Attention-based machine vision models and techniques for solar wind speed forecasting using solar EUV images

Edward E. J. Brown[1], Filip Svoboda[1], Nigel P. Meredith[2], Nicholas Lane[1], and Richard B. Horne[2]

[1]University of Cambridge
[2]British Antarctic Survey

November 22, 2022

Abstract

Extreme ultraviolet images taken by the Atmospheric Imaging Assembly on board the Solar Dynamics Observatory make it possible to use deep vision techniques to forecast solar wind speed - a difficult, high-impact, and unsolved problem. At a four day time horizon, this study uses attention-based models and a set of methodological improvements to deliver an 11.1% lower RMSE error and a 17.4% higher prediction correlation compared to the previous work testing on the period from 2010 to 2018. Our analysis shows that attention-based models combined with our pipeline consistently outperform convolutional alternatives. Our model has learned relationships between coronal holes' characteristics and the speed of their associated high speed streams, agreeing with empirical results. Our study finds a strong dependence of our best model on the position in the solar cycle, with the best performance occurring in the declining phase.

Attention-based machine vision models and techniques for solar wind speed forecasting using solar EUV images

Edward J. E. Brown [1,2,*], Filip Svoboda [1,*], Nigel P. Meredith[2], Nicholas Lane [1,3], Richard B. Horne[2]

[1]Department of Computer Science and Technology, University of Cambridge, Cambridge, UK
[2]British Antarctic Survey, Natural Environment Research Council, Cambridge, UK
[3]Samsung AI Center, Cambridge, UK
[*]equal contribution

Key Points:

- Attention-based machine vision models and methodological enhancements are developed to improve solar wind speed forecasts from solar images
- Attention-based architectures outperform convolutional models, motivating their use in future studies and production systems
- The models perform best in the declining phase of the solar cycle when activity is driven by coronal holes

Abstract

Extreme ultraviolet images taken by the Atmospheric Imaging Assembly on board the Solar Dynamics Observatory make it possible to use deep vision techniques to forecast solar wind speed - a difficult, high-impact, and unsolved problem. At a four day time horizon, this study uses attention-based models and a set of methodological improvements to deliver an 11.1% lower RMSE error and a 17.4% higher prediction correlation compared to the previous work testing on the period from 2010 to 2018. Our analysis shows that attention-based models combined with our pipeline consistently outperform convolutional alternatives. Our model has learned relationships between coronal holes' characteristics and the speed of their associated high speed streams, agreeing with empirical results. Our study finds a strong dependence of our best model on the position in the solar cycle, with the best performance occurring in the declining phase.

Plain language summary

Solar images contain rich information that can be used to forecast conditions at Earth. This study develops a robust methodology for processing solar images and trains machine learning models that can use them to predict the solar wind speed. Combined, these deliver a very significant 17.4% improvement in the correlation between the prediction and the ground truth over previous works. The models perform better during the quieter, declining phase of the solar cycle when the solar activity is driven by coronal holes. Finally, the trained models learn properties of coronal holes that agree with prior empirical studies.

1 Introduction

The solar wind is a stream of charged particles that is emitted from the upper atmosphere of the Sun. The speed, density, temperature and the magnitude and direction of the associated magnetic field of the solar wind are constantly varying affecting the way in which it ultimately interacts with the Earth's magnetosphere. High speed solar wind streams (HSS) emanating from coronal holes are particularly effective at coupling with the Earth's magnetosphere. The weak storms they produce tend to have long-lasting recovery phases which often result in prolonged and enhanced substorm activity (Tsurutani et al., 1995; Meredith et al., 2011). This results in repeated injections of suprathermal electrons into the inner magnetosphere and significant increases in the fluxes of relativistic electrons in the outer radiation belt, increasing the risk to satellites via surface charging and internal charging respectively (e.g., Borovsky and Denton (2006)). Indeed, it has been suggested that satellites at geostationary orbit are more likely to be at risk from an extreme HSS-driven storm than a Carrington type event (Horne et al., 2018). Furthermore, prolonged and enhanced substorm activity associated with HSS-driven storms results in increased thermospheric densities and satellite drag (Chen et al., 2012). Consequently, accurately forecasting the solar wind speed associated with coronal holes is very important for our modern society.

