

Solar Thematic Map Generation via Machine Learning

by
J. Marcus Hughes

Professor Jon Park, Advisor

A thesis submitted in partial fulfillment
of the requirements for the
Degree of Bachelor of Arts with Honors
in Computer Science

Williams College
Williamstown, Massachusetts

May 2, 2018

Contents

1	Introduction	10
1.1	Motivation	10
1.2	Goals	11
1.3	Organization	12
2	Solar Physics Background	13
2.1	Space Weather Phenomenon	13
2.2	Solar Structures	14
3	Related work	16
3.1	Synoptic Charts and Thematic Maps	16
3.2	Data sources	17
3.2.1	Raw imagery to classify	17
3.2.2	Labeled data for training	18
3.3	Solar machine learning	18
3.3.1	Unsupervised solar segmentation	18
3.3.2	Supervised solar segmentation	23
3.4	Earth Remote Sensing	28
3.5	General Computer Vision	29
3.5.1	Fully convolutional neural networks	29
3.5.2	Mask R-CNN	30
3.6	Summary	30
4	Data	32
4.1	Solar Imagery	32
4.1.1	Solar Ultraviolet Imager (SUVI) Data	32
4.1.2	Noise-gater procedures	33
4.1.3	H-alpha imagery	36
4.2	Labeled Imagery	36
4.2.1	Annotating Software	38
4.2.2	Analysis of labeled data	41
4.2.3	Case Study	45
4.2.4	Labeled database	46
5	Classifiers	48
5.1	Naive Bayesian Maximum Likelihood	48
5.2	Random Forest	49
5.3	Feed-forward Neural Network	50
5.4	Convolutional Neural Network	51

CONTENTS	3
6 Software package	52
6.1 Overview	52
6.2 Data structuring	52
6.3 Annotation GUI	54
6.4 Trainers	54
6.5 Classifiers	55
6.6 Filters	55
6.7 Evaluator	56
7 Experiments and Results	57
7.1 Evaluation approaches	57
7.1.1 Qualitative	57
7.1.2 Standard machine learning metrics	57
7.1.3 Confusion matrix	58
7.2 Baseline results	59
7.3 Normalization of data	60
7.4 Line ratios	60
7.5 Spatial features	60
7.6 H α inclusion	61
7.7 Hyperparameter selection	61
7.7.1 Random forest	61
7.7.2 Neural network	61
7.8 Random forest feature importance	61
7.9 Comparison to existing classifiers	61
8 Applications	67
8.1 Database building	67
8.2 Fractal dimension and class properties	67
9 Conclusion	68
9.1 Future work	68
9.1.1 Stability over solar cycles	68
9.1.2 Converting digitized synoptic charts	68
9.1.3 Prediction	69
9.2 Summary	70

List of Figures

1.1	Spatial extent of Carrington Event: As noted by Cliver and Svalgaard (2004) the Carrington event was observed to very low latitudes in the Americas. Closed circles represent overhead aurora; open circles represent visible aurora. The heavy curved line denotes the geomagnetic equator and the \oplus symbol indicates the anti-Sun point. The lowest geomagnetic latitude at which the storm was observed was Honolulu (not shown).	11
2.1	Parts of the Sun: This image indicates many of the parts of the Sun. The top cutaway portion shows the internal structure of the Sun, not discussed in this thesis. The bottom portion shows many of the features discussed here which comprise the solar atmosphere. The solar corona is the uppermost portion of the Sun which is millions of degrees Celsius. From this and the cooler and lower photosphere and chromosphere, the space weather emanates. Several different features discussed in Section 2.2 are shown, for example prominences, flares, and coronal holes [94].	14
2.2	Parker Spiral: The Parker Spiral explains how events on the Sun propagate outward. Events on one region of the Sun have a direct line along the Parker Spiral toward Earth while events on another portion will miss Earth [89]	15
3.1	Synoptic map example: Song et al. (2015) propose the solar synoptic map to include a labeled composite image with different wavelength images accompanying to provide a full image of solar activity [84].	17
3.2	Histogram segmentation: Olmedo et al.(2008) utilized histogram segmentation to identify coronal mass ejections. “The intensity profile along the angular axis showing the 1D projection of the CME image. Only positive pixels along the radial axis are used. This profile effectively indicates the angular positions of a CME when it is present.” [66]	20
3.3	First half of SDO classifiers [60]	22
3.4	Second half of SDO classifiers [60]	23
3.5	Performance of different methods: Revathy, Lekshmi, & Nayar (2005) compared the performance of different segmentation techniques in identifying active regions [74]. At left is the result of a histogram thresholding approach while at the right is fuzzy-based segmentation.	24
3.6	de Wit Segmentation: This figure indicates the power of quick, multi-spectral, supervised segmentation done by de Wit (2006), very similar to other work by Rigler et al.(2012) and De Visscher et al.(2015) [20]. The classes for this study are: “(1) Tenuous corona outside of the disk, in regions with open magnetic-field lines.(2) Dense corona outside of the disk. (3) Coronal holes. (4) Quiet sun, including the chromospheric network and regions inside the network boundaries. (5) Active regions on the disk” [20].	26

3.7	Fully convolutional neural network: This is an example architecture for a fully convolutional neural network (figure 1 from Shelhamer, Long, & Darrell (2016) [58]).	29
3.8	Mask regional convolutional neural networks: The top row was the existing state-of-the-art instance segmentation [57], an example of a FCN, compared to the bottom row of Mask R-CNN performance on the same scene. The overlaid coloration indicates the segmentation while bounding boxes indicate where Mask R-CNN evaluated these masks. Clearly, the Mask R-CNN produces more coherence classifications. In addition, it runs in less time with higher accuracy than the existing state-of-the-art systems and was consequently awarded the 2017 International Conference on Computer Vision best paper award.	30
4.1	Wavelengths: This diagram indicates which S UVI wavelengths are most helpful in identifying different space weather events [31].	33
4.2	Example S UVI images: These first-light images taken on January 29, 2017 exemplify S UVI’s response to different solar features in different passbands. For example, the large coronal hole (the darker boot-like shape in the lower portion of the Sun) has higher contrast from the background in 19.5 nm than 304 nm [31].	34
4.3	Effectiveness of noise-gating The upper left is a good image, no cleaning necessary. However, images like the upper right, dominated by shot noise, are typical for the 94 angstrom channel. This image is created by taking the image on the upper left and adding Poisson noise with a signal-to-noise ratio of 2. The algorithm still performs even if it is worse, although some artifacts appear. DeForest’s algorithm was applied to create the cleaned image on the bottom left. This can be compared to simply smoothing the image to decrease the noise as in the bottom right, a typical alternative procedure.	35
4.4	Example Hα image: This image is a typical H α view of the Sun from the GONG network. A prominence is visible on the bottom right, and a dark filament is visible in the center of the center of the Sun.	37
4.5	Startup of annotation software: The region on the left shows a preview image of the Sun. Initially, it is configured to show a three-color composite of the Sun, with channels scaled to show most features. The right region shows the labeling of the Sun.	40
4.6	Mistake for annotator Left and right are annotations for Group 16, at center, by Annotator A. As you can see the annotator appears to have forgotten to label bright regions in one case. Such errors could be identified manually and cleaned from the labeled set.	41
4.7	Typical inconsistency for annotator Left and right are annotations for Group 17, at center, by Annotator A. In this case, the annotator makes a typical kind of disagreement in the extent of boundaries. It is unclear which boundary is correct so both must be included in the labeled set.	42
4.8	Structured outer space discrepancy At left is the 30.4 nm image for Group 6. Very faintly in the upper left, some structure is visible in outer space as a CME moves outward. This is only seen in the 30.4 passband. Thus, if Annotator A had chosen to label this image they may or may not have missed this class.	45
4.9	Inter-Annotator disagreement Group 1 provides a key example for disagreement among annotators. Annotator A is shown at left while Annotator B is shown at right. In the center, is the three-color image of the Sun at that time. Annotator A identified smaller coronal holes than Annotator B and included a filament Annotator B did not. This filament is not seen in H α but does look rather strong in the EUV three-color. In general, they identifies the same features but with different boundaries.	46

4.10 An interesting case study from Group 13 For this group, Annotator A and Annotator B submitted double labeling and Annotator C also submitted a labeling. This group comes from 2017-09-08 during a solar flare. All labelings include the annotated flare in the bottom right of the image. However, there are many other differences discussed in Section 4.2.3.	47
5.1 An example random forest construction: This helpful graphic shows the general layout of how a random forest works. Multiple decision trees are evaluated to determine a majority class vote [65]	50
5.2 An example neural network construction: This helpful graphic shows the general layout of a feedforward neural network. The number of nodes and hidden layers may vary. The size of the input layer is determined by the input features, and the output size is determined by the classification task [18].	51
6.1 Code layout: A outline of the major classes in each module of <code>smachy</code>	53
7.1 An example confusion matrix: A confusion matrix is a method of showing label classification. On the vertical axis the correct label is shown while on the horizontal axis the annotator's label is shown. Then, looking at the cell defined shows the percentage of labels in that region. Ideally, the diagonal should be 100 indicating every feature was labeled with the correct label always. The confusion matrices presented in this thesis are normalized.	59
7.2 Baseline performance for Bayesian classifier	62
7.3 Baseline performance for random forest	63
7.4 Baseline performance for neural network	64
7.5 Degeneracy of filament and coronal hole spectrally with EUV images This histogram compares 19.5 nm pixel values for coronal holes, top, and filaments, bottom. They nearly overlap completely making it difficult for classification methods to key on sufficient themes for classification. All other EUV channels suffered from complete overlap. H α images differentiate filaments and coronal holes as shown in Figure 7.6. . .	65
7.6 Coronal hole and filament in Hα In H α images, filaments are dark strand structures while coronal holes have no apparent signature. This defining difference complements the degeneracy in EUV images and boosts classifier performance.	65
7.7 Hα inclusion experiment results. add relevant caption based on results . . .	66
9.1 Daily synoptic drawing Here are four synoptic charts drawn at NOAA ranging from 1974 in the upper left to 2018 in the bottom right. Even more recent synoptic charts vary in style because different forecasters have slightly different styles.	69

List of Tables

2.1	Solar phenomena: This is a short description of some of the solar events related to space weather.	15
4.1	List of event labels for HEK [42].	37
4.2	List of event labels for curated data gathered in this study.	38
4.3	Times for images used in the labeling image set with the number of seconds between the first and last image in the grouping.	39
4.4	Interpretation of Cohen's kappa coefficient for agreement [63]	42
4.5	Interpretation of Fleiss's kappa coefficient for agreement	43
4.6	Annotator A's consistency: This table details Annotator A's labelings for each class. The rows indicate one session and the columns another. Thus, column CH and row FIL with a value of 6400 indicate that in one session Annotator A labeled 6400 pixels as coronal hole and in the other session as filament. The classes have been abbreviated as follows: bright region is BR, coronal hole is CH, empty outer space is EOS, filament is FIL, flare is FLA, limb is limb, prominence is PR, quiet Sun is QS, structured outer space is SOS, and unlabeled is UL. As seen here, Annotator A did not use the structured outer space label and did leave some pixels unlabeled.	43
4.7	Annotator B's consistency: This table details Annotator B's labelings in the same format as Table 4.6.	44
4.8	Annotator consistency: For each annotator and image, Cohen's kappa for each class is shown. Annotator B is more consistent for each class. The overall agreement is skewed by the large number of outer space pixels. Annotator A has no agreement score for structured outer space because the label was disregarded.	44
4.9	Inter-annotator agreement: For each class the agreement between annotators on what pixels are declared that class is shown. Annotator A did not use the structured outer space class. Coronal hole is really the only non-automatic class that is reliable.	45
7.1	Baseline performance: When given only a single pixel view and EUV channels, here is the performance for each classifier..	60
7.2	Normalization coefficients for each channel	60

Abstract

The new Solar Ultraviolet Imager (SUVI) instruments aboard NOAA's GOES-R series satellites collect continuous, high-quality imagery of the Sun in six wavelengths. SUVI imagers produce at least one image every 10 seconds, or 8,640 images per day, considerably more data than observers can digest in real time. Over the projected 20-year lifetime of the four GOES-R series spacecraft, SUVI will provide critical imagery for space weather forecasters and produce an extensive but unwieldy archive. In order to condense the database into a dynamic and searchable form we have developed solar thematic maps, maps of the Sun with key features, such as coronal holes, flares, bright regions, quiet corona, and filaments, identified. Thematic maps will be used in NOAA's Space Weather Prediction Center to improve forecaster response time to solar events and generate several derivative products. Likewise, scientists use thematic maps to find observations of interest more easily and guide experiments.

Using a naive Bayesian classifier, a random forest, and a neural network, we create thematic maps in real-time. We created software to collect expert classifications of solar features based on SUVI images. Using this software, we compiled a database of expert classifications. These are used to establish experimental limits on human performance. Given new images, the classifiers assign each pixel the most appropriate label according to the trained distribution. Here we describe the software to collect expert training and the successes and limitations of the classifier. These results are promising and encourage future research into an ensemble classification approach and solar weather prediction with similar methods.

Acknowledgments

I'd like to thank Dan Seaton and Jon Darnel for guiding me during this project and supplying data and annotations. I'd like to thank Jay Pasachoff for his support in solar physics and discovery of this project. And, I'd like to thank Jon Park for helping me organize and embark on an independent thesis. I'd like to thank all the aforementioned individuals again for reading the thesis and providing feedback as well as the rest of the Williams College computer science faculty.

Chapter 1

Introduction

1.1 Motivation

From August 28th to September 5th, 1859, the night sky nearly all over the world blazed with auroral displays for hours. Sources reported, “there was another display of the Aurora last night so brilliant that at about one o'clock ordinary print could be read by light” (The New York Times, New York Herald, Washington Daily National Intelligencer, September 2, 3, 5, 1859) [30]. Normally, aurora, a visible consequence of material streaming (because of the substorms caused) in from the Sun and interacting Earth’s magnetic field, are confined to polar regions. However, there were observations in New Orleans and even as far south as Honolulu, shown in Figure 1.1 [17]. These displays were symptomatic of a massive solar event impacting Earth’s magnetic fields. Consequently, telegraphs ceased to work, and human operators reported burns and other injuries as small fires started [30]. Events like this, while rare, are fairly periodic, with a 12% probability of another within the decade [76].

On March 13th, 1989, another geomagnetic storm, less than a third of the strength of the 1859 Carrington event, disrupted power in Canada and the United States, resulting in blackouts for the majority of Quebec for nine hours; the cost for repair was \$6 billion [49]. A much smaller event occurred in 2003, known now as the Halloween event [64, 91]. In preparation, satellites were placed into a stable stand-by mode. Ultimately, only one satellite was damaged with most satellites unscathed and exceeding their nominal lifespan. An extreme Carrington event was observed pointed away from Earth in 2012 [25]. It is difficult to quantify the total cost; estimates of the damage a modern Carrington event would cause range from \$140 billion to \$3.4 trillion [25].

There are many much more frequent but less devastating solar events, discussed in Section 2.1 that can cause damage to power grids, satellites, Earth communications, astronauts, and many other sensitive systems. While they cannot be prevented, advanced warning allows for preparation that can mitigate the damage. Coronal mass ejections, one danger resulting when the Sun spews large amounts of charged material sometimes towards Earth, have been recorded to reach speeds of up to 2000 miles per hour, reaching Earth within the day [26]. Within minutes to hours of a solar flare, the ionospheric disturbance can interrupt radio communications [85]. For proper safety

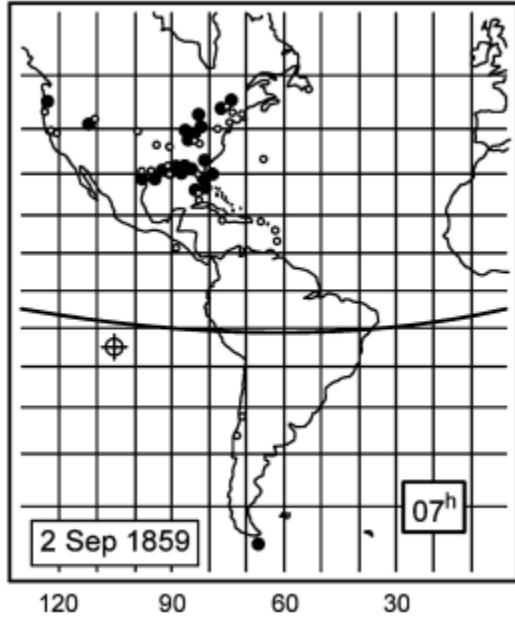


Figure 1.1: Spatial extent of Carrington Event: As noted by Cliver and Svalgaard (2004) the Carrington event was observed to very low latitudes in the Americas. Closed circles represent overhead aurora; open circles represent visible aurora. The heavy curved line denotes the geomagnetic equator and the \oplus symbol indicates the anti-Sun point. The lowest geomagnetic latitude at which the storm was observed was Honolulu (not shown).

protocols to be enacted, fast warning of an event is necessary. This thesis explores solar feature classification through modern machine learning approaches, specifically the creation of a computer vision systems utilizing satellite imagery to identify solar activity in real time for quick response. These classification systems allow for real-time warning of space weather events.

1.2 Goals

This thesis has several key goals as outlined here:

- There does not exist a curated database of human annotated solar images designed for solar machine learning representing multiple experts input. This thesis will present the first of its kind, comparing it to existing automated databases. In addition, this thesis will analyze the human labeling to understand agreement/disagreement between different annotators and consistency for each annotator. In order to create this database, the necessary labeling software will be created in this thesis.
- This thesis will provide a suite of modern machine learning approaches to solar image classification with a random forest, Bayesian, and neural network implementation. These will be compared to existing solar classification approaches as well as each other to characterize their

strengths, weaknesses, and overall performance.

- Ultimately, high-quality solar image classification opens up avenues of research for solar physics. Thus, a prototype solar feature database will be compiled from the images, indexing the images and allowing solar physicists to easily find interesting events. Further, a research application estimating the fractal dimension of active regions and properties of coronal hole will be presented.

1.3 Organization

Chapter 2 introduces the background information for solar physics. This includes a definition of the relevant solar structures to classify. Then, Chapter 3 documents prior work in solar image classification including both the unsupervised systems in Section 3.3.1 and supervised systems in Section 3.3.2. The author's original contributions begin in Chapter 4, after a description of the solar imagery, with an overview and analysis of the human annotated images. Chapter 5 details the classifiers tested in this approach through experiments in Chapter 7. Two applications of the solar classification are explored in Chapter 8: the labeled solar database in Section 8.1 and fractal dimension estimation for solar features in Section 8.2. Finally, Chapter 9 outlines the results of the entire project and potential future work.

Chapter 2

Solar Physics Background

As mentioned, space weather has dangerous and expensive consequences including harm to astronauts and satellites, destruction of power grids, and routine rerouting of intercontinental flights over the north pole. These different effects stem from a variety of solar events. These are outlined in this section with a description of the structure of the varying phenomena.

2.1 Space Weather Phenomenon

The Sun is responsible for many phenomena in daily life providing basic fundamentals such as heat and light to sustain life on Earth. In 1733, Jean Jacque d'Ortous de Mairan proposed an explanation for aurora: it is from solar gas entering Earth's atmosphere [52, p. 51]. Later in the 20th century, Birkeland argued that it is from Earth's magnetic field focusing electrons around Earth's poles [52, p. 51]. Today, the Sun's magnetic field is a recognized as the driver of many of its internal and visible properties: the cycle of sunspots and the coronal temperature. In 1958, Eugene Parker advanced a model for the Sun's impact on Earth via the solar wind, the pressure caused from outflowing solar material, with a theoretical model where the wind propagates as a spiral into interplanetary space from the Sun's rotation and radial flow [52, p. 62]. This is now referred to as the Parker Spiral in Figure 2.2.

The solar wind continuously is blowing millions of tons of material away from the Sun as a stream of rarefied plasma. When it reaches Earth's orbit, it is traveling at roughly 600 km/s with 10 million particles per cubic meter [52, p. 67]. As the Sun goes through an 11-year cycle, this stream of material shifts from stronger to weaker and back again. During the most energetic portion of the cycle, there are frequent sudden and rapid outbursts called solar flares. These release 10^{32} ergs [6], the equivalent to million of 100-megaton bombs exploding at the same time over less than ten minutes [52, p. 89]. These solar flares hurl high energy particles along the Parker Spiral into interplanetary space at nearly the speed of light. The strongest of these flares can cause "planet-wide radio blackouts and longlasting radiation storms" [52, p. 255].

Flares are often accompanied by coronal mass ejections (CMEs), although CMEs and flares can occur independently of each other. These eruptions send roughly 10 billion tons of solar material into

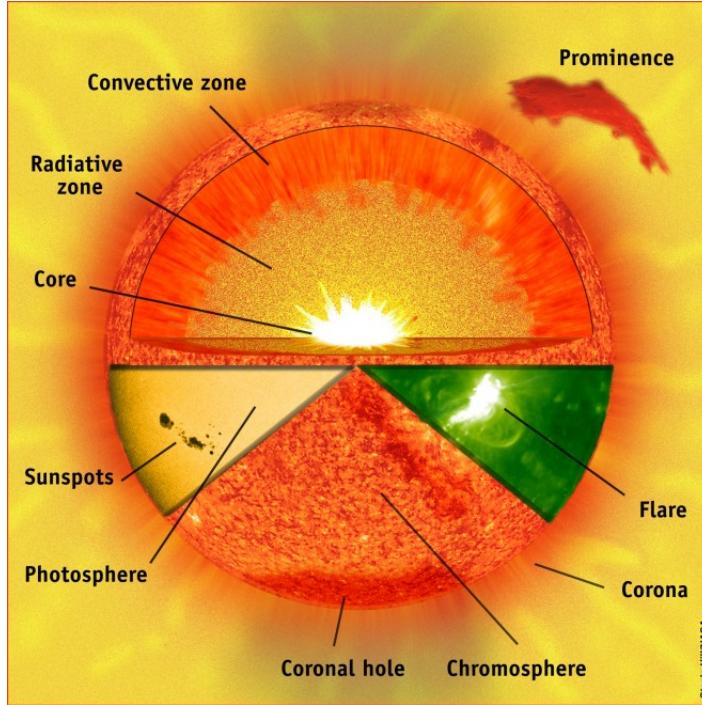


Figure 2.1: Parts of the Sun: This image indicates many of the parts of the Sun. The top cutaway portion shows the internal structure of the Sun, not discussed in this thesis. The bottom portion shows many of the features discussed here which comprise the solar atmosphere. The solar corona is the uppermost portion of the Sun which is millions of degrees Celsius. From this and the cooler and lower photosphere and chromosphere, the space weather emanates. Several different features discussed in Section 2.2 are shown, for example prominences, flares, and coronal holes [94].

interplanetary space. These have been reported to reach speeds in excess of 2000 miles per hour, reaching Earth within the day [26]. Smaller outbursts called filament eruptions can also release material towards Earth [52, p. 285]. Beyond these eruptions, there is variable strength solar wind constantly baraging Earth. Wind from coronal holes, discussed shortly, is faster and can be more of a constant threat.

