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

April 14, 2018

Contents

List of Figures

List of Tables

Abstract

The new Solar Ultraviolet Imager (SUVI) instruments aboard NOAAs 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 NOAAs 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.

Using an expert-trained, naive Bayesian classifier to label each pixel, 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, from which we could characterize the distribution of pixels associated with each theme. Given new images, the classifier assigns 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. The algorithm excellently identifies coronal holes but fails to consistently detect filaments and prominences. We compare the Bayesian classifier to an artificial neural netowrk, one of our attempts to overcome the aforementioned limitations. These results are very promising and encourage future research into an ensemble classification approach.

This abstract will be updated throughout the thesis process.

Acknowledgments

I'd like to thank Dan Seaton and Jon Darnel for guiding me during my summer work, the project that motivated this thesis. I'd like to thank Jay Pasachoff for taking me on a winter study travel course that helped me discover the REU which lead to said project. And, I'd like to thank Jon Park for helping me organize and embark on an independent thesis.

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 oclock ordinary print could be read by light" (The New York Times, New York Herald, Washington Daily National Intelligencer, September 2, 3, 5, 1859) [21]. Normally, aurora, a visible manifestation of material streaming 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 [13]. 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 [21]. Events like this, while rare, are fairly periodic, with a 12% probability of another within the decade [50].

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 [32]. A much smaller event occurred in 2003, known now as the Halloween event [42, 62]. 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 [16]. 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 [16].

There are many much more frequent but less devastating solar events, discussed in Section 2.1.2 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 [17]. Within minutes to hours of a solar flare, the ionospheric disturbance can interrupt radio communications [57]. For proper safety protocols to be enacted, fast warning of an event is necessary. This thesis explores solar feature

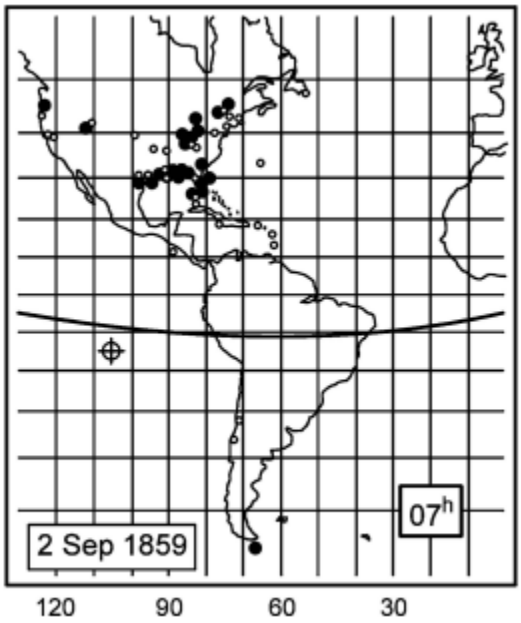

Figure 1.1: **Spatial extent of Carrington Event**: As noted by Cliver and Svalgaard (2004) the Carrington event was observed to very low latitudes in the Americas. Closed circles represent overhead aurora; open circles represent visible aurora. The heavy curved line denotes the geomagnetic equator and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

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. 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 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 in Section 2.1 and machine learning 2.2. This includes a definition of the relevant solar structures to classify. Then, Chapter 3 documents prior work in solar imge classification including both the unsupervised systems in Section 3.3.1 and superverised 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 6. Two applications of the solar classification are explored in Chapter 7: the labeled solar database in Section 7.1 and fractal dimension estimastion for solar features in Section 7.2. Finally, Chapter 8 outlines the results of the entire project and potential future work.

Chapter 2

Background

This chapter describes the background of solar weather as well as an introduction to machine learning used in this project.

2.1 Solar Physics

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. 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 [17]. It is critical to have early warning of such events to enact proper protections. The new Solar Ultraviolet Imager (SUVI), a camera aboard the National Oceanic and Atmospheric Administration's (NOAA) GOES-R weather satellite, responds to this need as it captures an image of the Sun every 10 seconds [54]. This flood of data cannot be digested by human forecasters quick enough, necessitating machine learning image classification to identify important events. Machine learning also allows for organized archival of the projected 20 years of imaging, allowing researchers to find relevant data to test new solar models. This is an interesting supervised machine learning problem because it operates on movies of multi-spectral, noisy images. SUVI by no means is the first satellite to explore this problem. This paper explores existing approaches to the astronomical problem as well as looking at parallel solutions in Earth remote sensing applications. Finally, it explores state-of-the-art computer vision research and new approaches to image segmentation and classification.

