Application and Interpretation of Deep Learning for Identifying Pre-emergence Magnetic Field Patterns

Dattaraj B. Dhuri\textsuperscript{1}, Shravan M. Hanasoge\textsuperscript{1}, Aaron C. Birch\textsuperscript{2}, and Hannah Schunker\textsuperscript{2,3}

\textsuperscript{1} Department of Astronomy and Astrophysics, Tata Institute of Fundamental Research, Mumbai, 400005, India; dattaraj.dhuri@tifr.res.in
\textsuperscript{2} Max-Planck-Institut für Sonnensystemforschung, Göttingen, Germany
\textsuperscript{3} School of Mathematical and Physical Sciences, The University of Newcastle, New South Wales, Australia

Received 2019 November 15; revised 2020 July 17; accepted 2020 September 9; published 2020 October 29

Abstract

Magnetic flux generated within the solar interior emerges to the surface, forming active regions (ARs) and sunspots. Flux emergence may trigger explosive events—such as flares and coronal mass ejections, and therefore understanding emergence is useful for space-weather forecasting. Evidence of any pre-emergence signatures will also shed light on subsurface processes responsible for emergence. In this paper, we present a first analysis of EARs from the Solar Dynamics Observatory/Helioseismic Emerging Active Regions dataset using deep convolutional neural networks (CNN) to characterize pre-emergence surface magnetic field properties. The trained CNN classifies between pre-emergence line-of-sight magnetograms and a control set of nonemergence magnetograms with a true skill statistic (TSS) score of approximately 85\% about 3\,hr prior to emergence and approximately 40\% about 24\,hr prior to emergence. Our results are better than a baseline classification TSS obtained using discriminant analysis (DA) of only the unsigned magnetic flux, although a multivariable DA produces TSS values consistent with the CNN. We develop a network-pruning algorithm to interpret the trained CNN and show that the CNN incorporates filters that respond positively as well as negatively to the unsigned magnetic flux of the magnetograms. Using synthetic magnetograms, we demonstrate that the CNN output is sensitive to the length scale of the magnetic regions, with small-scale and intense fields producing maximum CNN output and possibly a characteristic pre-emergence pattern. Given increasing popularity of deep learning, the techniques developed here to interpret the trained CNN—using network pruning and synthetic data—are relevant for future applications in solar and astrophysical data analysis.

Unified Astronomy Thesaurus concepts: Solar magnetic flux emergence (2000); Convolutional neural networks (1938); Solar active region magnetic fields (1975); Astronomy data analysis (1858)

1. Introduction

Magnetic flux in the Sun, generated by the dynamo operating within the interior, rises to the visible surface, the photosphere, and further into the solar atmosphere (Stein 2012; Cheung & Isobe 2014). The rising flux appears at the photosphere in the form of large-scale (∼10–100\,Mm) structures of concentrated magnetic flux (∼kG); this is known as sunspots and active regions (ARs). In the low-β solar atmosphere and corona, these flux tubes expand significantly and form gigantic loop structures rooted in the ARs. The rising magnetic flux also pumps nonpotential energy into the ARs that powers explosive events such as flares and coronal mass ejections (CMEs) that can lead to severe consequences for space weather (Shibata & Magara 2011; Eastwood et al. 2017). Knowledge of the flux-emergence mechanism in ARs is thus useful to gain warning time for space-weather forecasting. Additionally, a comprehensive picture of the solar dynamo includes processes responsible for the generation as well as transport of the magnetic flux (Cheung & Isobe 2014). Understanding the formation and evolution of ARs is therefore necessary to constrain solar dynamo models (Birch et al. 2012; Cameron et al. 2016; Schunker et al. 2016).

The physical mechanism that leads to the appearance of new ARs in the photosphere, henceforth referred to as “emergence,” is not fully known (Leka et al. 2012; Cheung & Isobe 2014; Schunker et al. 2016). Broadly, two categories of emergence scenarios are considered in the literature. In the first scenario, flux tubes may form deep in the convection zone and rise intact to emerge on the surface (Fan 2009). Alternatively, small-scale flux tubes, which may be generated within the bulk of the convection zone or in the near-surface layers, coagulate and form ARs on the surface (Brandenburg 2005; Brandenburg et al. 2014). Case studies for detecting subsurface pre-emergence (PE) signatures, e.g., using helioseismology (Birch et al. 2010), may not be successful because the signal is weak in comparison to the uncertainties on the detection. For an unambiguous detection of PE signatures, a statistical study of a large sample of EARs is therefore in order. A statistical study, such as that by Komm et al. (2009, 2011), is expected to capture common characteristics of EARs that may be universal to the emergence mechanism (Schunker et al. 2016). Leka et al. (2012), Birch et al. (2012), and Barnes et al. (2014) carried out an extensive survey of EARs using helioseismic holography. This is henceforth referred to as the LBB survey. Analyzing line-of-sight magnetograms obtained with the Michelson Doppler Imager (MDI) on board the Solar and Heliospheric Observatory (SOHO) and Global Oscillation Network Group (GONG) dopplergrams, the LBB survey detected weak converging flows and reduced travel-times to distinguish emerging active regions (EARs) from a set of control regions (CRs) that did not show emergence. The study also identified a PE magnetic field signal correlated with converging flows in the EAR population. Although a causal relationship between the converging flow and the PE magnetic field signal could not be established, the average unsigned radial magnetic field of EARs was demonstrated to be the leading discriminator between the EAR and CR populations,
yielding true skill statistics (TSS; Peirce skill score; Peirce 1884; Bobra & Couvidat 2015) of ~50% about 3 hr prior to the emergence. This warrants further investigation of PE surface magnetic field characteristics.

Motivated by the results of the LBB survey, here we perform comprehensive statistical analysis to attempt to uncover spatio-temporal patterns associated with PE surface magnetic fields. Surface magnetic field data of superior spatial and temporal resolution are available from the Helioseismic and Magnetic Imager (HMI; Schou et al. 2012) on board the Solar Dynamics Observatory (SDO; Pesnell et al. 2012). Following the LBB survey, Schunker et al. (2016) assembled the Helioseismic Emerging Active Region (SDO/HEAR) dataset, also comprising PE magnetograms of EARs and CRs. Although originally designed for measuring subsurface PE activity, the SDO/HEAR dataset is also useful to study surface magnetic field properties of EARs (Schunker et al. 2016, 2019). The SDO/HEAR dataset comprises about 200 samples of EARs and CRs each, with 60° × 60° spatial extent and tracked over a duration of up to one week before emergence. Detecting spatio-temporal correlations and patterns over such a large dataset of magnetograms is statistically challenging. With the advent of machine-learning (ML), however, advanced algorithms for analyzing and classifying images to detect complex correlations are available (Hastie et al. 2001). In particular, deep convolutional neural networks (CNNs) have proven immensely successful for visual recognition tasks, such as image and video classification, by accurately characterizing structural information from the images (Goodfellow et al. 2016; Krizhevsky et al. 2012; LeCun et al. 2015). Here, we use CNNs to discriminate between magnetograms from EAR and CR populations and in this process, uncover associated PE surface magnetic field characteristics.

The classification problem considered here can be naturally formulated as a supervised-learning (Hastie et al. 2001) problem, where the CNN is trained using the dataset of magnetograms labeled PE and nonemergence (NE). The CNN is trained to optimize the TSS score for classification of PE and NE magnetograms taken at different PE times. CNNs, and deep neural networks in general, may be trained for challenging tasks. However, it is difficult to interrogate the operation of neural networks into comprehensible components and thus interpret their performance. For the present work, the interpretation of the trained network is important for obtaining quantifiable information about PE surface magnetic field characteristics. Our analysis in this work therefore focuses on understanding the performance of the network using synthetic magnetograms and advanced techniques such as functionality-based network pruning (LeCun et al. 1990; Han et al. 2015; Qin et al. 2018; Frankle et al. 2019).

The paper is organized as follows. In Section 2 we describe the processing of EARs and CRs in the SDO/HEAR dataset for deep-learning analysis using a CNN. In Section 3 we explain the CNN architecture and training method. In Section 4 we describe in detail the results of the classification of PE and NE magnetograms and compare with earlier work using a discriminant analysis. In particular, we explain the working of the CNN using the network-pruning algorithm developed here to facilitate the CNN interpretation. We detail various statistical analyses performed to comprehend the CNN performance. We also explain the probing of the trained CNN using synthetic magnetograms. In Section 5 we summarize our findings.

2. Data

We use the SDO/HEAR dataset comprising PE line-of-sight magnetograms of ARs observed by SDO/HMI between 2010 May and 2014 July (Schunker et al. 2016). EARs in the dataset are selected to be emergences into a relatively quiet Sun to minimize the difficulty of distinguishing signatures of emerging flux from any preexisting magnetic field. Each EAR is paired with a CR that does not show emergence. Each CR is cospatial with the corresponding EAR in terms of latitude and distance from the central meridian, and is separated by no more than two solar rotation periods. The SDO/HEAR dataset, although originally prepared for a helioseismology survey, is also useful for studying PE surface magnetic field properties (Schunker et al. 2016). The dataset provides Postel-projected EAR and CR maps covering 60° × 60° around the emergence location. However, following the LBB survey, we restrict our analysis to the central 30° × 30° region. This region is projected onto a 256 × 256 pixel grid with a plate scale of 1.4 Mm per pixel. In the following, we explain in detail the problem of classifying between 10^5 × 10^6 PE and NE magnetogram segments subsampled from the original 30° × 30° magnetograms of EARs and CRs in the SDO/HEAR dataset (see Figure 1). We use the following definitions of the magnetograms referred to in this work (see Table 1):

1. EARs—Emerging active regions of size 30° × 30° taken from the SDO/HEAR dataset.
2. CRs—Control regions of spatial extent 30° × 30° that do not show emergence, taken from the SDO/HEAR dataset.
3. PEs—Pre-emergence magnetograms of spatial extent 10° × 10^6 randomly subsampled from the central regions of EARs, as shown in Figure 1 (red).
4. NEs—Nonemergence magnetograms of spatial extent 10° × 10^6 randomly subsampled from CRs and noncentral regions of EARs, as shown in Figure 1 (blue).

