SPACE SCIENCES

A deep learning virtual instrument for monitoring extreme UV solar spectral irradiance

Alexandre Szenicer[1,2]*†, David F. Fouhey[2,3]*†, Andres Munoz-Jaramillo[2,4], Paul J. Wright[2,5,6], Rajat Thomas[2,7], Richard Galvez[2,8], Meng Jin[2,9,10], Mark C. M. Cheung[2,5,9]

Measurements of the extreme ultraviolet (EUV) solar spectral irradiance (SSI) are essential for understanding drivers of space weather effects, such as radio blackouts, and aerodynamic drag on satellites during periods of enhanced solar activity. In this paper, we show how to learn a mapping from EUV narrowband images to spectral irradiance measurements using data from NASA's Solar Dynamics Observatory obtained between 2010 to 2014. We describe a protocol and baselines for measuring the performance of models. Our best performing machine learning (ML) model based on convolutional neural networks (CNNs) outperforms other ML models, and a differential emission measure (DEM) based approach, yielding average relative errors of under 4.6% (maximum error over emission lines) and more typically 1.6% (median). We also provide evidence that the proposed method is solving this mapping in a way that makes physical sense and by paying attention to magnetic structures known to drive EUV SSI variability.

INTRODUCTION

The extreme ultraviolet (EUV) radiation from the Sun is the dominant driver of the Earth's thermosphere/ionosphere system. During periods of elevated solar activity, enhanced solar EUV driving causes adverse space weather effects such as radio communication blackouts, increased aerodynamic drag on satellites in low-earth orbit, and scintillation of Global Navigation Satellite Systems (GNSS) signals (1–3). NASA's Solar Dynamics Observatory (SDO) (4) was launched in 2010 with the goal to help us understand solar activity and how their variations affect life on Earth. SDO carries three instruments. The EUV Variability Experiment (EVE) instrument suite (5) provides EUV solar spectral irradiance (SSI) measurements integrated over the full Sun. The spectrum measured by the Multiple EUV Grating Spectrograph A (MEGS-A) and MEGS-B modules of EVE contains emission lines from a variety of ions that exist at temperatures ranging from 7000 K to 9.3 MK. Of these, 39 lines can be extracted with relative ease and these define the level 2 lines data product. The Atmospheric Imaging Assembly (AIA) instrument (6) captures full-disk images of the Sun at 4096 × 4096 pixels in seven EUV channels, two UV channels, and one visible wavelength channel. The Helioseismic and Magnetic Imager (HMI) (7) delivers photospheric vector magnetograms and Dopplergrams of the full Sun at 4096 × 4096 pixels. The AIA and EVE instruments are two separate, although complementary, entities, and they are both crucial in their own right. EVE is meant to give us detailed information about the radiative budget of the Sun and how it affects the Earth's atmosphere, but without any spatial resolution. On the other hand,

AIA is meant to give us information about the structure of the corona, its density, temperature, and evolution, but with lower spectral resolution than EVE. Presently, AIA, HMI, and the MEGS-B module of EVE continue to co-observe the Sun. However, an electrical malfunction of the EVE MEGS-A channel compromised our ability to monitor EUV SSI in the 5- to 37-nm wavelength range, which, as shown in Fig. 1, contains roughly 60% of the solar irradiance in the EUV. The goal of this work is to fill this gap in measurement capabilities with a virtual replacement for MEGS-A. Toward this goal, we examine how well a differential emission measure (DEM)–based model and learning-based models can use narrowband AIA images to reproduce MEGS-A EUV SSI measurements.

This approach is feasible due to both a wealth of data from when AIA and EVE instruments co-observed the Sun and the common underlying solar dynamics that generate both sets of observations. Before it was taken offline, both the MEGS-A channel and the AIA instrument observed the Sun for roughly 4 years together, producing a set of paired AIA and MEGS-A observations from which we can learn a data-driven mapping. This mapping ought to exist because of three reasons:

1) The same population of solar plasma is expected to be responsible for emitting the EUV radiation observed by the two instruments.

2) As shown in Fig. 1, images taken by the AIA instrument have a significant overlap with the EUV spectra measured by EVE's MEGS-A module.

3) As shown in Fig. 1, each of the AIA filters has a different and overlapping response to radiation emitted by plasmas of different temperature, enabling the use of AIA filters as basis of a mapping between plasma temperature and irradiance.

There are currently two main approaches for using information contained in AIA images to reconstruct solar EUV irradiance. The first one uses AIA images to estimate the DEM of the solar corona (a combined estimate of the density and temperature of the coronal plasma along the line of sight) (8), in combination with the CHIANTI Atomic Database (9, 10) to produce a full integrated spectrum. The second uses an empirical segmentation and classification of pixels of AIA images into different types of solar structures, each with a different radiative output (11), which is also integrated to produce a complete spectrum. However, because of the relatively sparse wavelength coverage of AIA, simplifying physical assumptions used in both methods (i.e.,

[1]Department of Earth Sciences, University of Oxford, Oxford, UK. [2]Frontier Development Lab, Mountain View, California, USA. [3]Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor, Michigan, USA. [4]Southwest Research Institute, Boulder, Colorado, USA. [5]Hansen Experimental Physics Laboratory, Stanford University, Stanford, California, USA. [6]SUPA School of Physics and Astronomy, University of Glasgow, Glasgow, UK. [7]Department of Psychiatry, Academic Medical Centre, University of Amsterdam, Amsterdam, Netherlands. [8]Center for Data Science, Center for Cosmology and Particle Physics, New York University, New York, NY, USA. [9]Lockheed Martin Solar and Astrophysics Laboratory, Palo Alto, California, USA. [10]SETI Institute, Mountain View, California, USA.
*These authors contributed equally to this work.
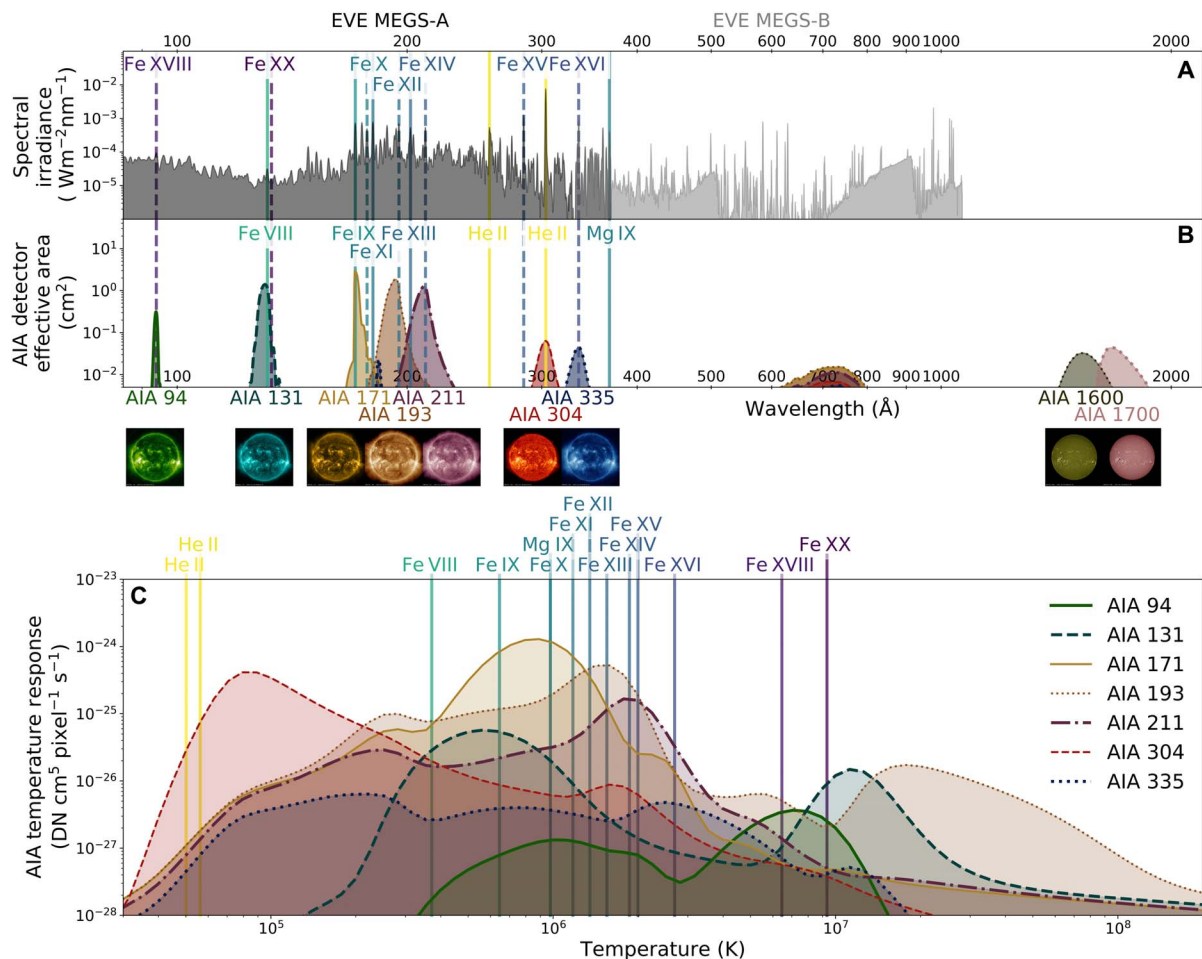†Corresponding author. Email: alexandre.szenicer@linacre.ox.ac.uk (A.S.); fouhey@umich.edu (D.F.F.)