Coronal holes are large dark areas on the Sun as seen in extreme ultraviolet (EUV) and soft X-ray images (Cranmer, 2009). They are regions of open magnetic field and cooler plasma, leading to the production of high speed solar wind streams. Coronal holes are long-lasting features that can persist from one solar rotation to the next, giving rise to a 27 day periodicity in the arrival of HSS at Earth. The occurrence rate of coronal holes peaks during the declining phase of the solar cycle (Burlaga & Lepping, 1977) and high speed streams observed at Earth during these intervals tend to be coronal-hole driven. The distribution of speeds in high speed streams associated with coronal holes ranges from 400 to 800 kms^{-1} (Kilpua et al., 2017). While these streams do not result in ma-

jor geomagnetic storms (Richardson et al., 2006), they have extensive recovery phases, typically lasting from 5–10 days, and, as a result, may deposit more energy in the magnetosphere than larger storms (Kozyra et al., 2006; Turner et al., 2006).

Coronal holes are not the only source of high speed solar wind at Earth. Coronal mass ejections (CMEs) also cause high speed solar wind, although not all CMEs are associated with high solar wind speeds (Kilpua et al., 2017). CME's are large explosions on the Sun that hurl vast amounts of plasma into space. The occurrence rate of CMEs peaks at solar maximum (St. Cyr et al., 2000) so that most periods of high solar wind speed observed during these periods tend to be CME-driven. The distribution of speeds in interplanetary coronal mass ejections (ICMEs) and sheath regions associated with CMEs on the Sun ranges from 250 to 950 kms^{-1} (Kilpua et al., 2017). Unlike coronal holes, CMEs are not associated with long lasting features on the Sun. In contrast they are best observed in coronagraph images where they appear as expanding shells of material.

In this study we build a machine learning model to use solar images to forecast the solar wind speed at Earth. This technique is expected to perform best when there are associated visible features on the Sun. The method is thus expected to work well for coronal holes, which are large features on the solar disk. In contrast, coronal mass ejections are barely noticeable within EUV images and so the ML model would not be expected to work well for these events.

The field of machine learning has built a lot of momentum over the last 10 years. This has largely been the result of improvements in algorithmic capability, availability of data, funding and hardware. Not to be overlooked though is the creation of field benchmarks like ImageNet (Deng et al., 2009) and open-source software such as PyTorch (Paszke et al., 2019) which dramatically shortened the development cycle in the field and greatly increased its standardization.

Deep (Machine) Learning excels where rich data exists in large quantities, because models with deep structures and therefore many parameters need to consume richly varied data sources to build complex internal representations of the data generating system. This is the essence of deep learning. Recently, curated solar image datasets have been created such as the SDOML dataset (Galvez et al., 2019) which contains images of the Sun taken at various EUV wavelengths. These data allow the rapid application of machine learning algorithms to consume solar images.

In this paper we use the EUV images taken by the Solar Dynamics Observatory (SDO) using the Atmospheric Image Assembly (AIA) (Lemen et al., 2011) to forecast the so-

lar wind speed at the Lagrangian L1 point. We present results for forecasting at a four day lag from a single 211 Å image - but this forecast could be used for any lag up to four days. We also explore the model's learned behaviour by examining relationships between the peak solar wind speed and the coronal hole area and intensity. Previous works and the datasets are presented in Sections 2 and 3 respectively. In Section 4, we discuss our general methodology and model architectures. Our results are presented and discussed in 5. Finally, our conclusions are summarised in Section 6.

2 Previous Works

The works of Wintoft and Lundstedt (1997) and Wintoft and Lundstedt (1999) were the first to use neural networks to forecast the solar wind speed. These are small, so-called fully connected, models that could learn non-linear relationships between a limited set of pre-computed feature inputs, such as the flux tube expansion factor, and the solar wind speed. More recently, similar studies were performed by D. D. Liu et al. (2011), Yang et al. (2018), Chandorkar et al. (2019), and Bailey et al. (2021) using similar non-image-based inputs to the models, albeit with more advanced models than the earlier works.

