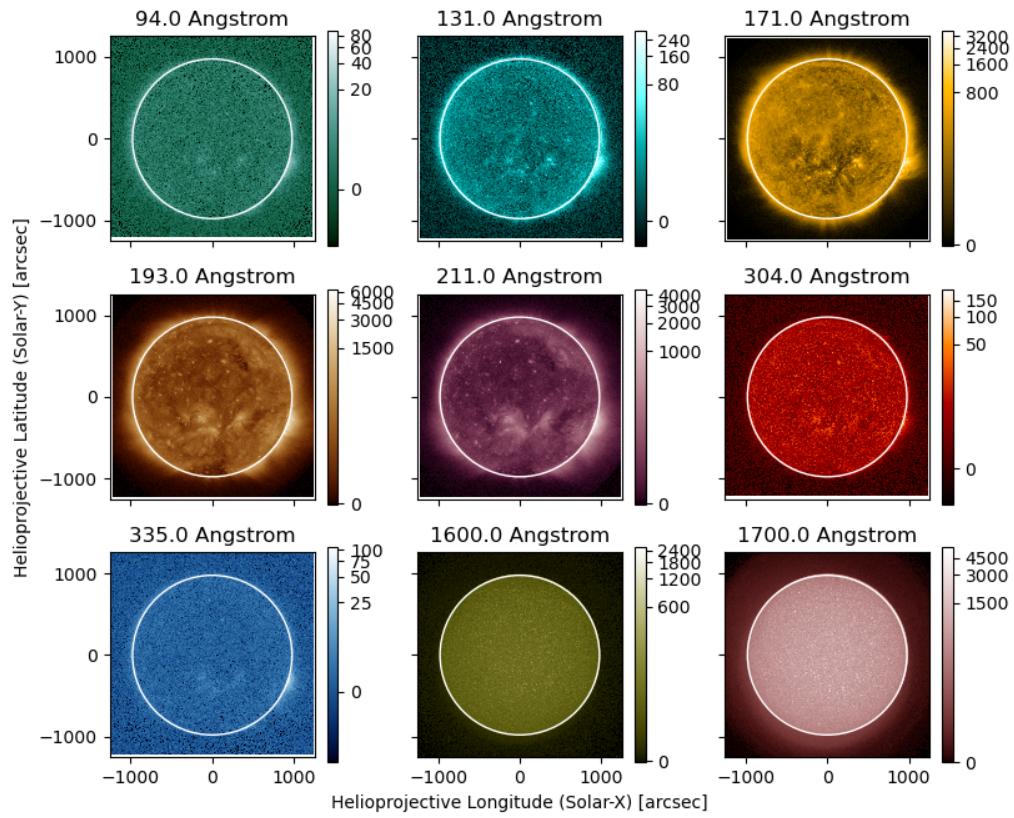


Figure 5: The Sun in each wavelength without an active flare. Note the much smaller colorbar scales than in the image above.

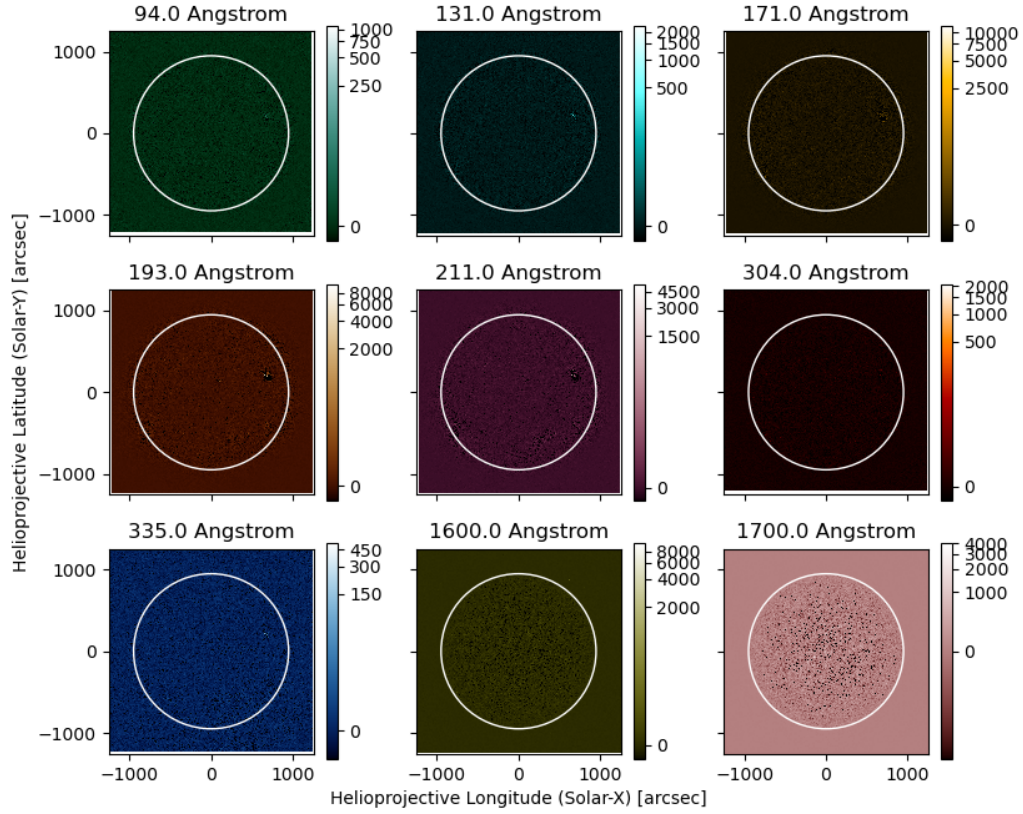


Figure 6: The same flare as in Figure 4 above, after subtracting a previous image from the current one. This process is sometimes called difference imaging, and is useful in seeing how an object changes over time, such as the appearance of a flare. These difference images were taken over a span of two minutes, but varying this time could optimize our detection process. The changing granulation pattern is especially apparent in the 1700 Å image, where the flare is completely drowned out.