Fig. 1. Spectral and temperature response of AIA and EVE. The spectra measured by the EVE Instrument (**A**) and spectral response of the narrowband images from the AIA instrument onboard SDO (**B**) overlap. Each of the 14 emission lines we recover in this project (shown in all panels as vertical lines) has a characteristic temperature associated with them (denoted using shades of color). Given that each of the AIA filters assembles light emitted by plasma of a wide range of temperatures (**C**), it is possible to combine their information to recover the part of the spectrum that used to be measured by the MEGS-A instrument, which is no longer operational. For these EVE spectra (A), the MEGS-A spectral region contains 60% of the total solar EUV irradiance. AIA images and EVE data from 5 January 2014.

the assumption that emission is formed under coronal equilibrium conditions and that elemental abundances are uniform in the solar corona), and the empirical segmentation of images, inaccuracies in the mapping from AIA to MEGS-A are expected. In this work, we examine how a data-driven approach can be used to provide an improved mapping between these two instruments.

We propose a data-driven approach that fuses two complementary models and takes advantage of recent developments in machine learning (ML). The first component is a linear model that transforms global summary statistics about the AIA input to produce the EVE MEGS-A output; this handles the bulk of the mapping, especially during quiet periods in solar activity. The second component is a convolutional neural network (CNN) model that uses AIA images to additively correct the output of this linear model, often during flaring periods. Both components are learned from data.

Given the fact that this is the first application of CNNs to the problem of EUV spectral irradiance reconstruction, one of the contributions of this paper is a proper protocol and baselines for this application. These include training and testing splits that can evaluate whether the method is generalizing as opposed to memorizing, and proper

baselines to examine and better understand the contributions of more advanced methods. We also report results from a DEM-based model that aims to directly reconstruct the DEM distribution, as well as simple statistical models for the mapping problem, which can serve as benchmarks that future, more complex methods of inferring EUV irradiance need to outperform.

Our proposed method is able to achieve strong performance with a median (across MEGS-A lines) average relative error of 1.6% and a maximum error (on Fe XVI) of 4.6%, outperforming both our baseline models: a linear model on engineered features and a physics method based on DEM inversions. Furthermore, while our proposed method improves the overall results, it is substantially better on channels that are spectrally far away from the AIA observations than our baselines, and in circumstances that violate the underlying assumptions of the DEM-based model. In addition, although ML models have a reputation for being indiscernible black boxes, we show evidence that our learned model is solving the task in a sensible way: A careful specification of our neural network architecture provides spatially resolved maps that suggest how the CNN is using information to infer irradiance, and these maps match our physical intuition.

RESULTS

Data setup and evaluation criteria

We train and evaluate our learned mappings using data from the period between 1 January 2011 and 26 May 2014, when the SDO/AIA imager and the SDO/EVE spectrograph co-observed the Sun. Our AIA data derive from (12), consisting of nine-channel AIA images at 6-m cadence (in total: 264K images) that have been corrected for angular resolution variation and instrument degradation. We further downsample these data spatially to 256 × 256. The EVE MEGS-A data are also obtained from (12), consisting of a 15-channel signal at 10-s cadence. The Fe XVI 361 Å emission line has too few observations (available less than 1% of the time compared to other lines) to learn a meaningful model, and thus, we work with the remaining 14 lines. We assign to each AIA observation the MEGS-A observation that is closest to it in time (mean/maximum time distance, 8.5 s/12 s). Thus, the problem becomes the mapping of a 256 × 256 × 9 measurement to a 1 × 1 × 14 measurement.

Because the Sun's dynamics at the level of total EUV irradiance move at a far slower rate than our observations, it is crucial to properly split the data into training (used for fitting model parameters), validation (for choosing model hyperparameters), and test (for evaluating prediction) sets. In particular, randomly splitting the data, commonly done in other settings, will yield overly optimistic performance estimates because access to one observation makes it easy to predict the subsequent observation. We therefore split our data year wise. We set aside years 2012 and 2013 as testing data, and use the remaining data, covering 2011 and 2014, for training and validation. We split this sequentially into 70% training (1 January 2011 to 18 December 2011) and 30% validation (18 December 2011 to 31 December 2011 and 1 January 2014 to 27 May 2014).

We evaluate predictions using the relative absolute error (i.e., if y is the ground truth observed irradiance and y_p is the predicted, we report $|y - y_p|/|y|$). This produces an error for every single observation and line. We summarize results into tables by taking the average over the observations, i.e., producing an average relative absolute error. We additionally report results only during flaring conditions. Rather than manually specifying flaring times, we use ground truth EVE Fe XX observations as a proxy variable for flare conditions and define all observations in the top 5% of Fe XX emissions across the whole dataset as flares.

Baseline models

To put our results into proper context, we first report the results of a number of simpler methods. We begin with an approach that models the physics of the problem, and then report the performance of a set of models that are data driven, but use only nonspatially resolved summary statistics about AIA measurements. These results additionally provide future experiments about MEGS-A nowcasting with proper baselines.