Upendran et al. (2020) was the first study aiming to forecast solar wind speed from solar EUV images using deep learning techniques. The work uses images from both 193 and 211 Angstrom wavelengths to forecast the solar wind speed at a one day resolution. Upendran uses GoogleNet (Szegedy et al., 2014), trained on the ImageNet dataset (Deng et al., 2009), as a feature extractor for each image. The extracted per-image features are then passed into an LSTM Recurrent Neural Network (Hochreiter & Schmidhuber, 1997) to produce the predicted solar wind speed. The study achieves a best performing model at a lag of 3 days and a history of 4 days, with a correlation of 0.55 and an RMSE 80.28 km/s. This study will build on this insightful initial work.

Next, Raju and Das (2021) proposed a smaller three-layer convolutional feature extractor, which they train on the 193 Angstrom wavelength solar EUV images. Their method targets a subtly different task than that of Upendran et al. (2020). While Upendran et al. (2020) forecast future solar wind speeds based on images at a fixed distance in the present, Raju and Das (2021) backcast current solar wind speed based on flexible-lag past images. Specifically, Raju and Das (2021) use the current solar wind speed to infer which past image was likely to have caused the recorded solar wind speed, and then pass this image into their model with the expectation that the model will be able to correctly reconstruct the observed solar wind speed. The key difference between the two approaches is that in the forecasting setup the model needs to use the image information to both infer the speed of the caused solar wind and make a judgement on whether the particle stream at the given speed will be geo-effective. In the Raju and Das (2021) backcast-

ing setup, the observed features are guaranteed to have been geo-effective, as the image was chosen based on this criterion, and thus the model needs to infer only the speed, not the geo-effectiveness too. Its task is thus made easier. When used in prediction it is assumed that the predicted solar wind speed will be used to infer the time when it will arrive at Earth. The difference becomes clearer when the models are to be deployed as solar wind speed predictors. Under the forecasting setup, today's images can be used to produce the predicted solar wind speed 4 days from now. In contrast, under the backcasting setup not all time stamps would receive a prediction. Indeed, the inference process by which images are paired with time stamps does not guarantee a unique prediction for each time stamp, and so some time stamps can be expected to receive multiple solar wind speed predictions, while others would get none. Furthermore, any error in the speed prediction will be significantly magnified when the time offset is made dependent on this prediction. Thus this model is not comparable to Upendran et al. (2020). Nevertheless, they provide results for a model specially trained at a fixed 4 day forecast horizon (their Table 4), with the year 2018 held out as a test set. They report 78.3 km/s RMSE and a prediction correlation of 0.55. This would be comparable to Upendran et al. (2020), except they provide no results for 2018 alone. Their test results are from across multiple years. Therefore, our study will compare to Upendran et al. (2020) for dates across an 8.5 year range and then run a separate training run to compare to Raju and Das (2021)'s fixed 4-day model, just evaluating on the year 2018.

3 Data

3.1 Solar Images

The image dataset consists of EUV images from NASA's Solar Dynamics Observatory (SDO) taken by the Atmospheric Imaging Assembly (AIA) (Lemen et al., 2011) that have been processed by performing various instrumental corrections, downsampled to usable spatial and temporal resolutions and synchronised both spatially and temporally to form the SDOML dataset (Galvez et al., 2019). The resulting dataset contains 8 and a half years of images every 6 minutes from June 2010 to December 2018. These images are monochromatic and the pixel values represent the intensity of light. This study uses the EUV images at 211 Angstroms.

3.2 Solar Wind Speed

The solar wind speed data are taken from the OMNIWeb service. Specifically, we use the solar wind speed, measured in km/s, at a 1 minute time resolution for the period of the SDOML dataset. The data come from WIND and the Advanced Composition Explorer (ACE) spacecraft, both positioned at the L1 point, about 1.5 million km from Earth.

The solar wind speed is highly auto-correlated with itself over hourly time periods and is still at 0.7 after 1 day. By four days, the correlation has dropped to negligible amounts. Notably, at 27 days, there is a spike in the auto-correlation. This is because the Sun has a synodic rotation period of approximately 27 days and some longer lasting features, such as coronal holes, come around again causing similar solar wind speed conditions at L1. This auto-correlation is important since it has implications for which images are included in training and test sets due to their dependence on each-other. This is further discussed in Section 4.1.7.