The SDO/HEAR dataset is compiled following several criteria to facilitate characterization of EARs and CRs using statistical analysis (Schunker et al. 2016). Only EARs are selected that (1) are visible in the continuum, (2) reach a total area of at least 10 μH or 30 Mm^2, and (3) are within ±50° of the central meridian at the time of emergence t_e, as defined by the National Oceanic and Atmospheric Administration (NOAA). Consistent with the LBB survey, the emergence time t_0 of an EAR is defined as the time when the absolute flux of an EAR is 10% of the maximum observed value within a 36 h interval following t_e^NOAA. Moreover, EARs are selected only when the absolute flux rises monotonically after the emergence from a low, steady value. Following these restrictions, 182 EAR samples between 2010 May and 2014 July are included in the SDO/HEAR dataset. CRs in the dataset are selected to match the spatial distribution of EARs. This is achieved by selecting a CR cospatial with each EAR and also from the identical phase of the solar cycle. The absolute difference between the average PE magnetic field within the central 10° × 10^2 of a CR and associated EAR is restricted to be less than 10 G or 1.4 × 10^12 Mx (see Figure 3 in Schunker et al. 2016). Importantly, the CRs do not show a marked increase in the absolute flux similar to emergence. CRs are assigned a mock-emergence time when their Stonyhurst coordinates (Thompson 2006) are identical to the corresponding EARs at the time of emergence t_0.
Figure 1. Subsampling PE (red) and NE (blue) magnetogram segments from the EARs (left) and CRs (right) in the SDO/HEAR dataset (Schunker et al. 2016). The $10^5 \times 10^5$ segments are randomly subsampled from the original $30^5 \times 30^5$ EARs and CRs. Subsample locations are randomly selected for each EAR/CR, thus they are different for different EARs and CRs. We constrain the subsampling to ensure that the emergence region (the central $5^5 \times 5^5$ of EARs) is fully included (fully excluded) within PE (NE) magnetogram segments. We subsample NE magnetogram segments from CRs as well as noncentral region of EARs. PE and NE segments are Postel projections of size $85 \times 85$ pixels, with a plate scale of 1.4 Mm per pixel, i.e., the same as the original EARs and CRs. We label PE and NE segments 1 and 0, respectively, for supervised learning using the deep neural network.

Table 1
The SDO/HMI Survey of EARs (SDO/HEAR) Dataset: We Consider EARs and CRs between 2010 May and 2014 July

| time before emergence (hr) | EARs ($30^5 \times 30^5$) | CRs ($30^5 \times 30^5$) | PE ($10^5 \times 10^5$) | NE ($10^5 \times 10^5$) |
|---------------------------|---------------------------|---------------------------|--------------------------|--------------------------|
|                           | (original)                | (original)                | (subsampled)             |                          |
| 3.2                       | 178                       | 180                       | 17800                    | 18080                    |
| 8.5                       | 174                       | 165                       | 17400                    | 17115                    |
| 13.9                      | 173                       | 169                       | 17300                    | 17269                    |
| 19.2                      | 172                       | 167                       | 17200                    | 17117                    |
| 24.5                      | 158                       | 158                       | 15800                    | 15958                    |

Note. We subsample PE and NE segments from original EARs and CRs to furnish sufficient data for training a deep neural network.

The EARs and CRs are tracked up to a consecutive two-week duration, up to one week on either side of emergence, at the Carrington rotation rate. The SDO/HMI line-of-sight magnetograms are available at a cadence of 45 s. The magnetograms are temporally averaged over intervals of 410.25 min (547 frames) and spaced by 320.25 min (427 frames) to smooth out temporal fluctuations. Thus, there is an overlap of 90 min (120 frames) between consecutive temporal-averaging intervals (see Figure 5 in Schunker et al. 2016). Consistent with the LBB survey, we only consider PE temporally averaged magnetograms up to a day prior to emergence. The number of EAR and CR samples considered in our analyses at different PE times are listed in Table 1. We include only those EARs and CRs for which magnetogram samples are available at all PE times, and therefore the number of EARs and CRs is not exactly equal.

Figure 2 shows time-averaged PE magnetograms of an example EAR and corresponding CR. As the emergence time approaches, we see the early appearance of a bipolar field within the central $10^5 \times 10^5$ of EARs. In addition, we see no obvious difference between EARs and CRs prior to emergence. The signature of the PE bipolar field becomes evident when we consider the average of all EARs and CRs in the dataset, shown in Figure 3, after accounting for Hale’s polarity law and Joy’s law (Schunker et al. 2016). To account for Hale’s law, we reverse the sign of the magnetic field for the magnetograms in the southern hemisphere. To account for Joy’s law, we flip the magnetograms in the southern hemisphere in the latitudinal direction, i.e., the $+y$ direction points away from the equator. For the average EAR, the PE field within the central $10^5 \times 10^5$ region may be distinctly seen, even up to a day prior to the emergence. As mentioned earlier, this PE magnetic field was a leading factor in the discriminant analysis performed in the LBB survey to differentiate the EAR and CR samples. The present analysis, using deep learning, is expected to identify this PE field as well as other correlated patterns.

As in the case of the LBB survey, the SDO/HEAR dataset used in this analysis is specifically designed for examination of PE signatures using helioseismology and surface magnetic field properties. The EARs and CRs are selected with a bias of prior knowledge of where emergence occurred. Therefore, our work is not an attempt of forecasting emergence. Rather, the focus is on characterizing EARs and CRs based on their surface magnetic field properties and understanding the process of emergence. The SDO/HEAR dataset is well suited for studying spatio-temporal characteristics of the PE surface magnetic field, as demonstrated in Schunker et al. (2016).

A major obstacle for our analysis is that the number of EAR samples is indeed small in the SDO/HEAR dataset for training a deep neural network. It is difficult to know the exact number of samples required for successfully training a deep neural network a priori. The number of parameters in such a network are $O(100,000) -$ many of these may be redundant, but for successful training, thousands of examples from each category (class) may be necessary (Goodfellow et al. 2016). In order to have a sufficient number of examples for training, we consider PE magnetograms of smaller size in our analysis. These are subsampled from the original $30^5 \times 30^5$ Postel-projected maps. By design, the emergence is restricted to the central $10^5 \times 10^5$ area of EARs in the SDO/HEAR dataset. Thus, we obtain magnetogram segments of $10^5 \times 10^5$ selected randomly from the original EAR and CR magnetograms. Because of the random selection, the subsampled segments are different geometric regions in different EARs and CRs. We constrain the random selection to ensure that the subsampled $10^5 \times 10^5$ segments either fully contain or fully exclude the emergence location, i.e., segments that contain the emergence location even partially are disallowed (see Figure 1). However, there is no minimum distance between two random PE or NE from an EAR or CR, i.e., there may be significant overlap. Note that the NE segments are subsampled from the CRs as well as from the noncentral region of the EARs. Moreover, the NE segments in the training set could include regions where flux does emerge and that were not classified as PE regions in the SDO/HEAR survey. This could increase the chances of the machine-learning algorithm returning false negatives. The PE and NE segments are now Postel-projected maps on $85 \times 85$ pixel grids. The plate scale of the magnetogram segments is 1.4 Mm per pixel, i.e., similar to EARs and CRs. We subsample exactly 100 PEs from the central part of each EAR and approximately 50 NEs each from the noncentral part of each EAR and from each CR. Table 1 lists the total number of PE and NE segments used for analysis. We train a deep neural network, a machine-learning algorithm, to classify between these PE and NE segments labeled class 1 and class 0, respectively.
3. Methods

Machine-learning is a set of algorithms that is applied to analyze patterns and correlations in large, high-dimensional datasets without being explicitly programmed for it. Here, we apply supervised learning, where the machine is trained using a known set of input-output pairs from a part of the available data (Hastie et al. 2001). PE and NE magnetograms, at a given PE time $t$, are inputs, and class labels 1 and 0, respectively, are the desired outputs (see Figure 4). The machine parameters, i.e., weights and biases (see Appendix A), are tuned during the training process when the machine learns the probability distribution of the inputs. The trained machine is used to make predictions on the remaining part of the data, which is called validation or test data. We use deep CNN as our method of choice for the classification of PE and NE magnetograms (Krizhevsky et al. 2012; LeCun et al. 2015; Goodfellow et al. 2016). CNNs have been used extensively for analyses of images and proven widely successful.

The CNN consists of layers of neurons with 2D convolutional filters, particularly designed for analyzing images. A convolutional filter of $K \times K$ neurons slides over input magnetograms (Figure 5), transforming a $K \times K$-pixel subregion at a time according to the neuron activation function (Equation (A1)). Unlike typical image classification problems, the magnetograms contain inputs with both positive and negative values, representing opposite magnetic polarities. Therefore, we use the identity activation function for CNN filters, i.e., every neuron in the CNN filter outputs $y = \sum_i w_i x_i + b$, where $x$ are inputs to the neuron with weights $w$, and $b$ is the bias of the filter (Appendix A). The identity activation function ensures that positive and negative (magnetic field) values from the input are treated equally by the network. In addition to the convolutional layers, max-pooling layers are also used in CNNs. A max-pooling layer of size $M \times M$ extracts the maximum value from an $M \times M$-pixel subregion, thus downSampling the input by a factor of $M$. A combination of convolutional and pooling layers, placed at increasingly deeper levels, is sensitive to features of increasing length scales of the original image. We use a CNN architecture based on VGGNet (Simonyan & Zisserman 2014), which uses $3 \times 3$ convolutional filters and $2 \times 2$ max-pooling filters (see Figure 6). The VGGNet architecture includes fully connected (FC) layers to connect the final convolutional layer to the output. To facilitate interpretation of the trained CNN by analysis of the features learned by convolutional filters in the final layer, we incorporate a CNN without any FC layers, i.e., a fully convolutional neural network (Shelhamer et al. 2017). The five-layer deep CNN that we use reduces the original $85 \times 85$ pixel input to $6 \times 6$ pixels via max-pooling operations.