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. 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 [17]. It is critical to have early warning of such events to enact proper protections. The new Solar Ultraviolet Imager (SUVI), a camera aboard the National Oceanic and Atmospheric Administration's (NOAA) GOES-R weather satellite, responds to this

need as it captures an image of the Sun every 10 seconds [54]. This flood of data cannot be digested by human forecasters quick enough, necessitating machine learning image classification to identify important events. Machine learning also allows for organized archival of the projected 20 years of imaging, allowing researchers to find relevant data to test new solar models. This is an interesting supervised machine learning problem because it operates on movies of multi-spectral, noisy images. SUVI by no means is the first satellite to explore this problem. This paper explores existing approaches to the astronomical problem as well as looking at parallel solutions in Earth remote sensing applications. Finally, it explores state-of-the-art computer vision research and new approaches to image segmentation and classification.

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 material high in the solar atmosphere by magnetic arches
Prominences	A filament observed off the disk of the Sun
Coronal holes	Regions 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 on the surface of the Sun thought to be precursors to flares
Quiet corona	Parts of the Sun devoid of activity

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

Earth is situated inside the Sun's tremendous magnetic field. Features on the Sun evolve and expose Earth to new environmental conditions. The interaction between the Sun's magnetic field and Earth is called space weather.

2.1.1 Solar Structures

2.1.2 Space Weather Phenomenon

2.1.3 Space Weather Mitigation

2.2 Machine Learning

2.2.1 Machine Learning Experiment Design

2.2.2 Classifiers

2.2.3 Computer Vision

Chapter 3

Related work

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 as it rotates on its axis every 25 days. This results in magentic field lines in a spiral pattern called the "Parker Spiral." As such, an ejection of material from the Sun on the eastern side of the Sun will reach Earth quicker than an ejection from the western 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 reponsible for synoptic maps' invention Song et al. (2015) describe the necessary components of synoptic charts [56]. 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 [56]. 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 NOOA'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 [67]. 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].

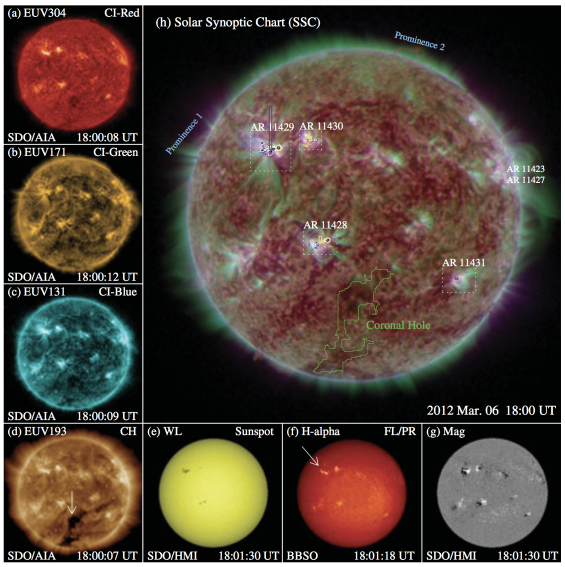

Figure 3.1: **Synoptic map example**: Song et al. (2015) propose the solar synoptic map to include a labeled composite image with different wavelength images accompanying to provide a full image of solar activity [56].

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 [58]. Unsupervised approaches to the problem use this tool and others like it to fetch Hα, ultraviolet, x-ray, white light, magnetic, and other forms of solar images. They utilize from the Solar Dynamics Observatory, the Solar and Heliospheric Observatory, Solar Terrestrial Relations Observatory, Transition Region And Coronal Explorer, the Global Oscillation Network

Group, 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.

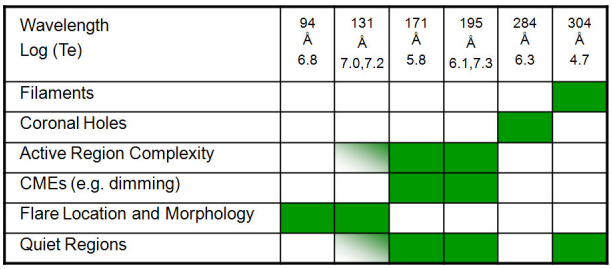

Figure 3.2: **Wavelengths**: This diagram indicates which SUVI wavelengths are most helpful in identifying different space weather events [54].

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 [52]. 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 [4]. This dataset and tool was later expanded to the full SDO observing database [53].