4 Methodology

4.1 Methodological Improvements

Here we discuss changes in our methodology to the only previous work, (Upendran et al., 2020), covering all the date ranges available from the SDOML dataset.

4.1.1 Image pre-processing

The EUV images at their provided resolution are too large to practically process on standard computing hardware. Previous works elected to down-sample the full 512 by 512 pixel image to 224 by 224 by max pooling. Instead, we take a 300 by 300 pixel square who's corners are approximately at the edges of the solar disk, and then down sample this cropped image to the desired 224 by 224 image size. This results in lower loss of information content in the relevant section of the Sun because 1) the cropped solar poles are unlikely to be geo-effective, 2) the cropped features at the western limb take about 7 days to be geo-effective and so are outside of the max 4 day forecasting horizon, 3) this allowed us to down-sample the central, relevant, portion of the image less aggressively. Figure 1 shows an example of our cropping technique.

Regarding scaling the cropped image images, the same method as used in Upendran et al. (2020) is employed by clipping the pixels to have values between a minimum of 25 and a maximum of 2500 and taking the natural logarithm. However, after this we rely on a batchnorm layer to learn an optimal scaling, as opposed to fixing it (further detailed in Section 4.2).

4.1.2 Sampling frequency

We replace the previously used daily sampling resolution with a 30 minute schedule, because solar wind speeds can change significantly even on a 30 minute time scale.

4.1.3 Carrington rotation

The Sun rotates on average every 27.28 days as viewed from Earth, this is one Carrington rotation(Ridpath, 2012). As such, the solar features that affected the solar wind speed

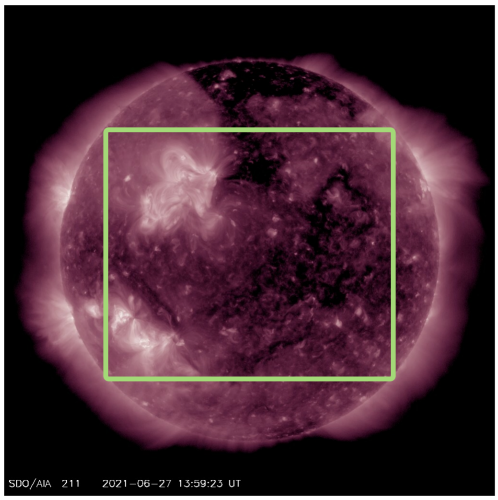


Figure 1: SDO AIA 211Å image taken on 2021-06-27 (Lemen et al., 2011)

at a given point come back approximately 27 days later and produce similar effects. Thus, the solar wind speed is also auto-correlated at the Carrington rotation periodicity with a value of 0.42 at 27 days. As this value is available to all forecasters operating at lower than 27 days forecast horizon, it should be used as an input to our models.

4.1.4 North-south augmentation

We augment the dataset by randomly flipping the training images north to south, as features, such as coronal holes, produce a similar increase in solar wind speed regardless of which side of the solar equator they are on. Although it is not claimed these are valid physical suns.