During training, error in the predicted output, as measured by a loss function, is minimized by tuning the weights and biases of the network via stochastic gradient descent (see Appendix A). Here, we use the binary cross-entropy loss function $L_{CE}$ (Equation (A2)). Training a neural network also involves fine-tuning the network hyperparameters. The hyperparameters associated with the stochastic gradient descent for training include the learning rate $lr$ and the batch size $N_{BS}$. The learning rate determines the step of the gradient descent. Too low a learning rate slows the training down, and too high a learning rate diverges the loss function value $L_{CE}$ (Equation (A2); Hastie et al. 2001). The batch size determines the number of training examples considered for the gradient descent at each iteration. We tune the learning rate, batch size, and depth of the CNN (number of convolutional layers) to optimize the classification of the PE and NE magnetograms as measured by TSS, see below.

Subsequent to the final convolutional layer, we perform a $6 \times 6$ max-pooling operation to reduce the output to the one neuron in the final layer. We choose the activation of the output neuron as a sigmoid function that yields a value between 0 and 1 (Hastie et al. 2001). This corresponds to the probability of the input magnetogram containing the emergence. The predicted classes are labeled positive (1) and negative (0) by thresholding the CNN output at 0.5. The CNN output thus falls in one of the following categories:

1. True positives (TPs)—Subpopulation of emerging magnetic field segments (PE) that are accurately classified as positive (1).
2. True negatives (TNs)—Subpopulation of nonemerging magnetic field segments (NE) that are accurately classified as negative (0).
The outputs for the subregions shaded in orange and blue are each 3\(\times\)3 pixel subregion from the input in order. Assuming that the bias \(b\) of the filter is 0, the identity activation of the neurons yields the output \(\sum w_i x_i\) for each subregion. The outputs for the subregions shaded in orange and blue are highlighted.

3. False positives (FPs)—Subpopulation of nonemerging magnetic field segments (NE) that are inaccurately classified as positive (1).

4. False negatives (FNs)—Subpopulation of emerging magnetic field segments (PE) that are inaccurately classified as negative (0).

Note that the FPs categorized here as such may still include true emergences that do not fall under the SDO/HEAR criteria.

Following the LBB survey, we use the Peirce skill score or TSS to measure the machine performance (Peirce 1884; Barnes et al. 2014; Bobra & Couvidat 2015). The TSS is defined as \(\text{TSS} = TP/(TP + FN) - FP/(FP + TN)\). The TSS is 0 for both random and unskilled predictions and 1 for the perfect classification.

4. Results

4.1. Training

We train the CNN (Figure 6) using supervised learning (Hastie et al. 2001) with PE and NE magnetograms as inputs and labels 1 and 0, respectively, as outputs. Typically, the available data are split into three categories—training, validation, and testing. The training data are used to determine weights and biases of the network, and validation data are used to tune the network hyperparameters (Section 3). The trained machine is then used to make predictions on the test data. A well-trained network is able to reproduce the high accuracy obtained during training on test data.

Because the number of EARs and CRs available are limited (see Table 1), we use a fivefold cross-validation for tuning network hyperparameters instead of a dedicated testing set (Figure 7; Hastie et al. 2001). The EARs and CRs are each randomly split into five (approximately) equal parts. We use PE and NE magnetogram segments from the four parts for training and the remaining part for validation. Note that the subsampled PE and NE magnetogram segments from an EAR or CR are included in either the training data or the validation data, but not both. We perform the training five times, with EARs and CRs from five different parts used for validation. After each training, we obtain TSS using predictions on PE and NE magnetograms from validation data. Because PE and NE magnetograms are subsampled randomly from EARs and CRs, PE and NE segments obtained from an EAR or CR contain overlapping regions. Therefore, while obtaining TSS, only nonoverlapping segments are considered. There are nine such nonoverlapping \(10^\circ \times 10^\circ\) segments from each EAR and CR (which are \(30^\circ \times 30^\circ\), i.e., 17 NE segments for every PE segments. Thus the classification problem considered here is class-imbalanced. The network hyperparameters are tuned to obtain the maximum average TSS value over a fivefold cross-validation. TSS is a good metric for such class-imbalanced problems.

We set up the CNN using the Python deep-learning library keras\(^4\). At the beginning of training, weights are initialized using the Glorot uniform initializer (Glorot & Bengio 2010) and biases are initialized as zeros. The input magnetograms are standardized, i.e., the mean is subtracted and divided by the

\(^4\) https://keras.io
standard deviation. Note that the mean and standard deviation used for standardization is calculated over all PE and NE magnetogram samples. This limits the operational range of the pixel values, representing the magnetic field.

We obtain optimum training by setting the learning rate $\eta = 1.0 \times 10^{-7}$ and the batch size $N_{bs} = 32$. We observe that the performance of CNN gradually improves with increasing depth (Table 2). The max-pooling operation downsamples feature maps obtained in each convolution layer by half. Therefore we limit the depth of the CNN to five layers to yield a final layer feature maps (the final convolution layer outputs) of $6 \times 6$ pixel grid. Each pixel from the final layer feature map corresponds to approximately $20 \times 20 \text{Mm}^2$ of the input PE and NE magnetograms, which is sufficiently large to incorporate the PE bipolar magnetic field pattern (Schunker et al. 2019).

Using the set of PE and NE magnetograms available at different PE times (see Table 1), we train the CNN to yield a maximum average fivefold cross-validation TSS. We monitor training and validation losses at each iteration to ensure that they decrease monotonically and converge as training progresses, indicating that the network is well trained and does not suffer from overfitting. (Figure 8).

The fivefold cross-validation TSS obtained using the CNN trained with the set of magnetograms taken at different PE times are listed in Table 3. The TSS is approximately 40% at 24.5 hr before emergence, comparable to the LBB survey results using the average unsigned radial magnetic field of EARs and CRs as discriminator. The TSS increases as emergence approaches, yielding approximately 85% at 3.2 hr before emergence, which is significantly higher than the approximately 53% obtained in the LBB survey.
Table 3
Mean Fivefold Cross-validation TSS to Classify PE and NE Magnetograms at Different PE Times Obtained Using the CNN (Figure 6) with Five Hidden Convolutional Layers

| time before emergence (hr) | fivefold cross-validation TSS (%) |
|---------------------------|----------------------------------|
| −3.2                      | 84.57 ± 6.40                     |
| −8.5                      | 61.97 ± 7.54                     |
| −13.9                     | 48.68 ± 5.45                     |
| −19.2                     | 40.80 ± 7.25                     |
| −24.5                     | 43.30 ± 6.93                     |

Note. The 1σ error is quoted.

4.2. Discriminant Analysis of Unsigned Line-of-sight Magnetic Flux

The LBB survey used a nonparametric discriminant analysis of the radial magnetic field \( \mathbf{B} \), of \( 30^\circ \times 30^\circ \) EARS (CRs), averaged over 45.5 Mm about the emergence (center) location. The radial magnetic field was obtained from the line-of-sight SOHO/MDI data using a potential-field model. The present work is concerned with the classification of \( 10^8 \times 10^8 \) line-of-sight PE and NE magnetogram segments, obtained from EARS and CRs, using a CNN. We obtain a baseline to compare the performance of the CNN using a nonparametric discriminant analysis, similar to the LBB survey, of the total unsigned line-of-sight magnetic flux \( MTOT = \sum |B| \) of PE and NE magnetogram segments. MTOT is a well-defined keyword in the SDO database (Bobra et al. 2014). The MTOT measure used here is slightly different in that we do not require membership of a coherent magnetic field structure, but rather more simply all of the magnetic field within a specific area of \( 10^9 \times 10^9 \).

For the discriminant analysis, we estimate the probability density of MTOT of the PE and NE magnetogram segments using the Epanechnikov kernel (Silverman 1986; Barnes et al. 2014). We choose the kernel-smoothing parameter value that is optimum for a normal distribution. We estimate the probability density for magnetograms in the training data. As shown in the left panel in Figure 9, we find the discriminant boundaries at which the estimated probability densities of MTOT of PE and NE magnetograms are equal. These are used to classify magnetograms in the validation data, and a TSS is obtained. Similar to the CNN, this process is performed for the five cross-validation sets.

In the left panel of Figure 9, the distributions of MTOT for PE and NE magnetogram segments are well separated. For a PE time \( −3.2 \), we find discriminant boundaries at \( 7.9 \times 10^{20} \) Mx and \( 20.2 \times 10^{20} \) Mx and at \( 0.2 \times 10^{20} \) Mx and \( 2.7 \times 10^{20} \) Mx. The discriminant analysis yields a cross-validation TSS of 65.8% ± 4.5%. Repeating the analysis, we obtain the TSS to classify PE and NE magnetogram segments at different PE times. The right panel in Figure 9 shows that the CNN outperforms the discriminant analyses using the unsigned magnetic field \( |\mathbf{B}| \) or MTOT, most significantly, for PE times \( −3.2 \) hr and \( −8.5 \) hr.