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 [2, 64, 22, 25]. Due to the lack of consistent evaluation, one cannot directly claim this for solar segmentation.

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 sunpots in Hα images [14]. Since sunpots 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 [14]. 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 sunpots. Qu et al.(2005) ave a similar system that identifies filaments [46].

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 [18]. This work is based off more generic computer vision region growing by Gonzalez & Woods (2002) [20]. After calibrating, removing dust, and sharpening the solar Hα images [65], 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 [18]. 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 Hα images. Other region growing methods include: Benkhalil et al.(2006) which used ionized calcium, Hα, and extreme ultraviolet imagery to grow active region boundaries [7], Higgins et al. (2011) which combined magnetograms, image differencing, and region growing to identify and track emerging active regions [24], and McAteer et al. (2005) which used full-disk magnetograms to identify magnetically significant regions and characterize their flare potential [41].

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 3.3. 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 [43].

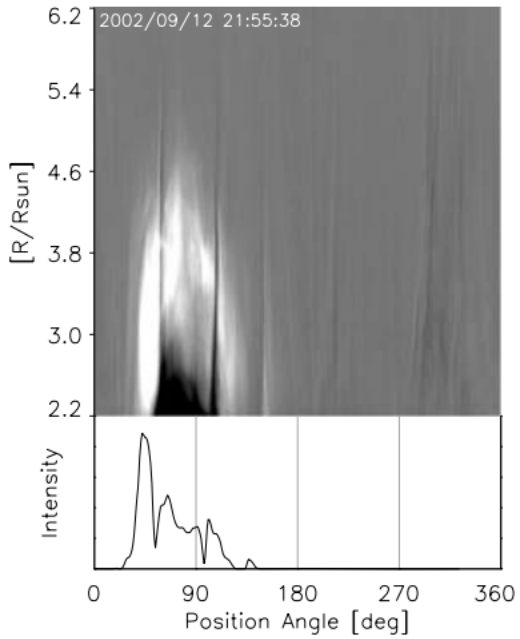

Figure 3.3: **Histogram segmentation**: Olmedo et al.(2008) utilized histogram segmentation to identify coronal mass ejections. "The intensity profile along the angular axis showing the 1D projection of the CME image. Only positive pixels along the radial axis are used. This profile effectively indicates the angular positions of a CME when it is present." [43]

Other histogram based methods include Preminger et al.(2001) which used contrast ratios between ionized calcium and magnetograms to identify sunspots and faculae [44].

Hybrid

Some approaches bridge between region-based and edge-based techinques. One key example is 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) [61]. 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 [8], 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 [27]. Thus, the fuzzy c-means requirement of membership summing to one is relaxed to form of possibilistic c-means [28]. SPoCA is a further modification that incorporates weighting by spatial extent such that neighboring pixels should be assigned similar labels [5, 6].

Barra et al. (2009) used SPoCA to both segment and track features on the Sun [6]. 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 [6].

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 [10]. 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 [39]. All of the component parts can be seen in Figures 3.4 and 3.5.

The SDO suite suite is a comprehensive approach to classification that utilizes all of the tech-

Science Target	**Flares**	**Coronal Dimming**	**Magnetic Feature Tracking**	**Filaments**
S/W Module Name	Flare Detective	Dimming Detector	SWAMIS	AAFDCC
Triggered or Continuous?	Triggered	Triggered	Continuous	Continuous
Trigger Source Data	AIA 193	AIA 193		
Trigger Source Module(s)	Self	Self		
Binned Image for Trigger	16 X 16	512 X 512		
Cadence for Trigger	Full	5-6 minutes		
Source Data	All AIA Channels	AIA	HMI Time Averaged Magnetograms	Global Hi-res Ha Network
Source Data Binning	None	512 X 512	None	Full; some smoothing
Source Data Cadence	Full	5-6 minutes	6 minutes	Full (~ 4/day eventually)
S/W Module Data	Flare Detective	Dimming Detector	SWAMIS	AAFDCC
Programming Language	IDL	IDL	Perl/PDL	IDL and C
Module Location	LMSAL	LMSAL	LMSAL	SAO
Computing Requirement		20 minutes CPU/day	8-16 CPUs to operate at 4x realtime	Most < 1 min CPU/Image
Responsible Team Member	Paolo Grigis	Meredith Wills-Davey	Craig DeForest	Pietro Bernasconi
Originating Organization	SAO	SAO	SwRI	JHU-APL
Output Data Volume		10 kB/event, 5 MB/movie	120 MB/day nominal	~ 1.5 MB per Ha image
First Pipeline Installation	June 2010	July 2010	November 2010	March 2010
Heritage	EIT, TRACE, RHESSI	New	MDI	GBO H-alpha