4.1.5 Single image versus sequence

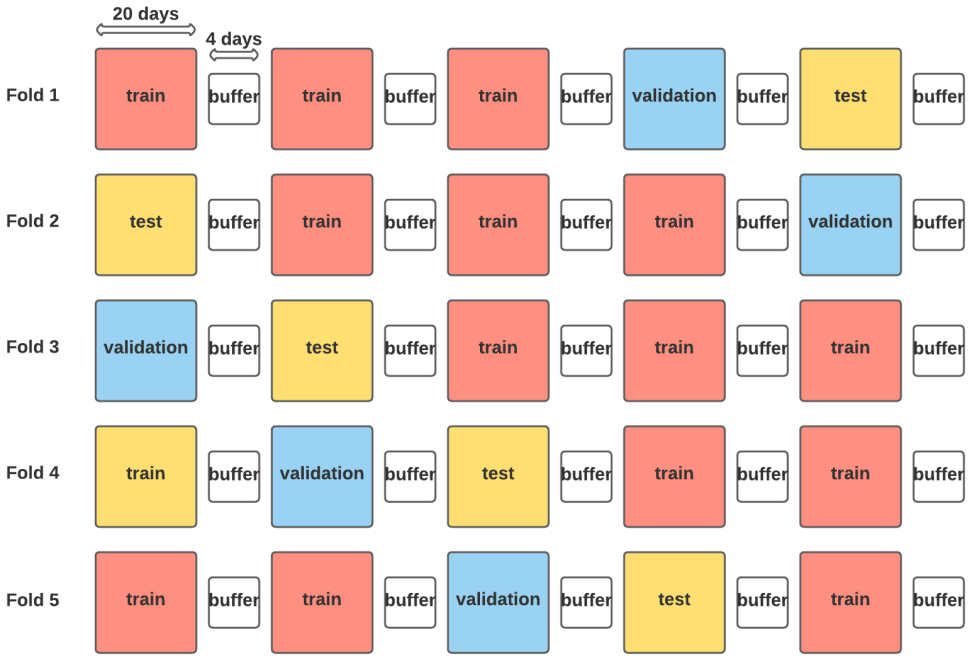
The previous work relies on a convolutional feature extractor pre-trained on ImageNet in combination with an LSTM cell and a fully connected layer (Upendran et al., 2020). Up to 4 images were sequentially passed through the convolutions. Separate for each image, the model's activations at multiple layers were extracted, concatenated, and passed into the LSTM as individual time steps. The convolutions remained parametrized by the weights obtained on ImageNet and only the other layers' parameters were trained. The high auto-correlation of solar images is likely to, again, exaggerate the model's multi-collinearity in hidden features while providing little additional context. Thus we replaced the LSTM feeding into a fully connected output layer with two consecutive fully connected layers.

4.1.6 Feature extractor re-training

This study will use pre-trained vision models at the core of the model architecture (see Section 4.2 for more details). Rather than to use the fixed pre-trained ImageNet weights, the model will be initialised with these weights but they will not be fixed. This we believe to be strictly necessary due to the wide gap between the EUV and the ImageNet datasets.

4.1.7 Training, validation and test sets

For this study, 5-fold cross-validation is employed to evaluate the models. Solar wind speed is auto-correlated up to a period of about 4 days. For the period of June 2010 to December 2018, the auto-correlation is as high as 0.70 at one day. This means that if timestamps are too close to each-other between training, validation and test sets, it is not a fair reflection of the performance of a model, since the Sun has not changed much in for example 30 minutes. Furthermore, this will mean that the model overfits on the validation sets, meaning they will not generalise as well. In order to create more independent training and test sets, a method similar to that used in Upendran et al. (2020) is employed whereby the timestamps from 2010-2018 are split into chunks of 20 days. However, a buffer period of 4 days between each chunk is thrown out to ensure the independence of the training, validation and test sets. It is noted that this throws out approximately one fifth of all the data. However, this is justified to ensure the independence of datasets while also covering as many parts of the solar cycle as possible. These chunks are then put into training, validation and test buckets. This process is repeated 5 times to ensure that each 20 day chunk serves a turn in the training set 3 times, the validation set once and the test set once. This creates 5 folds of training, validation and test. For each fold, a model is trained on the training set and evaluated on the validation set for 100 epochs (1 epoch is a full pass over the data). The model is saved every epoch. The version of the model that performs best on the validation set is the final model. This final model is then applied and evaluated on the unseen test set. Figure 2a shows the training sets in orange, the validation sets in blue and the test sets in yellow. White buffer sets of 4 days are included between the 20 day chunks.



(a) 5-fold cross validation with buffer data thrown out. Pattern is repeated across the May 2010 to December 2018 range.

(b) Dataset split with 2018 as hold-out test set for comparison with Raju and Das (2021)

Figure 2: Training, Validation and Test sets

4.2 Model Architectures

For this study, the architectures for the different models will follow the format in Figure 3. The image will pass through a batch norm layer that will rescale it. Then it is passed into the candidate architecture, be it a CNN or a vision transformer. The outputs from this model as well as the solar wind speed from one Carrington rotation ago are then passed into two final consecutive non-linear projections that produce the model's solar wind speed prediction.