2.2 Solar Structures

Figure 2.1 showed some of the different parts of the Sun. In Table 2.1 there is a listing of some of the more prominent solar features with a brief description of them.

Active regions will often be called bright regions in this work. Bright regions is a broader term used in space weather forecasting while active regions is the more technical description.

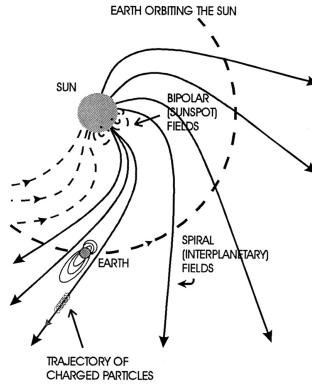


Figure 2.2: **Parker Spiral:** The Parker Spiral explains how events on the Sun propagate outward. Events on one region of the Sun have a direct line along the Parker Spiral toward Earth while events on another portion will miss Earth [89]

Name	Description
Active regions	Complexes of brighter and darker regions in UV observations caused by the solar magnetic field piercing through the solar atmosphere
Filaments	A suspension of cool material high in the solar atmosphere by the magnetic field
Prominences	A filament observed off the disk of the Sun
Coronal holes	Regions that appear dark in EUV observations where the magnetic field is open with no clear reconnection back into the Sun which allow fast outflow of material
Flares	A sudden brightening on the Sun
Coronal mass ejection	An event when the Sun dispels mass and charged particles, sometimes towards Earth
Sigmoids	S-shaped structures, a special subclass of filaments, on the surface of the Sun thought to be precursors to flares because of their instability from high shear
Quiet corona	Parts of the Sun devoid of strong magnetic fields

Table 2.1: Solar phenomena: This is a short description of some of the solar events related to space weather.

Chapter 3

Related work

This chapter surveys some of the underlying work on solar image classification.

3.1 Synoptic Charts and Thematic Maps

When classifying space weather phenomena, it is important to understand its context: where it is on the solar disk, when it developed, and its strength. The Sun has a consistent radial outflow of material at approximately 400 km/s except for regions of fast 700 km/s wind largely caused by coronal holes, as it rotates on its axis every 27.5 days. Unlike Earth's rotation, the Sun undergoes differential rotation with a slower 25 day rotational period at the equator. Differential rotation tangles magnetic field lines that can lead in sudden reconnection and release of energy. The rotation causes magentic field lines to propogate outward from the Sun in a spiral pattern called the "Parker Spiral." As such, solar wind from coronal holes or other energetic releases on the western side of the Sun will reach Earth quicker than from the eastern side of the Sun. Thus, knowing the both the location of the event and the time of the event is very important when considering when the impacts will be felt on Earth. A detailed understanding of the strength of the event and exigent conditions is required to estimate the impact on Earth when it does arrive.

The necessary information can be summarized in a solar synoptic chart. Solar synoptic charts (also referred to as solar thematic maps) detail the solar activity at any given moment using a labeled image of the Sun. While not responsible for synoptic maps' invention Song et al. (2015) describe the necessary components of synoptic charts [84]. The synoptic chart must be produced in real-time so that space weather forecasters can read and respond as needed. It must be quantitative when describing observations and object boundaries so that the information can be used in other follow-up systems such as expert validation and database generation. Finally, it must be comprehensive, providing more than sufficient information and easy to examine images of the Sun at various important solar atmospheric heights and temperatures. Based upon a literature study, Song et al. (2015) argued that magnetogram and extreme ultraviolet (EUV) imagery are most valuable for general solar event classification [84]. They created a database of 1586 space weather papers and investigated which types of solar phenomena are correlated to different wavelengths in modern research. After

analyzing these trends for active regions, coronal holes, filaments/prominences, flares, and coronal mass ejections, they found that for all the categories 87.4 % used magnetograms and 59 % used extreme ultraviolet images. Thus, these two types of data should be featured most prominently on any synoptic charts.

At the moment, reliable synoptic charts used in forecasting are predominantly human drawn. There are existing automatic classifiers, but they often only detail one type of feature. Space weather forecasters at NOAA's Space Weather Prediction Center (SWPC) still hand draw synoptic maps daily, outlining magnetic field lines, coronal holes, flares, filaments/prominences, and plages. Historical maps are available in PDF format back until 1972. In a future project, this could serve as an interesting source of labeled data, especially for the difficult task of finding magnetic neutral lines. Zheng et al. (2016) utilized similar synoptic drawings from Yunnan Observatory to extract text annotations about sunspots with a convolutional neural network [98]. Some observatories are moving toward automated feature classification. For example, when the person in charge of synoptic maps at the Meudon Observatory was set to retire, they implemented a filament classifier and tracker [1].

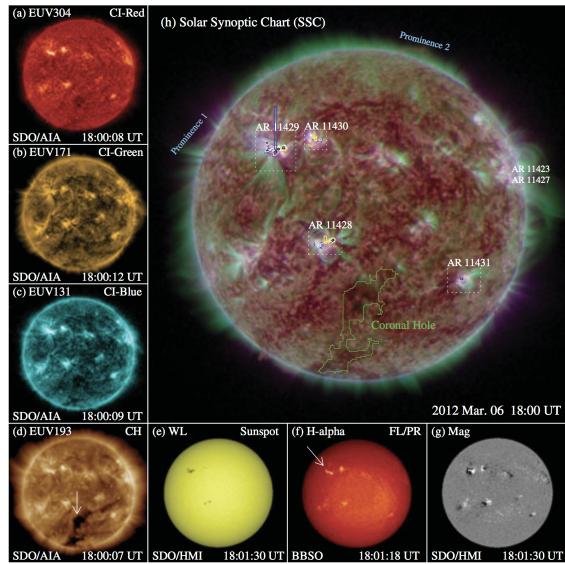


Figure 3.1: **Synoptic map example:** Song et al. (2015) propose the solar synoptic map to include a labeled composite image with different wavelength images accompanying to provide a full image of solar activity [84].

3.2 Data sources

3.2.1 Raw imagery to classify

For solar image classification many types and sources of images can be utilized. While brand new imagers like the Solar Ultraviolet Imager (SUVI) do not yet have public data repositories, Sunpy,

a python package for solar physics, provides an interface for the Virtual Solar Observatory's large database of solar images [86]. (SUVI's data release is May 4th, 2018, and will subsequently be available through the VSO.) Unsupervised approaches to the problem use this tool and others like it to fetch H α , ultraviolet, x-ray, white light, magnetic, and other forms of solar images. They utilize from the Solar Dynamics Observatory [68, 55], the Solar and Heliospheric Observatory [23, 22], Solar Terrestrial Relations Observatory [24, 40], Transition Region And Coronal Explorer [34], the Global Oscillation Network Group[38, 37], and many other sources. There is continuous imagery of the Sun at many wavelengths and heights; raw data is generally not a limitation. Instead, the limitation is incorporating raw imagery into a classifier or obtaining labeled data for a supervised trainer.

3.2.2 Labeled data for training

The supervised techniques discussed in Section 3.3.2 require labeled data to train their classifiers. A research group formerly at Montana State University and now at Georgia State University has collated large amounts of images with labeling from unsupervised classifiers. The original dataset in 2013 comprised over 15,000 images with 24,000 events observed in the first half of 2012 by the Solar Dynamics Observatory (SDO). Using the SDO unsupervised classifiers, small grid regions of the image were labeled active region, coronal hole, filament, flare, sigmoid, and sunspot. Each grid region is also statistically analyzed and assigned an entropy, mean, standard deviation, fractal dimension, skewness, kurtosis, uniformity, relative smoothness, contrast, and directionality measures [78]. This data was later incorporated into a database tool that allows the user to identify an example image and query the full dataset for similar events [7]. This dataset and tool was later expanded to the full SDO observing database [79].

3.3 Solar machine learning

Solar machine learning can be divided into two categories: unsupervised and supervised. Unsupervised techniques do not required human input of labeled images but instead often run on rules; this approach is much more common in astronomy. Supervised techniques are often more flexible and outperform unsupervised techniques in other related fields [4, 95, 32, 41].

3.3.1 Unsupervised solar segmentation

Unsupervised solar segmentation can be broken into many approaches: edge-based algorithms, region-based algorithms, hybrid algorithms, and artificial intelligence approaches. The first three categories are more image-processing techniques solely while artificial intelligence approaches are more generic examples of clustering, support-vector machines, and other tools. These can be used in tandem with the pure image processing techniques.

Edge-based algorithms

Edge-based techniques utilize discontinuities and identify different features utilizing boundaries. Curto, Blanca, & Martinez (2008) employed edge-based unsupervised detection when identifying sunspots in H α images [19]. Since sunspots have crisp boundaries their edges can be used to quickly identify them. Curto, Blanca, & Martinez (2008) used morphological operations to emphasize these boundaries: erosion, dilation, opening, closing, and the top hat transformation. Erosion shrinks bright regions by removing boundary pixels while dilation grows them. Dilation will also fill holes in features. Opening is an erosion followed by a dilation while closing is dilation followed by erosion. Both closing and opening smooth the image: opening fills shape holes, whereas closing breaks wide lines and erases thin lines [19]. The top hat transformation subtracts the original image by the closing of the image. It results in an image showing only the erased parts. By stringing together an empirically determined set of operations, they were able to reliably identify sunspots. Qu et al.(2005) ave a similar system that identifies filaments [72].

Region-based methods

Region based approaches include histogram segmentations, clustering/thresholding, and region-growing approaches. Fuller, Aboudarham, and Bentley (2005) implemented a filament classifier using region growing [27]. This work is based off more generic computer vision region growing by Gonzalez & Woods (2002) [29]. After calibrating, removing dust, and sharpening the solar H α images [96], seed pixels are chosen for region growing using a thresholding technique. Only the dimmest pixels are chosen since they should be at the center of filaments. For multiple iterations, the region grows adding new pixels that are connected to seeds and follow the mean and standard deviation of the neighborhood and consequently are similar to the seed. Finally, a morphological closing operation is applied to remove any holes and make the filaments smoother. After this, the center line of the filament can be determined and characterized using a combination of convolutions, dilations, and erosions. This characterization makes it easier to track filament evolution and measure their length. Ultimately, this technique produced 1149 filaments compared to a human labeled 1232 filaments [27]. This resulted in missing 10% of the filaments in an image. Roughly 5% of the detections were false positives, keying on sunspots instead of filaments because there was no spatial requirement for filaments to be long and skinny. Thus, they could be confused for sunspots which are also dark in H α images. Other region growing methods include: Benkhil et al.(2006) which used ionized calcium, H α , and extreme ultraviolet imagery to grow active region boundaries [10], Higgins et al. (2011) which combined magnetograms, image differencing, and region growing to identify and track emerging active regions [36], and McAteer et al. (2005) which used full-disk magnetograms to identify magnetically significant regions and characterize their flare potential [62].

Instead of growing regions, one can identify significant features by looking at the histogram of their intensities in various wavelengths. Olmedo et al. (2008) designed such a system to identify coronal mass ejections. The intensity in solar images can be plotted as a histogram as a function of position angle as shown in Figure 3.2. A threshold is used to determine what is a significant event in the histogram. If portions of the histogram exceed this, they are declared a region in

the image and grouped together. Some region growing is also used in this approach. Ultimately, they were able to recover about 75% of the human identified coronal mass ejections in a 12-month period. Interestingly, they found an equal number of small coronal events that had been overlooked by humans, often weaker but creating an interesting new population for scientific research and space weather awareness [66].

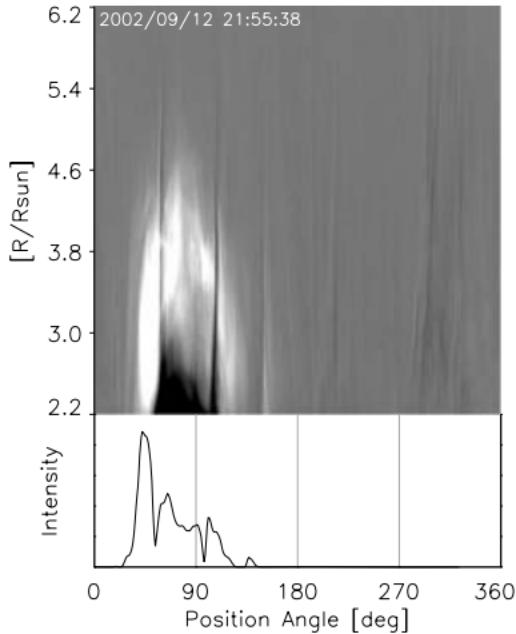


Figure 3.2: **Histogram segmentation:** Olmedo et al.(2008) utilized histogram segmentation to identify coronal mass ejections. “The intensity profile along the angular axis showing the 1D projection of the CME image. Only positive pixels along the radial axis are used. This profile effectively indicates the angular positions of a CME when it is present.” [66]

Other histogram based methods include Preminger et al.(2001) which used contrast ratios between ionized calcium and magnetograms to identify sunspots and faculae [69].

Hybrid

Some approaches bridge between region-based and edge-based techniques. One key example is Spatial Possibilistic Classification Algorithm (SPoCA), a fuzzy classifier that uses multiwavelength extreme ultraviolet observations to assign multiple classes: quiet Sun, active region, and coronal hole. SPoCA is more accurately a suite that implements three types of fuzzy clustering algorithms tailored to the segmentation of solar coronal EUV images: Fuzzy C-means (FCM), Possibilistic C-means (PCM) algorithm, and Spatial Possibilistic Clustering Algorithm (SPoCA) [90]. It differentiates itself from other techniques because of its reliance on fuzzy logic. Barra et al. (2008) note that the use of fuzzy logic allows the algorithm to overcome noise in the images and the scientific definitions

of solar features while meeting many needs. Often, unsupervised solar segmentation techniques are developed with a specific research question and thus are very restricted. For example, one classifier may focus only on the brightest cores of active regions to study energy transfer while another may be designed to catch active regions as they form and thus catch many weak active regions. SPoCA is generic enough to overcome this conflict.

SPoCA works by utilizing the assumptions of fuzzy logic. Each pixel has membership in all of the different classes used in the thematic map. The sum of all the memberships for a pixel must be one, i.e. the pixel membership indicates how much the pixel matches that theme. In fuzzy c-means [11], a generalization of K-means clustering to fuzzy logic, the variance within a cluster, all the pixels labeled a theme, is minimized. This approaches is very susceptible to the noise in astronomical images though [44]. Thus, the fuzzy c-means requirement of membership summing to one is relaxed to form of probabilistic c-means [45]. SPoCA is a further modification that incorporates weighting by spatial extent such that neighboring pixels should be assigned similar labels [8, 9].

Barra et al. (2009) used SPoCA to both segment and track features on the Sun [9]. They exercised the power of SPoCA to perform two very different experiments. First, they tracked the biggest active region for a month, quantifying its size, average intensity, fractal dimension, and other parameters for scientific inquiry. Then, they identified and tracked coronal bright points, a feature not initially intended in SPoCA. These short lived (< 2 days) bright regions have some spectral similarity to active regions but are much smaller and can appear within coronal holes. They impact the structure and dynamics of the solar corona. During the study, Barra et al. (2009) tracked their counts, north/south asymmetry, intensity fluctuations, and other parameters [9].

There are other types of unsupervised classifiers that do not fit nicely into the two main categories. For example, Bratsolis & Sigelle (1998) utilized mean field fast annealing to segment sunspots [14]. The approach uses simulated annealing to minimize the classification into q classes. Each pixel is assigned a label. This labeling has an energy described by mean field theory and the Potts interaction between pixels. Essentially, this approach attempts to find the most meaningful classification. It excels over histogram methods which often are not granular enough to separate regions of different activity in the sunspot.

Example system: Solar Dynamics Observatory

Many unsupervised approaches that only deal with one class at a time can be chained together to create all the necessary data for a thematic map. The Solar Dynamics Observatory (SDO) satellite mission produces 1.5 TB of imagery per day in multiple ultraviolet wavelengths, a magnetogram, and other data channels. To deal with this influx of data, teams of researchers developed classifiers that identified specific classes of features [60]. All of the component parts can be seen in Figures 3.3 and 3.4.

The SDO suite is a comprehensive approach to classification that utilizes all of the techniques mentioned thus far and more. For features that are beyond the individual classifiers' scope, a trainable module is employed. While this is a supervised system, it is mentioned here to emphasize the difference between supervised and unsupervised approaches. A user can identify a specific type

Science Target S/W Module Name	Flares Flare Detective	Coronal Dimming Dimming Detector	Magnetic Feature Tracking SWAMIS	Filaments AAFDCC
Triggered or Continuous? Trigger Source Data Trigger Source Module(s) Binned Image for Trigger Cadence for Trigger	Triggered AIA 193 Self 16 X 16 Full	Triggered AIA 193 Self 512 X 512 5-6 minutes	Continuous	Continuous
Source Data Source Data Binning Source Data Cadence	All AIA Channels None Full	AIA 512 X 512 5-6 minutes	HMI Time Averaged Magnetograms None 6 minutes	Global Hi-res H α Network Full; some smoothing Full (~ 4/day eventually)
S/W Module Data Programming Language Module Location Computing Requirement Responsible Team Member Originating Organization Output Data Volume First Pipeline Installation	Flare Detective IDL LMSAL Paolo Grigis SAO June 2010	Dimming Detector IDL LMSAL 20 minutes CPU/day Meredith Wills-Davey SAO 10 kB/event, 5 MB/movie July 2010	SWAMIS Perl/PDL LMSAL 8-16 CPUs to operate at 4x realtime Craig DeForest SwRI 120 MB/day nominal November 2010	AAFDCC IDL and C SAO Most < 1 min CPU/Image Pietro Bernasconi JHU-APL ~ 1.5 MB per H α image March 2010
Heritage	EIT, TRACE, RHESSI	New	MDI	GBO H-alpha
Science Target SPoCA	Active Regions	Sigmoids Sigmoid Sniffer	PIL Mapping PIL Finder	CMEs CME Detector/Tracker
Triggered or Continuous? Trigger Source Data Trigger Source Module(s) Binned Image for Trigger Cadence for Trigger	Continuous AIA 211 Self Full Resolution 10 min	Triggered AIA 211 Self Full Resolution 10 min	Continuous	Triggered LASCO C2 and C3 Flares, Dimmings, Sigmoids Full Resolution Full LASCO Cadence
Source Data Source Data Binning Source Data Cadence	AIA 171/195, 211/335, 94 None 15 minutes	AIA 94, 131 None 10-20 s	HMI Magnetograms Depends on science specs ~5 min	LASCO C2 and C3 None; front smoothing Full (~1 hr)
S/W Module Data Programming Language Module Location Computing Requirement Responsible Team Member Originating Organization Output Data Volume Scheduled Installation	SPoCA C++ LMSAL 30 sec/pair of full-res AIA Veronique Delouille ROB 6 kB/event October 2010	Sigmoid Sniffer IDL LMSAL 10 min CPU/image Nour-Eddine Raouafi JHU-APL 5 MB/day November 2010	PIL Finder IDL SAO Standard desktop Alexander Engell SAO ~ 2 Mb/day December 2010	CME Detector/Tracker IDL SAO 30 minutes/event Meredith Wills-Davey SAO 1-8 Mb/day Spring 2011
Heritage	EIT	SXT, SXI, XRT	Kitt Peak, SOLIS	LASCO

Figure 3.3: First half of SDO classifiers [60]

of feature they are interested in by identifying them within an image. These are placed in feature vectors of 12 texture parameters (for example mean, entropy, uniformity). These train either a support vector machine or a C4.5 decision tree and will then identify similar features from the rest of the database [50]. The SDO suite lacks the ability to combine all the classifications into a single thematic map.

Comparisons

As shown in Figure 3.5, different methods produce often similar but still different results. By computing the fractal dimension over the segmented image, they were able to characterize the difference between fuzzy clustering, region growing, iterative thresholding, and histogram thresholding. They found that depending on the height of the active region in the solar atmosphere it was segmented differently. In general, the fuzzy-based and histogram approaches outperformed the others. They propose that using longer wavelength ultraviolet images tends to larger area active regions.