Differential emission measure model

To illustrate the need for a data-driven solution, we first analyze the results of a model that uses only a physical model. In particular, we transform each set of six EUV images (i.e., all AIA EUV channels except the 30.4 nm) into DEM maps (8) spanning 18 temperature bins from log T = 5.5 to 7.2 K, at $\Delta \log T$ = 0.1. This is done with the assumption that the emitting plasma is optically thin, that the emission is formed under coronal equilibrium conditions, and that elemental abundances are uniform in the solar corona. With the same assumptions, the DEM solutions are then used to compute the emission of EUV lines observed by EVE MEGS-A. Once the pixel-by-

pixel EUV line emissions have been obtained, we integrate them spatially to obtain the line irradiances (as reported by MEGS-A). The specific inversion method used here is the neural DEM (DeepEM) method outlined in (13). This implementation is able to provide solutions where the underlying Linear Programming (LP) solver of standard basis pursuit DEM inversion technique (8) fails to converge on a satisfactory solution. The DeepEM method was trained on data from 2011, mirroring our models.

Channel summary statistics models

We expect that many of the MEGS-A observations can be explained empirically by the total intensity of the AIA images because AIA observes data at nearby wavelengths to the MEGS-A spectrum (see Fig. 1). To demonstrate this, we use features that are the result of averaging spatially, i.e., producing a single feature per-AIA channel, and then fit a series of models to these summary statistics features.

As features, we use the average AIA data count and standard deviation of each AIA image (AIA_μ and AIA_σ, both nine-dimensional feature vectors), i.e., $AIA_{\mu,k} = \frac{1}{256^2} \sum_{i=1,j=1}^{256,256} AIA_{i,j,k}$ and similarly for AIA_σ. The first feature captures the total irradiance, and the second feature reveals the extent to which that irradiance is constant across the Sun. While AIA_σ is nonlinear with respect to the original AIA image, AIA_μ is linear and any linear model fit to it demonstrates to what extent a properly parameterized linear model explains the observed MEGS-A data.

We report results for a number of models. All model parameters were fit with the training set, and relevant hyperparameters were fit on the validation set (see Materials and Methods for more precise details). (i) We begin with the most basic model, a least-squares fit to the average AIA image, or L2-Mean, which represents the most basic data-driven approach to this problem. (ii) Our most complex and effective method is Huber-Mean-Std, or finding a linear model on AIA_μ and AIA_σ features that minimizes the robust Huber loss; regularization and Huber parameters were optimized on the validation set. While it has few parameters, this model already involves both feature and model engineering. (iii) To evaluate the importance of the loss function compared to standard least squares, we also report results for L2-Mean-Std, or the same model, but fit to minimize a squared Euclidean distance. (iv) Similarly, to evaluate whether the nonlinear standard deviation features help, we report Huber-Mean or fitting the parameters only to AIA_μ. Note that because the mean is a linear operator, this evaluates how well a properly fit linear model can explain performance.

We report absolute relative error for the above methods in Table 1. Before discussing the results in detail, we point out that some of the MEGS-A channels should be readily predicted from AIA observations, while others should not: Some MEGS-A channels are spectrally close to AIA observations.

The DEM-based model obtains strong performance on a number of lines, with 6/14 lines predicted with under 3% error. Its performance on the He II 304 Å is far worse, because the plasma is optically thick for this line and because it forms in nonequilibrium conditions (14). For these reasons, AIA DEM inversions do not even use data from the He II 304 Å channel. With this in mind, it is not a surprise that the pure DEM-based model gives predicted He II 304 Å irradiances that are a factor 37 smaller than the corresponding MEGS-A measurement. The metrics for this line quoted in Table 1 have been adjusted to account for this huge discrepancy.

Some other lines, such as Fe XX (which is sensitive to plasma at about 10 million K), have discrepancies of ~10% during quiet conditions,

but the average error can increase to 60% during flares. This may be due to a number of reasons. First, saturation of the AIA charge-coupled device (CCD) detectors during flares can result in bleeding of signal along columns of pixels that overlap with the flaring site. Second, automatic exposure control during flaring times (which aim to mitigate CCD saturation) can degrade the signal-to-noise of AIA images. Third, there is evidence that coronal plasma in the quiescent state may have different elemental abundances than plasma that have been impulsively evaporated to fill flare loops (15).

A properly fit linear model (Huber-Mean) results in better performance on most lines compared to the physics model and strong performance on many lines overall. Many emission lines (Fe VIII, Fe IX, Fe XII, Fe XIII, and He II) can be explained well by a linear transformation of the average of the AIA observation, with average relative errors close to 1% and as low as 0.7%. While the linear transformation is simple, choosing the right loss function is important: The model fit with Huber regression consistently improves over L2.

While a linear model applied to the average of the AIA observation does well on many lines, it performs poorly on many others, including Fe XX, Fe XV, Fe XVI, and Mg IX. Adding standard deviations as a feature gives a proxy variable for the presence of flares, and thus, Huber-Mean-Std method improves performance on these flare sensitive lines, reducing, e.g., Fe XVI prediction from 10.5% relative error to 7.1%. This performance gain on flare-sensitive lines comes at the cost of modest increases in error for the lines that are modeled

well by a linear transformation of the AIA data: The larger capacity models start fitting to noise to better explain the training data. Nonetheless, during flaring conditions, performance degrades: The Fe XX emission prediction error is 440% of the error seen over all conditions.

The best performing linear model (Huber-Mean-Std) is outperformed on some lines (Fe XV, Fe XVI, and Mg IX) by the physics model, and one might naturally wonder whether the linear model is simply underpowered. We experimented with two standard ways of adding expressiveness to the model: adding pairwise interaction features between all the terms (16) and applying gradient boosting regression trees (17), a standard nonlinear model. Both failed to substantially improve upon performance of the Huber loss on Mean + Std model, with average errors of 2.5 and 2.8%, respectively, as compared to 2.5%. Additional information, it seems, should come either from additional physics knowledge to provide a better fitting model (as used in the DEM-based model) or by examining the spatial information in the image (e.g., actually looking at the flare rather than inferring their presence via features).

Convolutional neural networks

Having demonstrated what appear to be the limits of models that consider only first and second moment summary statistics of AIA data, we now present results from models that examine them spatially, in particular CNNs. These models build a single function that is

Table 1. Per-line average relative error for the methods evaluated in this paper. We provide summary statistics for different models for all lines, where the values are percentages of the average relative error defined in the "Data setup and evaluation criteria" section in Results. The top part of the table indicates results averaged over the entire test set, whereas for the bottom part only data points where the Fe XX line is in the 95th percentile are kept, to focus on flare conditions. Numbers in bold indicate the best result within the given column.