In all cases the models are trained in their entirety on the EUV data. That is, after their parameters are initialized using either random, or when available, pre-set weights the algorithm iteratively updates them with the goal of incrementally decreasing the mean squared error of its prediction.

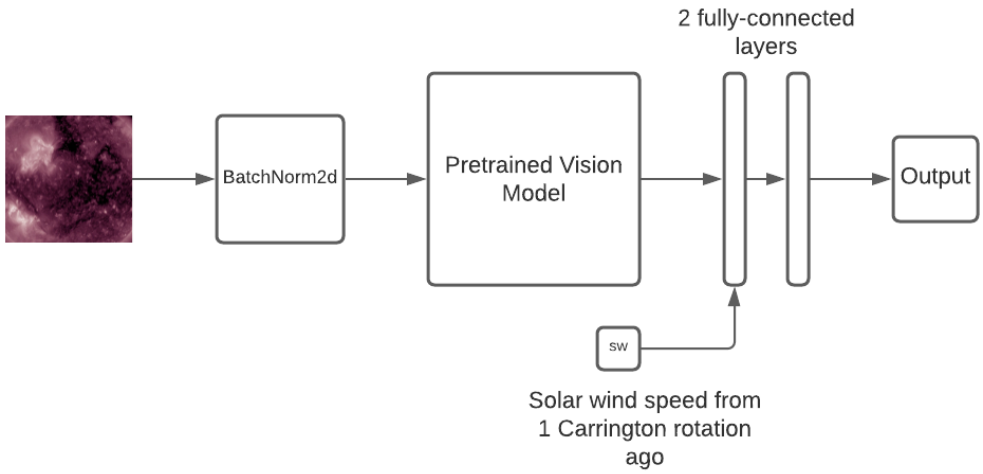


Figure 3: The Solar model architecture

4.2.1 Benchmark CNN-based models

In general every deep model can be seen as a layered composition of non-linear projections, each forming a separate layer. Model inputs, solar images in our case, can be seen as the zero-th layer, while, model outputs, the predicted solar wind speed, can be treated as the last layer. Each layer in between is a non-linear projection that receives inputs from the preceding layer, and that outputs its value to the next layer. Commonly, several layers are grouped into modules and used as a type of meta-layer. Modern architectures are defined by the features that build on and expand this basic structure.

Previous work used convolutional models in the forecasting of solar wind, (Upendran et al. (2020); Raju and Das (2021)). These models are designed to process images, each of which has three dimensions - the height, the width, and the number of channels. A standard colour image has 3 channels: red, green, and blue. Convolutions are operations that split the image into a grid of patches and then use a three dimensional kernel to compute weighted averages per each patch. The same kernel is used on each patch and the averages it produces become the pixel values the layer outputs. Multiple kernels may be

employed, in which case their outputs are treated as separate channels of the outputted image.

GoogleNet, also known as InceptionNet v1, is the convolutional architecture at the heart of Upendran et al. (2020)'s work. It is a convolutional architecture that replaces layers with modules. Each module computes several, rather than just one convolution. These are computed in parallel, and are meant to complement each other. The desired effect is to make the model's computation more parallelizable, thus faster, while improving the model's ability to fit complex patters in the data (Szegedy et al., 2014).

InceptionNet v2 is a second generation and a refinement of the GoogleNet. The architecture builds on GoogleNet's inception modules by decomposing their convolutions serially. Specifically, more computationally expensive, that is larger-kernel convolutions, are replaced by a series of much cheaper smaller-kernel convolutions carried out one after the other. The desired effect is to make the working set of this algorithm smaller, while further improving the model's capacity, i.e. its ability to fit complex data patterns (Szegedy et al., 2016).

ResNet is a predecessor of GoogleNet. ResNet's modules consist of two consecutive convolutions, and a so called residual connection. The residual connection is a bypass that circumvents the two convolutions. In effect this results in a block that outputs both its convolution's output as well as the original inputs to the block. This trick helps to propagate the training gradients through the network, mitigating the vanishing gradient problem. The architecture was the first one to breach the 20 layer depth ceiling (He et al., 2016).