The CNN yields a \( \sim 20\% \) higher TSS to classify PE and NE magnetograms than the baseline classification that uses the discriminant analysis of MTOT. In the following, we attempt to interpret the CNN performance to understand the information learned to classify PE and NE magnetograms.

4.3. The Mapping between the MTOT and the CNN Output

Given an input PE or NE magnetogram segment, the trained CNN outputs a value between 0 and 1. The predicted binary class labels 0 and 1 are obtained by thresholding the output at 0.5. The original CNN output can be interpreted as the probability of the input magnetogram segment showing emergence after a certain time. Thus, the trained CNN maps the surface magnetic field (input) to the probability of emergence (output). Patterns common to the magnetograms for which the CNN output is \( y ≃ 0 \) are weakly correlated with emergence and patterns common to the magnetograms for which the CNN output \( y \approx 1 \) are strongly correlated with emergence.

We visually inspect patterns common to the PE and NE magnetogram segments by arranging the magnetograms according to corresponding CNN output into six bins centered at CNN output values \( y = \{1/12, 3/12, 5/12, 7/12, 9/12, 11/12\} \), bounded by \( y \pm \Delta/2 \), where the bin width is \( \Delta = 1/6 \). The left panel of Figure 10 shows representative magnetograms (taken at \( −3.2 \) hr prior to emergence) with accurate predicted class labels from the six CNN output bins. We see that the CNN output is low \( y ≃ 0 \) (top row) for the magnetograms with very low as well as very high value of MTOT. With the CNN output progressively increasing for the magnetograms in the subsequent bins, the corresponding value of MTOT either progressively increases from the low value or progressively decreases from the high value. The magnetograms that yield \( y ≃ 1 \) (bottom row) are the magnetograms for which the MTOT value lies within an intermediate range. The high value of TSS \( ∼ 85\% \) to classify PE and NE magnetogram segments using the CNN implies that the distribution of MTOT of magnetograms within this intermediate range closely matches the distribution of MTOT for the PE samples.

The right panel of Figure 10 similarly shows magnetogram samples with accurately predicted class labels arranged in the six bins of CNN output. Overall, there are significantly fewer such magnetogram segments compared to the number of magnetograms with accurately predicted class labels (Figure 11). The PE magnetogram segments that yield \( y \approx 0 \) are relatively unclean emergences, i.e., the segments emerge in the vicinity of a strong preexisting magnetic field. The NE magnetogram segments that yield \( y \approx 1 \) show a distinct bipolar magnetic field structure similar to that of the PE active region (Figure 3). Only a few NE magnetogram segments with moderately high MTOT also yield a high CNN output \( y \approx 1 \).

Figure 10 clearly shows that there is a wide range of MTOT values for many of the CNN output bins. Therefore we analyze MTOT of PE and NE samples from each category of the CNN output bins. We only consider a subpopulation of PE and NE magnetograms between MTOT of \( 4 \times 10^{20} \) Mx and \( 16 \times 10^{20} \) Mx and discard a small number of PE and NE samples with very high and very low values of MTOT (see Appendix B). The left panel of Figure 11 shows the distribution of PE and NE magnetograms binned according to the corresponding CNN output. The PE distribution clusters at \( y \approx 1 \), and the NE distribution clusters at \( y \approx 0 \).

We calculate the average value \( \langle MTOT \rangle_{PE/NE} = (1/A) \sum_{bin} \langle MTOT \rangle_{PE/NE} \binom{PE}{NE} \), from each bin of the CNN output. The right panel of Figure 11 shows the variation of the mean value of \( \langle MTOT \rangle_{PE/NE} \) with CNN output. PE segments with low CNN output \( y \approx 0.1 \) have high \( \langle MTOT \rangle_{PE} = 12 \times 10^{20} \) Mx, and NE segments with low CNN
output $y \sim 0.1$ have low $\langle \text{MTOT}\rangle_{\text{NE}} \sim 7 \times 10^{20}\text{Mx}$. As the CNN output increases, $\langle \text{MTOT}\rangle_{\text{PE}}$ decreases and $\langle \text{MTOT}\rangle_{\text{NE}}$ increases. The PE and NE samples that produce output $y > 0.5$ fall within an intermediate range of $\langle \text{MTOT}\rangle$ between $9 \times 10^{20}\text{Mx} - 11 \times 10^{20}\text{Mx}$. Thus, MTOT is an important factor contributing to the CNN performance such that magnetograms with the CNN output $y \geq 0.5$ have an average MTOT value within an intermediate range (between $9 \times 10^{20}\text{Mx} - 11 \times 10^{20}\text{Mx}$). As shown in the left panel of Figure 9, the population fraction of the PE magnetograms peaks within the MTOT range of $9 \times 10^{20}\text{Mx} - 11 \times 10^{20}\text{Mx}$. Therefore, $\langle \text{MTOT}\rangle$ values in the right panel of Figure 11 are affected by the original distribution of PE and NE magnetograms with respect to MTOT. An explicit analysis of the CNN output as a function of MTOT shows that

Figure 9. Left: discriminant analysis of PE and NE magnetogram segments, taken 3.2 hr before emergence, based on the $\text{MTOT} = \sum |B_{\text{Rad}}| \text{dA}$. The TSS achieved is $\sim 65\%$, significantly lower than the TSS obtained using the CNN. Dashed lines show the discriminant boundaries at which PE and NE population densities are equal. Right: comparison of TSS achieved to classify PE and NE magnetogram segments using the CNN (this work) and the discriminant analysis using MTOT of PE and NE magnetograms taken at different PE times. Also shown are the results reported in the LBB survey (Barnes et al. 2014) to classify EARs and CRs using the discriminant analysis applied to average the unsigned radial magnetic field. The CNN outperforms the discriminant-analysis classification, most significantly for magnetograms at PE times $-3.2$ hr and $-8.5$ hr. The $1\sigma$ error bars are shown.

Figure 10. Categorizing input PE and NE magnetogram samples (3.2 h before emergence) into bins corresponding to the CNN output values. For each bin, three samples with a minimum MTOT (left) and three samples with maximum MTOT (right) are shown. The bins are arranged such that the corresponding CNN output increases from top to bottom. Left: Magnetograms with accurately predicted class labels. Accurately classified PE magnetograms are true positives (TPs), and accurately classified NE magnetograms are true negatives (TNs). Right: Magnetograms with inaccurately predicted class labels. Inaccurately classified PE magnetograms are false negatives (FNs), and inaccurately classified NE magnetograms are false positives (FPs). The magnetogram plots are saturated at 150 G (white) and $-150$ G (black). For TN ($y \sim 0$) magnetograms, the MTOT value can be very low as well as very high. For TP ($y \sim 1$) magnetograms, MTOT lies in an intermediate range.

The Astrophysical Journal, 903:27 (18pp), 2020 November 1 Dhuri et al.
Figure 11. Statistical analysis of the MTOT of PE and NE magnetograms taken 3.2 hr before emergence. The PE and NE samples are distributed in bins according to the corresponding CNN output. The bins of CNN output are centered at \( y = \{1/12, 3/12, 5/12, 7/12, 9/12, 11/12\} \) and bounded by \( y = \Delta/2 \) where \( \Delta = 1/6 \). Left: The population density of PE and NE samples binned by the CNN output. For a significant majority of the PE samples \( y \approx 1 \), and for a significant majority of the NE samples \( y \approx 0 \). Right: The average (MTOT) calculated over PE and NE samples from each bin of the CNN output. The PE and NE samples for which the CNN output \( y \leq 0.5 \) are labeled nonemerging and \( y > 0.5 \) are labeled emerging. The \( \langle \text{MTOT}\rangle_{\text{PE}} \) decreases from a high value as the CNN output increases, whereas \( \langle \text{MTOT}\rangle_{\text{NE}} \) increases from low value as the CNN output increases. Thus, NE samples with low CNN output \( y \approx 0 \) have low values of \( \langle \text{MTOT}\rangle_{\text{NE}} \), and PE samples with low CNN output \( y \approx 0 \) have high values of \( \langle \text{MTOT}\rangle_{\text{NE}} \). For PE and NE samples with high CNN output \( y \approx 1 \), \( \langle \text{MTOT}\rangle_{\text{PE}} \) and \( \langle \text{MTOT}\rangle_{\text{NE}} \) values are in the intermediate range. The 3\( \sigma \) error bars are shown.

Figure 12. Filter contributions (maximum positive value times the filter weight, Equation (C1)) of the final convolutional layer in the CNN trained with the pruning algorithm (Appendix C). The PE and NE samples are categorized into five bins according to the MTOT. The average filter contribution \( \langle F_i \rangle_{\text{bim}} \) corresponding to PE and NE samples from each bin and each filter \( i \) is plotted. The contributions from filters 1 and 2 increase with growing MTOT, whereas filters 3 and 4 show decreasing contribution with increasing MTOT. Thus, the CNN incorporates filters that respond positively and negatively to increasing MTOT. The 20\( \sigma \) error bars are shown.

4.4. Outputs of the Filters in the Final Convolutional Layer

We analyze the response of the four convolutional filters in the final layer of the pruned CNN with respect to MTOT. We divide the subpopulation of PE and NE samples (Appendix B) into five bins of equal widths. For the PE and NE samples in each bin, we calculate the filter contribution \( F_i \) (maximum positive value multiplied by the filter weight, Equation (C1)) for each of the four filters in the pruned CNN. We obtain the average value of the filter contribution \( \langle F_i \rangle_{\text{bim}} \) for each bin for each filter \( i \). Figure 12 shows filters 1 and 2, which correspond to surface magnetic field patterns that are correlated with emergence, \( \langle F_i \rangle_{\text{bim}} \) increases with increasing MTOT. Similarly, for filters 3 and 4, which correspond to surface magnetic field patterns that are anticorrelated with emergence, \( \langle F_i \rangle_{\text{bim}} \) decreases with increasing MTOT. Thus, the trained CNN incorporates filters that respond both positively and negatively to MTOT of the PE and NE magnetograms. The magnitude of the filter contributions \( \langle F_i \rangle_{\text{bim}} \) increases monotonically with increasing MTOT.