Science Target	**Active Regions**	**Sigmoids**	**PIL Mapping**	**CMEs**
S/W Module Name	SPoCA	Sigmoid Sniffer	PIL Finder	CME Detector/Tracker
Triggered or Continuous?	Continuous	Triggered	Continuous	Triggered
Trigger Source Data		AIA 211		LASCO C2 and C3
Trigger Source Module(s)		Self		Flares, Dimmings, Sigmoids
Binned Image for Trigger		Full Resolution		Full Resolution
Cadence for Trigger		10 min		Full LASCO Cadence
Source Data	AIA 171/195, 211/335, 94	AIA 94, 131	HMI Magnetograms	LASCO C2 and C3
Source Data Binning	None	None	Depends on science specs	None; front smoothing
Source Data Cadence	15 minutes	10-20 s	~5 min	Full (~1 hr)
S/W Module Data	SPoCA	Sigmoid Sniffer	PIL Finder	CME Detector/Tracker
Programming Language	C++	IDL	IDL	IDL
Module Location	LMSAL	LMSAL	SAO	SAO
Computing Requirement	30 sec/pair of full-res AIA	10 min CPU/image	Standard desktop	30 minutes/event
Responsible Team Member	Veronique Delouille	Nour-Eddine Raouafi	Alexander Engell	Meredith Wills-Davey
Originating Organization	ROB	JHU-APL	SAO	SAO
Output Data Volume	6 kB/event	5 MB/day	~ 2 Mb/day	1-8 Mb/day
Scheduled Installation	October 2010	November 2010	December 2010	Spring 2011
Heritage	EIT	SXT, SXI, XRT	Kitt Peak, SOLIS	LASCO

Figure 3.4: **First half of SDO classifiers** [39]

niques 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 of feature they are interested in by identifing them within an image. These are placed in feature vectors of 12 texture paramters (e.g. 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 [33]. The SDO suite lacks the ability to combine all the classifications into a single thematic map.

Comparisons

As shown in Figure 3.6, 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.

Science Target	**Coronal Holes**	**X-ray Bright Points**	**Sunspots**	**Global NLFFFs**
S/W Module Name	SPoCA	BP Finder	SWAMIS	Optimization Code for Full Disk
Triggered or Continuous?	Continuous	Continuous	Continuous	Continuous
Trigger Source Data				
Trigger Source Module(s)				
Binned Image for Trigger				
Cadence for Trigger				
Source Data	AIA 171/195, 211/335, 94	AIA 171, 195, 211	HMI Magnetograms	HMI+AIA for comparison
Source Data Binning	None	None (TBC)	None	Binned to 256x256 for ARs
Source Data Cadence	15 minutes	Full	TBD (5-60 min)	~5 min
S/W Module Data	SPoCA	BP Finder	SWAMIS	Optimization Code
Programming Language	C++	IDL	PDL	C
Module Location	LMSAL	SAO	LMSAL	MPS
Computing Requirement	Same as for Active Regions		< 1 CPU	2 hours/vector-magnetogram
Responsible Team Member	Veronique Delouille	Steve Saar	Craig DeForest	Thomas Wiegelmann
Originating Organization	ROB	SAO	SwRI	Max Planck Sonnenforschung
Output Data Volume	6 kB/event	600 kB/10 minutes	< 1 MB/day	256x256x256 datacube
First Pipeline Installation	January 2011	February 2011	Spring 2011	When HMI VMG's become routine
Heritage	EIT	EIT	MDI	SOT