Line and log T	He II 4.7 303 Å	He II 4.8 256 Å	Fe VIII 5.6 131 Å	Fe IX 5.8 171 Å	Fe X 6.0 177 Å	Mg IX 6.0 368 Å	Fe XI 6.1 180 Å	Fe XII 6.1 195 Å	Fe XIII 6.2 202 Å	Fe XIV 6.3 211 Å	Fe XV 6.3 284 Å	Fe XVI 6.4 335 Å	Fe XVIII 6.8 93 Å	Fe XX 7.0 132 Å	Min	Mean	Median	Max
Overall																		
Physics	8.66	6.09	3.53	2.76	2.86	3.58	2.66	2.62	2.92	2.78	4.90	5.69	13.07	13.39	2.62	5.39	3.55	13.39
L2-Mean	2.70	2.48	1.95	1.53	1.95	3.80	1.76	1.63	2.05	**1.97**	4.28	10.03	1.78	11.71	1.53	3.54	2.01	11.71
L2-Mean-Std	2.48	2.40	2.72	2.15	2.77	4.56	2.30	2.30	2.63	3.34	3.81	7.48	2.53	10.85	2.15	3.74	2.67	10.85
Huber-Mean	**1.15**	1.36	1.70	**0.68**	2.00	4.90	2.36	1.83	1.80	2.00	5.63	10.51	1.74	3.55	0.68	2.94	1.92	10.51
Huber-Mean-Std	1.27	1.74	**1.21**	0.79	2.04	4.71	1.92	1.39	**1.33**	2.44	4.92	7.12	1.44	2.67	0.79	2.50	1.83	7.12
Linear + CNN	1.31	**1.24**	1.67	1.95	**1.19**	3.21	1.15	**1.00**	1.50	2.40	**3.72**	**4.61**	1.28	**2.07**	1.00	2.02	1.59	4.61
Flare conditions (Fe XX 95 percentile)																		
Physics	18.18	15.48	16.57	3.87	1.91	**3.23**	2.33	2.60	4.80	2.91	4.41	4.83	66.20	64.00	1.91	15.09	4.60	66.20
L2-Mean	2.16	2.25	2.99	2.07	3.02	5.06	2.84	2.57	2.94	2.93	**3.20**	5.98	4.36	11.52	2.07	3.85	2.97	11.52
L2-Mean-Std	4.18	4.64	5.35	3.71	4.79	7.02	3.96	4.13	4.18	5.06	5.62	5.70	5.20	11.09	3.71	5.33	4.93	11.09
Huber-Mean	**1.58**	1.47	2.15	**0.98**	1.96	4.59	1.72	**1.27**	1.25	**1.76**	4.31	8.45	2.64	17.91	0.98	3.72	1.86	17.91
Huber-Mean-Std	2.05	2.39	1.81	1.24	3.07	4.82	2.36	1.94	2.82	2.71	4.50	5.23	4.04	10.67	1.24	3.55	2.77	10.67
Linear + CNN	1.89	1.55	**1.69**	1.79	**1.48**	3.32	**1.30**	1.41	2.30	2.15	3.27	**3.85**	**2.03**	7.55	1.30	2.54	1.96	7.55

optimized end to end, consisting of interleaved convolutions (that aggregate spatial features) and nonlinearities (that enable the composition of these convolutions to learn nonlinear functions). In the process, the model discovers features that, when extracted from the image, help make good inferences.

These models face large learning challenges when applied to the AIA to MEGS-A nowcasting problem. First, there is very little supervision available for learning the mapping: Only 80K data points are available for training and are densely sampled in time, causing correlations and decreasing the effective number of data points. Second, a great deal, but not all, of the mapping can be explained by a linear transformation of basic summary statistics of the AIA irradiance, and the CNN must capture this linear mapping while also incorporating other features visible in the images. While the CNN could, in principle, compute the average AIA intensity via a set of properly designed filters, the CNN's parameterization and random initialization in practice might encourage it to converge to a more complex solution (e.g., consisting of a series of basis expansions followed by a linear model).

We avoid this problem by training a model that explicitly corrects the predictions of the linear model, as shown in the bold components in Fig. 2. We fit a linear model to the training set, using mean and standard deviation features to represent each AIA observation and minimizing Huber loss, and then compute the residual, $y_{diff} = y - y_{linear}$, between ground truth EVE irradiance and the linear model's prediction on the training set. This y_{diff} serves as the learning target for the CNN. We adjust the input of the CNN to compensate for this: The CNN is predicting a variable that has had, in some sense, the information about its average image removed and must then focus on the spatial details. We accordingly compute the per-channel mean of the image and subtract it, yielding an image that is, per channel, each pixel's deviation from the mean.

At inference time, the network produces an estimate of y_{diff}, which is added to the prediction from the linear model. The inference time procedure is thus computing per-channel statistics over the input and making a prediction using the linear model, using this per-channel mean of the AIA image to convert the AIA image to a deviation-from-the-mean image, and making a correction prediction via a CNN.

As a side benefit, this method also uses the linear model to explain a great deal of the variability in the output, which, we will show later, has a number of advantages. One such advantage is that the linear model can also be turned into a spatial model (i.e., produce a per-pixel prediction), which we demonstrate in a subsequent section.

CNN model configurations
Because of the size of the data (recall the data are both <80K observations and samples are highly non-iid), we experimented primarily with smaller networks. We obtained strong performance with a shallow network derived from AlexNet (18) with approximately 1 million parameters. We could obtain similar performance from more standard networks, including an 18-layer ResNet (19) (11.7 million parameters), but found that the high capacity of the model coupled with limited variability of the data led to severe overfitting even with regularization. In contrast, the shallower approach worked more consistently. All networks were trained from random initialization (see implementation details in Materials and Methods) to minimize the Huber loss. While transfer learning typically improves results, early experiments yielded negative results when initializing with ILSVRC-pretrained models (20) [by replicating the RGB filters three times in the channel dimension to create nine-channel filters similar to (21)].

Our proposed method, which we denote ANet3, consists of three convolutional layers derived from AlexNet, followed by average pooling, and a linear map. In particular, we use the convolutional blocks from (18) (i.e., convolutions, ReLUs, and max-pools). We adapt the input channel count to match AIA and append a batch normalization layer after each convolution because batchnorm has been shown to be empirically effective. After three convolutions, the feature map is averaged just as in ResNet (19), passed through dropout for regularization, and then linearly transformed into the EVE prediction.

Model results
We report results in Table 1. The method obtains the lowest overall mean error. This error, however, conflates lines for which a linear transformation of the total AIA-observed irradiance suffices and lines for which a linear transformation is a poor model. The proposed model improves on the best of either Huber-Mean and Huber-Mean-Std on 9/14 lines, with especially strong performance gains on modeling Fe XX, Fe XV, Fe XVI, and Mg IX, achieving a 26% error reduction on Fe XVI. This trend continues under flaring conditions, where the CNN reduces the error rate on Fe XX emission prediction from 10.7% (Huber-Mean-Std) to 7.6%. At the same time, as was the case with adding standard deviation features, the increased model capacity leads to performance losses on lines that are well described by a linear model: In particular, lines like Fe IX, Fe XIV, and He II are slightly degraded by the use of a CNN.

The methods used are far shallower than typical approaches used in learning-based computer vision, and one might reasonably wonder whether a more off-the-shelf approach might work. Here, we report a number of additional experiments that help justify our approach. Our early efforts tried a standard ResNet18 (19) trained to directly map from AIA to MEGS-A. We trained these using standard parameters and ones in adjacent orders of magnitude, but the networks overfit quickly, even with regularization, obtaining minimum validation loss in very early epochs and producing poor results. We cannot rule out the possibility that different parameter settings are needed on this problem, but the number of parameters vastly outnumbers the number of independent data points that would point to the problem being that the network is simply far too parameterized.