To further illustrate the operation of the convolutional filters in the pruned CNN, we visually inspect the output of each filter for a few representative PE and NE samples. Figure 13 shows select examples of outputs of each convolutional filter, binned by the normalized contributions (Equation (C1)) from the filter. Filters 1 and 2 contribute positively toward the CNN output. (The minimum) MTOT of the magnetogram samples also

The number of filters in the final layer during training to only a few filters that are easier to interpret. This process is known as network pruning (LeCun et al. 1990; Han et al. 2015; Qin et al. 2018; Frankle et al. 2019).

We develop a network-pruning algorithm to iteratively reduce the final convolutional layer filters from \( N = 256 \) to \( N = 4 \). The pruning algorithm is based on identifying top \( N/4 \) filters that are maximally correlated and top \( N/4 \) filters that are maximally anticorrelated with the surface magnetic field patterns associated with emergence and removing the remaining relatively neutral \( N/2 \) filters (Appendix C). The four filters in the final convolutional layer at the end of the training are the dominant four filters that contribute to the CNN output, two corresponding each to surface magnetic field pattern that is correlated and anticorrelated with emergence.
increases with filter contribution. We see that the filter output is maximum for the prominent magnetic flux regions in the input magnetograms, yielding positive filter output. Filters 3 and 4 contribute negatively to the machine prediction. Moreover, the (minimum) MTOT of the magnetogram samples decreases as the filter contribution increases. Filter 3 produces positive and negative outputs corresponding to magnetic flux areas. Filter 4 produces negative output corresponding to magnetic flux areas of the input magnetograms. Information about the polarity of the magnetic flux areas may not be necessary to train a CNN that performs comparably to one trained on data including the field (see Appendix F). These filter outputs also spatially conform to the edges of the magnetic regions (see Figure 14).
4.5. CNN Output and the Length Scale of Magnetic Regions

This analysis shows that the outputs of the convolutional filters in the final layer strongly depend on the MTOT of PE and NE regions. MTOT is also an important factor deciding the final CNN output, although other factors are important for the classification (see Appendices E and G). The value of MTOT depends on the size of the magnetic regions and on the magnetic field intensity. We use synthetic magnetograms to explicitly test the dependence of the CNN output on both these factors.

We use circular synthetic bipolar magnetic regions with uniform magnetic field intensity, as shown in the left panel of Figure 15, to probe the CNN output. The right panel of Figure 15 shows the pruned CNN output for synthetic bipoles of different magnetic field intensity and length scales (radius). We see that the CNN output for a synthetic bipole of a given size generally increases with increasing value of the magnetic field intensity. The CNN output saturates for a magnetic field value of \( \sim 1000 \) G and falls sharply to 0 beyond the length scale of \( \sim 30 \) Mm. For the 100 G bipole, the CNN output peaks \((y \sim 0.9)\) at a length scale of 15 Mm and falls sharply to 0 beyond a length scale of \( \sim 30 \) Mm. For the 10 G bipole, the CNN output is consistently low \((y \sim 0.2)\). Thus, the CNN associates magnetic bipoles of small length scales and intense fields with emergence, according to Figure 3.

Similar to Figure 13, Figure 14 shows outputs of the filters of the pruned CNN for a synthetic bipole. As discussed earlier, the filter outputs spatially coincide with the edges of the magnetic regions. Figure 16 shows contributions (maximum positive value times the filter weight, Equation (C1)) of the filters of the pruned CNN as a function of the synthetic bipole length scale for a 100 G uniform magnetic field.

Figure 15 shows the pruned CNN output for synthetic bipoles of different magnetic field intensity and length scales (radius). We see that the CNN output for a synthetic bipole of a given size generally increases with increasing value of the magnetic field intensity. The CNN output saturates for a magnetic field value of \( \sim 1000 \) G and falls sharply to 0 beyond the length scale of \( \sim 30 \) Mm. For the 100 G bipole, the CNN output peaks \((y \sim 0.9)\) at a length scale of 15 Mm and falls sharply to 0 beyond a length scale of \( \sim 30 \) Mm. For the 10 G bipole, the CNN output is consistently low \((y \sim 0.2)\). Thus, the CNN associates magnetic bipoles of small length scales and intense fields with emergence, according to Figure 3.

Figure 17 shows the CNN output for synthetic bipoles for CNNs with an increasing number of convolutional layers. The correlation between the CNN output and length scale of the synthetic bipoles exists for a CNN with more than two convolutional layers. The characteristic length scale beyond which the CNN output sharply drops increases with increasing number of CNN layers.

Figure 14. Outputs of filters in the final convolutional layer of the pruned CNN for a synthetic bipole of 100 G uniform field and radius 25 Mm (Figure 15). Filter outputs are saturated at 1 (red) and –1 (blue). 100 G (white) and –100 G (black) contours from the synthetic bipole are shown on top of the filter outputs. Filter outputs conform to the edges of the positive- and negative-polarity regions.

Figure 15. Left: circular synthetic bipole of uniform magnetic field strength used to probe the trained CNN. Right: output of the pruned CNN as a function of (uniform) intensity of the magnetic field and synthetic bipole length scale (radius). For 100 G and 1000 G, there is a characteristic length scale for the CNN beyond which the output sharply drops to 0.

Figure 16. Filter contributions \( F_i \) (maximum positive value times the filter weight, Equation (C1)) of the pruned CNN as a function of the synthetic bipole (Figure 15) length scale (radius) for a 100 G uniform magnetic field.

Figure 17. The CNN output for synthetic bipoles for CNNs with an increasing number of convolutional layers. The correlation between the CNN output and length scale of the synthetic bipoles exists for a CNN with more than two convolutional layers. The characteristic length scale beyond which the CNN output sharply drops increases with increasing number of CNN layers.
output decreases to 0 is therefore a cumulative result of the contributions from all filters.

Figure 17 shows the dependence of the characteristic length scale for the CNN output on the depth of the CNN (without pruning filters in the final layer). A correlation between the length scale of the magnetic regions and the CNN output forms a CNN with more than two layers, which results in an increased length scale for the CNN output on the depth of the CNN.

The characteristic length scale continues to increase for the CNN with layer 6 (∼40 Mm), after which it falls abruptly (∼25 Mm). This is possibly a result of the severe reduction of the input magnetograms (size 2 × 2 and 1 × 1 for layers 7 and 8, respectively) via max-pooling. Thus, the characteristic length scale depends on downsampling the input as a result of increasing number of CNN layers as well as on the size of the max-pooling kernel.

4.6. CNN Visualizations

The analysis has thus far focused on deriving quantitative information about the surface magnetic field patterns that the CNN uses to discriminate between PE and NE magnetogram samples. The approach is specific to the problem at hand and has provided sufficient information to understand the operation of the trained CNN. The ML literature also provides us with tools for qualitative interpretation or visual explanations of deep neural networks (Simonyan et al. 2013; Zeiler & Fergus 2014; Selvaraju et al. 2017). These tools, such as saliency maps (Simonyan et al. 2013) and gradient-based class-activation maps (Selvaraju et al. 2017), make use of the gradient information that is backpropagated (Hastie et al. 2001) from the output layer to the hidden and input layers.

Here, we generate visualizations of the trained CNN using a simpler technique known as occlusion maps (Zeiler & Fergus 2014). Occlusion maps are obtained by measuring changes in the output when different patches in the input are masked. We use 25 × 25 pixel mask for input PE and NE magnetogram segments to generate occlusion maps. The mask size is chosen to be large enough to cause a significant difference in the CNN output after occlusion. For a given PE or NE sample, the occlusion map is of the same size as the magnetogram and initialized with the uniform value of the predicted class label $Y_{\text{pred}} = 1$ or 0 for the sample. A new predicted class label $Y_{\text{mask}}$ is obtained after masking or occluding a 25 × 25 pixel patch in the input image. The values at the corresponding pixels in the occlusion map are updated to $Y_{\text{pred}} - Y_{\text{mask}}$. Thus, the positive (negative) values in the occlusion map represent regions in the input image that are correlated (anticorrelated) with emergence as seen by the CNN.

For a comprehensive visualization of occlusion maps corresponding to all PE and/or NE segments from an EAR or CR, we recombine subsampled magnetograms, corresponding CNN outputs and occlusion maps (see Appendix D). Figure 18 shows the CNN outputs and occlusion maps for two CR and two EAR samples. These samples are taken from 3.2 hr before emergence. The MTOT values of 85 × 85 pixel segments including various regions of interest are quoted in Table 4.

![Figure 18](image-url) Visual depictions from the trained CNN using occlusion maps (Zeiler & Fergus 2014), obtained by systematically masking patches in the input and noting the change in the CNN output. The positive (negative) pixels in the occlusion map indicate regions in the input that are correlated (anticorrelated) with emergence as seen by the CNN. The CNN output and occlusion maps, calculated using Equations (D1) and (D2), respectively, are shown for two CRs (top) and two EARs (bottom) taken from 3.2 hr before emergence. The magnetic field in EARS and CRs is saturated at 150 G (white) and −150 G (black). The output is saturated at 0 (blue) and 1 (red). The occlusion map is calibrated at −1 (blue) and 1 (red). Note that the emergence takes place within the central $10^5 \times 10^5$ region of the EARS. Magnetic regions of interest, classified as true positives (TP, red), true negatives (TN, blue), and false positives (FP, black) are encircled.