Caballero & Aranda (2013) conducted an independent comparison of unsupervised techniques for active regions [15]. Using 6000 images from SOHO in 195 angstroms, they initially segmented

Science Target S/W Module Name	Coronal Holes SPoCA	X-ray Bright Points BP Finder	Sunspots SWAMIS	Global NLFFs Optimization Code for Full Disk
Triggered or Continuous? Trigger Source Data Trigger Source Module(s) Binned Image for Trigger Cadence for Trigger	Continuous	Continuous	Continuous	Continuous
Source Data Source Data Binning Source Data Cadence	AIA 171/195, 211/335, 94 None 15 minutes	AIA 171, 195, 211 None (TBC) Full	HMI Magnetograms None TBD (5-60 min)	HMI+AIA for comparison Binned to 256x256 for ARs ~5 min
S/W Module Data Programming Language Module Location Computing Requirement Responsible Team Member Originating Organization Output Data Volume First Pipeline Installation	SPoCA C++ LMSAL	BP Finder IDL SAO	SWAMIS PDL LMSAL < 1 CPU Steve Saar SAO 600 kB/10 minutes January 2011	Optimization Code C MPS 2 hours/vector-magnetogram Thomas Wiegelmans Max Planck Sonnenforschung 256x256x256 datacube When HMI VMG's become routine
Heritage	EIT	EIT	MDI	SOT
Science Target S/W Module Name	Jets Jet Detector	Oscillations Oscillation Finder	"EIT Waves" EIT Wave Tracker	Trainable Feature Recognition
Triggered or Continuous? Trigger Source Data Trigger Source Module(s) Binned Image for Trigger Cadence for Trigger	Triggered TBD AIA Channels XRBP & CH 1024x1024 (TBC) 1 min	Triggered AIA 193 Flares, CMEs, Dimmings N/A N/A	Triggered AIA 193 Dimmings 1024x1024 Full	Continuous
Source Data Source Data Binning Source Data Cadence	AIA 193 None 1 min	Appropriate AIA Channels None None	AIA 193 None; smoothing applied Full	All AIA Channels, some HMI Data 128 X 128 subimages Cadence TBD
S/W Module Data Programming Language Module Location Computing Requirement Responsible Team Member Originating Organization Output Data Volume First Pipeline Installation	Jet Detector IDL SAO	Oscillation Finder IDL SAO	EIT Wave Tracker IDL SAO Very small James McAteer New Mexico State Univ.	Trainable Feature Recognition C/IDL SAO Rafal Angryk MSU 0.1% of image volume December 2010 for image vectors
Heritage	XRT	TRACE, EIT	TRACE	TRACE

Figure 3.4: Second half of SDO classifiers [60]

the images using region growing techniques. Then, the different independent regions were clustered together into units using either partition approaches or a hierarchical classification. They found that the hierarchical classification, the idea that nearby regions should be more strongly connected, produced more human-like clusters. However, this approach requires exponential time complexity and results in a hierarchy instead of a simple division into clusters.

For a much longer comparison of many different techniques see Aschwanden (2010) [5].

3.3.2 Supervised solar segmentation

There are relatively few supervised solar segmentation approaches but the existing approaches are very promising and tend to perform on a broader set of classes and image types. Thus, a bit more detail is provided in describing each approach.

Maximum likelihood estimation

Rigler et al. (2012) proposed a preliminary naive Bayesian classifier for SVDI images using the Solar Dynamics Observatory (SDO) Advanced Imaging Array (AIA) observations for testing [75].

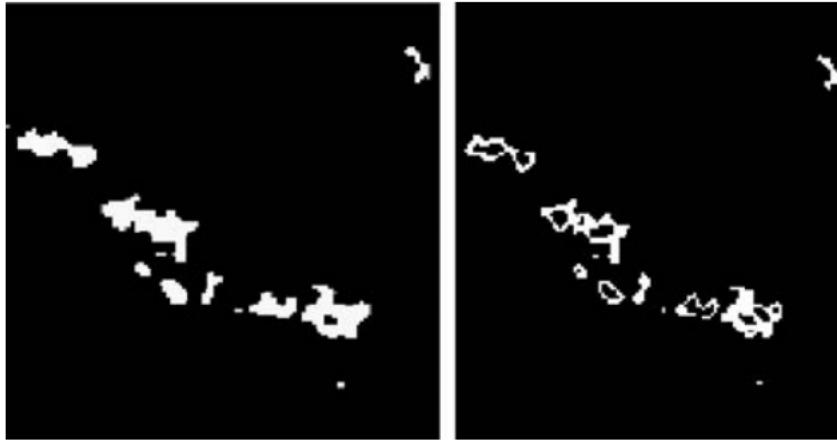


Figure 3.5: **Performance of different methods:** Revathy, Lekshmi, & Nayar (2005) compared the performance of different segmentation techniques in identifying active regions [74]. At left is the result of a histogram thresholding approach while at the right is fuzzy-based segmentation.

Their work focused on eight classes: flare, prominence, active region, quiet corona (off-disk), quiet corona (on-disk), coronal hole (off-disk), coronal hole (on-disk), and outer space. They were able to achieve an average accuracy of 86%. This was calculated by training the classifier on a set of data then testing it on a classified image that was not used for training. The tabulation was done strictly on a pixel by pixel basis, not checking if the error was coherent or random noise. Prominence was the most problematic class with only 41% of prominence pixels being classified correctly. They were often misclassified as off-disk quiet corona or on-disk coronal hole.

The naive Bayesian approach works by classifying every pixel into one of n classes using multi-spectral ultraviolet images. Therefore, a pixel, a spatial element at (i,j) corresponds to h channels and can be described as a vector:

$$x_{(i,j)} = \begin{bmatrix} x_1 & x_2 & \dots & x_h \end{bmatrix}^T$$

The approach is assign each $x_{(i,j)}$ pixel a label w_k from the set of classes W . This approach employs Bayes' Theorem:

$$P(w_k|x_{(i,j)}) = \frac{P(x_{(i,j)}|w_k)P(w_k)}{P(x_{(i,j)})}$$

Since $P(x_{(i,j)})$ is not a function of the label classification, it can be ignored.

$$P(w_j|x_{(i,j)}) \propto P(x_{(i,j)})P(w_k)$$

Rigler et al. (2012) simplify this even further by stating that “if there is no a priori reason to believe a pixel should be assigned label w_k , $P(w_k)$ can be assumed to be drawn from a uniform

distribution”. Thus,

$$P(w_k|x_{(i,j)}) \propto P(x_{(i,j)}|w_k)$$

This approach is the maximum likelihood solution to this problem.

For training they simplify each class into a multivariate normal, i.e. for each potential label there is a archetypal example pixel and all pixels with that label should be distributed normally about it. This is overly constraining if any given class has multiple distinct modes with respect to the selected data. The multivariate distribution for class w_k is characterized by a mean vector μ_k and covariance matrix C_k which are calculated as:

$$\mu_k = \frac{\sum_{x \in W_k} x}{|W_k|}$$

where W_k is the collection of pixels with label w_k . Similarly,

$$C_k = \frac{\sum_{x \in W_k} [x - \mu_k] \times [x - \mu_k]^T}{|W_k|}$$

These mean vectors and covariance matrices characterize the class. Given this characterization for class w_k one can calculate the conditional probability of a pixel $x_{(i,j)}$ having label w_k :

$$P(x_{(i,j)}|w_k) = \frac{1}{\sqrt{(2\pi)^h} \sqrt{|C_k|}} \exp \left(\frac{-1}{2} \times (x_{(i,j)} - \mu_k)^T \times C_k^{-1} \times (x_{(i,j)} - \mu_k) \right)$$

Thus, the pixel is assigned the class that maximizes this probability.

Since each pixel is treated separately, any noise in single pixels or in the image as a whole can result in a noisy classification where a pixel class does not agree with its neighboring pixels as expected. Therefore, Rigler et al. (2012) propose a smoothness prior be enforced so that a pixel’s labeling relies on its neighbors. This can be enforced by iteratively calculating the thematic map, calculating a smoothed map relying on neighbor probabilities, and repeating until convergence using simulated annealing, maximizing posterior marginals, or iterated conditional modes as proposed by Tso and Mather (2009) [88].

Their results were promising with high accuracies and maps that generally coherent. However, their results are concerning because some statistics and accuracy measurements come from running the classifier on training data, providing no indication on how the classifier would perform on unfamiliar, real-world examples.

Rigler et al.(2012) built upon earlier work by de Wit (2006) who suggested the Bayesian approach [20]. After decreasing the noise and normalizing the intensity in each image, de Wit (2006) instead used only four ultraviolet wavelengths and projected them into a three-dimensional parameter space using singular-value decomposition [20]. Thus, de Wit (2006) ran a naive Bayesian classifier on these transformed parameters instead of the higher dimensional wavelengths. Figure 3.6 shows an example result by de Wit (2006), emphasizing the coherence of this segmentation without any forced smoothness [20]. This approach ran in near real-time, taking only a few minutes to classify every pixel.

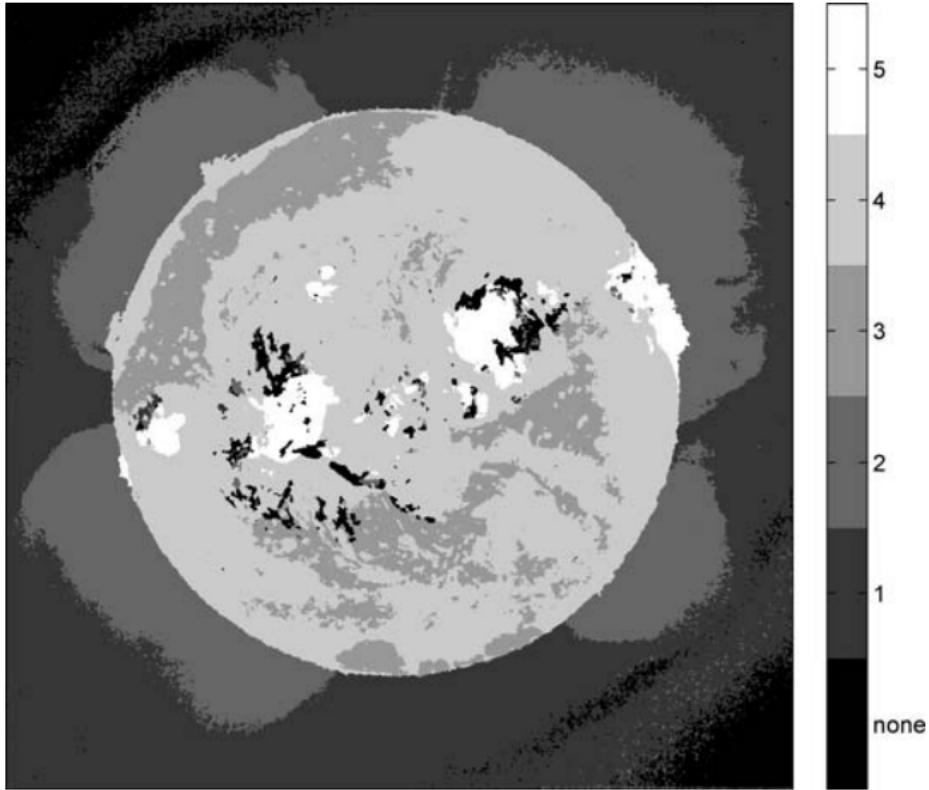


Figure 3.6: **de Wit Segmentation:** This figure indicates the power of quick, multi-spectral, supervised segmentation done by de Wit (2006), very similar to other work by Rigler et al.(2012) and De Visscher et al.(2015) [20]. The classes for this study are: “(1) Tenuous corona outside of the disk, in regions with open magnetic-field lines.(2) Dense corona outside of the disk. (3) Coronal holes. (4) Quiet sun, including the chromospheric network and regions inside the network boundaries. (5) Active regions on the disk” [20].

Maximum A Posteriori

De Visscher et al. (2015) improve upon Rigler et al. (2012) by recognizing that the classes are not equally likely (for example the majority of the Sun is covered by quiet corona at any given time) and that crisp segmentation procedures are too limiting compared to fuzzy segmentation [92]. Further, they incorporate spatial information by letting the probability of a class rely on both the intensity and latitude, assuming these are statistically independent variables so $P((x, L(x))|w_k) \approx P(x|w_k)P(L(x)|w_k)$ where $L(x)$ is the latitude of the pixel and w_k is a class. This addresses the observable fact that coronal holes tend to form at solar poles, high latitudes, while active regions form near the equator. Unlike Rigler et al. (2009), De Visscher et al. (2015) use only one wavelength of light in their dataset: the 19.3 nm SDO-AIA channel. Additionally, they only assign three classes: active region, coronal hole, and quiet sun. This makes it difficult to directly compare their results because the troubling classes Rigler et al.(2009) observed are not included. Therefore, the probability

they are maximizing is instead:

$$P(x|w_k)P(L(x)|w_k)P(w_k)$$

That is the product of the probability of pixel x given its class, the probability of the latitude given the class, and the probability of the class. While doing this, they have assumed a fuzzy segmentation that allows for degrees of membership in each class. This approach allows estimation of how certain different classes are under different conditions. For example, they confirmed that wrongly classified pixels were classified as half one class and half another class and often on the object boundaries. In addition, they were able to pinpoint that there is disagreement within the gold standard database. Upon second examination, they find that the human assigned labels may actually be wrong in this region, highlighting a key problem in solar segmentation: there is no clear universally agreed upon definition of some of the classes so no human classification can be accepted as completely correct. Even if a human standard does exist, it is sometimes still impossible to discern between classes due to degeneracy in the observational parameters. Ultimately, they report average 94% accuracy over all classes.

In addition to accuracy with a single frame, De Visscher et al. (2015) establish a criteria for accurate segmentation into large-scale features [92]:

- “Stable segmentations on short timescales in the absence of major solar activity”
- Consistent and smooth trends and classifications over longer periods of time
- Consistency with human drawn maps

While there is no quantified method for the first two criteria, they are highly relevant and often not addressed in other approaches explicitly.

Neural Networks

Deep convolutional neural networks have proven to be very skilled in classifying and segmenting in various contexts [87, 88, 46]. Convolutional neural networks are specifically designed for image data such as solar images. Each layer extracts local features from an image using a kernel which are combined in intermediate pooling layers. This allows for a robust classification with respect to distortions or noise in the images. Activation functions allow only significant features to influence the final classification. Deep learning approaches are advantageous because they automate feature selection by weighting input data according to their training algorithms instead of having a scientist develop detailed rules about what data components indicate which classes.

Kucuk et al.(2017) applied the first convolutional neural network to solar imagery for classification [47]. While classifying over a finer granularity of classes they were able to achieve an average 70% accuracy across each class. This convolutional neural network approach outperformed the only other published neural network solar segmentation found during this review. Zharkova & Schetinin (2005) employed a feed-forward neural network with two hidden neurons and one output neuron to identify solar filaments at 82% accuracy [97]. This result is not directly comparable since it only classified one type of feature. However, it illustrated the power of neural networks in solar

images. Filament classification by classical image processing techniques is often confused by the highly variable background between different parts of a filament and from filament to filament. The artificial neural network was able to flexibly learn many patterns and more accurately identify filaments. At the time, it was only outperformed by a region-growing approach [27].

Up until now, classification has only been discussed in a spatial domain. However, flares and coronal mass ejections have a temporal components. They are by definition changing features. Borda et al. (2002) implemented a simple neural network consisting of two layers (not including input): a hidden layer of nonlinear neurons and an output layer of one linear neuron [13]. Given optical H α images, it identifies solar flares in real-time. It operates on 7 input features: mean image brightness, standard deviation of the brightness, the pixel of maximum brightness change between images, absolute brightness of pixel with maximum change, radial position of that pixel, variation of mean brightness between two images (to characterize possible weather influences), and the contrast between the pixel with largest change in brightness and its neighbors. Given 124 test events, fewer than 5% were misidentified. (The paper does clarify false positives versus false negatives.) Accounting for normal operations, this would be a misclassification every few days. There has been limited solar time series neural networks beyond this, but it establishes a baseline system for future architectures and generalizations to other feature types.

3.4 Earth Remote Sensing

Earth remote sensing of multiwavelength features has many more applications and a longer availability of data and thus has advanced further than solar machine learning techniques.

I could not find an example of random forests for solar image classification. However, they are routinely used in Earth remote sensing classifications. Random forests are an ensemble of tree classifiers. To classify a new feature vector, the input vector is classified with each tree in the forest, and the forest chooses the classification having the most votes over all the trees in the forest. Random forests have many advantages: high accuracy compared to current algorithms, efficient implementation on large data sets, and an easily storable data structure for future use [28]. Lowe & Kulkarni (2015) used a random forest to identify terrain type in hyperspectral images, nearly simultaneous images of Earth composed of tens to hundreds of narrow passbands for each pixel [59]. For this application, the random forest had 96.25 % accuracy compared to neural network's 76.87%, support vector machines 86.88%, and maximum likelihood's 83.11% [59]. This high performance for random forests with this type of problem is not uncommon [70, 77, 16, 48].

Similarly, neural networks have a rich tradition in Earth remote sensing. Lee & Kwon (2017) developed a 9 layer convolutional neural network, both wider and deeper than state-of-the-art methods for this problem, to classify land types in Earth remote sensing [54]. This network achieves over 95% accuracy in nearly every class. It is specially designed for spectral-spatial data and explores neighborhood relationships in a more optimized fashion than previous networks by allowing for multi-scale examination.

Li et al. (2014) present a comprehensive review of Earth remote sensing classification techniques based on spatial techniques [56]. They detail the usage of K-means, ISODATA, SOM, hierarchical

clustering, Maximum likelihood, Minimum distance-to-means, Mahalanobis distance, Parallelepiped, k-nearest Neighbors, artificial neural network, classification tree, random forests, support vector machine, genetic algorithms, Fuzzy classification, neural networks, regression modeling, regression tree analysis, spectral mixture analysis, fuzzy-spectral mixture analysis, and image segmentation and object-based image analysis techniques in Earth remote sensing.

3.5 General Computer Vision

Given sufficiently large datasets, cutting edge computer vision research can be applied to the solar segmentation problem. The solar segmentation problem and producing thematic maps is an example of semantic segmentation, a well studied problem in computer vision. However, it is difficult to apply these techniques to solar image segmentation because they require large labeled datasets that are currently non-existent in solar imaging.

3.5.1 Fully convolutional neural networks

The notion of extending convolutional neural networks to do dense prediction, effectively creating a thematic map, was first proposed by Matan et al. (1991) [61] to extend the LeNet convolutional neural network [53] for handwritten digit recognition. Shelhamer, Long, & Darrell (2016) presented a new implementation, the fully convolutional neural network (FCN), that takes arbitrarily sized input and creates a similarly sized semantic segmentation [58]. During their development, they thoroughly describe convolutional neural network for semantic segmentation up until 2016 [58].

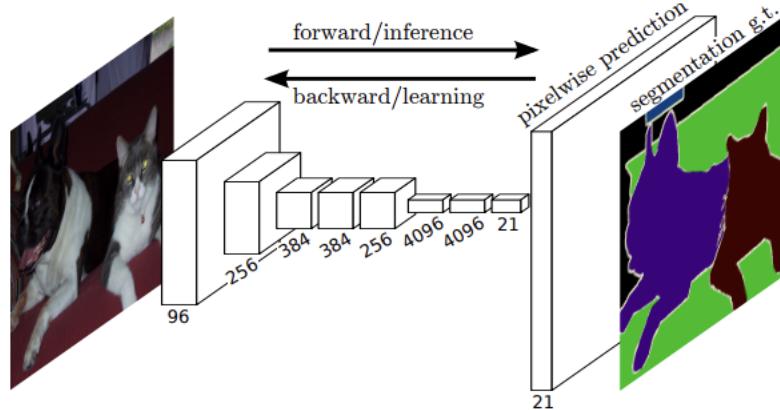


Figure 3.7: **Fully convolutional neural network:** This is an example architecture for a fully convolutional neural network (figure 1 from Shelhamer, Long, & Darrell (2016) [58]).

A FCN works by using existing full neural networks in a convolutional fashion. Then, the last steps of the existing network are removed so that it cannot make a classification for the entire input image at that time. Instead, a pixelwise prediction layer is added using deconvolution and striding

over the input image. Without specialized refinement, the FCN can then create a dense output map. At the time, the FCN gave 20% improvement over state-of-the-art semantic segmentation in a shorter inference time.

3.5.2 Mask R-CNN

He, Gkioxari, Dollár, & Girshick developed Mask Regional Convolutional Neural Network (Mask R-CNN), an extension of Faster R-CNN [73] that creates another form of a semantic segmentation [35]. Technically, Mask R-CNN is a type of FCN [58]. Faster R-CNN worked by quickly determining bounding boxes for various objects in a scene. Mask R-CNN extends this by determining a pixel mask for each bounding box in parallel. Thus, it can distinguish both the type of an object and between neighboring objects in a scene.

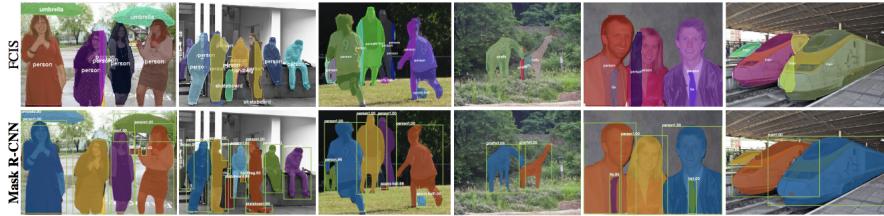


Figure 3.8: Mask regional convolutional neural networks: The top row was the existing state-of-the-art instance segmentation [57], an example of a FCN, compared to the bottom row of Mask R-CNN performance on the same scene. The overlaid coloration indicates the segmentation while bounding boxes indicate where Mask R-CNN evaluated these masks. Clearly, the Mask R-CNN produces more coherence classifications. In addition, it runs in less time with higher accuracy than the existing state-of-the-art systems and was consequently awarded the 2017 International Conference on Computer Vision best paper award.

The Mask R-CNN design was tested in several domains: cityscapes, human pose estimation, common objects. In all cases, it outperformed existing systems. With modification, this type of architecture solves the problem of solar segmentation.

3.6 Summary

Solar image segmentation is well motivated by space weather and archival concerns. Up until now, there have been numerous unsupervised approaches, too many to completely document in this paper. These approaches employ a variety of computer vision techniques but are often limited to only determining membership in one class, for example determining which pixels are filament. They can be used in tandem to create solar thematic maps that label all classes on the Sun in an ensemble classifier. Settling disputes between the independent classifiers can be difficult and running them all in parallel can be costly. Instead, machine learning techniques can be applied to label the entire image at one time. A few supervised approaches have been developed and perform well. However, there is limited quantification and even more limited consistent quantification of that performance making

it difficult to compare systems. By looking to Earth remote sensing and state-of-the-art computer vision machine learning approaches, solar image segmentation can be advanced in a systematic and measurable fashion.

Chapter 4

Data

This chapter first describes the raw solar data utilized in this work. It then details the processing done to use the images in the project. Finally, it describes the annotation process for the images and analyzes the results.