We additionally show the distribution of predicted-versus-actual irradiances in Fig. 3 for Fe XX emissions. The DEM inversion approach systematically overestimates irradiance. The least-squares fit to mean AIA data count produces a substantially better fit but still cannot accurately model emissions. Changing the loss function for fitting the model and adding additional features produces a much tighter fit, and as seen in Fig. 3 (and the 20% relative decrease in error), using a CNN even further reduces errors.

While the goal of this study is to reproduce the 14 lines of EUV SSI, we performed an additional experiment with the same method on total irradiance to test the generality of the method. Specifically, we trained the same linear + CNN model to predict the SSI over the entire MEGS-A range as an additional output. This model performs similarly well, obtaining a 1.22% relative error on the total irradiance channel.

Opening the black box: Analysis of models
In addition to looking at the predictive abilities of the model, we can analyze the models to evaluate how they are solving the problem. We report analysis of both a channel summary statistic model and the full model. Both suggest that the learning models are, generally, solving the problem in the correct way.
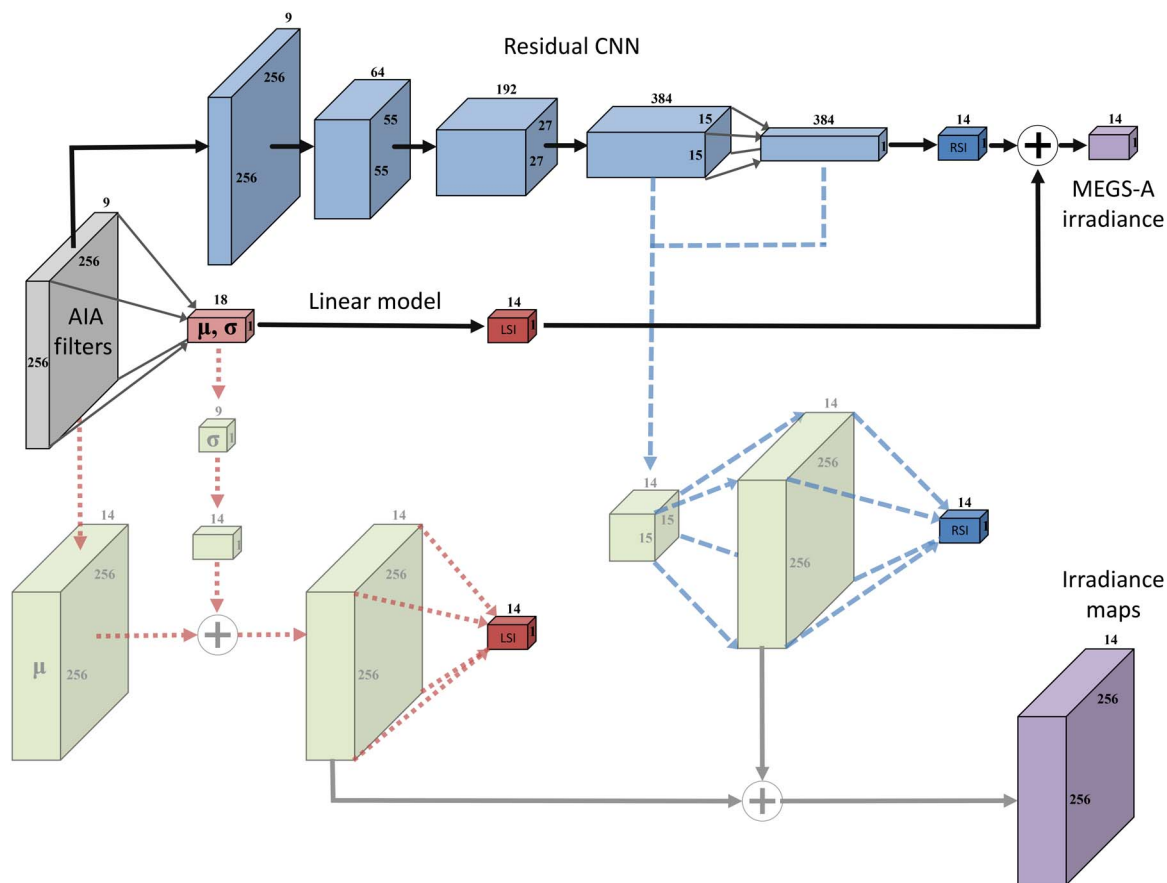
Fig. 2. Proposed neural network architecture. After computing summary statistics of the input AIA images and making a prediction via a linear model, a CNN makes a prediction that corrects this linear model. The combined linear + CNN model is shown in bold colors and arrows. The numbers attached to the boxes denote the sizes of the representations of the data as they goes through the network, e.g., the first block annotated with 256, 256, and 9 represents an input of 256 × 256 pixels and 9 channels. We produce spatially resolved maps in units of irradiance to validate how the CNN is operating (see Fig. 5) by rearranging commutative operations in the last two layers of the CNN (blue dashed path) and the linear model (red dashed path). These operations yield identical outputs as the original (bold) model (illustrated with faint vertical lines), but recasting them this way enable the diagnosis of the model's operation.

One of the simplest ways to visualize a model is to examine its weights. In Fig. 4, we show visualizations of the linear weights for the model using Huber loss on Mean + Std AIA features. As expected, the largest contributor for each emission line tends to be the closest AIA filter. This takes advantage of the direct overlap between most of the MEGS-A emission lines and at least one AIA filter wavelength response (the only exceptions being He II, Fe XV, and Mg IX). However, many of the emission lines use information from more than one AIA filter, as well as information about the variability of the irradiance to adjust their predictions. A careful comparison between each of the emission line weights in Fig. 3 with the AIA filters' temperature response of Fig. 1C shows that our model is using the AIA filters' temperature response as a basis for the reconstruction of EUV SSI. Fe XVIII is a good example of this behavior: Fig. 1B shows that this line is centered on the AIA 94 filter, but Fig. 4 shows clear contributions from the 131, 193, 211, and 335 channels. All these channels can be seen to have similar values of response at the temperature of the Fe XVIII line in Fig. 1C. This means that our model is essentially working the way the DEM-based models do (i.e., performing an internal assessment of the density and temperatures of the solar plasma to estimate irradiance) while using wavelength proximity to

enhance the recovery of the MEGS-A emission lines using the AIA image filters.

In the case of the combined linear + CNN model, as done in (22) and similar to (23), it is possible to change the order of the averaging and linear combination operations in the later stages of the CNN and linear models to produce spatially resolved irradiance maps that can be used as a diagnostic of the operation of the model. For example, the CNN concludes by taking a 15 × 15 384-channel image, averaging it spatially to a 384-dimensional feature vector, and applying a linear transformation to produce the final 14-dimensional EVE predictions. This is equivalent to a model in which the linear transformation is applied per pixel in the 15 × 15 384-channel image (producing a 15 × 15 14-channel image), followed by spatial averaging to a 14-dimensional result. To get a sense of how the CNN is solving the task, we can examine this preaveraged 14-dimensional prediction.