Table 4

| MTOT (10^{20} Mx) | MTOT (10^{20} Mx) |
|-------------------|-------------------|
| TP1 10.8          | TN1 5.2           |
| TP2 15.2          | TN2 23.3          |
| FP1 7.2           | TN3 6.5           |
| FP2 55.2          | TN4 6.5           |
| FP3 8.6           | TN5 6.7           |

Figure 18 shows the CNN outputs and occlusion maps for two CRs (top) and two EARs (bottom). The MTOT values of 85 × 85 pixel segments including various regions of interest are quoted in Table 4.

![Figure 17](image-url) The MTOT of 85 × 85 pixel Segments Including Various Regions of Interests Labeled in Figure 18.
Table 4. Note that overall, MTOT values of TN regions are low, with the exception of TN2, for which the MTOT value is high. Moreover, overall, the MTOT values of TP and FP regions are intermediate, with the exception of FP2, for which the MTOT value is very high. This is largely consistent with our earlier analyses. Regions TP1, TP2, FP1, FP2, and FP3 are highlighted positively in the occlusion map. Regions FP1 and FP3 show small-scale bipolar field patterns, which are an early sign of emergence, similar to regions TP1 and TP2. Regions TN2, TN3, and TN4 are highlighted negatively in the occlusion map, whereas regions TN1 and TN5 are neutral in the occlusion map. These regions include bipoles, which may also be a factor for the CNN classification (Appendix E). The CNN output $y \sim 1.0$ for FP2 suggests that there may be other factors in addition to the MTOT and bipolarity for the CNN classification (see Appendix G). These factors, however, are not easily interpreted.

5. Summary

We perform a statistical analysis of EARs in the SDO/HEAR dataset (Schunker et al. 2016) using CNNs. The trained CNN described here successfully classifies PE and NE magnetograms with a TSS of $\sim 85\%$, 3.2 hr before emergence, which is significantly better than the TSS obtained in the LBB survey (Birch et al. 2012; Leka et al. 2012; Barnes et al. 2014). The TSS score decreases to $\sim 40\%$, 24.5 hr before emergence, similar to the LBB survey.

The discriminant analysis in the LBB survey (Barnes et al. 2014) showed that the PE surface magnetic field is a leading factor in distinguishing between EARs and CRs (Figure 3). We therefore perform a discriminant analysis of PE and NE magnetograms using the MTOT $\equiv \sum |B_{LOS}| \, dA$ to establish a baseline for comparing the CNN performance. We find that the maximum TSS achieved using the MTOT for classification of the PE and NE magnetograms is $\sim 65\%$, 3.2 hr before emergence. Thus, the trained CNN performs better than the baseline model of discriminant analysis using only the MTOT.

Through visual inspection of the PE and NE magnetograms, binned by the corresponding CNN output, we identify the MTOT as an important attribute of magnetograms for the CNN classification. We find that the CNN produces maximum output $y \sim 1$ for magnetograms with an MTOT in an intermediate range. Magnetograms with very low as well as very high values of MTOT yield a low CNN output $y \sim 0$. Furthermore, statistical analysis of the MTOT of PE and NE magnetograms, also binned by CNN output, reveals that the average value of the MTOT for magnetograms that are predicted as emerging by the CNN lies between $9 \times 10^{20}$ Mx–$11 \times 10^{20}$ Mx. These findings are validated by visual insights into the operation of the CNN obtained using occlusion maps (Zeiler & Fergus 2014). The TSS for classification decreases for PE and NE magnetograms from earlier PE times as the difference between distributions of the MTOT for the PE and NE samples diminishes (Figure 3). Explicit analysis of the CNN output as a function of MTOT shows that there are other factors in addition to MTOT that contribute to the CNN performance (Appendix G). These factors may be associated with length scales, field intensities, and the geometry of the PE small-scale bipolar shown in Figure 3.

The fully CNN (Figure 6) for this work was specifically chosen to facilitate subsequent assessment of the contribution of the convolutional filters to the final layer. Still, all filters in the final layer $N = 256$ are difficult to analyze in detail to understand the specific patterns learned. We therefore develop a network-pruning algorithm to systematically remove unimportant filters from the final layer during training and retain only the most important (four) filters (Appendix C). From visual inspection as well as statistical analysis of filter outputs, we find that the CNN incorporates convolutional filters that are sensitive to the MTOT of PE and NE magnetograms. Some of these filters yield positive output in response to increasing magnetic flux, whereas others yield a negative output. The CNN output is the sum total of outputs from all filters. A multivariable discriminant analysis of the MTOT and a measure of dipole moments of magnetograms yields a classification TSS of 75.6% ± 3.3% (3.2 h before emergence), which is comparable to the CNN (Appendix E).

Using synthetic bipolar magnetograms of circular shape, we demonstrate that a characteristic length scale of magnetic regions exists beyond which the CNN output drops to 0 regardless of the magnetic field intensity. The characteristic length scale is also a cumulative result of contributions from all filters and is an artifact of the number of layers in the CNN and max-pooling operation. Explicit dependence on these factors remains to be explored.

The CNN output peaks for small-scale ($\sim 30$ Mm) and intense ($\sim 100$ G) synthetic bipoles. Our analysis does not conclusively show that the bipolarity of magnetograms is an important factor for the CNN classification. Interestingly, a CNN trained on PE and NE segments without any polarity information also yields a comparable value of the classification TSS (Appendix F), altogether this suggests that the CNN identifies small-scale and (unipolar or bipolar) intense fields as a characteristic PE pattern. A time-dependent analysis is necessary to shed light on the formation of such small-scale fields by either rising intact from deep within the convection zone (Fan 2009) or by forming over a less localized area near the surface (Brandenburg 2005). Furthermore, Birch et al. (2019) showed that the emerging flux interacts with the supergranulation pattern, and therefore emergence locations are correlated with supergranulation. The network is highlighted in PE and NE magnetograms because magnetic flux tends to collect in supergranular lanes. Thus, the CNN performance here may also be dependent on factors associated with the supergranulation. These remain to be explored.

It is to be noted that the SDO/HEAR dataset considered here contains limited ($\sim 200$) independent samples of EARs and CRs. The quantity of data required for training the CNN to gain deeper and concrete insights about PE patterns may not be known in advance. However, the limited dataset hinders the CNN from learning the convolutional filters associated with more and general complex patterns than just the total unsigned PE magnetic field, e.g., small-scale PE bipoles. The SDO/HMI instrument provides terabytes of high-resolution, high-cadence magnetic field data every day. With these data, it may be possible to construct a large enough dataset, with more general selection criteria compared to the SDO/HEAR dataset, to study emergence. Such a dataset is also likely to conform to real-time emergence scenarios and therefore may also be analyzed with a focus on forecasting emergence. Alternatively, it may be possible to train the CNN convolutional filters to exactly detect the PE bipolar magnetic field and correlated patterns using additional constraints on convolutional filters (Bayar & Stamm 2018) and/or to learn the emergence location along
with a probability (Zhao et al. 2018). A machine that successfully completes a task like this will also be useful for the temporal evolution analysis of the PE. These analyses are deferred to future work. With Petabyte-scale astronomical datasets expected in the upcoming decade, ML is expected to become increasingly relevant in analyzing data (Baron 2019). Insights obtained from this work may therefore be useful for training and interpreting deep neural networks in future solar and astrophysical data-analysis applications.

D.B.D., S.M.H., A.C.B., and H.S. designed the research. D.B.D. performed the data analysis. All authors contributed to the interpretation of the results. D.B.D. wrote the manuscript with contributions from all authors.

S.M.H acknowledges funding support from the Max-Planck partner group program. The HMI data used here are courtesy of NASA/SDO and the HMI science team. We acknowledge partial support from the European Research Council Synergy Grant WHOLE SUN #810218. Observations are courtesy of NASA/SDO and the HMI science teams. We thank the anonymous referee for their comments and suggestions that helped improve the analysis and clarity of the paper.

Appendix A
Neural Networks

A deep neural network consists of many layers of neurons, as shown in Figure 19. The $i^{th}$ neuron in the $n^{th}$ layer performs the following generic operation:

$$x_i^n = f\left(\sum_{j=1}^{N} w_{ij} x_j^{n-1} + b_i^n\right).$$

(A1)

Here, $x_j^{n-1}$ are the outputs from the neurons in the $(n - 1)^{th}$ layer (the zeroth layer is the input layer), $w_{ij}$ is the weight for the $i^{th}$ input from the $(n - 1)^{th}$ layer, $b_i^n$ is termed the bias, and $f$ is the activation function for the neuron. The activation applied on the final layer produces the output $y_{pred}$. The error in the predicted output is determined by a loss function $L(y, y_{pred})$. During training, weights and biases of the neurons in the network are determined via stochastic gradient descent to minimize the loss $L(y, y_{pred})$. Here, we use binary cross-entropy (CE) as the loss function, which is a popular choice for classification problems (Hastie et al. 2001). It is defined as

$$L_{CE}(y, y_{pred}) = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log (y_i^{pred}) + (1 - y_i) \log (1 - y_i^{pred})],$$

(A2)

where $N$ is the batch size, i.e., the number of examples considered in the stochastic gradient descent. The training process involves many cycles of iterations (epochs) over the entire training set, and the loss function $L_{CE}$ converges gradually during training.

Appendix B
Subpopulation of PE and NE Samples

The attribute of the PE and NE magnetograms that stands out for the CNN classification is the total unsigned line-of-sight magnetic flux, MTOT (Section 4). Therefore, we analyze MTOT of PE and NE samples from each category of the CNN output bins. The distribution of PE and NE samples according to MTOT is shown in Figure 20. For the statistical analysis binned according to MTOT (Figure 13), it is desirable that each bin contains a sufficient number of samples for meaningful statistics. We therefore select samples with an MTOT value between $4 \times 10^{20}$ Mx and $16 \times 10^{20}$ Mx (both inclusive). For each bin within this MTOT range, the number of available samples is large, 1500.