4.1 Solar Imagery

Ultraviolet observations of the Sun are optimal for seeing both dynamic, high temperature solar eruptions and cooler, more stable coronal holes. There are a few ultraviolet solar imagers potentially suitable for this investigation: the Atmospheric Imaging Assembly (AIA) aboard the *Solar Dynamics Observatory* (SDO) spacecraft [55], the Extreme-ultraviolet Imaging telescope aboard the *Solar and Heliospheric Observatory* (SOHO) spacecraft [22], and the Sun Watcher with Active Pixels and Image Processing (SWAP) aboard the *Project for On-Board Autonomy 2* spacecraft [33, 80]. This project instead uses data from the new Solar Ultraviolet Imager (SUVI) aboard the *Geostationary Operational Environmental Satellite 16* (GOES-16) and GOES-17 operated by the United States National Oceanic and Atmospheric Administration (NOAA) [81, 82]. SDO, SOHO, and SUVI share many of the same wavelength passbands, but SUVI has a wider field of view allowing for classification of events further off the Sun. SWAP has the same field of view as SUVI but only one passband that is very similar to SUVI's 17.1 nm band. SUVI has a slightly higher spatial resolution than SWAP. In addition, GOES-16 is the first four commissioned satellites in the R series, all of which will have SUVI instruments. Thus, the classification system and database built here will be used for at least 20 years. The classification techniques applied here could be applied, with modification, to any of the aforementioned satellites, including to archived data.

4.1.1 Solar Ultraviolet Imager (SUVI) Data

SUVI observes in six different wavelength passbands (and their corresponding coronal EUV emission line) using filters and multilayer mirrors sensitive to specific wavelengths: 9.4 nm (Fe XVIII), 13.1 nm (Fe XXI), 17.1 nm (Fe IX/X), 19.5 nm, (Fe XII), 28.4 nm (Fe XV), and 30.4 nm (He II)

Each wavelength is most sensitive to a specific wavelength and solar feature as detailed in Figure 4.1.

	94 Å 6.8	131 Å 7.0,7.2	171 Å 5.8	195 Å 6.1,7.3	284 Å 6.3	304 Å 4.7
Wavelength						
Log (Te)						
Filaments						
Coronal Holes						
Active Region Complexity						
CMEs (e.g. dimming)						
Flare Location and Morphology						
Quiet Regions						

Figure 4.1: **Wavelengths:** This diagram indicates which SUIV wavelengths are most helpful in identifying different space weather events [31].

However, these passbands overlap with other spectral lines that form at different temperatures than the predominant one listed above. This can cause potential confusion between channels during classification. The qualitatively apparent in the first-light images from SUIV in Figure 4.2.

Each image is composed of 1280×1280 32-bit pixels. The individual pixels are 2.5×2.5 arcseconds on a side resulting in a 53.3×53.3 arcminute total field of view horizontally [82]. The field of view is slightly larger along the diagonal but vignetting in some of the image corners in some passbands renders these regions unusable. Ultimately, SUIV can observe out to 1.6-2.3 solar radii (depending whether on the horizontal or diagonal).

One image, in a single passband, is taken every 10 seconds. The observation sequence insures that every passband is observed at least once every four minutes. In addition, SUIV captures a short exposure time and a long exposure time image in each passband. During energetic and bright events like flares, the detector would saturate at the longer exposure time needed for high contrast images during times without energetic events. These two exposure times are combined to create high-dynamic-range composite images, which a two to three orders of magnitude increased dynamic range than a single exposure time. In addition, SUIV has anti-blooming circuitry so that saturation in the long exposure image does not deteriorate quality in the composite image. This compositing will allow better classification of very bright active regions and flares than if another data set were used. Finally, the composite images are aligned so that the Sun is centered in each.

4.1.2 Noise-gater procedures

The SUIV 94 angstroms channel is riddled with shot noise that degrades the signal quality. Shot noise is statistical noise caused by uncertainty in counting photons received by a camera sensor that follows the Poisson distribution and is proportional to the square root of the number of received

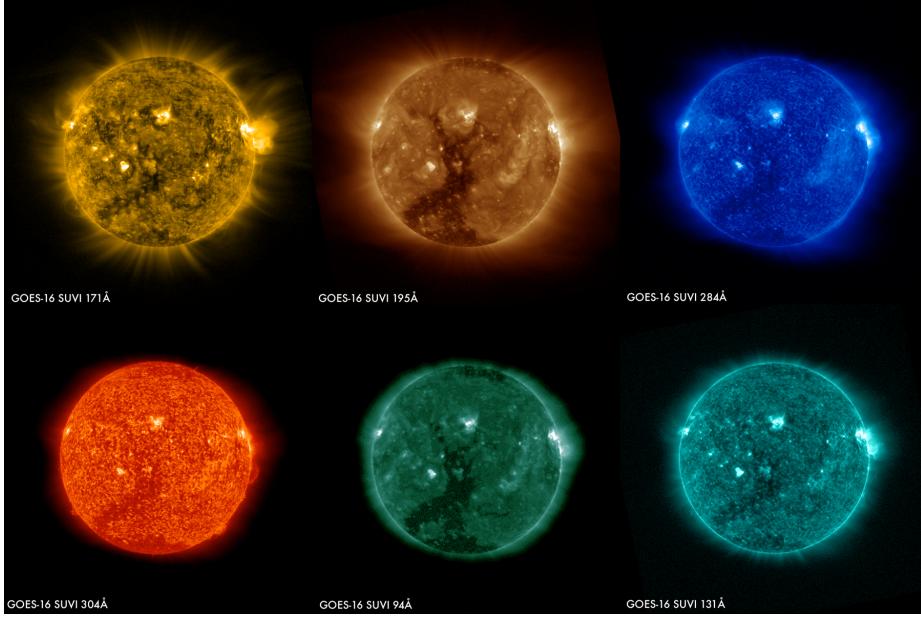


Figure 4.2: Example S UVI images: These first-light images taken on January 29, 2017 exemplify S UVI’s response to different solar features in different passbands. For example, the large coronal hole (the darker boot-like shape in the lower portion of the Sun) has higher contrast from the background in 19.5 nm than 304 nm [31].

photons. Since the 94 angstrom channel is so short and consequently high energy, there are many fewer photons emitted by the Sun at this wavelength resulting in a larger shot noise problem in this channel than others. The shot noise buries the signal and could potentially make it more difficult for the machine learning classifiers to cleanly identify the solar classes. Deforest (2017) proposed a method using localized Fourier transforms to characterize and remove the noise from images with specific application to extreme ultraviolet images. The effectiveness of noise-gating is shown in Figure 4.3.

The noise-gating algorithm was developed by [21] for use in extreme ultraviolet solar images, medical imaging, and other applications. The term noise-gating stems from the approach used in audio noise reduction and functions similarly: local adaptive filters in Fourier domain distinguish between coherent spatial structure and background noise.

Beginning with an image sequence, described as a mapping from coordinates in the horizontal and vertical spatial dimensions and a time dimension, i.e. $\mathbb{Z}^3 \rightarrow \mathbb{R}$, the noise is modeled as:

$$Im(x, y, t) = Im_0(x, y, t) + N_a(x, y, t) + N_s(x, y, t) + N_{\text{other}}(x, y, t)$$

The observed image sequence $Im(x, y, t)$ is a combination of Im_0 , the idealized, noise-free image; N_a , background noise that is independent of the signal of Im_0 ; N_s , shot noise that is dependent on the signal of Im_0 ; and N_{other} , all other noise sources. For this application, N_a and N_{other} are ignored because N_s dominates them. N_s is then approximated with a normal distribution $G(x, y, t)$

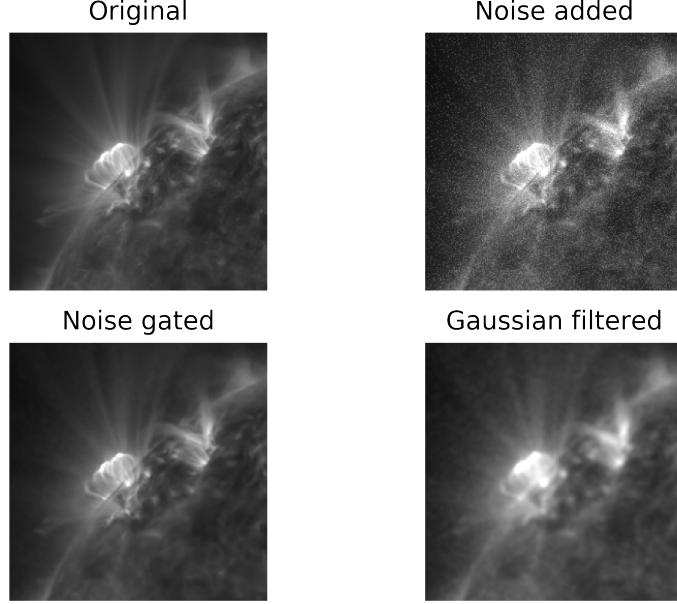


Figure 4.3: Effectiveness of noise-gating The upper left is a good image, no cleaning necessary. However, images like the upper right, dominated by shot noise, are typical for the 94 angstrom channel. This image is created by taking the image on the upper left and adding Poisson noise with a signal-to-noise ratio of 2. The algorithm still performs even if it is worse, although some artifacts appear. DeForest’s algorithm was applied to create the cleaned image on the bottom left. This can be compared to simply smoothing the image to decrease the noise as in the bottom right, a typical alternative procedure.

and instrumental constant α :

$$N_s(x, y, t) \approx \alpha G(x, y, t) \sqrt{\text{Im}_0(x, y, t)}$$

This is transformed into Fourier space as:

$$|N'_s(k_x, k_y, \omega)| \approx \beta(k_x, k_y, \omega) \sum_{x,y,t} \sqrt{\text{Im}(x, y, t)}$$

β is spectrum that is characteristic of the imaging instrument and remains constant across images. The image is broken into many small neighborhoods $\text{Im}_i(x, y, t)$. Since the image is dominated by shot noise, which should be fairly uniform in Fourier space, the noise distribution is approximated by:

$$\beta_{\text{approx}}(k_x, k_y, \omega) = \text{median}_i \left(\frac{|\text{Im}'_i(k_x, k_y, \omega)|}{\overline{\text{Im}}_i} \right)$$

Then, the threshold to filter out noise is set for each image section independently by:

$$T_i(k_x, k_y, \omega) = 3N'_i(k_x, k_y, \omega)$$

with a gate filter:

$$F'_{i,\text{gate}}(k_x, k_y, \omega) = \begin{cases} 0 & \text{if } \text{Im}'_i(k_x, k_y, \omega) < T_i(k_x, k_y, \omega) \\ 1 & \text{otherwise} \end{cases}$$

. This results in a cleaned image $\text{Im}'_i(k_x, k_y, k_\omega)F'_i(k_x, k_y, k_z)$ which can be mapped back into image space. The Fourier space is apodized by a Hanning window to avoid any vignetting. The full details can be found in the original paper [21].

I experimented with noise-gating and implemented it in Python, which I have made available on GitHub. Unfortunately, the temporal sparsity of the images in this thesis means it does not provide significant improvement and was not applied to the images. However, it could provide improved classification certainty in near-real-time application.

4.1.3 H-alpha imagery

As discussed in Section 7.6, it was impossible to differentiate between all labeled classes using only ultraviolet imagery. Thus, H α images, a narrow red passband at 656.28 nm caused by a hydrogen electron falling from the third energy level to the second, were included. They are particularly helpful in observing solar prominences and filaments. These images were gathered using the Virtual Solar Observatory [39], an online compilation of images of the Sun from various sources, using SunPy [86, 43], a Python toolkit for solar physics. In particular, H α images from the Global Oscillation Network Group (GONG) were used since they have continuous time coverage because of their seven globally distributed observing sites in Learmonth, Australia; Udaipur, India; El Teide, Spain; Cerro Tololo, Chile; Tuscon, Arizona, United States; Big Bear, California, United States; and Mauna Loa, Hawaii, United States. Each image is 2048×2048 pixels with each pixel being one arcsecond on a side. These are rescaled to SUII using a linear transformation that scales the 1 arcsecond per pixel H α down to SUII's plate scale of 2.5 arcseconds per pixel. The image was shifted so that the Sun remained centered.

4.2 Labeled Imagery

Since this study is using supervised machine learning, labeled images of the Sun are needed. There does not exist a verified sample of human labeled solar events for more than one category of event. The closest system is the Heliophysics Event Knowledgebase (HEK) [42] which combines data mining and computer vision with data visualization techniques to create a database of labeled events. These are sometimes verified by humans but can be problematic, especially with multiple entries per event to sift through. Each event is given one of their designated labels as shown in Table 4.1.

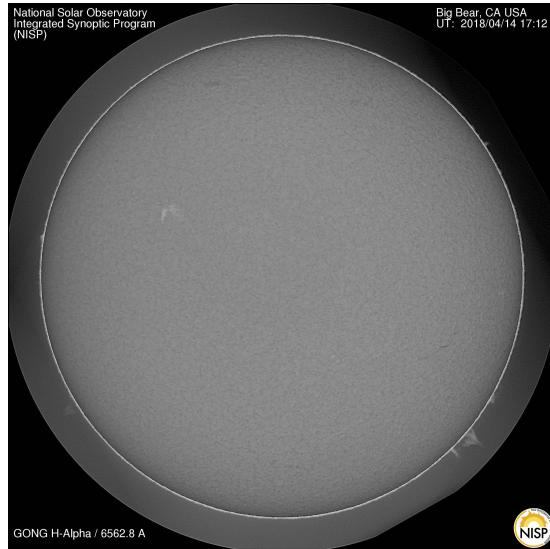


Figure 4.4: **Example H α image:** This image is a typical H α view of the Sun from the GONG network. A prominence is visible on the bottom right, and a dark filament is visible in the center of the Sun.

Event Class	Description
Active Region	Solar Active Region
Coronal Mass Ejection	Ejection of material from the solar corona
Coronal Dimming	A large-scale reduction in EUV emission
Coronal Jet	A jet-like object observed in the low corona
Coronal Wave	EIT or Morton waves spanning a large fraction of the solar disk
Emerging Flux	Regions of new magnetic flux in the solar photosphere
Filament	Solar Filament or Prominence
Filament Eruption	A sudden launching of a filament into the corona
Filament Activation	A sudden change in a filament without launching
Flare	Solar Flare
Loop	Magnetic loops typically traced out using coronal imagery
Oscillation	A region with oscillating coronal field lines
Sigmoid	S-shaped regions seen in soft X rays; indicator for flares
Spray	Surge Sudden or sustained intrusion of chromospheric material well into the corona
Sunspot	Sunspots on the solar disk
Plage	Bright areas associated with active regions
Other	Something that could not be classified good candidate for further research
Nothing	Reported Used to indicate that the particular data were examined, but had nothing noteworthy to the observer

Table 4.1: List of event labels for HEK [42].

For this project, a small curated set of solar events was produced using three solar physics experts. Four annotators were asked however only one (Annotator B) provided a complete annotation of every

image. One (Annotator A) provided a nearly complete sample. Another (Annotator C) provided only an image. The last withdrew from the project. They used an abbreviated set of solar event categories as detailed in Table 4.2.

Event Class	Description
Bright Region	Solar Active Region
Coronal Hole	Dimmer region in EUV where magnetic field lines are open
Filament	Solar Filament
Flare	Solar Flare
Prominence	Solar prominence
Limb	Edge of solar disk in EUV
Structured outer space	region off the disk with structure
Unstructured outer space	region off the disk with no structure
Unlabeled	region where no label was given or confidence was especially low
Quiet Sun	region on disk with no particularly interesting structures

Table 4.2: List of event labels for curated data gathered in this study.

Twenty-seven image groupings of the SUIVI six-band imagery were used for the labeled data as shown in Table 4.3. Each group consists of one image from each SUIVI band, an H α image, and any needed derived images used as features in the machine learning. These images were selected because they were spread over the entirety of SUIVI's operational lifetime and capture a variety of solar phenomena.

4.2.1 Annotating Software

In general, a solar annotation tool requires the following features:

- automatic solar coordinate alignment
- an interface to select a labeling class
- utility to indicate which pixels should be labeled
- ability to edit already annotated images
- navigation tools to move throughout the image

To aid annotators, I created a very simple annotation tool in Tkinter that met these requirements and more. This tool is very much a work in progress and has iteratively been improved based on user feedback. There are currently plans to move the annotation tool into a web interface to avoid a cumbersome install process. (This is a response to users having problems installing necessary dependencies because of the wide variety of machine operating systems and environments.)

A user of the tool must first configure a reference database. A reference database is an HDF5 file that coordinates the image paths and groups. This database contains meta-data about when the database was compiled, what channels it includes, and the master path to the images. Its main purpose is to coordinate image groupings. For every image grouping, there is an entry which contains a name and paths to each passband image. This structuring allows users to classify based

Group Number	Date	Time	Span of time (seconds)
0	2017-04-01	00:02:19	123
1	2017-04-15	00:01:49	150
2	2017-05-15	00:01:20	220
3	2017-05-20	00:02:07	121
4	2017-06-01	00:03:09	210
5	2017-06-15	00:02:57	120
6	2017-06-19	06:02:09	220
7	2017-07-01	00:02:18	120
8	2017-07-15	00:02:07	120
9	2017-07-28	05:02:17	122
10	2017-08-01	00:02:36	120
11	2017-08-20	00:01:07	160
12	2017-09-01	00:01:59	150
13	2017-09-08	00:01:58	180
14	2017-09-15	00:02:31	130
15	2017-10-01	12:01:56	150
16	2017-10-15	00:02:02	200
17	2017-11-02	00:02:07	210
18	2017-11-15	00:01:47	190
19	2017-11-30	00:01:51	150
20	2017-12-15	00:02:31	220
21	2018-01-01	00:03:04	210
22	2018-01-15	00:02:57	130
23	2018-02-01	00:02:54	200
24	2018-02-15	00:01:35	180
25	2018-03-01	00:01:57	220
26	2018-03-03	00:02:15	200

Table 4.3: Times for images used in the labeling image set with the number of seconds between the first and last image in the grouping.

on a shared data repository instead of having individual copies. The database can automatically be generated from a directory containing S UVI images with a provided script. In the case of this thesis, the reference database and passband images were compiled and given to the users to minimize their work.

With this database set up, the annotator calls the software from the commandline by specifying where the reference database is located and where to save labeled images. To make this easier, the users were given a parameterless version of this call which was configured to run based on the zipped reference database and images given to them.

Upon startup, the software will look at its configuration file and determine how many times each image should be classified. As discussed in Section 4.2.2, each annotator was asked to label every group twice to determine how consistent they were. If there are still images left to label, the software loads one using the reference database. Using meta-data in the images, it solves for where the Sun should be located in each image. It is configured with an option to automatically classify the entire disk of the Sun as quiet Sun, the limb of the Sun, and everything else as unstructured outer space. The location of these classes is known and do not need classification. The annotator can then draw

on top of these labels by dragging their mouse. A lasso tool is used to determine the region enclosed by their drawing and given their selected label.

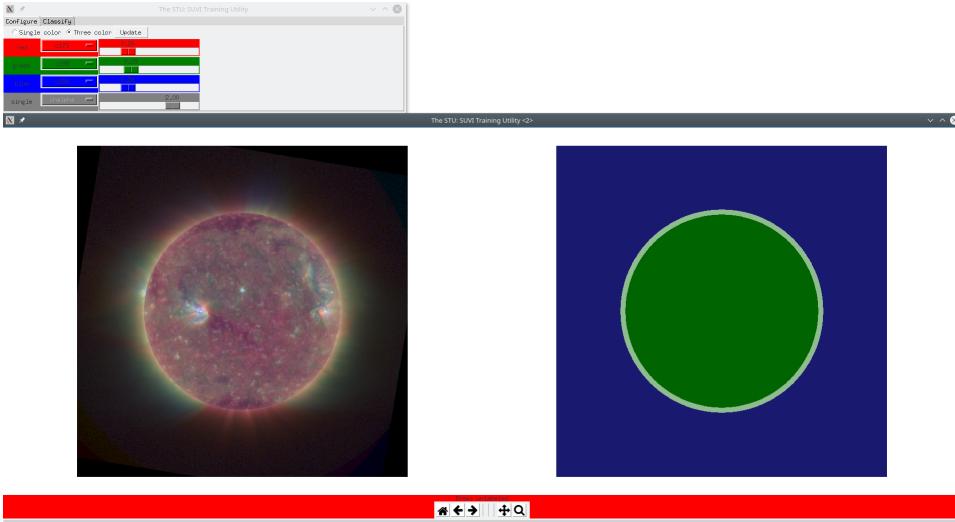


Figure 4.5: Startup of annotation software: The region on the left shows a preview image of the Sun. Initially, it is configured to show a three-color composite of the Sun, with channels scaled to show most features. The right region shows the labeling of the Sun.

In a single static image, it is difficult to label all the features because they are on different spatial, brightness, and spectral scales. Pan and zoom tools were included to allow the user to easily navigate about an image. In the configure panel, the annotator can adjust the left, preview solar image. They can choose to either see an RGB image that combines three different passbands or look at a single passband in a greyscale image. If they select the RGB they can assign the colors to their desired passband as well as what scaling factor to use. The scaling factor is included because the brightness values between different passbands differ. Without the adjustable scaling factor, the brightest passband will always dominate and more subtle detail cannot be seen. A single color option is included because some features may only be definitively visible in one passband, for example filaments in H α .

Iterated improvements

While this software is not complicated, it was not the first approach. In the very first approach, the image was automatically segmented into super-pixels using the Simple Linear Iterative Clustering (SLIC) segmentation algorithm [2]. Drawing boundaries is more time consuming than simply clicking an region of the image. SLIC was used in the hopes of producing more labeled images. However, it systematically over/under segmented specific classes, making their boundaries consistently too large or small, and did not consistently segment filaments. It instead lumped the structure into a larger superpixel with other classifications.