As shown in the faint components of Fig. 2, we apply this technique to our best model by splitting the final prediction into the sum of three linear models applied to: standard deviation AIA features (linear model), average AIA features (linear model), and the average of the last convolution layer (CNN). The first computation remains a
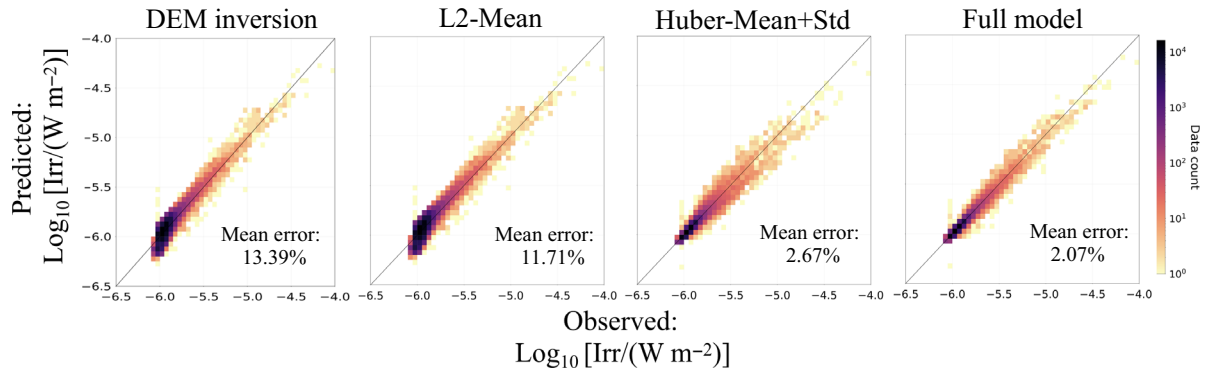
Fe XX (log T = 6.97) emission prediction

Fig. 3. Emission prediction of Fe XX line for several models. We plot a histogram of the results of the prediction on the test set for different model implementations, with a log scale color bar. The closer the points are scattered around the line, the better the predictions are when compared to observations. We can see that the data-driven models greatly improve upon the DEM inversion, and the full model that includes the CNN further increases accuracy, especially on the less frequent, higher amplitude flares.
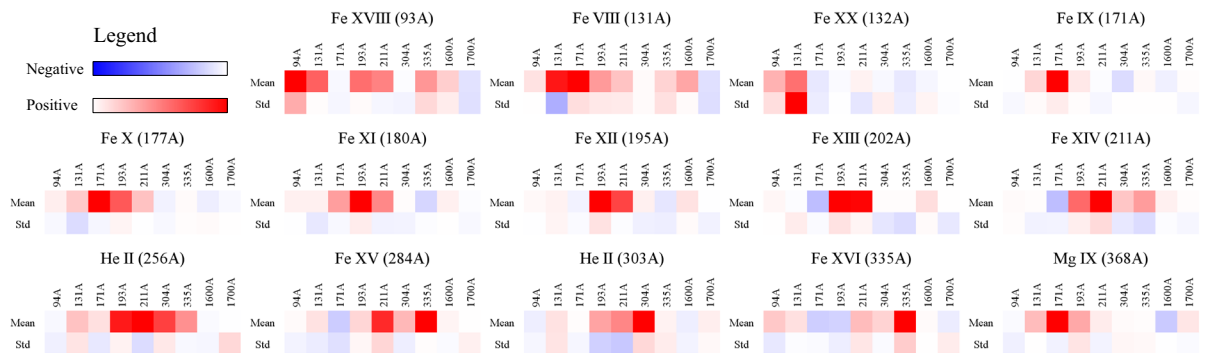
Fig. 4. Linear weights for the Huber Mean + Std model per line. We visualize heatmaps, where red indicates positive weights and blue indicates negative weights with intensity proportional to weight. The first row is features that are the average (i.e., total irradiance); the second is features that are the standard deviation (i.e., variance of irradiance). Most MEGS-A lines are primarily a function of nearby AIA observations (e.g., Fe IX is overwhelmingly just a rescaling of the AIA data). Other lines make use of the standard deviation features (e.g., Fe XX is primarily driven by variance in 131 Å).

per-channel scalar, and we rearrange the latter two, yielding spatially resolved feature maps at 15 × 15 and 256 × 256, respectively. We produce a final spatially resolved map by bilinearly upsampling all components of the 15 × 15 images to 256 × 256 and summing all the feature activation maps in the last convolutional layer. When averaged, this spatial map produces the same results as the original computation graph, but spatially resolved, as can be seen in Fig. 5. These maps look remarkably like one would expect from a spectroscopic imager with bright patches that match the location of active magnetic regions, surrounded by a darker quiet Sun, and even brighter flaring regions. This is evidence that our combined model is paying attention to the right sources of EUV irradiance in a way that makes physical sense. This opens tantalizing opportunities to further constrain them, validate them, and exploit their physicality in future work, showing the potential that deep learning has to enhance the scientific output of solar data.

DISCUSSION

We have implemented an approach for nowcasting MEGS-A spectral irradiance measurements from AIA images, achieving a median error of 1.6% across MEGS-A lines and outperforming a variety of

alternate ML approaches and a DEM-based approach. In particular, our approach produces substantially better performance on a variety of lines that are not well reconstructed by our DEM model (e.g., the flare-sensitive Fe XX line). In addition to quantitatively characterizing the performance of the method, our analysis has produced evidence that the model is solving the AIA to MEGS-A mapping in a physically sensible way. Although it is too early to claim that our approach can produce spatially resolved maps of EUV irradiance based on AIA and EVE data, the results are promising enough that they are worth pursuing. In future work, we plan to integrate the data-driven and physically based techniques to produce models that, by design, solve the problem in a physically valid way.

In this work, we have demonstrated that we can leverage 4 years' worth of AIA and MEGS-A to create a virtual MEGS-A that will be available as long as AIA is functioning. This showcases the incredible potential that CNNs, and deep learning in general, have for heralding a new age of virtual instruments. Other important examples of this potential include the estimation of photospheric velocities based on continuum images (24) and the assembly of superresolution magnetograms based on magneto-hydrodynamic simulations and photospheric magnetograms (25). It is important to stress that these virtual instruments will not be substitutes for their hardware counterparts, as it

AIA EUV channel inputs
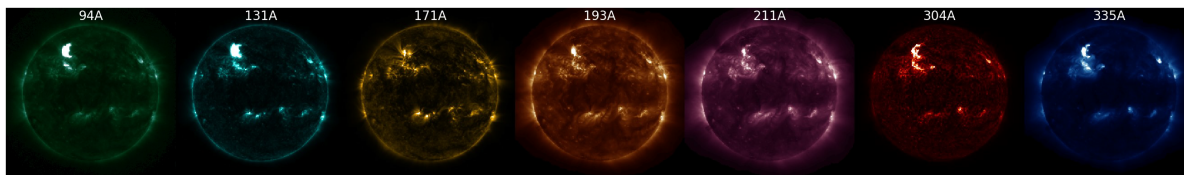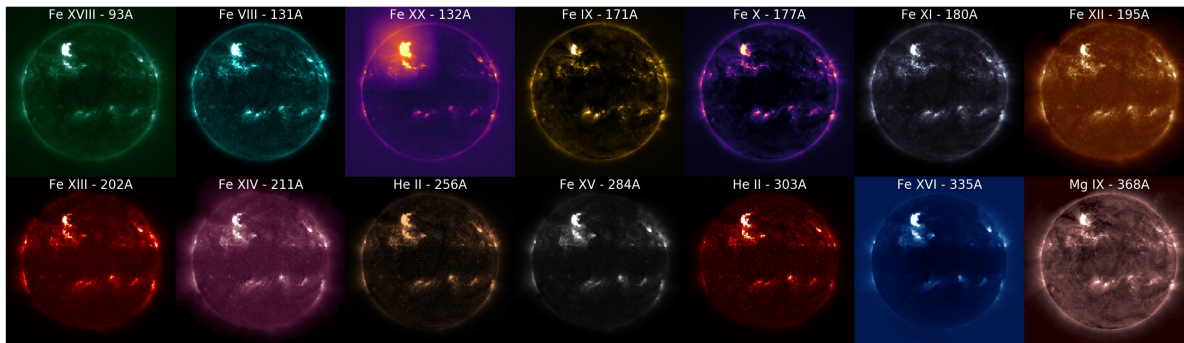