Appendix C
Network Pruning

Neural network pruning was primarily proposed to reduce the computation and memory required for tasks such as vision and speech recognition, natural language processing, on embedded mobile applications (Han et al. 2015). The pruned network, significantly reduced in size compared to the original, is also easier to interpret (Frankle et al. 2019). A straightforward approach to network pruning is to identify important neuron connections during the training step and remove all connections with weights below a threshold value (Han et al. 2015). This process may be repeated in an iterative manner.

Figure 19. A two-layer deep fully connected (FC) neural network with four inputs and one output.

Figure 20. Distribution of PE and NE samples taken at 3.2 hr before emergence according to the MTOT. We select samples with $4.0 \times 10^{20}$ Mx $\leq$ MTOT $\leq 16.0 \times 10^{20}$ Mx for the statistical analysis.
slightly higher learning rate, the CNN is not retrained subsequent to the training, yielding the trained CNN with four convolutional layers. The pruning is visible as discontinuities in the training and validation losses. The number of CNN filters in the final convolutional layer at each training stage are indicated below the curve for that corresponding stage. During training to obtain a compressed trained network. In the functionality-based approach, one can remove those neurons from the network that are functionally similar to the other neurons (Qin et al. 2018).

Here, we perform pruning of the final convolutional layer of the CNN in an iterative manner to reduce the number of filters from 256 to 4. Our pruning strategy is based on identifying the convolutional filters that maximally contribute toward the network output. In the CNN (Figure 6), the final convolutional layer is connected to the output layer through a $6 \times 6$ max-pooling layer. The max-pooling layer picks out the pixel with the maximum value from the $6 \times 6$ output of a convolutional filter. For a given input, the contribution of each filter in the final layer is thus given by

$$ F_i = \max(O_i)w_i, $$

where $O_i$ is the $6 \times 6$ output of the $i$th convolutional filter in the final convolutional layer, and $w_i$ is the corresponding weight. The output of the CNN is obtained by applying a sigmoid function (Hastie et al. 2001) to the sum of the output of each convolutional filter,

$$ y = \text{sigmoid} \left( \sum_{i=1}^{N} F_i + b \right), $$

where $b$ is the bias of the output neuron, and $N$ is the number of convolutional filters in the final layer. The output $y$, a value between 0 and 1, is a measure of the probability of the input magnetogram showing emergence at a certain time $t$. The contribution of CNN filters $F_i$ which correspond to the surface magnetic field patterns correlated (anticorrelated) with emergence, increases (decreases) the CNN output $y$.

We develop a pruning algorithm to retain the top four filters while maintaining prediction accuracy—two each corresponding to patterns correlated and anticorrelated with emergence. We initiate the training with $N = 256$ filters in the final convolutional layer and perform the pruning operation during training, which is repeatedly applied every few epochs. At each pruning step, half the number of filters associated with patterns that are minimally correlated and anticorrelated with emergence are removed.

We identify the filters to be pruned in the following manner. At each pruning step during training, we calculate the total contribution $S_{i}^{TP}$ of each convolution filter $i$ for all samples $n^{TP}$ predicted as true positives (TPs) from the training data,

$$ S_{i}^{TP} = \sum_{j=1}^{n^{TP}} F_i. $$

Similarly, we calculate the total contribution $S_{i}^{TN}$ for all samples predicted as true negatives (TNs) from the training data. Finally, we calculate net positive contribution $S_{i}^{net}$ of each convolutional filter $i$ for all accurately classified samples from the training data,

$$ S_{i}^{net} = S_{i}^{TP} - S_{i}^{TN}. $$

For a convolutional filter associated with patterns maximally correlated with emergence $S_{i}^{net} \gg 0$. Likewise, for a convolutional filter associated with patterns maximally anticorrelated with emergence $S_{i}^{net} \ll 0$. Thus, at each pruning step, we retain $N/4$ filters with maximum $S_{i}^{net}$ value and $N/4$ filters with minimum $S_{i}^{net}$ value. The remaining $N/2$ filters, which are relatively neutral to the patterns associated with emergence, are removed. The network is trained further for a predefined number of epochs, after which it is pruned again to remove half the number of neutral filters. This is repeated until only four filters remain in the final layer—the dominant two corresponding each to patterns correlated and anticorrelated with emergence. The step-by-step algorithm is as follows:
Table 5

Mean Fivefold Cross-validation TSS to Classify Uniform-polarity PE and NE Magnetograms at Different PE Times Obtained Using the CNN (Figure 6) with Five Hidden Convolutional Layers

| time before emergence (h) | fivefold cross-validation TSS (%) |
|---------------------------|----------------------------------|
| −3.2                      | 88.20 ± 4.78                    |
| −8.5                      | 60.07 ± 1.53                    |
| −13.9                     | 50.66 ± 7.32                    |
| −19.2                     | 43.89 ± 7.30                    |
| −24.5                     | 40.68 ± 2.03                    |

Note. The 1σ error is quoted.

1. Initialize the CNN (Figure 6) with the final convolutional layer filters \( N = 256 \).
2. Initialize a function for the number of training epochs \( E(N) \) as a function of the number of the final convolutional layer filters.
3. Train the network for epochs \( E(N) \).
4. Using the trained network, obtain TP samples \( n^{TP} \) and TN samples \( n^{TN} \) of all PE and NE magnetograms in the training data.
5. Calculate the net contribution of each of the \( i \) filters \( S_{ij}^{net} \) (Equation (C4)) toward the CNN output for all accurately predicted PE and NE samples (TPs and TNs) in the training data.
6. Identify \( N/4 \) filters with maximum (positive) \( S_{ij}^{net} \) and \( N/4 \) filters with minimum (negative) \( S_{ij}^{net} \). Remove the remaining \( N/2 \) filters.
7. If \( N = 4 \), i.e., after pruning, stop and output the trained CNN model. Else go to 3.

Using this algorithm, we train the CNN and iteratively prune the final convolution filters from \( N = 256 \) to \( N = 4 \). Training and validation losses as training progresses are shown in Figure 21. We see that the losses converge in an identical manner, with discontinuities encountered at each pruning stage. The number of filters in the final convolutional layer is shown below the curve corresponding to each pruning stage. At each pruning step, half the convolutional filters are removed according to the pruning algorithm. At the end of the final epoch, pruning is carried out for the last time, reducing the convolution filters from \( N = 8 \) to \( N = 4 \). The network is not retrained subsequently.

The CNN, with the final convolution layer pruned to \( N = 4 \) filters, yields \( \text{TSS} = 82.50\% \pm 4.52\% \) to classify PE and NE samples (−3.2 h before emergence), which is comparable to the original trained CNN with \( N = 256 \) filters (Table 3). The CNN obtained from pruning a random assortment of \( N/2 \) filters in the final convolution layer at each step yields \( \text{TSS} = 34.71\% \pm 28.86\% \). This validates the algorithm developed for identifying filters in the final convolutional layer that are correlated and anticorrelated with surface magnetic field patterns associated with emergence. We analyze the top four filters that are retained after the pruning to interpret the performance of the CNN for classification of the PE and NE magnetograms (Section 4). Note that in this case, the CNN with \( N = 4 \) filters in the final convolutional layer from the outset and trained for an identical number of epochs yields \( \text{TSS} = 78.38\% \pm 6.60\% \), which is also comparable to the original trained CNN. In general, the number of optimum filters (neurons) in a CNN (deep neural network) is not known in advance, therefore the functionality-based pruning approach developed here is useful for future applications as well. Further pruning the CNN to \( N = 2 \) filters in the final layer yields a mean TSS of 68.9% with a significantly larger standard deviation of 17.1%.

Appendix D

Recombining Subsampled PE and NE Magnetograms

The \( 10^4 \times 10^4 \) PE and NE magnetograms used for training the CNN are randomly subsampled from \( 30^4 \times 30^4 \) EARS and CRs. Hence, PE and/or NE magnetograms subsampled from an EAR or CR may contain overlapping regions (see Section 2). These PE and/or NE segments subsampled from an EAR or CR may be recombined to recover the original EAR or CR in the following manner. Let the \((i,j)\) pixel from an EAR or CR be included in \( C_{ij} \) number of PE and/or NE magnetogram samples from the EAR or CR. We obtain the net CNN output corresponding to each pixel of the EAR or CR as

\[
y_{ij} = (1/C_{ij}) \sum_{k=1}^{C_{ij}} y_{ij}^k, \quad (D1)
\]

where \( y_{ij}^k \) is the CNN output corresponding to \( k - th \) PE and/or NE sample from the EAR or CR. Similarly, the occlusion map \( A \) for each EAR or CR pixel is obtained as

\[
A_{ij} = (1/C_{ij}) \sum_{k=1}^{C_{ij}} A_{ij}^k, \quad (D2)
\]

where \( A_{ij}^k \) is the occlusion map corresponding to the \( k - th \) PE and/or NE sample from the EAR or CR. By recombining PE and/or NE magnetogram segments into the original EAR or CR, we comprehensively inspect outputs and occlusion maps of all PE and/or NE magnetograms that are part of the respective EAR or CR.

Appendix E

Multivariable Discriminant Analysis of the MTOT and Dipole Moment

Through visual inspection and statistical analysis of accurately classified PE and NE magnetograms, we find that MTOT is an important factor for the CNN classification. As noted in Figure 2, the appearance of a small-scale bipolar field is an early sign of emergence. Therefore bipolarity of the magnetograms may also be an important factor for the CNN classification.