Science Target	**Jets**	**Oscillations**	**"EIT Waves"**	**Trainable Feature Recognition**
S/W Module Name	Jet Detector	Oscillation Finder	EIT Wave Tracker	
Triggered or Continuous?	Triggered	Triggered	Triggered	Continuous
Trigger Source Data	TBD AIA Channels	AIA 193	AIA 193	
Trigger Source Module(s)	XRBP & CH	Flares, CMEs, Dimmings	Dimmings	
Binned Image for Trigger	1024x1024 (TBC)	N/A	1024x1024	
Cadence for Trigger	1 min	N/A	Full	
Source Data	AIA 193	Appropriate AIA Channels	AIA 193	All AIA Channels, some HMI Data
Source Data Binning	None	None	None; smoothing applied	128 X 128 subimages
Source Data Cadence	1 min	None	Full	Cadence TBD
S/W Module Data	Jet Detector	Oscillation Finder	EIT Wave Tracker	Trainable Feature Recognition
Programming Language	IDL	IDL	IDL	C/IDL
Module Location	SAO	SAO	SAO	SAO
Computing Requirement	15 seconds per event	60 minutes CPU/event	Very small	
Responsible Team Member	Antonia Savcheva	James McAteer	Meredith Wills-Davey	Rafal Angryk
Originating Organization	SAO	New Mexico State Univ.	SAO	MSU
Output Data Volume	90 MB/jet		2 Mb/event	0.1% of image volume
First Pipeline Installation	February 2011	Spring 2011	Spring 2011	December 2010 for image vectors
Heritage	XRT	TRACE, EIT	TRACE	TRACE

Figure 3.5: **Second half of SDO classifiers** [39]

Caballero & Aranda (2013) conducted an independent comparison of unsupervised techniques for active regions [11]. Using 6000 images from SOHO in 195 angstroms, they initially segmented the images using region growing techinques. Then, the different independent regions were clustered together into units using either partition approaches or a hierarchial classification. They found that the hiearchial 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) [3].

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.

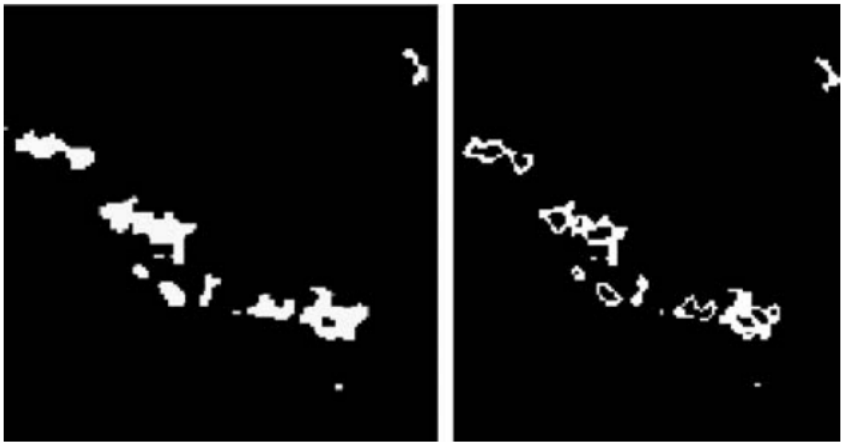

Figure 3.6: **Performance of different methods**: Revathy, Lekshmi, & Nayar (2005) compared the performance of different segmentation techniques in identifying active regions [48]. At left is the result of a histogram thresholding approach while at the right is fuzzy-based segmentation.

Maximum likelihood estimation

Rigler et al. (2012) proposed a preliminary naive Bayesian classifier for SUVI images using the Solar Dyanmics Observatory (SDO) Advanced Imaging Array (AIA) observations for testing [49]. 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 accuracyof 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,

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

. 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:

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

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) [60].

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 [15]. 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 [15]. Thus, de Wit(2006) ran a naive Bayesian classifier on these transformed parameters instead of the higher dimensional wavelengths. Figure 3.7 shows an example result by de Wit(2006), emphasizing the coherence of this segmentation without any forced smoothness [15]. This approach ran in near real-time, taking only a few minutes to classify every

pixel.