This approach would render the annotated database full of biases that would be difficult to

correct. Effectively, the machine learning approaches used here would approximate the segmentation algorithm instead a gold-standard human definition. Thus, SLIC was abandoned and annotators denoted boundaries with a set of points that were individually clicked. Eventually, the current drawing approach was reached. It most effectively simulates the act of hand-drawing labeled maps by space weather forecasters used at NOAA today. The current version of the annotation software is available at the author's Github.

4.2.2 Analysis of labeled data

Since this is the first digitally collected dataset of human solar annotations for multiple classes, it brings the opportunity to establish baseline metrics for human performance on the task of solar image segmentation.

Self-consistency of annotators

Each annotator was asked to label each image group twice. They were not informed when they were relabeling but instead shown the images in a random order. This allows for a check on human self-consistency. If the task were trivial, then the annotators would be expected to have perfect consistency. However, this is not the case.

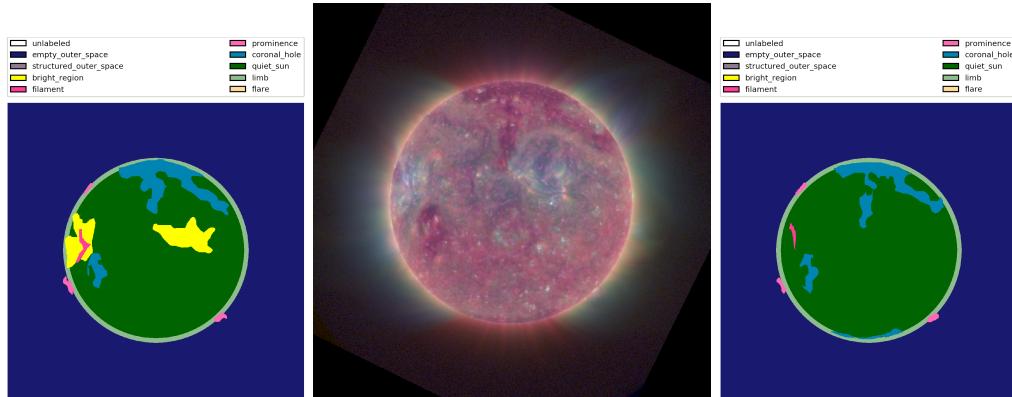


Figure 4.6: **Mistake for annotator** Left and right are annotations for Group 16, at center, by Annotator A. As you can see the annotator appears to have forgotten to label bright regions in one case. Such errors could be identified manually and cleaned from the labeled set.

The agreement can be quantified using Cohen's kappa or Fleiss's kappa [51]. Cohen's kappa can be used when there are two raters or in this case two annotations of one solar image group by the same annotator. Fleiss's kappa is used when more than two raters exist or the raters are not deliberately chosen but selected from a population at random.

Cohen's kappa is defined as

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

where p_o is the relative agreement among raters and p_e is the hypothetical chance agreement. p_o

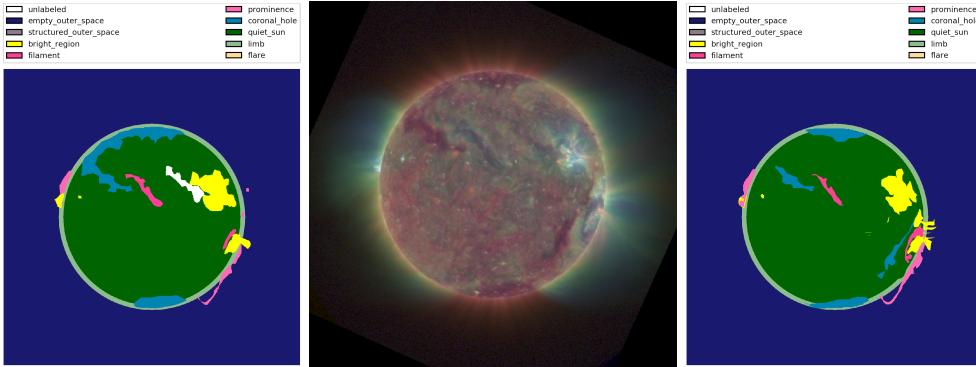


Figure 4.7: **Typical inconsistency for annotator** Left and right are annotations for Group 17, at center, by Annotator A. In this case, the annotator makes a typical kind of disagreement in the extent of boundaries. It is unclear which boundary is correct so both must be included in the labeled set.

is thus analogous to accuracy and in this case is interpreted as the number of pixels the annotator labeled the same between sessions. p_e provides context for p_o is calculated in the binary label class as the sum of the random chance of labeling true p_t and the random chance of labeling false p_f . In turn, p_t is the percentage of pixels labeled with a given class out of the total for each trial and p_f is similarly defined. With more categories, it is the sum of chance of labeling a category as something other than agreed between sessions. Cohen's kappa can be interpreted as shown in Table 4.4 [63].

κ range	Agreement interpretation	% of data that are reliable
0-.20	None	0-4%
.21-.39	Minimal	4-15%
.40-.59	Weak	15-35%
.60-.79	Moderate	35-63%
.80-.90	Strong	64-81%
> .90	Almost Perfect	82-100%

Table 4.4: Interpretation of Cohen's kappa coefficient for agreement [63]

Fleiss's kappa is defined as

$$\kappa = \frac{\overline{P} - \overline{P}_e}{1 - \overline{P}_e}$$

where $\overline{P} - \overline{P}_e$ quantifies actual observed agreement above chance and $1 - \overline{P}_e$ quantifies attainable agreement above chance. These terms are calculated where N is the number of pixels, n is the number of ratings per pixel, and k is the number of classes each pixel can be assigned. Each pixel is index $i = 1, 2, \dots, N$ and each class is indexed $j = 1, 2, \dots, k$. Then, n_{ij} is the number of annotators who labeled the i -th pixel with class j .

$$\begin{aligned}
p_j &= \frac{1}{Nn} \sum_{i=1}^N n_{ij} \\
P_i &= \frac{1}{n(n-1)} \left(\sum_{j=1}^k n_{ij}^2 - n_{ij} \right) \\
\bar{P} &= \frac{1}{N} \sum_{i=1}^N P_i \\
\bar{P}_e &= \sum_{j=1}^k p_j^2
\end{aligned}$$

Landis and Koch (1977) proposed the scale in Table 4.5 for Fleiss's kappa [51].

κ range	Interpretation
< 0	No agreement
0.0 - 0.19	Poor agreement
0.20 - 0.39	Fair agreement
0.40 - 0.59	Moderate agreement
0.60 - 0.79	Substantial agreement
0.80 - 1.00	Almost perfect agreement

Table 4.5: Interpretation of Fleiss's kappa coefficient for agreement

Since each annotator labeled twice, Cohen's kappa was used here as shown in Table 4.8.

A	BR	CH	EOS	FIL	FLA	limb	PR	QS	SOS	UL	Sum
BR	115923	0	4051	550	351	2363	1051	25379	0	0	149668
CH	25	138006	222	4526	0	12361	0	44810	0	0	199950
EOS	5758	11	6829470	0	0	0	2369	0	0	0	6837608
FIL	815	6400	183	8415	0	548	1412	4105	0	0	21878
FLA	567	0	0	0	231	19	0	0	0	0	817
limb	3585	6258	0	100	0	229637	2441	334	0	0	242355
PR	98	0	6627	0	0	6311	13417	4	0	0	26457
QS	58507	47494	0	9850	14	0	0	2229489	0	0	2345354
SOS	0	0	0	0	0	0	0	0	0	0	0
UL	1	0	0	0	0	0	0	6312	0	0	6313
Sum	185279	198169	6840553	23441	596	251239	20690	2310433	0	0	

Table 4.6: Annotator A's consistency: This table details Annotator A's labelings for each class. The rows indicate one session and the columns another. Thus, column CH and row FIL with a value of 6400 indicate that in one session Annotator A labeled 6400 pixels as coronal hole and in the other session as filament. The classes have been abbreviated as follows: bright region is BR, coronal hole is CH, empty outer space is EOS, filament is FIL, flare is FLA, limb is limb, prominence is PR, quiet Sun is QS, structured outer space is SOS, and unlabeled is UL. As seen here, Annotator A did not use the structured outer space label and did leave some pixels unlabeled.

From Table 4.8, it is clear that labeling filaments is very difficult. Oddly, flares, which were

B	BR	CH	EOS	FIL	FLA	limb	PR	QS	SOS	UL	Sum
BR	80107	605	3094	280	94	2474	141	41576	0	0	128371
CH	227	319404	2613	0	0	19069	0	58172	0	0	399485
EOS	1954	447	10256024	0	0	0	8498	0	150	0	10267073
FIL	3101	0	0	12366	0	8	186	13238	0	0	28899
FLA	570	0	0	0	573	0	0	31	0	0	1174
limb	1012	15889	0	0	0	331215	10019	0	0	0	358135
PR	0	92	3677	0	0	5229	29701	273	0	0	38972
QS	30045	73408	0	7160	0	0	518	3411629	0	0	3522760
SOS	0	0	141	0	0	0	0	0	590	0	731
UL	0	0	0	0	0	0	0	0	0	0	0
Sum	117016	409845	10265549	19806	667	357995	49063	3524919	740	0	

Table 4.7: Annotator B’s consistency: This table details Annotator B’s labelings in the same format as Table 4.6.

Class	Annotator A	Annotator B
Empty outer space	0.996	0.997
Structure outer space	N/A	0.802
Bright region	0.687	0.650
Filament	0.370	0.507
Prominence	0.568	0.674
Coronal hole	0.687	0.783
Quiet Sun	0.945	0.958
Limb	0.929	0.923
Flare	0.327	0.622
Overall	0.941	0.956

Table 4.8: Annotator consistency: For each annotator and image, Cohen’s kappa for each class is shown. Annotator B is more consistent for each class. The overall agreement is skewed by the large number of outer space pixels. Annotator A has no agreement score for structured outer space because the label was disregarded.

anticipated to be simple were inconsistent as well. However, by looking at Tables 4.6 and 4.7 it is clear that the inconsistency is not in identifying them but drawing the appropriate boundary. Since there are so few pixels, bleeding over into the surrounding bright region or being too limited in the boundary has a huge skew. Classes that are fairly automatic like empty outer space and limb, where suggestion is provided in the annotating software, have high agreement.

Agreement among annotators

Further, there is no universally agreed upon definition for each labeling class. Some cases are obvious, for example outer space. However, there are times where it is unclear which label to apply, for example a filament versus a coronal hole, or where exactly the boundary of the label should be drawn, for example which portions consist of active regions. The labeled database gives data-driven insight on where solar physics experts disagree. The disagreement may be amplified if experts from different research institutions or very different research focus are selected.

In this labeling, Annotator A did not use the structured outer space class. However, it is a rare

class, and they chose to skip the image group that Annotator B labeled substantial structured outer space.

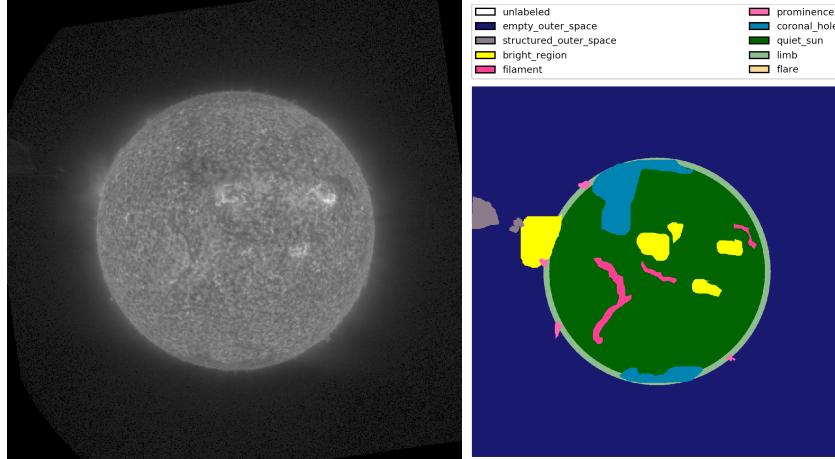


Figure 4.8: **Structured outer space discrepancy** At left is the 30.4 nm image for Group 6. Very faintly in the upper left, some structure is visible in outer space as a CME moves outward. This is only seen in the 30.4 passband. Thus, if Annotator A had chosen to label this image they may or may not have missed this class.

A much more typical example of disagreement is shown in Figure 4.10.

This disagreement can be quantified with Cohen's kappa in Table 4.9.

Class	Inter-annotator agreement
Empty outer space	0.992
Structured outer space	0.0
Bright region	0.471
Filament	0.380
Prominence	0.432
Coronal Hole	0.703
Quiet Sun	0.935
Limb	0.902
Flare	0.366

Table 4.9: Inter-annotator agreement: For each class the agreement between annotators on what pixels are declared that class is shown. Annotator A did not use the structured outer space class. Coronal hole is really the only non-automatic class that is reliable.

A third annotator submitted only a handful of images, however they provide an interesting case study in Section 4.2.3.

4.2.3 Case Study

Some of the annotator disagreement noted in Section 4.2.2 and Table 4.9 for bright regions stems from the size of bright regions labeled. Annotator A tends to draw a larger boundary including the

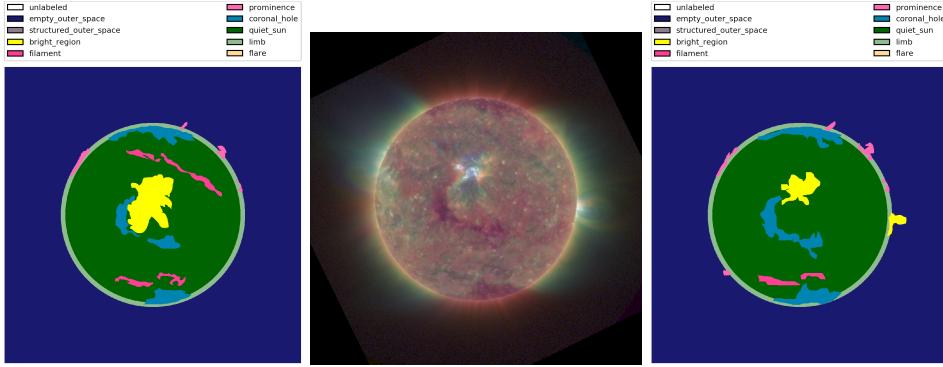


Figure 4.9: Inter-Annotator disagreement Group 1 provides a key example for disagreement among annotators. Annotator A is shown at left while Annotator B is shown at right. In the center, is the three-color image of the Sun at that time. Annotator A identified smaller coronal holes than Annotator B and included a filament Annotator B did not. This filament is not seen in H α but does look rather strong in the EUV three-color. In general, they identifies the same features but with different boundaries.

cascades of loops. Annotator B was a bit more conservative and only included the brighter regions of the cascades. Annotator C was the most conservative and only included the bright core. Along those same lines, Annotator C left a gap between the flare and surrounding bright region. The upper left bright regions were only sometimes separated even though a clear gap divides them. It appears Annotator C was also uncertain about the southernly coronal hole and chose to leave it unlabeled. However, they did identify another coronal hole that was not labeled by any others. The southern filament was only identified half the time. Annotators A and C agreed that the upper left dark region was a coronal hole while Annotator B decided it was filament. Upon further examination there is no filament visible in H α here so it likely is coronal hole. This may be why Annotator B also left it as quiet Sun. Prominences were very inconsistently labeled. This likely stems from them only being visible in 30.4 nm with special attention. Thus, a quick labeling may overlook some. Annotator A identified a spicule as a prominence in trial one. This potentially argues for another category for spicules or other "extra" features.

Overall, there are a few types of disagreements:

- Uncertainty in data for labeling: coronal holes and filaments are often unclear.
- Boundary: It is unclear where the edge of a feature should be drawn and this creates disagreement.
- Omission: Sometimes, one annotator completely skips a clearly identifiable feature another annotator sees or even worse completely skips labeling a class by mistake.

4.2.4 Labeled database

Ideally, the labeled database would be compiled with input from all annotators. However, only Annotators A and B consistently submitted labeled images. Therefore, there is no third vote to

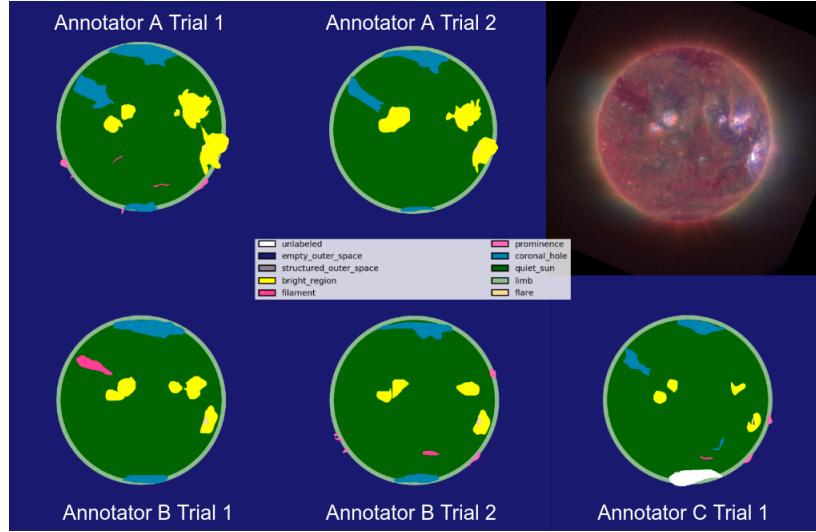


Figure 4.10: **An interesting case study from Group 13** For this group, Annotator A and Annotator B submitted double labeling and Annotator C also submitted a labeling. This group comes from 2017-09-08 during a solar flare. All labelings include the annotated flare in the bottom right of the image. However, there are many other differences discussed in Section 4.2.3.

break ties when they have disagreements. Picking a label at random for each pixel when there is a disagreement would result in a very pixelated image, for example when it is unclear whether a feature is a filament or coronal hole the feature would be a mess of pixels switching between coronal hole and filament with quiet Sun on boundary disagreements. This would create a poor gold standard database.

Instead, when there are multiple annotators or trials for an image, one is selected at random as the gold standard. This incorporates definitions from all annotators and should spread out and hopefully eliminate systematic biases by preferencing one annotator's definition. The database is currently available upon request but will subsequently be made available online at the author's website.

The database was divided in half, a training set and a labeling set, for the machine learning goals. For more details, see Section 7.1.

Chapter 5

Classifiers

In best conditions, a deep learning approach such as a convolutional neural network would be used because in other computer vision applications they have greatly outperformed other machine learning techniques. However, they require large amounts of training data. Therefore, this chapter introduces the machine learning techniques used.

Instead of treating the images as a collection of individual pixels as described in the related work (Chapter 3) or in a convolutional sense, these classifiers use neighborhoods of multichannel pixels, an compromise between the two approaches. The neighborhood of neighborhood of (x_0, y_0) is $\{(x_i, y_i) | x_0 - w \leq x_i \leq x_0 + w, y_0 - h \leq y_i \leq y_0 + h\}$ where w and h are the number of pixels on either side of the reference pixel (x_0, y_0) used in the horizontal and vertical dimension respectively. Each multichannel pixel is a $m \times 1$ -dimension column vector where each entry corresponds to a spectral passband or derived channel. The derived channels are ratios of spectral passbands or spatial information but could be extended to any kind of value represented as a matrix of values the same size as the input images, in this case 1280x1280 pixels. The goal is to assign each multichannel pixel a label from Table 4.2.

5.1 Naive Bayesian Maximum Likelihood

This approach mimics the maximum likelihood estimation of [75] described in Section 3.3.2. It is modified for pixel neighborhoods. If $w = h = 0$, this model is identical to [75].

As before, every pixel is classified into exactly one class using multispectral ultraviolet images, $H\alpha$, and other derived channels as features. A pixel is a member of a pixel neighborhood that is used in classifying. Therefore, a pixel at (i, j) is the reference pixel to the neighborhood $\{(i, j) | x_0 - j \leq x_i \leq x_0 + j, y_0 - k \leq y_i \leq y_0 + k\}$ where each neighborhood pixel has c channels and can be described as a vector:

$$x_{(i,j)} = [x_1 \quad x_2 \quad \dots \quad x_c]^T$$

Thus, the neighborhood of pixels is generalized as a $(2w + 1) \times (2h + 1) \times c$ dimensional matrix, the amalgamation of all neighborhood pixels. For each $x_{(i,j)}$ pixel a label w_k from the set of classes W .

This approach employs Bayes' Theorem:

$$P(w_k|x_{(i,j)}) = \frac{P(x_{(i,j)}|w_k)P(w_k)}{P(x_{(i,j)})}$$

Since $P(x_{(i,j)})$ is not a function of the label classification, it can be ignored.

$$P(w_k|x_{(i,j)}) \propto P(x_{(i,j)}|w_k)P(w_k)$$

Following [75]:

$$P(w_k|x_{(i,j)}) \propto P(x_{(i,j)}|w_k)$$

This approach is the maximum likelihood solution to this problem.

Each class is modeled as a multivariate normal distribution, i.e. for each potential label there is an archetypal example pixel and all pixels with that label should be distributed normally about it. This is overly constraining if any given class has multiple distinct modes with respect to the selected data. However, this can be solved adding more subclasses, for example diving the bright regions into more types. Alternatively, this section could be replaced with a Gaussian mixture approach. The multivariate distribution for class w_k is characterized by a mean vector μ_k and covariance matrix C_k which are calculated as:

$$\mu_k = \frac{\sum_{x \in W_k} x}{|W_k|}$$

where W_k is the collection of pixel neighborhoods where the reference pixel has label w_k . Similarly,

$$C_k = \frac{\sum_{x \in W_k} [x - \mu_k] \times [x - \mu_k]^T}{|W_k|}$$

These mean vectors and covariance matrices characterize the class. Given this characterization for class w_k one can calculate the conditional probability of a pixel $x_{(i,j)}$ having label w_k :

$$P(x_{(i,j)}|w_k) = \frac{1}{\sqrt{(2\pi)^h} \sqrt{|C_k|}} \exp \left(\frac{-1}{2} \times (x_{(i,j)} - \mu_k)^T \times C_k^{-1} \times (x_{(i,j)} - \mu_k) \right)$$

Thus, the label that maximizes this probability is assigned for the reference pixel.