Spatially rearranged MEGS-A predictions

Fig. 5. Spatially rearranged predictions. Results from the model after it has been rearranged to produce spatial results. Global (i.e., image-sized 1 × 1) features are due to linear model applied to standard deviation features; low-frequency (i.e., blocky 15 × 15) features are due to the CNN, which has low spatial resolution; high-frequency details are due to the linear model on average AIA features (i.e., 256 × 256). The linear model has learned a largely correct model mapping AIA to MEGS-A, which is corrected, especially during flaring events, by the CNN.

is clear that this work would have been impossible without the data taken by AIA and the MEGS-A/EVE instruments. However, these virtual instruments will leverage existing and historic scientific instruments to yield similar levels of scientific data products as hardware missions currently do.

MATERIALS AND METHODS

Throughout, we used supervised ML methods. A complete introduction is beyond the scope of this paper and can be found in (26), but to provide context, we will briefly introduce a few key concepts. Supervised learning aims to perform function approximation given a set of N training examples of inputs $X_i : X_i \in \mathcal{X}$ and desired outputs $y_i : y_i \in \mathcal{Y}$. In our case, each X_i may be the 256 × 256 × 9 AIA input$\in R^{256 \times 256 \times 9}$ or a 19-dimensional summary statistic feature vector$\in R^{19}$, and each y_i is a MEGS-A line. Each of our methods is a function parameterized by θ, or $f(X_i; \theta) : \mathcal{X} \rightarrow \mathcal{Y}$. We fit these functions by optimizing the parameters to minimize empirical risk for a given loss function L [which measures how close $f(X_i; \theta)$ is to y_i], $\theta^* = \text{argmin}_\theta \frac{1}{N} \Sigma_{i=1}^N L(f(X_i; \theta), y_i)$, or, in other words, θ^* is the set of parameters minimizing the average loss incurred on the training set. How this is optimized depends on the form of each X_i, the form of f, and the loss function L, but generally, a standard optimization technique such as gradient descent was used.

Some of our initial models are linear regression. For these, we modeled $f(X_i; \theta)$ as $f(X_i; \theta) = \Sigma_{j=1}^D \theta_j X_{ij} = \theta^T X_i$, where D is the dimensionality of the input (typically including a constant term to act as a bias). If we set the loss function to $L(f(X_i; \theta), y) = (f(X_i; \theta) - y)^2$, the resulting minimization problem is ordinary least squares and can be solved analytically both with and without Tikhonov/ridge regression regularization. For more complex loss functions (e.g., Huber),

we can optimize the loss function with gradient descent: Starting with an initial point, we take the gradient of the loss (plus any regularization) with respect to θ and iteratively update θ to follow the negative gradient.

Our full model incorporates a CNN, a complete introduction to which can be found in (27). We have seen that the linear model represents the output of $f(X_i; \theta)$ as a linear or affine transformation of the input. Similarly, a CNN represents the output of $f(X_i; \theta)$ as a transformation of the input, this time involving multiple layers of convolution or linear filtering, interleaved with nonlinearities, resulting in a complex nonlinear function built of smaller simple components. The parameter vector θ is then all of the parameters of the convolutions. While the resulting empirical risk minimization problem of minimizing the loss with respect to θ is nonconvex, the local minima of the networks tend to perform well in practice across most problem domains.

The advantage of this model is that the desired output variables are often linear transformations of the input variables and the network can discover a transformation from inputs to a feature space, which is related by a linear transformation to the outputs. Without this sort of model, it is common to need a heuristic nonlinear preprocessing step (e.g., in our work for the baseline models, taking the standard deviation of the input). In the case of CNNs, this transformation is discovered via the data to minimize the final objective.

Models

Here, were report a number of implementation details about the models:

Channel summary statistics model

As is common practice in ML, we appended 1 as a feature before fitting all linear models to add a bias. This results in a 10-dimensional feature

vector for AIA mean models and a 19-dimensional one for mean and standard deviation models.

L2 regression

This can be solved geometrically and has no hyperparameters.

Huber regression

We fit a model using Scikit Learn's SGDRegressor per EVE channel, using a Huber loss and L2 regularization. We standardized the variables (subtracting mean and dividing by standard deviation from the train set) and used the validation set to pick the regularization strength and Huber delta parameter.

Gradient boosting regression trees

We fit a model with Scikit Learn's Gradient Boosting Regressor, minimizing Huber loss to set weights and fitting 100 trees. We used the validation set to pick the Huber delta parameter, tree depth, and minimum number of samples per leaf. We found the default optimization parameters (e.g., convergence criteria) to work poorly with EVE's small magnitude and multiplied the target variables by a factor of 10^5.

Convolutional neural net

We used a modified AlexNet (18) containing three convolutional layers, followed by average pooling and an affine (i.e., linear + bias) transformation. Let $C(k,n,s,p)$ denote a convolutional layer with n filters of size $k \times k$, applied with stride s and padding p; BN denote batchnorm; R denote a ReLU; and $MP(k,s)$ denote max-pooling over a $k \times k$ grid with stride s. The convolutional block is then $C(11,64,4,2)$, BN, R, $MP(3,2)$, $C(5,192,1,2)$, BN, R, $MP(3,2)$, $C(3,384,1,1)$, BN, R. This is then followed by average pooling, dropout ($P = 0.5$), and a fully connected layer with 14 outputs. We used dropout due to the small amount of data.

The network is trained following standard practices from a random initialization with stochastic gradient descent. We used pytorch defaults for initialization, except for convolutional layers, which were initialized with the Kaiming normal method. After initialization, we trained the network with stochastic gradient descent with Nesterov momentum (28) and weight decay of 10^{-9} for 24 epochs. We started with a learning rate of 0.1 and multiplied it by 0.1 every 8 epochs. We applied gradient clipping (to have maximum norm 0.5), which we found anecdotally to improve performance.

Properly scaling the inputs is important. We divided each channel of an input AIA image with an average computed over the training set (and divided by this same value at test time). We treated the residual EVE similarly, dividing by the average residual over the training set, but we also multiplied it by 10^2 to make the order of magnitude of the target values match our network's initial outputs.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/5/10/eaaw6548/DC1

Fig. S1. Emission prediction of Fe XVIII line for several models.
Fig. S2. Emission prediction of He II line for several models.
Fig. S3. Emission prediction of Fe XI line for several models.