We calculated a baseline TSS using a nonparametric discriminant analysis of only the MTOT. We extend the baseline model to include a measure of “dipole moment” of the magnetograms (Wilson 1994; Illarionov et al. 2015), calculated as

\[
\text{DM} = \left| \sum_{i} B_i (r_i - r_0) \right|, \quad (E1)
\]

where \( B_i \) is the magnetic field of the \( i \)th pixel located at \( r_i, r_0 \) is a reference pixel, and the sum is over all pixels in the magnetogram. We estimate probability density of PE and NE magnetograms as a function of two variables—MTOT and DM —using the Epanechnikov kernel (Silverman 1986; Barnes et al. 2014). Similar to the nonparametric discriminant analysis
of MTOT, the smoothing parameter for the density estimation is chosen to be optimum for a normal distribution. For each cross-validation set, the probability density for PE and NE magnetograms is estimated using the training data. The trained density estimator is then used to obtain the probability score for PE and NE magnetograms in the validation data, and a classification label is obtained to calculate the TSS. The cross-validation TSS for the multivariable discriminant analysis for different pre-emergence times thus obtained is plotted in Figure 22. We see that the multivariable discriminant analysis using MTOT and DM is comparable to the CNN classification results within the error bars.

Appendix F
CNN Classification of Uniform Positive-polarity PE and NE Magnetograms

Our analysis shows that the CNN classification depends on MTOT and outperforms the discriminant analysis of only MTOT. We also show that the discriminant analysis of MTOT and DM—a measure of bipolarity—is comparable to the CNN classification. This suggests that the bipolarity of magnetograms may also be an important factor for the CNN classification. To validate this, we train the CNN (Figure 6) with uniform positive-polarity PE and NE magnetograms, i.e. the absolute value of the line-of-sight field. We find that this CNN yields a TSS comparable to the CNN trained on the original PE and NE magnetograms (see Table 5). This means that the polarity information in data is not necessary for the CNN to yield a classification performance that is superior to the discriminant analysis of only MTOT.

To understand the operation of the CNN trained on uniform-polarity data, we train a CNN with $N = 4$ filters from the outset in the final layer and visualize the filter outputs in the final layer corresponding to a synthetic magnetogram with a uniform field similar to that in Figure 14. The CNN with $N = 4$ filters in the final layer yields a TSS of 71.75% ± 8.68%. The final layer filter outputs are shown in Figure 23. We find that the filter outputs conform to the edges of magnetic flux regions and are both positive and negative. Because the synthetic magnetogram contains no information about the polarity, the positive and negative filter outputs do not necessarily correspond to the polarity of the magnetic flux regions in Figure 23 and also in Figure 14. The filter outputs are a result of complex operations performed by the CNN in the hidden layers and may not be easily interpreted.

Figure 23. Outputs of filters in the final convolutional layer of the CNN, with $N = 4$ filters in the final layer, trained on uniform-polarity PE and NE segments. The input is a synthetic magnetogram with 100 G uniform-polarity field and radius 25 Mm.
Figure 24. The average CNN output ($y$) of PE and NE magnetogram samples binned by the unsigned line-of-sight magnetic flux MTOT. PE and NE samples considered are from MTOT range of 0–5 × 10$^{20}$ Mx. ($y_{PE}$) is consistently lower than 0.2. ($y_{NE}$) is greater than 0.5 for an intermediate range of MTOT between 5 × 10$^{20}$–20 × 10$^{20}$ Mx. 5σ error bars are shown.

Appendix G

**CNN Output versus MTOT**

Figure 11 shows that the average unsigned line-of-sight magnetic flux (MTOT) of PE and NE samples binned by the CNN output lies in an intermediate range of $9 \times 10^{20} - 11 \times 10^{20}$ Mx for samples with the CNN output $y \sim 1$. While this indicates a dependence of the CNN output on MTOT, the actual values of MTOT of samples with $y \sim 1$ may not all lie in the intermediate range and may be higher as well as lower. To explicitly study the dependence of the CNN output on MTOT, we consider the average CNN output of PE and NE samples binned by MTOT. We consider samples from a slightly wider range of MTOT of 0–25 × 10$^{20}$ Mx than considered earlier (Appendix B). The average CNN output ($y$) is plotted in Figure 24. While ($y_{NE}$) is consistently low <+0.2 and mostly independent of MTOT ($y_{PE}$) is >+0.5 for samples with MTOT values between 5 × 10$^{20}$–20 × 10$^{20}$ Mx and <+0.5 for samples with lower or higher MTOT values. Importantly, ($y_{PE}$) is consistently and significantly higher than ($y_{NE}$) regardless of MTOT. This suggests that the CNN output depends on other important factors in addition to MTOT.

**References**

Barnes, G., Birch, A. C., Leka, K. D., & Braun, D. C. 2014, ApJ, 786, 19
Barton, D. 2019, arXiv:1904.07248
Bayar, B., & Stamm, M. C. 2018, IEEE Transactions on Information Forensics and Security, 13, 2691
Birch, A. C., Braun, D. C., & Fan, Y. 2010, ApJL, 723, L190
Birch, A. C., Braun, D. C., Leka, K. D., Barnes, G., & Javorkin, B. 2012, ApJ, 762, 131
Birch, A. C., Schunker, H., Braun, D. C., & Gizon, L. 2019, A&A, 628, A37
Bobra, M. G., & Couvidat, S. 2015, ApJ, 798, 135
Bobra, M. G., Sun, X., Hoeksema, J. T., et al. 2014, SoPh, 289, 3549
Brandenburg, A. 2005, ApJ, 625, 539
Brandenburg, A., Gressel, O., Jabbari, S., Kleeroin, N., & Rogachevskii, I. 2014, A&A, 562, A53
Cameron, R., Dikpati, M., & Brandenburg, A. 2016, SSRv, 210, 367
Cheung, M. C. M., & Isobe, H. 2014, LRSP, 11, 3
Eastwood, J. P., Biffis, E., Haggood, M. A., et al. 2017, Risk A, 37, 206
Fan, Y. 2009, LRSP, 6, 4
Frankle, J., Dziugaite, G. K., Roy, D. M., & Carbin, M. 2019, arXiv:1903.01611
Glorot, X., & Bengio, Y. 2010, in Proc. Machine Learning Research, Vol. 9, Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, ed. Y. W. Teh & M. Titterington (Sardinia, Italy: FMLR), 249, http://proceedings.mlr.press/v9/glorot10a.html
Goodfellow, I., Bengio, Y., & Courville, A. 2016, Deep Learning (Cambridge: MIT Press)
Han, S., Pool, J., Tran, J., & Dally, W. J. 2015, in NIPS’15: Proc. 28th Int. Conf. on Neural Information Processing Systems, Learning both Weights and Connections from Efficient Neural Networks, ed. C. Cortes et al. (Red Hook, NY: Curran Associates), 1135
Hastie, T., Tibshirani, R., & Friedman, J. 2001, The Elements of Statistical Learning, Springer Series in Statistics (Berlin: Springer)
Illarionov, E., Tlatov, A., & Sokoloff, D. 2015, SoPh, 290, 351
Komm, R., Howe, R., & Hill, F. 2009, SoPh, 258, 13
Komm, R., Howe, R., & Hill, F. 2011, SoPh, 268, 407
Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2012, in Advances in Neural Information Processing Systems (NIPS) 25 (Red Hook, NY: Curran Associates, Inc), 1097
LeCun, Y., Bengio, Y., & Hinton, G. 2015, Natur, 521, 436, EP
LeCun, Y., Denker, J. S., & Solla, S. A. 1990, in Advances in Neural Information Processing Systems 2, Optimal Brain Damage, ed. D. S. Touretzky (Burlington, MA: Morgan Kaufmann), 598
Leka, K. D., Barnes, G., Birch, A. C., et al. 2012, ApJ, 762, 130
Peirce, C. S. 1884, Sci, 4, 453
Peissel, W. D., Thompson, B. J., & Chamberlin, P. C. 2012, SoPh, 275, 3
Qin, Z., Yu, F., Liu, C., & Chen, X. 2018, AIMS, 1, 149
Schou, J., Scherrer, P. H., Bush, R. I., et al. 2012, SoPh, 275, 229
Schunker, H., Birch, A. C., Cameron, R. H., et al. 2019, A&A, 625, A53
Schunker, H., Braun, D. C., Birch, A. C., Burston, R. B., & Gizon, L. 2016, A&A, 595, A107
Selvaraju, R. R., Cogswell, M., Das, A., et al. 2017, in Proc. 2017 IEEE International Conference on Computer Vision (ICCV), Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization (Piscataway, NJ: IEEE), 618
Shelhamer, E., Long, J., & Darrell, T. 2017, IEEE Trans. Pattern Anal. Mach. Intell., 39, 640
Shibata, K., & Magara, T. 2011, LRSP, 8, 6
Silverman, B. W. 1986, Density Estimation for Statistics and Data Analysis (London: Chapman Hall)
Simonyan, K., Vedaldi, A., & Zisserman, A. 2013, arXiv:1312.6034
Simonyan, K., & Zisserman, A. 2014, arXiv:1409.1556
Stein, R. F. 2012, LRSP, 9, 4
Thompson, W. T. 2006, A&A, 449, 791
Wilson, P. R. 1994, The Origin of the Large-scale Fields Cambridge Astrophysics (Cambridge: Cambridge Univ. Press), 142
Zeiler, M., & Fergus, R. 2014, Lecture Notes in Computer Science, Visualizing and Understanding Convolutional Networks (1st ed.; Cham: Springer), 818
Zhao, Z.-Q., Zheng, P., Xu, S.-t., & Wu, X. 2018, arXiv:1807.05511