No smoothing is done in this version as in Rigler et al. (2012) [75].

5.2 Random Forest

The random forest is an ensemble of decision tree classifiers and are used in a variety of Earth remote sensing applications [70, 77, 16, 48, 59]. Solar image segmentation is very similar to remote Earth sensing, just with a different subject matter and often fewer passbands. In this application, an ensemble of D decision trees with maximum depth Q are used to form the random forest **change to optimal**. The empirically best depth and number is determined in Section 7.7.1. The random forest is given the same types of input as the maximum likelihood classifier: the EUV channels,

the H α channel, and derived channels (spatial, line ratios). The random forest implementation in Scikit-Learn is used [67].

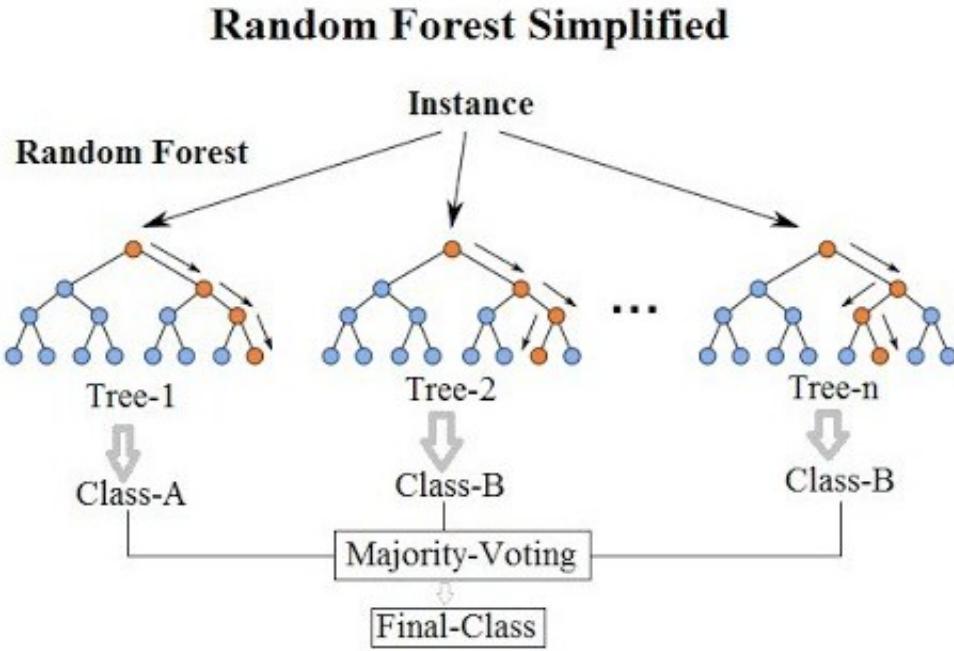


Figure 5.1: An example random forest construction: This helpful graphic shows the general layout of how a random forest works. Multiple decision trees are evaluated to determine a majority class vote [65]

Another benefit of the random forest is its ability to indicate what features are most important for classification. This is briefly explored in Section 7.8.

5.3 Feed-forward Neural Network

An artificial neural network is a network of nodes and weights. The neural network takes inputs with a node for each input type. These inputs are transformed and passed onto layers of intermediate nodes. Each node has an activation function. The inputs are combined with different weights through a transfer function. The combined input is then compared via an activation function to an activation threshold. If the combined input is strong enough the node activates and passes a transformed value along to subsequent nodes in the next layer. These propagate forward to an output layer that provides the labeled class.

A simple fully-connected feed forward network is employed here consisting of three densely connected hidden layers with an input layer with similar neighborhood inputs and an output layer for classes. The hidden layers had 16, 16, and 8 nodes each with relu activation.

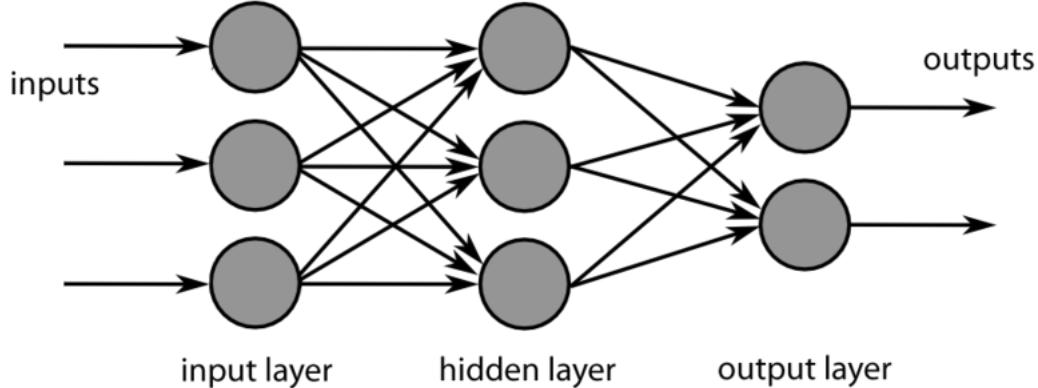


Figure 5.2: **An example neural network construction:** This helpful graphic shows the general layout of a feedforward neural network. The number of nodes and hidden layers may vary. The size of the input layer is determined by the input features, and the output size is determined by the classification task [18].

5.4 Convolutional Neural Network

A convolutional neural network was attempted with the expectation of poor results because of insufficient training. The system was based on the Fully Convolutional Neural Network approach by Long, Shelhamer, & Darrell (2015) [58], a method to take arbitrarily sized image input and create similarly sized image output with classification labels. The specific implementation used was a modification of Sarath Shekkizar's Tensorflow implementation [83]. This approach was initialized with VGG-16 pretrained model, although it is unclear if this training helps for solar image segmentation as well.

Chapter 6

Software package

This chapter details how the software is designed and how future features can be included. The codebase is available at the author's Github.

6.1 Overview

The software is available via Github in a Python package called "Smachy" for *solar machine learning in Python*. It consists of a few modules to partition the code: `gui`, `filter`, `train`, `classify`, `evaluate`, and `config`. The `gui` module contains the annotation software. The `filter` module provides support for creating derivative images such as entropy images and ratios of EUV images. The `train` module organizes the training objects for the various classes and provides a `TrainingLoader` class for ease of use. The `classify` module performs solar image classification using the aforementioned classifiers. Evaluation is done through the `evaluate` module and provides support for performance metrics and confusion matrices. Finally, the `config` module defines global configuration parameters such as label mapping. This will eventually be moved to a json configuration file.

6.2 Data structuring

Flexible Image Transport System or FITS is the standard image format for astronomical and solar images. The input solar images as well as any derived images are stored in this format. It has the benefit of providing convenient support for varied meta-data. For example, the annotated images have additional meta-data that indicates which images they were labeled using. All the channel names used in annotating are listed in a pipe-delimited entry under the keyword `CHANNELS`. Then, each channel name has an associated keyword with a relative path to the precise image used, in example `C131` might map to `131/NCEI_SUVI_COMPOSITE_131_G16_20170908_000148_v_0_2_0.fits`. The annotated image is an image of integer values where each integer corresponds to a label. The order of these integers is stored in the meta-data under the `CLASSES` keyword in case it is ever changed within the software. For example, the `CLASSES` entry may be:

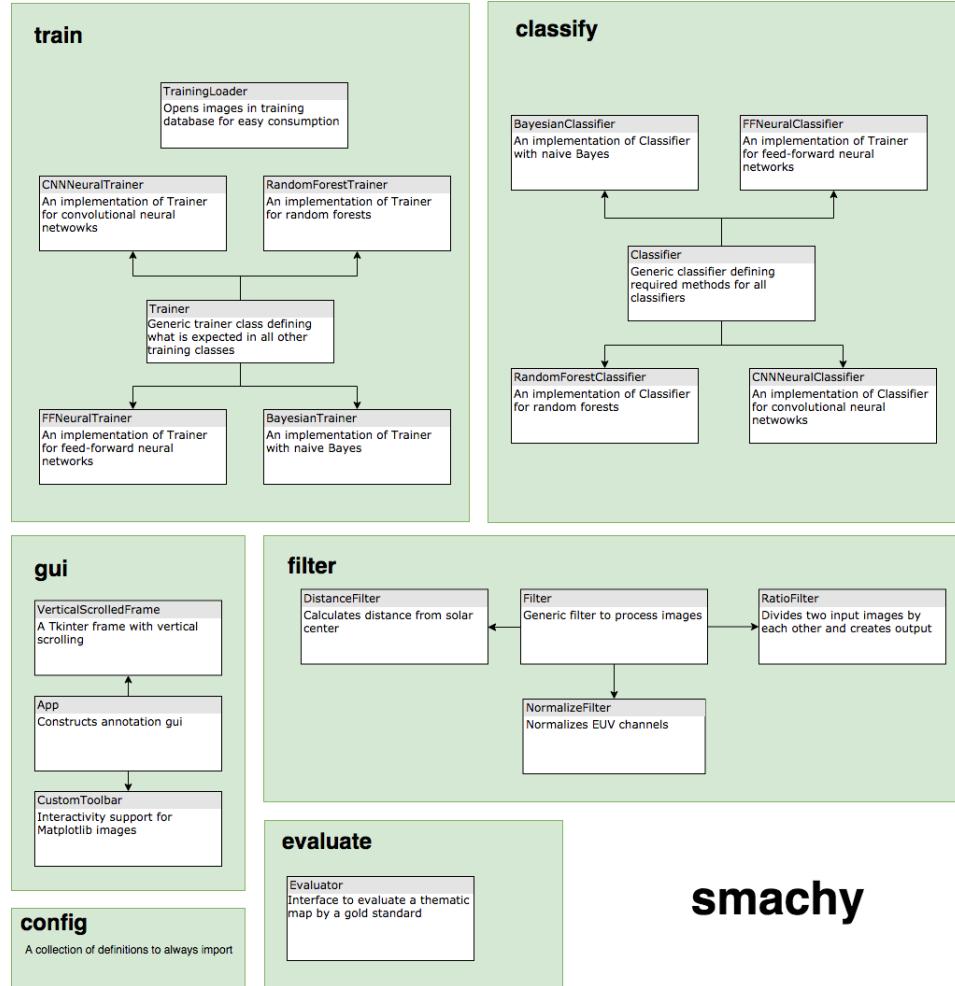


Figure 6.1: **Code layout**: A outline of the major classes in each module of **smachy**

```
unlabeled|empty_outer_space|structured\outer_space\bright_region|
filament|prominence|coronal_hole|quiet_sun|limb|flare
```

This can be split on the pipe and indicates that values of 0 map to `unlabeled`, values of 1 map to `empty_outer_space`, values of 2 map to `structured_outer_space`, and so forth sequentially.

Thus, an annotated image can be opened to recover which solar images are associated with a multichannel image group. Alternatively, the group database, an HDF5 database, also includes this data. The group database is generated first and handed to various portions of the software so it knows which images to open for analysis or annotation. This database is often called `groups.h5`. It retains when it was calculated as well as the path to the directory of images. Then, the `groups` entry contains a dictionary of each group. A group is given an unique designator starting with "g" based upon the observation time. For example, "g20170401000218" indicates the group is for April 1, 2017 at 00:02:18. The entry for the group then contains the relative paths for images associated

with it:

```
'g20170401000218':
{'files':
 {'c131': '131/NCEI_SUVI_COMPOSITE_131_G16_20170401_000143_v_0_2_0.fits',
 'c171': '171/NCEI_SUVI_COMPOSITE_171_G16_20170401_000303_v_0_2_0.fits',
 'c195': '195/NCEI_SUVI_COMPOSITE_195_G16_20170401_000213_v_0_2_0.fits',
 'c284': '284/NCEI_SUVI_COMPOSITE_284_G16_20170401_000333_v_0_2_0.fits',
 'c304': '304/NCEI_SUVI_COMPOSITE_304_G16_20170401_000133_v_0_2_0.fits',
 'c94': '94/NCEI_SUVI_COMPOSITE_94_G16_20170401_000233_v_0_2_0.fits',
 'chalpha': 'halpha/gong_6567_20170401_000130.fits'},
'timespread': 123.0},
```

This groups database can be automatically generated by using the provided `build_file_databases.py` script with commandline argument of directory path to images. Then, each channel should be separated into subdirectories of the desired channel name. The script will create n groups where n is the smallest number of images in any channel. It will then greedily pick other images such that the group has the smallest spread over observation time.

6.3 Annotation GUI

The annotation GUI's functionality is described in Section 4.2.1. It is implemented in Tkinter in Python. In hindsight, this was not the best decision because Tkinter appears differently on different operating systems. Even more importantly, since it was part of the `smachy` package it required installation of various dependencies to run. Annotators were given installation scripts and instructions to use virtual environments, but it still was ultimately more complicated than a more portable, stand-alone Python program, a very portable Java JAR, or a web application.

To make matters even easier, two execution scripts for the GUI are provided. The first `run_gui.py` requires very specific input regarding which image group to annotate, where to save, and other flags that can be accessed through the commandline help menu. Since annotators would be using the same directory structuring, an easier script `easy_gui.py` was provided. It defaulted all the parameters and could be run without designating any flags. It expects to be run in a directory containing a `groups.h5`, a directory of images, and a directory called `labeled` of annotated images. It will run until every group in the groups database has the requested number of annotated images.

6.4 Trainers

The `train` module implements a generic `Trainer` with some helper methods and abstract methods required to be implemented in any specific `Trainer`. A `Trainer` takes input of a list of annotated images, a dictionary of the paths to images associated with the annotated images, list of labels to classify on, and a list of image channels to utilize. Determining these parameters is made easy with the `TrainingLoader` class. The `Trainer` then provides helper functions to access pixels associated

with any label. Any specific `Trainer` must implement a `train()` method and a `make_classifier()` method. Train does the specific training routine for that kind of classifier and sets up the appropriate data structures. `make_classifier()` calls the associated `Classifier` object and returns it with the trained parameters.

Training is done with the appropriate `train_*.py` script, for example `train_bayesian.py` for the Bayesian model. The required inputs and outputs are described in the commandline help accessible with the `-h` flag. To do training en masse, `train_all.sh` runs all the specific training scripts with specified inputs.

Thus, to add a new trainer inherit the generic `Trainer` and implement a specific training script in the same fashion as one of the existing scripts.

6.5 Classifiers

The implementations of the classifiers was done in a polymorphic fashion so that they can be exchanged easily when running and evaluating performance. Each classifier is designed to include the following methods:

- `classify_pixel`: Given a single pixel's feature set, classify it
- `classify_image`: Iterates over an entire image and performs classification
- `save`: saves a classifier to a Python serialized file
- `load`: loads a classifier from a Python serialized file

The parent `Classifier` class also provides helper methods to load essential data.

Similar to training, there are specific classifier scripts for running and a `classify_all.sh` to run them all. To add a new `Classifier` inherit the generic one and provide model specific classification methods as well as a way to save and load. Then, provide a classification script, and it should run.

6.6 Filters

More complicated features than raw data values are desirable. The `filter` module provides support for generating output images. The main `Filter` class provides support for opening and saving an image/derived image as well as a general purpose `process` method. This `process` method utilizes the `_filter` method to process a single group in a group database. The `filter` method is implemented separately in every subclass. For example, the `RatioFilter` implements a `_filter` method that takes a group and then passes each channel pair to a `_filter_one` method that actually performs the division to scale an image. All the implemented filters can be called in the provided `run_filters.py` script.

6.7 Evaluator

To support consistent and convenient analysis, the `Evaluator` code in the `evaluate` module will compare a set of computer classified images to the gold standard annotated images. It will in turn generate confusion matrices and other performance metrics for analysis. It is accessible through the `run_evaluate.py` script.

Chapter 7

Experiments and Results

This chapter describes various experiments performed on the data and the conclusions derived while working toward the optimal machine learning solar image classifier, most importantly it explores feature selection. It begins by introducing the methods of evaluation in Section 7.1 and then proceeds through the experiments.

7.1 Evaluation approaches

The database of annotated images was broken into two categories: a training set and a testing set. Fourteen images were used in training and thirteen were used for testing. These were randomly selected so that no observational time was biased. The classifiers in the experiments are trained solely on the images in the training set and only exposed to the testing set for evaluation.

7.1.1 Qualitative

For each experiment, the generated thematic maps from the trained classifiers are qualitatively compared to the testing images. Several qualities are inspected:

- Are all classes expected in this image present? If not, what classes are missing or overly present?
- How do the boundaries of each region compare to the testing image?
- Is the classification noisy with neighboring pixels varying?
- Is the limb correctly aligned?
- Are there any other peculiarities?

7.1.2 Standard machine learning metrics

In addition to qualitative inspection, the testing set and classifier generated results are compared quantitatively. For each class several statistics are calculated based on true positive count

(TP), true negative count (TN), false positive count (FP), and false negative count (FN). For a given class label, TP is the number of pixels classified with that class and the gold standard database agreed. On the other hand, FP is the number of pixels classified with that class when the gold standard database disagreed. Similarly, TN is number of times the classifier correctly did not assign the class and FN is when the classifier labeled a class other than the designated class. From these values, several metrics of performance are derived. One of the more widely known metrics is accuracy, how often a classifier is correct:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

However, accuracy is not as helpful when classes are skewed, for example flares are very uncommon so never using the label of flare trivially optimizes accuracy. Other metrics are more revealing in this case. Precision is defined as:

$$\text{precision} = \frac{TP}{TP + FP}$$

It measures the ability of the classifier to not label a pixel the given class when it is not. Recall is another metric that characterizes the classifier's ability to identify all pixels of a given class:

$$\text{recall} = \frac{TP}{TP + FN}$$

Both precision and recall are important but are often anti-correlated. The f1 score combines them:

$$f1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

7.1.3 Confusion matrix

The aforementioned metrics do not indicate exactly how the classifier fails. For example, the accuracy for coronal holes might be low, but it isn't clear what the coronal pixels are being classified as instead. A confusion matrix like the example in Figure 7.1 helps remedy this.

The confusion matrices in this thesis are presented normalized such that each row should sum to 100 (± 1 for rounding). Confusion matrices might typically be presented as raw counts, but it is not beneficial in this case since some classes greatly outnumber others. So, in this matrix, prominences are correctly labeled for 5% of prominence pixels in the gold standard database. They are often labeled as limb instead, 74% of the time and sometimes as empty outer space 16%. This is not too unexpected since prominences are on the limb. While the columns cannot be summed, they do provide some insight. For example, the predicted label of limb comes from prominence and limb. It is important to note that limb is correctly classified 81% of the time, so the 74% between prominence and limb should not be interpreted to mean limb is equally likely to be labeled as prominence and limb.

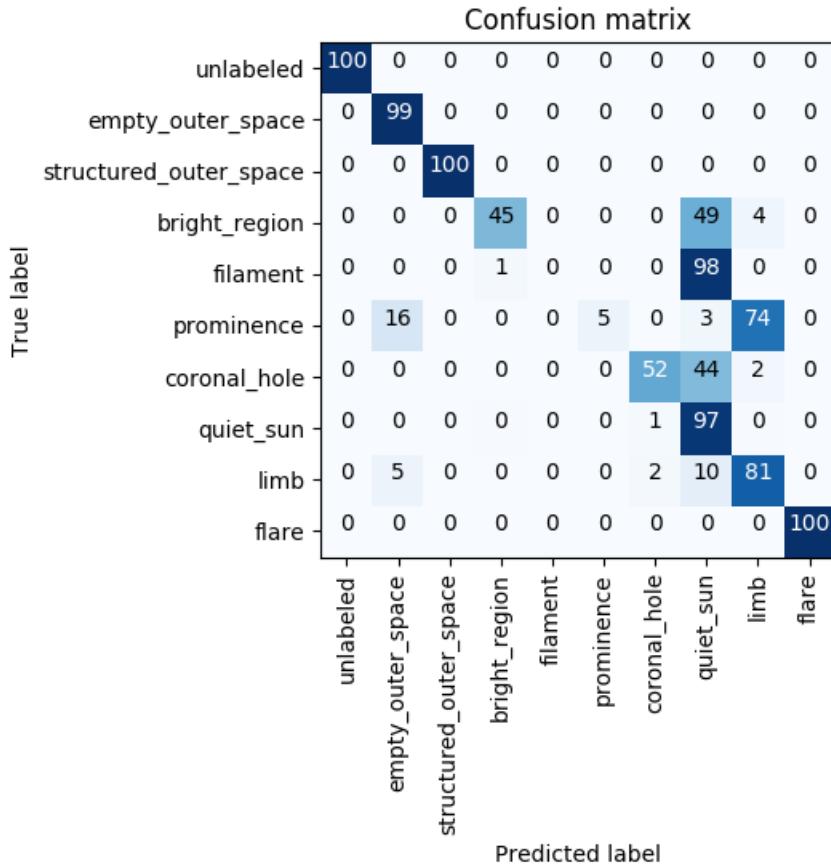


Figure 7.1: **An example confusion matrix:** A confusion matrix is a method of showing label classification. On the vertical axis the correct label is shown while on the horizontal axis the annotator's label is shown. Then, looking at the cell defined shows the percentage of labels in that region. Ideally, the diagonal should be 100 indicating every feature was labeled with the correct label always. The confusion matrices presented in this thesis are normalized.

7.2 Baseline results

To provide a baseline, the classifiers were first run with a neighborhood size of 0 meaning that classification only looked at a pixel at a time using only the EUV imagery from SUI. For each classifier, a typical example of the classification is shown:

As shown in Figure 7.2, the Bayesian classifier created a very noisy result with pixels even in outer space being misclassified. The baseline random forest classifier failed to identify large filaments on the disk of the Sun that were clearly identified in all training sets. However, it did a reasonable job with coronal holes and identified more structure in bright regions than human annotators recorded. The neural network was very similar to random forest results.

Class	Bayesian	Random Forest	Neural network
Empty outer space			
Structured outer space			
Bright region			
Filament			
Prominence			
Coronal Hole			
Quiet Sun			
Limb			
Flare			

Table 7.1: Baseline performance: When given only a single pixel view and EUV channels, here is the performance for each classifier..