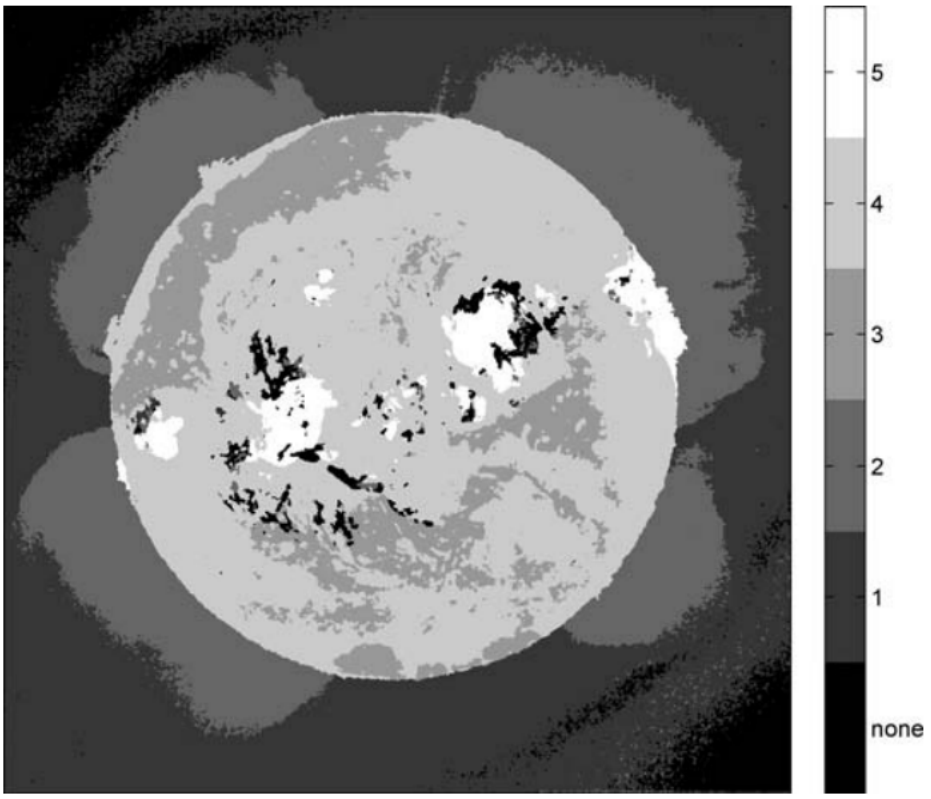

Figure 3.7: **de Wit Segmentation**: This figure indicates the power of quick, multi-spectral, supervised segmentation done by de Wit (2006), very similar to other work by Rigler et al.(2012) and Visscher et al.(2015) [15]. The classes for this study are: "(1) Tenuous corona outside of the disk, in regions 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" [15].

Maximum A Posteriori

Visscher et al. (2015) improve upon Rigler et al. (2012)'s result by recognizing that the classes are not equally likely (e.g. 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 [63]. 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), 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:

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

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 adition, they were able to pinpoint that there is disagreement with the gold standard reference. 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, Visscher et al.(2015) establish a criteria for accurate segmentation into large-scale features [63]:

- "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 [59, 60, 29]. 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 [30]. 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 [66]. 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 [18].

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 [9]. Given optical Hα 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 not make clear about false positives and 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.

This author 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 [19]. Lowe & Kulkarni (2015) used a random forest to identify terrain type in hyperspectral images [38]. 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% [38]. This high performance for random forests with this type of problem is not uncommon [45, 51, 12, 31].

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 [35]. 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 that previous networks by allowing for multi-scale examination.

Li et al. (2014) present a comprehensive review of Earth remote sensing classification techinques based on spatial techniques [36]. 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

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.

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) [40] to extend the LeNet convolutional neural network [34] 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 [55]. During their development, they thoroughly describe convolutional neural network for semantic segmentation up until 2016 [55].

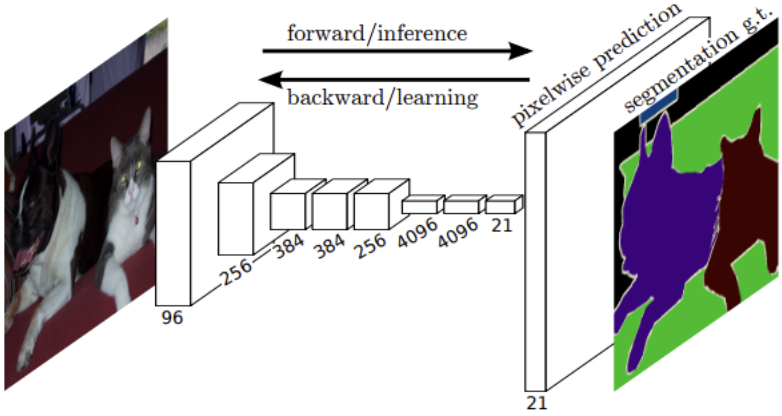

Figure 3.8: **Fully convolutional neural network**: This is an example architecture for a fully convolutional neural network (figure 1 from Shelhamer, Long, & Darrell (2016) [55]).

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, & Girschick developed Mask Regional Convolutional Neural Network (Mask R-CNN), an extension of Faster R-CNN [47] that creates another form of a semantic segmentation [23]. Technically, Mask R-CNN is a type of FCN [55]. Faster R-CNN worked by quckly 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.