REFERENCES AND NOTES

1. J. Lean, Solar ultraviolet irradiance variations: A review. *J. Geophys. Res.* **92**, 839–868 (1987).
2. W. Kent Tobiska, T. Woods, F. Eparvier, R. Viereck, L. Floyd, D. Bouwer, G. Rottman, O. R. White, The SOLAR2000 empirical solar irradiance model and forecast tool. *J. Atmos. Sol. Terr. Phys.* **62**, 1233–1250 (2000).
3. S. C. Solomon, T. N. Woods, L. V. Didkovsky, J. T. Emmert, L. Qian, Anomalously low solar extreme-ultraviolet irradiance and thermospheric density during solar minimum. *Geophys. Res. Lett.* **37**, L16103 (2010).
4. W. D. Pesnell, B. J. Thompson, P. C. Chamberlin, The solar dynamics observatory (SDO). *Sol. Phys.* **275**, 3–15 (2012).
5. T. N. Woods, F. G. Eparvier, R. Hock, A. R. Jones, D. Woodraska, D. Judge, L. Didkovsky, J. Lean, J. Mariska, H. Warren, D. McMullin, P. Chamberlin, G. Berthiaume, S. Bailey, T. Fuller-Rowell, J. Sojka, W. K. Tobiska, R. Viereck, Extreme ultraviolet Variability Experiment (EVE) on the Solar Dynamics Observatory (SDO): Overview of science objectives, instrument design, data products, and model developments. *Sol. Phys.* **275**, 115–143 (2012).
6. J. R. Lemen, A. M. Title, D. J. Akin, P. F. Boerner, C. Chou, J. F. Drake, D. W. Duncan, C. G. Edwards, F. M. Friedlaender, G. F. Heyman, N. E. Hurlburt, N. L. Katz, G. D. Kushner, M. Levay, R. W. Lindgren, D. P. Mathur, E. L. McFeaters, S. Mitchell, R. A. Rehse, C. J. Schrijver, L. A. Springer, R. A. Stern, T. D. Tarbell, J.-P. Wuelser, C. J. Wolfson, C. Yanari, J. A. Bookbinder, P. N. Cheimets, D. Caldwell, E. E. Deluca, R. Gates, L. Golub, S. Park, W. A. Podgorski, R. I. Bush, P. H. Scherrer, M. A. Gummin, P. Smith, G. Auker, P. Jerram, P. Pool, R. Soufli, D. L. Windt, S. Beardsley, M. Clapp, J. Lang, N. Waltham, The Atmospheric Imaging Assembly (AIA) on the Solar Dynamics Observatory (SDO). *Sol. Phys.* **275**, 17–40 (2012).
7. P. H. Scherrer, J. Schou, R. I. Bush, A. G. Kosovichev, R. S. Bogart, J. T. Hoeksema, Y. Liu, T. L. Duvall, J. Zhao, A. M. Title, C. J. Schrijver, T. D. Tarbell, S. Tomczyk, The Helioseismic & Magnetic Imager (HMI) Investigation for the Solar Dynamics Observatory (SDO). *Sol. Phys.* **275**, 207–227 (2012).
8. M. C. M. Cheung, P. Boerner, C. J. Schrijver, P. Testa, F. Chen, H. Peter, A. Malanushenko, Thermal diagnostics with the atmospheric imaging assembly on board the solar dynamics observatory: A validated method for differential emission measure inversions. *Astrophys. J.* **807**, 143 (2015).
9. K. P. Dere, E. Landi, H. E. Mason, B. C. Monsignori Fossi, P. R. Young, CHIANTI—An atomic database for emission lines. *Astron. Astrophys. Suppl. Ser.* **125**, 149–173 (1997).
10. E. Landi, K. J. H. Phillips, CHIANTI—An atomic database for emission lines. VIII. Comparison with solar flare spectra from the Solar Maximum Mission flat crystal spectrometer. *Astrophys. J. Suppl. Ser.* **166**, 421–440 (2006).
11. J. M. Fontenla, M. Codrescu, M. Fedrizzi, T. Fuller-Rowell, F. Hill, E. Landi, T. Woods, Five years of synthesis of solar spectral irradiance from SDID/SISA and SDO/AIA images. *Astrophys. J.* **834**, 54 (2017).
12. R. Galvez, D. Fouhey, M. Jin, A. Szenicer, A. Muñoz-Jaramillo, M. C. M. Cheung, P. J. Wright, M. G. Brobra, Y. Liu, J. Mason, R. Thomas, A machine learning catalog prepared from the NASA solar dynamics observatory mission. *Astrophys. J. Suppl. Ser.* **242**, 7 (2019).
13. P. J. Wright, M. C. M. Cheung, R. Thomas, R. Galvez, A. Szenicer, M. Jin, A. Muñoz-Jaramillo, D. Fouhey, *Machine Learning, Statistics, and Data Mining for Heliophysics*, M. G. Bobra, J. Mason, Eds. (2018), chap. 4; https://doi.org/10.5281/zenodo.1412824.
14. T. P. Golding, J. Leenaarts, M. Carlsson, Non-equilibrium helium ionization in an MHD simulation of the solar atmosphere. *Astrophys. J.* **817**, 2 (2016).
15. H. P. Warren, Measurement of absolute abundances in solar flares. *Astrophys. J.* **786**, L1 (2014).
16. T. Hastie, R. Tibshirani, J. H. Friedman, *The Elements of Statistical Learning* (Springer, 2009).
17. J. H. Friedman, Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **29**, 1189–1232 (2001).
18. A. Krizhevsky. One weird trick for parallelizing convolutional neural networks. arXiv:1404.5997 [cs.NE] (23 April 2014).
19. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2016).
20. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, L. Fei-Fei, ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **115**, 211–252 (2015).
21. J Carreira, A Zisserman, Quo vadis, action recognition? A new model and the kinetics dataset, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2018).
22. J. T. Barron, P. Arbeláez, S. V. E. Keränen, M. D. Biggin, D. W. Knowles, J. Malik, Volumetric semantic segmentation using pyramid context features, in *IEEE International Conference on Computer Vision (ICCV)* (IEEE, 2013).
23. B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2016).
24. A. Asensio Ramos, I. S. Requerey, N. Vitas, DeepVel: Deep learning for the estimation of horizontal velocities at the solar surface. *Astron. Astrophys.* **604**, A11 (2017).
25. C. J. Díaz Baso, A. Asensio Ramos, Enhancing SDO/HMI images using deep learning. *Astron. Astrophys.* **614**, A5 (2018).
26. T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning* (Springer, ed. 2, 2008).
27. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning* (MIT Press, 2016).
28. I. Sutskever, J. Martens, G. Dahl, G. Hinton, On the importance of initialization and momentum in deep learning. *Proc. Mach. Learn. Res.* **28**, 1139–1147 (2013).

Citation: A. Szenicer, D. F. Fouhey, A. Munoz-Jaramillo, P. J. Wright, R. Thomas, R. Galvez, M. Jin, M. C. M. Cheung, A deep learning virtual instrument for monitoring extreme UV solar spectral irradiance. *Sci. Adv.* **5**, eaaw6548 (2019).