7.3 Normalization of data

Since each EUV channel is not equally bright, the Bayesian classifier may be skewed. The high values in one channel can effectively wash out subtle value differences in another potentially. Thus, the EUV channels were replaced with their normalized counterparts. Normalization was done to the disk of the Sun by first calculating where the disk of the Sun should be in each image. The solar disk location was determined by using orbital information instead of edge detection. For each class, the median pixel value on the disk was calculated over all images and designated the normalization coefficient. Each channel must be normalized to the same value over all images or else when comparing an image without a flare/bright region to one with the actual pixel value loses context. Without consistent coefficients, a pixel of medium value in the flare image would be many times brighter than a pixel of the same value in the unflared image. The calculated coefficients are in Table 7.2.

Channel	Normalization coefficient
94	0.036344
131	0.072936
171	0.675868
195	0.782299
284	0.964538
304	7.265509
H α	2888.6

Table 7.2: Normalization coefficients for each channel

7.4 Line ratios

7.5 Spatial features

copy from notes

7.6 H α inclusion

As seen in Figure 7.5, filaments and coronal holes cannot be differentiated using spectral information in the EUV channels of SUIVI. Spatial approaches may discern between them because filaments are long, skinny strands while coronal holes are typically more round and extended. Alternatively, H α spectral information easily differentiates between coronal holes and filaments as shown in Figure 7.6.

In this experiment, the suite of classifiers are first trained with only EUV channels. Then, they are trained with both EUV and H α channels. As shown in Figure 7.7 the [insert results discussion here](#).

7.7 Hyperparameter selection

7.7.1 Random forest

7.7.2 Neural network

7.8 Random forest feature importance

7.9 Comparison to existing classifiers

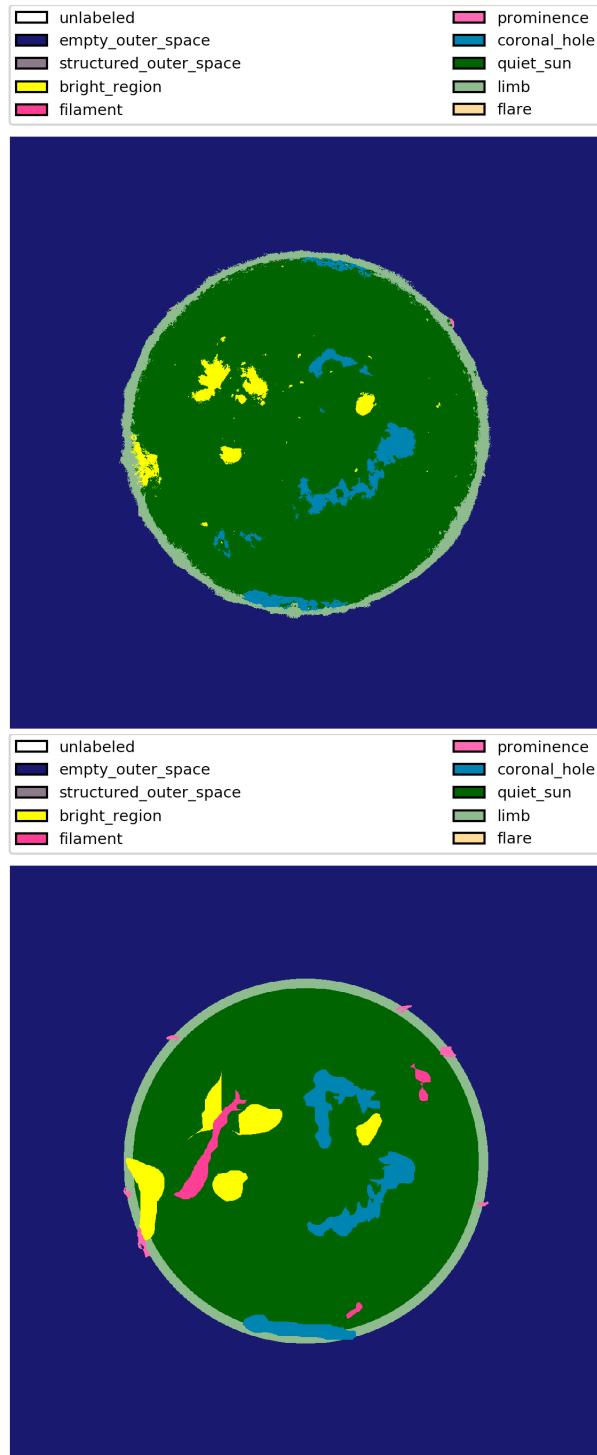


Figure 7.2: Baseline performance for Bayesian classifier

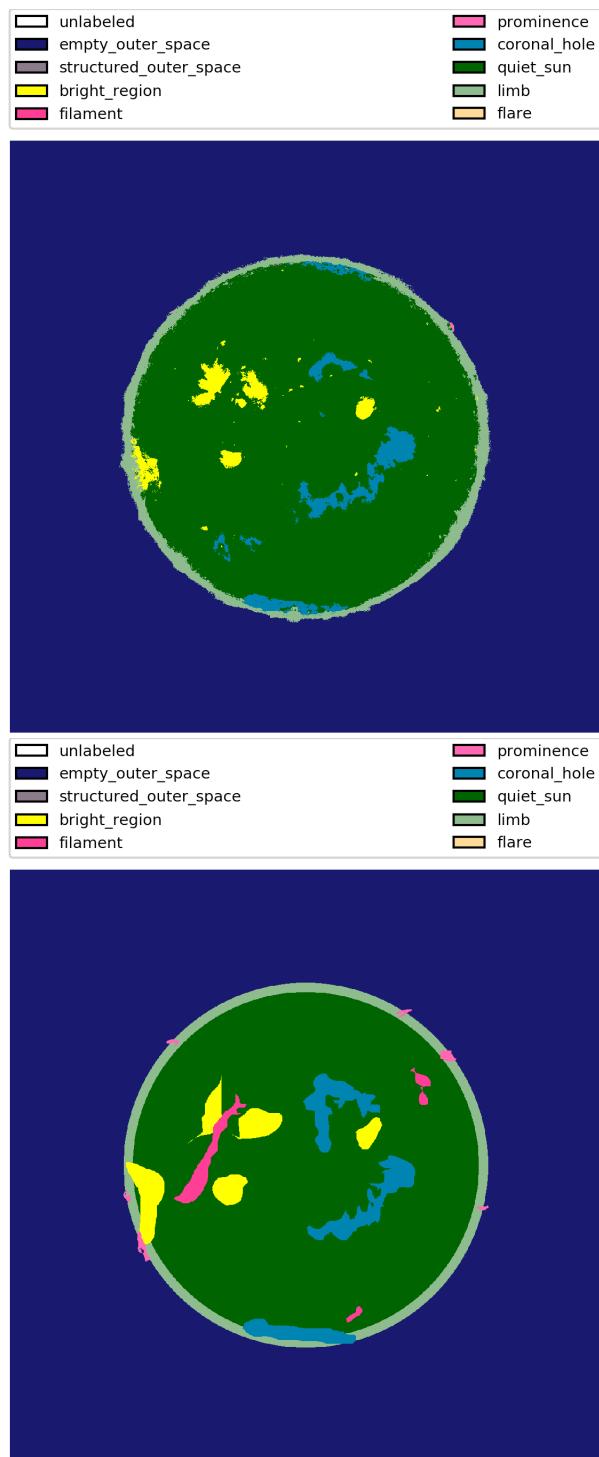


Figure 7.3: Baseline performance for random forest

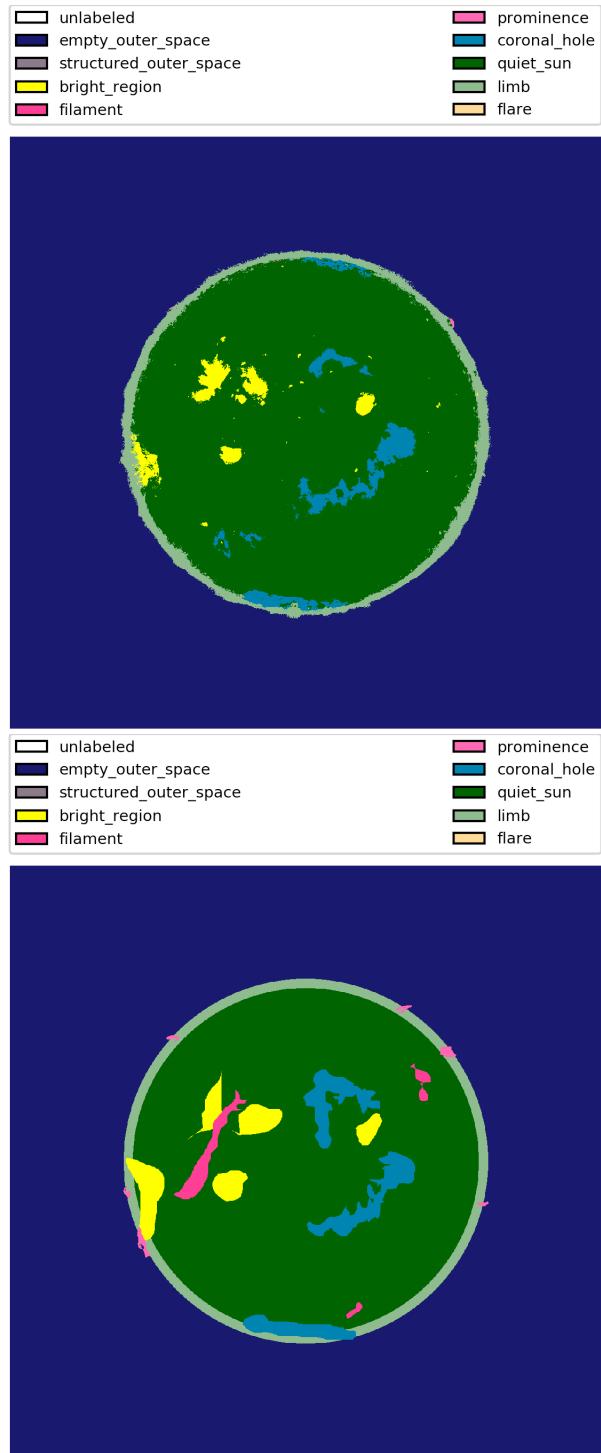


Figure 7.4: Baseline performance for neural network



Figure 7.5: **Degeneracy of filament and coronal hole spectrally with EUV images** This histogram compares 19.5 nm pixel values for coronal holes, top, and filaments, bottom. They nearly overlap completely making it difficult for classification methods to key on sufficient themes for classification. All other EUV channels suffered from complete overlap. H α images differentiate filaments and coronal holes as shown in Figure 7.6.



Figure 7.6: **Coronal hole and filament in H α** In H α images, filaments are dark strand structures while coronal holes have no apparent signature. This defining difference complements the degeneracy in EUV images and boosts classifier performance.

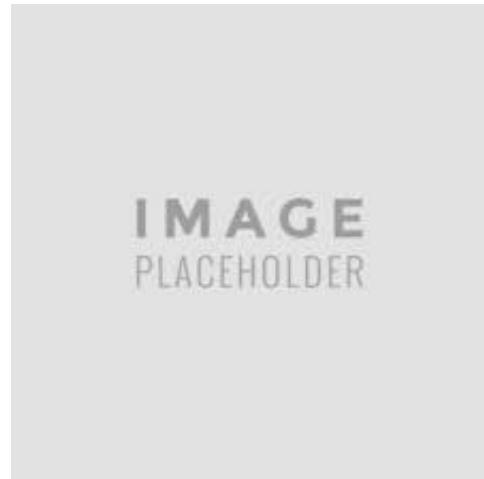


Figure 7.7: H α inclusion experiment results. **add relevant caption based on results**

Chapter 8

Applications

Solar image segmentation is helpful for space weather forecasting but is not the end goal for research. Solar physicists can utilize it as a tool in research applications. To aid in finding relevant data, a database of observations with key statistical features can be compiled from thematic maps as described in 8.1. Alternatively, the segmentation can be used directly as a boundary as illustrated in an exploration of fractal dimension and class properties in Section 8.2.

8.1 Database building

copy from notes

8.2 Fractal dimension and class properties

copy from fractals paper

Chapter 9

Conclusion

This chapter provides an overview of potential future work and research questions as well as a summary of the results.

9.1 Future work

There are many avenues for future inquiry stemming from this project. A few are presented here.

9.1.1 Stability over solar cycles

The Sun goes through 11-year cycles where behavior changes drastically. During the "solar minimum" phase, where the Sun is now, coronal holes are much more prominent and extend further down to the solar equator, there are fewer flares and bright regions. As solar maximum approaches the distribution and characteristics of the classes will change. This may require new training data. It also may require the addition of a solar cycle feature that links data to when in a solar cycle it was taken. One way to study this is by looking at satellite with a longer history, such as SDO, or waiting until SUIV has a long history of observations.

9.1.2 Converting digitized synoptic charts

Since obtaining annotated images was among the most difficult portions of this project, it would be nice to utilize as much existing annotation as possible. Space weather forecasters at NOAA create thematic maps daily by hand and have done so since 1972. These are digitized into a PDF format.

Despite their accessibility, they are difficult to use as a training database and propose a project all to themselves. Figure 9.1 illustrates the variety of drawing styles. While they are drawn on the same template (earlier templates are slightly different but anything within satellite observation interest would be consistent), the annotations vary and present a challenge for computer vision recognition. The color of red for prominences and flares can change. In the bottom right, the neutral magnetic lines do not extend to the solar limb, likely because of an uncertainty in their location, while in other

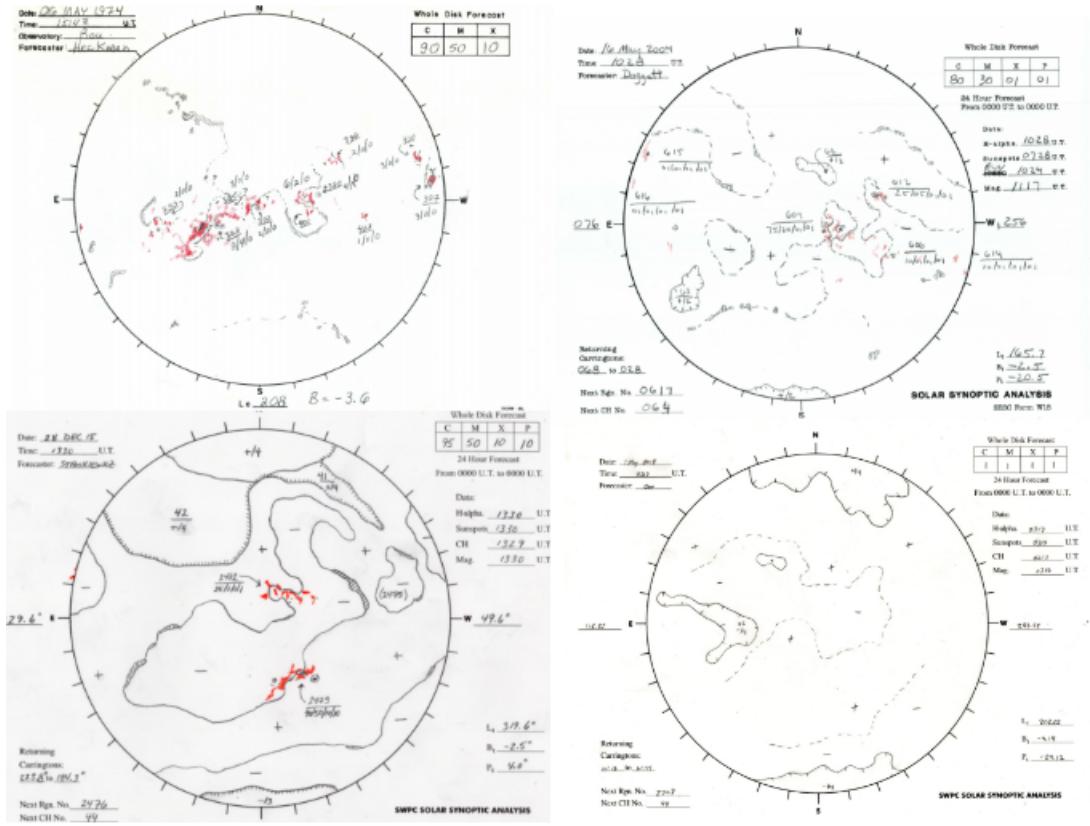


Figure 9.1: **Daily synoptic drawing** Here are four synoptic charts drawn at NOAA ranging from 1974 in the upper left to 2018 in the bottom right. Even more recent synoptic charts vary in style because different forecasters have slightly different styles.

figures they do. Coronal holes, indicated by boundaries with lines perpendicular to the boundary, are complicated since hatchings are drawn at variable widths and determining the interior of the the coronal hole relies upon that. Accompanying annotations are sometimes drawn next to features and other times drawn further away and linked by an arrow. Finally, these thematic maps do not detail features off the disk (except sometimes prominences) and do not include all the features (filaments) discussed in this thesis. If these challenges were overcome, the level of detail and care in the drawings makes them ideal for the labeled database.

In the meantime, a more realistic goal is for forecasters to make a daily drawing using the annotation software. This is being pursued through a more accessible online tool with a funding proposal in the process of submission.

9.1.3 Prediction

Potentially the most interesting and beneficial avenue for development is transitioning the classification paradigm to a prediction paradigm. There are many systems that predict solar flares

within one to two days with fair success with machine learning techniques such as Bayesian methods [93] and support vector machines [12]. However, these systems are still far from satisfactory: "To date, various systems and models designed to predict the occurrence of solar flares have made significant progress, but the achieved prediction performances are far from what is required by operations teams" [3]. Attempts at using CNNs have not been as successful, potentially because of insufficient data [71]. However, that model is based off of sunspot counts instead of imagery. Using the approaches presented here with an added time domain and solar flare prediction label could be very productive. The annotated data set in this thesis was too sparse for an investigation into prediction.

9.2 Summary

Bibliography

- [1] ABOUDARHAM, J., SCHOLL, I., FULLER, N., FOUESNEAU, M., GALAMETZ, M., GONON, F., MAIRE, A., AND LEROY, Y. Automation of Meudon Synoptic Maps. In *The Physics of Chromospheric Plasmas* (May 2007), P. Heinzel, I. Dorotovič, and R. J. Rutten, Eds., vol. 368 of *Astronomical Society of the Pacific Conference Series*, p. Heinzel.
- [2] ACHANTA, R., SHAJI, A., SMITH, K., LUCCHI, A., FUJIA, P., AND SÜSSTRUNK, S. Slic superpixels. *EPFL Technical Report*, 149300 (2010), 15.
- [3] AHMED, O. W., QAHWAJI, R., COLAK, T., HIGGINS, P. A., GALLAGHER, P. T., AND BLOOMFIELD, D. S. Solar flare prediction using advanced feature extraction, machine learning, and feature selection. *Solar Physics* 283, 1 (Mar 2013), 157–175.
- [4] ANZANELLO, M. J., ORTIZ, R. S., LIMBERGER, R., AND MARIOTTI, K. Performance of some supervised and unsupervised multivariate techniques for grouping authentic and unauthentic viagra and cialis. *Egyptian Journal of Forensic Sciences* 4, 3 (2014), 83 – 89.
- [5] ASCHWANDEN, M. J. Image Processing Techniques and Feature Recognition in Solar Physics. *Solar Physics* 262 (Apr. 2010), 235–275.
- [6] AULANIER, G., DÉMOULIN, P., SCHRIJVER, C. J., JANVIER, M., PARIAT, E., AND SCHMIEDER, B. The standard flare model in three dimensions. II. Upper limit on solar flare energy. *Astronomy and Astrophysics* 549 (Jan. 2013), A66.
- [7] BANDA, J. M., AND ANGRYK, R. A. *Large-Scale Region-Based Multimedia Retrieval for Solar Images*. Springer International Publishing, Cham, 2014, pp. 649–661.
- [8] BARRA, V., DELOUILLE, V., AND HOCHÉDEZ, J.-F. Segmentation of extreme ultraviolet solar images via multichannel fuzzy clustering. *Advances in Space Research* (2008), 917–925.
- [9] BARRA, V., DELOUILLE, V., AND HOCHÉDEZ, J.-F. Segmentation, tracking and characterization of solar features from eit solar corona images. *SCIA 2009, LNCS 5575* (2009), 199–208.
- [10] BENKHALIL, A., ZHARKOVA, V., ZHARKOV, S., AND IPSON, S. Active region detection and verification with the solar feature catalogue. *Solar Physics* 235, 1 (2006), 87–106.
- [11] BEZDEK, J. C. *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media, 2013.

- [12] BOBRA, M. G., AND COUVIDAT, S. Solar flare prediction using sdo/hmi vector magnetic field data with a machine-learning algorithm. *The Astrophysical Journal* 798, 2 (2015), 135.
- [13] BORDA, R. A. F., MININNI, P. D., MANDRINI, C. H., GÓMEZ, D. O., BAUER, O. H., AND ROVIRA, M. G. Automatic solar flare detection using neural network techniques. *Solar Physics* 206, 2 (2002), 347–357.
- [14] BRATSOLIS, E., AND SIGELLE, M. Solar image segmentation by use of mean field fast annealing. *Astronomy and Astrophysics Supplement Series* 131, 2 (1998), 371–375.
- [15] CABALLERO, C., AND ARANDA, M. A comparative study of clustering methods for active region detection in solar euv images. *Solar Physics* 283, 2 (2013), 691–717.
- [16] CLARK, M. L. Mapping land cover with hyperspectral and multispectral satellites using machine learning and spectral mixture analysis. In *Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International* (2016), IEEE, pp. 513–516.
- [17] CLIVER, E. W., AND SVALGAARD, L. The 1859 solar–terrestrial disturbance and the current limits of extreme space weather activity. *Solar Physics* 224, 1 (Oct 2004), 407–422.
- [18] COMMONS, W. File:multilayerneuralnetworkbigger english.png — wikimedia commons, the free media repository, 2014. [Online; accessed 2-May-2018].
- [19] CURTO, J., BLANCA, M., AND MARTÍNEZ, E. Automatic sunspots detection on full-disk solar images using mathematical morphology. *Solar Physics* 250, 2 (2008), 411–429.
- [20] DE WIT, T. D. Fast Segmentation of Solar Extreme Ultraviolet Images. *Solar Physics* 239 (Dec. 2006), 519–530.
- [21] DEFOREST, C. E. Noise-gating to Clean Astrophysical Image Data. *Astrophysical Journal* 838 (Apr. 2017), 155.
- [22] DELABOUDINIÈRE, J. P., ARTZNER, G. E., BRUNAUD, J., GABRIEL, A. H., HOCHÉDEZ, J. F., MILLIER, F., SONG, X. Y., AU, B., DERE, K. P., HOWARD, R. A., KREPLIN, R., MICHELS, D. J., MOSES, J. D., DEFISE, J. M., JAMAR, C., ROCHUS, P., CHAUVINEAU, J. P., MARIOGE, J. P., CATURA, R. C., LEMEN, J. R., SHING, L., STERN, R. A., GURMAN, J. B., NEUPERT, W. M., MAUCHERAT, A., CLETTE, F., CUGNON, P., AND VAN DESSEL, E. L. Eit: Extreme-ultraviolet imaging telescope for the soho mission. *Solar Physics* 162, 1 (Dec 1995), 291–312.
- [23] DOMINGO, V., FLECK, B., AND POLAND, A. I. Soho: The solar and heliospheric observatory. *Space Science Reviews* 72, 1 (Apr 1995), 81–84.
- [24] DRIESMAN, A., HYNES, S., AND CANCRO, G. The stereo observatory. *Space Science Reviews* 136, 1 (Apr 2008), 17–44.