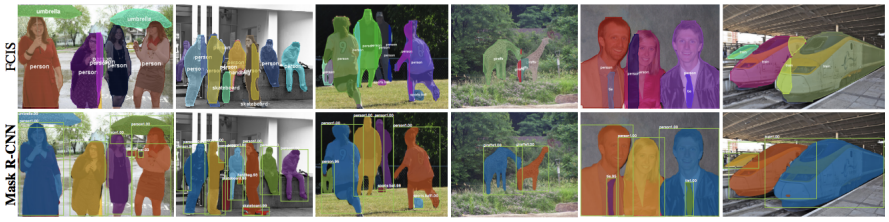

Figure 3.9: **Mask regional convolutional neural networks**: The top row was the existing state-of-the-art instance segmentation [37], an example of a FCN, compared to the bottom row of Mask R-CNN performance on the same scene. The overlaid coloration indicates the segmentation while bounding boxes indicate where Mask R-CNN evaluated these masks. Clearlly, 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, e.g. 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

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) [26] 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.

For this project, a small curated set of solar events was produced using three solar physics experts. They used an abbreviated set of solar event categories as detailed in Table 4.2.

Twenty-seven image groupings of the SUVI six-band imagery were used for the labeled data as shown in Table 4.3. Each group consists of one image from each SUVI band, an h-alpha image, and any needed derived images used as features in the machine learning.

4.2.1 Analysis of labeled data

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 [26].

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.

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.

Chapter 5

Classifiers

5.1 Naive Bayesian Maximum Likelihood

5.2 Random Forest

5.3 Feed-forward Neural Network

5.4 Convolutional Neural Network

Chapter 6

Experiments and Results

6.1 Labeled imagery analysis

6.2 Evaluation approaches

6.2.1 Standard machine learning metrics

6.2.2 Confusion matrix

6.3 Data noise levels

6.3.1 Noise-gater procedures

Some SUVI channels, specifically the 94 angstroms channel, are riddled with shot noise that degrades the signal quality. This 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 6.1.

6.4 Normalization of data

6.5 Spatial features

6.6 HALPHA inclusion

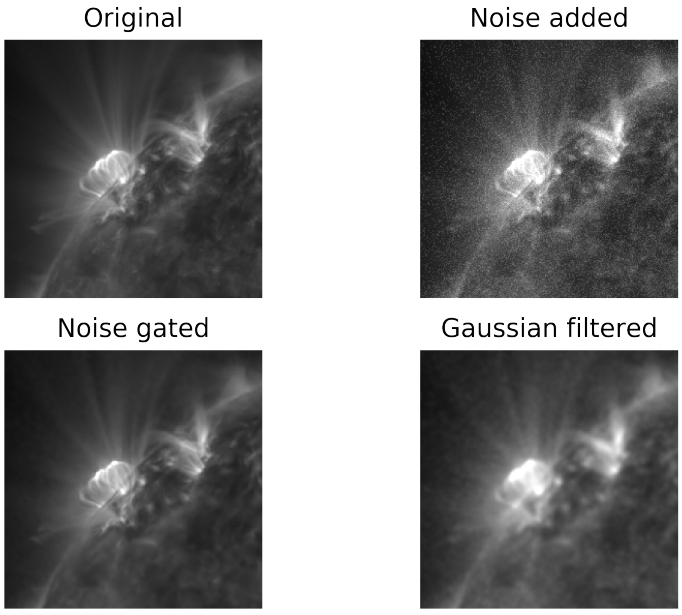

Figure 6.1: **Effectiveness of noise-gating** The upper left is a good image, no cleaning necessary. However, images like the upper right, dominated by shot noise, are typical for the 94 angstrom channel. This image is created by taking the image on the upper left and adding Poisson noise with a signal-to-noise ratio of 2. The algorithm still performs even if it's 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.

Chapter 7

Applications

7.1 Database building

7.2 Fractal dimension and class properties

Chapter 8

Conclusion

8.1 Summary

8.2 Future work

8.2.1 Stability over solar cycles

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] 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.

[3] Aschwanden, M. J. Image Processing Techniques and Feature Recognition in Solar Physics. *Solar Physics 262* (Apr. 2010), 235–275.

[4] Banda, J. M., and Angryk, R. A. *Large-Scale Region-Based Multimedia Retrieval for Solar Images*. Springer International Publishing, Cham, 2014, pp. 649–661.