- [25] EASTWOOD, J. P., BIFFIS, E., HAPGOOD, M. A., GREEN, L., BISI, M. M., BENTLEY, R. D., WICKS, R., MCKINNELL, L., GIBBS, M., AND BURNETT, C. The economic impact of space weather: Where do we stand? *Risk Analysis* 37, 2 (2017), 206–218.
- [26] FOX, K. C. Nasa stereo observes one of the fastest cmes on record. https://www.nasa.gov/mission_pages/stereo/news/fast-cme.html, Aug. 2012. Accessed: 2017-12-07.
- [27] FULLER, N., ABOUDARHAM, J., AND BENTLEY, R. D. Filament Recognition and Image Cleaning on Meudon H α Spectroheliograms. *Solar Physics* 227 (Mar. 2005), 61–73.
- [28] GHOSE, M., PRADHAN, R., AND GHOSE, S. S. Decision tree classification of remotely sensed satellite data using spectral separability matrix. *International Journal of Advanced Computer Science and Applications* 1, 5 (2010), 93–101.
- [29] GONZALEZ, R. C., AND WOODS, R. E. In *Digital Image Processing, 2nd edn.* (2002), Prentice-Hall, Inc., p. 613.
- [30] GREEN, J. L., BOARDSEN, S., ODENWALD, S., HUMBLE, J., AND PAZAMICKAS, K. A. Eyewitness reports of the great auroral storm of 1859. *Advances in Space Research* 38, 2 (2006), 145 – 154. The Great Historical Geomagnetic Storm of 1859: A Modern Look.
- [31] GROUP, G.-R. Instruments: Solar ultraviolet imager (suvi). <https://www.goes-r.gov/spacesegment/suvi.html>, 2018.
- [32] GUERRA, L., MCGARRY, L. M., ROBLES, V., BIELZA, C., LARRANAGA, P., AND YUSTE, R. Comparison between supervised and unsupervised classifications of neuronal cell types: A case study. *Developmental Neurobiology* 71, 1 (Jan. 2011), 71–82.
- [33] HALAIN, J.-P., BERGHMANS, D., SEATON, D. B., NICULA, B., DE GROOF, A., MIERLA, M., MAZZOLI, A., DEFISE, J.-M., AND ROCHUS, P. The swap euv imaging telescope. part ii: In-flight performance and calibration. *Solar Physics* 286, 1 (Aug 2013), 67–91.
- [34] HANDY, B. N., ACTON, L. W., KANKELBORG, C. C., WOLFSON, C. J., AKIN, D. J., BRUNER, M. E., CARAVALHO, R., CATURA, R. C., CHEVALIER, R., DUNCAN, D. W., EDWARDS, C. G., FEINSTEIN, C. N., FREELAND, S. L., FRIEDLAENDER, F. M., HOFFMANN, C. H., HURLBURT, N. E., JURCEVICH, B. K., KATZ, N. L., KELLY, G. A., LEMEN, J. R., LEVAY, M., LINDGREN, R. W., MATHUR, D. P., MEYER, S. B., MORRISON, S. J., MORRISON, M. D., NIGHTINGALE, R. W., POPE, T. P., REHSE, R. A., SCHRIJVER, C. J., SHINE, R. A., SHING, L., STRONG, K. T., TARBELL, T. D., TITLE, A. M., TORGERSON, D. D., GOLUB, L., BOOKBINDER, J. A., CALDWELL, D., CHEIMETS, P. N., DAVIS, W. N., DELUCA, E. E., McMULLEN, R. A., WARREN, H. P., AMATO, D., FISHER, R., MALDONADO, H., AND PARKINSON, C. The transition region and coronal explorer. *Solar Physics* 187 (July 1999), 229–260.
- [35] HE, K., GKIOXARI, G., DOLLÁR, P., AND GIRSHICK, R. B. Mask R-CNN. *CoRR abs/1703.06870* (2017).

- [36] HIGGINS, P. A., GALLAGHER, P. T., MCATEER, R. J., AND BLOOMFIELD, D. S. Solar magnetic feature detection and tracking for space weather monitoring. *Advances in Space Research* 47, 12 (2011), 2105–2117.
- [37] HILL, F., FISCHER, G., FORGACH, S., GRIER, J., LEIBACHER, J. W., JONES, H. P., JONES, P. B., KUPKE, R., STEBBINS, R. T., CLAY, D. W., INGRAM, R. E. L., LIBBRECHT, K. G., ZIRIN, H., ULRICHI, R. K., WEBSTERI, L., HIEDA, L. S., LABONTE, B. J., LU, W. M. T., SOUSA, E. M., GARCIA, C. J., YASUKAWA, E. A., KENNEWELL, J. A., COLE, D. G., ZHEN, H., SU-MIN, X., BHATNAGAR, A., AMBASTHA, A., AL-KHASHLAN, A. S., ABDUL-SAMAD, M.-S., BENKHALDOUN, Z., KADIRI, S., SÁNCHEZ, F., PALLÉ, P. L., DUHALDE, O., SOLIS, H., SAÁ, O., AND GONZÁLEZ, R. The Global Oscillation Network Group site survey, 2: Results. *Solar Physics* 152 (July 1994), 351–379.
- [38] HILL, F., FISCHER, G., GRIER, J., LEIBACHER, J. W., JONES, H. B., JONES, P. P., KUPKE, R., AND STEBBINS, R. T. The Global Oscillation Network Group site survey. 1: Data collection and analysis methods. *Solar Physics* 152 (July 1994), 321–349.
- [39] HILL, F., MARTENS, P., YOSHIMURA, K., GURMAN, J., HOURCLÉ, J., DIMITOGLOU, G., SUÁREZ-SOLÁ, I., WAMPLER, S., REARDON, K., DAVEY, A., BOGART, R. S., AND TIAN, K. Q. The Virtual Solar Observatory: A Resource for International Heliophysics Research. *Earth Moon and Planets* 104 (Apr. 2009), 315–330.
- [40] HOWARD, R. A., MOSES, J. D., VOURLIDAS, A., NEWMARK, J. S., SOCKER, D. G., PLUNKETT, S. P., KORENDYKE, C. M., COOK, J. W., HURLEY, A., DAVILA, J. M., THOMPSON, W. T., ST CYR, O. C., MENTZELL, E., MEHALICK, K., LEMEN, J. R., WUELTER, J. P., DUNCAN, D. W., TARBELL, T. D., WOLFSON, C. J., MOORE, A., HARRISON, R. A., WALTHAM, N. R., LANG, J., DAVIS, C. J., EYLES, C. J., MAPSON-MENARD, H., SIMNETT, G. M., HALAIN, J. P., DEFISE, J. M., MAZY, E., ROCHUS, P., MERCIER, R., RAVET, M. F., DELMOTTE, F., AUCHERE, F., DELABOUDINIÈRE, J. P., BOTHMER, V., DEUTSCH, W., WANG, D., RICH, N., COOPER, S., STEPHENS, V., MAAHS, G., BAUGH, R., McMULLIN, D., AND CARTER, T. Sun earth connection coronal and heliospheric investigation (secchi). *Space Science Reviews* 136, 1 (May 2008), 67.
- [41] HUANG, Q., XIA, X., AND LO, D. Supervised vs unsupervised models: A holistic look at effort-aware just-in-time defect prediction. In *Software Maintenance and Evolution (ICSME)*.
- [42] HURLBURT, N., CHEUNG, M., SCHRIJVER, C., CHANG, L., FREELAND, S., GREEN, S., HECK, C., JAFFEY, A., KOBASHI, A., SCHIFF, D., SERAFIN, J., SEGUIN, R., SLATER, G., SOMANI, A., AND TIMMONS, R. Heliophysics event knowledgebase for the solar dynamics observatory (sdo) and beyond. *Solar Physics* 275, 1 (Jan 2012), 67–78.
- [43] KEITH HUGHITT, STEVEN CHRISTE, F. M. M. E. S. M. J. I. A. S. D. P.-S., ET AL. SunPy: Open source solar data analysis tools for python, 2012–.

- [44] KRISHNAPURAM, R., AND KELLER, J. M. A possibilistic approach to clustering. *IEEE transactions on fuzzy systems* 1, 2 (1993), 98–110.
- [45] KRISHNAPURAM, R., AND KELLER, J. M. The possibilistic c-means algorithm: insights and recommendations. *IEEE transactions on Fuzzy Systems* 4, 3 (1996), 385–393.
- [46] KRIZHEVSKY, A., SUTSKEVER, I., AND HINTON, G. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems* 25, 1106–1114.
- [47] KUCUK, A., BANDA, J. M., AND ANGRYK, R. A. Solar event classification using deep convolutional neural networks. *ICAISC 2017, Part1, LNAI 10245* (2017), 118–130.
- [48] KULKARNI, A. D., AND SHRESTHA, A. Multispectral image analysis using decision trees. *International Journal of Advanced Computer Science and Applications* 8, 6 (2017).
- [49] LA TORRE, F. C.-D., GONZLEZ-TREJO, J. I., REAL-RAMREZ, C. A., AND HOYOS-REYES, L. F. Fractal dimension algorithms and their application to time series associated with natural phenomena. *Journal of Physics: Conference Series* 475, 1 (2013), 012002.
- [50] LAMB, R., ANGRYK, R., AND MARTIENS, P. An example based image retrieval system for the trace repository. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on* (2008), IEEE, pp. 1–4.
- [51] LANDIS, J. R., AND KOCH, G. G. The measurement of observer agreement for categorical data. *Biometrics* 33, 1, 159–174.
- [52] LANG, K. *The Sun from Space*. Astronomy and Astrophysics Library. Springer Berlin Heidelberg, 2008.
- [53] LECUN, Y., BOSER, B., DENKER, J. S., HENDERSON, D., HOWARD, R. E., HUBBARD, W., AND JACKEL, L. D. Backpropagation applied to handwritten zip code recognition. *Neural computation* 1, 4 (1989), 541–551.
- [54] LEE, H., AND KWON, H. Going deeper with contextual cnn for hyperspectral image classification. *IEEE Transactions on Image Processing* 26, 10 (2017), 4843–4855.
- [55] LEMEN, J. R., TITLE, A. M., AKIN, D. J., BOERNER, P. F., CHOU, C., DRAKE, J. F., DUNCAN, D. W., EDWARDS, C. G., FRIEGLAENDER, F. M., HEYMAN, G. F., HURLBURT, N. E., KATZ, N. L., KUSHNER, G. D., LEVAY, M., LINDGREN, R. W., MATHUR, D. P., McFEATERS, E. L., MITCHELL, S., REHSE, R. A., SCHRIJVER, C. J., SPRINGER, L. A., STERN, R. A., TARBELL, T. D., WUELTER, J.-P., WOLFSON, C. J., YANARI, C., BOOK-BINDER, J. A., CHEIMETS, P. N., CALDWELL, D., DELUCA, E. E., GATES, R., GOLUB, L., PARK, S., PODGORSKI, W. A., BUSH, R. I., SCHERRER, P. H., GUMMIN, M. A., SMITH, P., AUKER, G., JERRAM, P., POOL, P., SOUFLI, R., WINDT, D. L., BEARDSLEY, S., CLAPP, M., LANG, J., AND WALTHAM, N. The atmospheric imaging assembly (aia) on the solar dynamics observatory (sdo). *Solar Physics* 275, 1 (Jan 2012), 17–40.

- [56] LI, M., ZANG, S., ZHANG, B., LI, S., AND WU, C. A review of remote sensing image classification techniques: The role of spatio-contextual information. *European Journal of Remote Sensing* 47, 1 (2014), 389–411.
- [57] LI, Y., QI, H., DAI, J., JI, X., AND WEI, Y. Fully convolutional instance-aware semantic segmentation. *arXiv preprint arXiv:1611.07709* (2016).
- [58] LONG, J., SHELHAMER, E., AND DARRELL, T. Fully convolutional networks for semantic segmentation. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2015), pp. 3431–3440.
- [59] LOWE, B., AND KULKARNI, A. Multispectral image analysis using random forest. *International Journal on Soft Computing* 6, 1 (2015).
- [60] MARTENS, P., ATTRILL, G., DAVEY, A., ENGELL, A., FARID, S., GRIGIS, P., KASPER, J., KORRECK, K., SAAR, S., SAVCHEVA, A., ET AL. Computer vision for the solar dynamics observatory (sdo). *Solar Physics* 275, 1-2 (2012), 79–113.
- [61] MATAN, O., BURGES, C. J., LECUN, Y., AND DENKER, J. S. Multi-digit recognition using a space displacement neural network. In *Advances in neural information processing systems* (1992), pp. 488–495.
- [62] MCATEER, R. J., GALLAGHER, P. T., IRELAND, J., AND YOUNG, C. A. Automated boundary-extraction and region-growing techniques applied to solar magnetograms. *Solar Physics* 228, 1-2 (2005), 55–66.
- [63] MCHUGH, M. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 22, 3 (2012), 276282.
- [64] MULLER, C. The carrington solar flares of 1859: Consequences on life. *Origins of Life and Evolution of the Biosphere* 44, 3 (2014), 185–195.
- [65] NUGGETS, K. Random forests(r), explained. <https://www.kdnuggets.com/2017/10/random-forests-explained.html>.
- [66] OLMEDO, O., ZHANG, J., WECHSLER, H., POLAND, A., AND BORNE, K. Automatic detection and tracking of coronal mass ejections in coronagraph time series. *Solar Physics* 248, 2 (2008), 485–499.
- [67] PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COURNAPEAU, D., BRUCHER, M., PERROT, M., AND DUCHESNAY, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
- [68] PESNELL, W. D., THOMPSON, B. J., AND CHAMBERLIN, P. C. The Solar Dynamics Observatory (SDO). *Solar Physics* 275 (Jan. 2012), 3–15.

- [69] PREMINGER, D. G., WALTON, S. R., AND CHAPMAN, G. A. Solar feature identification using contrasts and contiguity. *Solar physics* 202, 1 (2001), 53–62.
- [70] PUISSANT, A., ROUGIER, S., AND STUMPF, A. Object-oriented mapping of urban trees using random forest classifiers. *International Journal of Applied Earth Observation and Geoinformation* 26 (2014), 235–245.
- [71] QAHWAJI, R., AND COLAK, T. Automatic short-term solar flare prediction using machine learning and sunspot associations. *Solar Physics* 241, 1 (Mar 2007), 195–211.
- [72] QU, M., SHIH, F. Y., JING, J., AND WANG, H. Automatic solar filament detection using image processing techniques. *Solar Physics* 228, 1 (2005), 119–135.
- [73] REN, S., HE, K., GIRSHICK, R., AND SUN, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- [74] REVATHY, K., LEKSHMI, S., AND NAYAR, S. R. P. Fractal-Based Fuzzy Technique For Detection Of Active Regions From Solar Images. *Solar Physics* 228 (May 2005), 43–53.
- [75] RIGLER, J. E., HILL, S. M., REINARD, A. A., AND STEENBURGH, R. A. Solar thematic maps for space weather operations. *Space Weather* 10 (2012).
- [76] RILEY, P. On the probability of occurrence of extreme space weather events. *Space Weather* 10, 2 (Feb 2012), 1–12.
- [77] SALAS, E. A. L., BOYKIN, K. G., AND VALDEZ, R. Multispectral and texture feature application in image-object analysis of summer vegetation in eastern tajikistan pamirs. *Remote Sensing* 8, 1 (2016), 78.
- [78] SCHUH, M., ANGRYK, R., PILLAI, K., BANDA, J., AND MARTENS, P. A large-scale solar image dataset with labeled event regions. In *2013 IEEE International Conference on Image Processing* (Sept. 2013), IEEE.
- [79] SCHUH, M. A., ANGRYK, R. A., AND MARTENS, P. C. A large-scale dataset of solar event reports from automated feature recognition modules. *Journal of Space Weather and Space Climate* 6, 27 (May 2016), A22.
- [80] SEATON, D. B., BERGHMANS, D., NICULA, B., HALAIN, J.-P., DE GROOF, A., THIBERT, T., BLOOMFIELD, D. S., RAFTERY, C. L., GALLAGHER, P. T., AUCHÈRE, F., DEFISE, J.-M., D’HUYS, E., LECAT, J.-H., MAZY, E., ROCHUS, P., ROSSI, L., SCHÜHLE, U., SLEMZIN, V., YALIM, M. S., AND ZENDER, J. The swap euv imaging telescope part i: Instrument overview and pre-flight testing. *Solar Physics* 286, 1 (Aug 2013), 43–65.
- [81] SEATON, D. B., DARNE, J., HILL, S. M., EDWARDS, C., MATHUR, D., SABOLISH, D., SEQUIN, R., MILLER SHAW, M., SHING, L., SLATER, G. L., AND GOPAL, V. First Results from the Solar Ultraviolet Imager on GOES-16. In *AAS/Solar Physics Division Meeting* (Aug. 2017), vol. 48 of *AAS/Solar Physics Division Meeting*, p. 305.01.

- [82] SEATON, D. B., AND DARTEL, J. M. Observations of an Eruptive Solar Flare in the Extended EUV Solar Corona. *Astrophysical Journal Letters* 852 (Jan. 2018), L9.
- [83] SHEKKIZAR, S. Fcn.tensorflow. <https://github.com/shekkizh/FCN.tensorflow>, 2017.
- [84] SONG, Q., WANG, J.-S., FEND, X.-S., AND ZHANG, X.-X. The design of solar synoptic chart for space weather forecast. *Solar and Stellar Flares and their Effects on Planets Proceedings IAU Symposium No. 320* (2015).
- [85] SPACE WEATHER OPERATIONS, R., AND FORCE, M. T. National space weather action plan. *United States Department of Commerce* (2015).
- [86] SUNPY COMMUNITY, T., MUMFORD, S. J., CHRISTE, S., PÉREZ-SUÁREZ, D., IRELAND, J., SHIH, A. Y., INGLIS, A. R., LIEDTKE, S., HEWETT, R. J., MAYER, F., HUGHITT, K., FREIJ, N., MESZAROS, T., BENNETT, S. M., MALOCHA, M., EVANS, J., AGRAWAL, A., LEONARD, A. J., ROBITAILLE, T. P., MAMPAEY, B., IVÁN CAMPOS-ROZO, J., AND KIRK, M. S. SunPy Python for solar physics. *Computational Science and Discovery* 8, 1 (Jan. 2015), 014009.
- [87] SZEGEDY, C., LIU, W., JIA, Y., SERMANET, P., REED, S., ANGUELOV, D., ERHAN, D., VANHOUCKE, V., AND RABINOVICH, A. Going deeper with convolutions. In *IEEE Conference on Computer Vision and Pattern Recognition* (June 2015), IEEE, pp. 1–9.
- [88] TSO, B., AND MATHER, P. M. Classification methods for remotely sensed data.
- [89] VALLEE, J. P. Spiral-shaped interplanetary magnetic field, 1998.
- [90] VERBEECK, C., DELOUILLE, V., MAMPAEY, B., AND DE VISSCHER, R. The spoca-suite: Software for extraction, characterization, and tracking of active regions and coronal holes on euv images. *Astronomy & Astrophysics* 561 (2014), A29.
- [91] VILJANEN, A., MYLLYS, M., AND NEVANLINNA, H. Russian geomagnetic recordings in 18501862 compared to modern observations. A11.
- [92] VISSCHER, R., DELOUILLE, V., DUPONT, P., AND DELEDALLE, C.-A. Supervised classification of solar features using prior information. *J. Space Weather Space Clim.* 5, A34 (2015).
- [93] WHEATLAND, M. S. A bayesian approach to solar flare prediction. *The Astrophysical Journal* 609, 2 (2004), 1134.
- [94] YOUNG, A. Solar structure. <http://www.thesuntoday.org/the-sun/solar-structure/>, 2018.
- [95] YU, G. W., AND ZEMEL, R. S. Comparing supervised vs. unsupervised image segmentation methods.
- [96] ZHARKOVA, V. V., IPSON, S. S., ZHARKOV, S. I., BENKHALIL, A., ABOUDARHAM, J., AND BENTLEY, R. D. A full-disk image standardisation of the synoptic solar observations at the Meudon Observatory. *Solar Physics* 214 (May 2003), 89–105.

- [97] ZHARKOVA, V. V., AND SCHETININ, V. Filament recognition in solar images with the neural network technique. *Solar Physics* 228, 1 (May 2005), 137–148.
- [98] ZHENG, S., ZENG, X., LIN, G., ZHAO, C., FENG, Y., TAO, J., ZHU, D., AND XIONG, L. Sunspot drawings handwritten character recognition method based on deep learning. *New Astronomy* 45 (2016), 54–59.