[5] Barra, V., Delouille, V., and Hochedez, J.-F. Segmentation of extreme ultraviolet solar images via multichannel fuzzy clustering. *Advances in Space Research* (2008), 917–925.

[6] Barra, V., Delouille, V., and Hochedez, J.-F. Segmentation, tracking and characterization of solar features from eit solar corona images. *SCIA 2009, LNCS 5575* (2009), 199–208.

[7] 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.

[8] Bezdek, J. C. *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media, 2013.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] de Wit, T. D. Fast Segmentation of Solar Extreme Ultraviolet Images. *Solar Physics 239* (Dec. 2006), 519–530.

[16] 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.

[17] 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.

[18] Fuller, N., Aboudarham, J., and Bentley, R. D. Filament Recognition and Image Cleaning on Meudon Hα Spectroheliograms. *Solar Physics 227* (Mar. 2005), 61–73.

[19] 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.

[20] Gonzalez, R. C., and Woods, R. E. In *Digital Image Processing, 2nd edn.* (2002), Prentice-Hall, Inc., p. 613.

[21] 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.

[22] 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.

[23] He, K., Gkioxari, G., Dollár, P., and Girshick, R. B. Mask R-CNN. *CoRR abs/1703.06870* (2017).

[24] 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.

[25] 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).*

[26] 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 forthesolar dynamics observatory (sdo) and beyond. *Solar Physics 275*, 1 (Jan 2012), 67–78.

[27] Krishnapuram, R., and Keller, J. M. A possibilistic approach to clustering. *IEEE transactions on fuzzy systems 1*, 2 (1993), 98–110.

[28] Krishnapuram, R., and Keller, J. M. The possibilistic c-means algorithm: insights and recommendations. *IEEE transactions on Fuzzy Systems 4*, 3 (1996), 385–393.

[29] Krizhevsky, A., Sutskever, I., and Hinton, G. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems 25*, 1106–1114.

[30] 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.

[31] Kulkarni, A. D., and Shrestha, A. Multispectral image analysis using decision trees.

[32] 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.

[33] 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.

[34] 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.

[35] Lee, H., and Kwon, H. Going deeper with contextual cnn for hyperspectral image classification. *IEEE Transactions on Image Processing 26*, 10 (2017), 4843–4855.

[36] 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.

[37] Li, Y., Qi, H., Dai, J., Ji, X., and Wei, Y. Fully convolutional instance-aware semantic segmentation. *arXiv preprint arXiv:1611.07709* (2016).

[38] Lowe, B., and Kulkarni, A. Multispectral image analysis using random forest. *International Journal on Soft Computing 6*, 1 (2015).

[39] 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.

[40] 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.

[41] 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.

[42] Muller, C. The carrington solar flares of 1859: Consequences on life. *Origins of Life and Evolution of the Biosphere 44*, 3 (2014), 185–195.

[43] 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.

[44] Preminger, D. G., Walton, S. R., and Chapman, G. A. Solar feature identification using contrasts and contiguity. *Solar physics 202*, 1 (2001), 53–62.

[45] 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.

[46] 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.

[47] 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*.

[48] 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.

[49] Rigler, J. E., Hill, S. M., Reinard, A. A., and Steenburgh, R. A. Solar thematic maps for space weather operations. *Space Weather 10* (2012).

[50] Riley, P. On the probability of occurrence of extreme space weather events. *Space Weather 10*, 2 (Feb 2012), 1–12.

[51] 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.

[52] 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.

[53] 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.

[54] Seaton, D. B., Darnel, 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.

[55] Shelhamer, E., Long, J., and Darrell, T. Fully convolutional networks for semantic segmentation. In *IEEE Transactions on Pattern Analysis and Machine Intelligence* (May 2016), vol. 39, IEEE, pp. 640–651.

[56] 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).

[57] Space Weather Operations, R., and Force, M. T. National space weather action plan. *United States Department of Commerce* (2015).

[58] 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.

[59] 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.

[60] Tso, B., and Mather, P. M. Classification methods for remotely sensed data.

[61] 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.

[62] Viljanen, A., Myllys, M., and Nevanlinna, H. Russian geomagnetic recordings in 18501862 compared to modern observations. A11.

[63] 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).

[64] Yu, G. W., and Zemel, R. S. Comparing supervised vs. unsupervised image segmentation methods.

[65] 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.

[66] Zharkova, V. V., and Schetinin, V. Filament recognition in solar images with the neural network technique. *Solar Physics 228*, 1 (May 2005), 137–148.

[67] 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.