DenseNet is a generalization of ResNet that adds multiple residual connections to each module. The beginning of a block of convolutions, is connected not only to the output of that same module, but also to the outputs of all modules down-stream from it (Huang et al., 2017).

4.2.2 Attention-based Models

This paper proposes using attention, rather than convolution, as the core model feature. Attention is a deep learning mechanic that, rather than learn a weight per each input pixel or a patch of pixels, learns a method for generating these weights from the input data. Consequently, the models can weight each patch based on what its position is and what the rest of the image depicts (Vaswani et al., 2017). In contrast, convolutions are designed to analyze each patch of each input image using the same kernel of weights, regardless of what the image depicts outside of the patch and what its position is. Formally,

convolutions enforce transition invariance, while attention models do not. Transition invariance in computer vision is achieved when the model maintains the same output even if the objects in the image are moved around.

Attention's ability to judge each image patch in the context of its position in the image and the contents of the rest of the image is critical for making sound solar wind speed predictions from the EUV data. First, the attention mechanism allows the model to assign higher importance to features on the Sun's surface if they appear in the equatorial region. Moreover, the model is able to learn to distinguish between situations when an active region interferes with a coronal hole, and when it does not. The weights it places on the patches of the image with the coronal hole in it will depend not only on its position in the image, but also on whether the model identified an interference from an active region. In contrast, convolution-based models were designed to identify an object anywhere in the input image field. Therefore, they place equal weight on each image patch as they process it using the same fixed-weight convolution kernel. It was assumed that multiple layers of convolutions would learn increasingly complex representations by deriving higher-layer features from simple lower-layer ones. Recently, however, it was shown that convolutional models do not recognize complex features, instead they aggregate low level texture features from across the input image and then make their prediction based on which texture prevails in the input image (Geirhos et al., 2018). Consequently, attention-based models will make better and more theory-sensible predictions as it, for example, will account for and internalize the higher importance of features in the equatorial region and the interference of active regions with coronal holes while convolution will fail to do so.

The *Vision Transformer* was the first transformer architecture successfully used in image recognition (Dosovitskiy et al., 2020). The architecture combines large image patches with the attention mechanism. Each patch is first individually passed through a linear projection, then the attention mechanism applies context-derived weights on each. The result is then passed into two consecutive non-linear projections, sometimes called fully connected layers, before being outputted. An important point of comparison is the size of the model's patches. While all benchmark models only consider patches of no more than 5x5 pixels, our Vision Transformer works with patches of 16x16. This is meant to allow it a larger receptive field and to steer clear of focusing on textures.

The *Transformer in Transformer* follows the same general architecture as the original Vision Transformer, the crucial difference is that the linear projection at the beginning of the outer transformer is replaced by an inner transformer that is modelled as a smaller

version of the same original Vision Transformer (Han et al., 2021). Therefore, the input image is first split into 16 by 16 patches. Each of these patches is then passed into the inner Vision Transformer, as if they were images in their own right. This splits them into smaller (4 x 4) patches still, derives the attention weight for each sub-patch based on the rest of each patch, and outputs the processed image back to the outer transformer. The outer transformer then uses these processed patches to derive its attention weights per each patch based on what the rest of the full image's processed patches are like. Then the outer transformer uses two consecutive non-linear projections to produce the final output.

The *Swin Transformer* is similar to the Vision Transformer except it builds hierarchical feature maps by merging image patches, as opposed to treating image patches separately as in the Vision Transformer (Z. Liu et al., 2021). The idea is that the model is able to treat features on different scales, whereas the vanilla vision transformer is limiting itself to a predetermined patch size. Furthermore, a feature of the algorithmic construction is a linear scale in computational complexity based on image size.

These pre-trained attention-based models, as well as the benchmark CNN models, all accept three-channel RGB images normally. In order to use these powerful models, the solar images have to be repeated 3 times to form the three channels. Normally, one would use the advised normalisation schedule from the papers that produced these models. In this case however, since the models are not RGB in the first case, it was decided that an initial batch norm layer is applied before the model, so that the best normalisation schedule can be learned and not fixed. The reported RMSE and Correlation is then averaged over the five folds and reported.

$$\text{RMSE} = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}\Big(x_i - y_i\Big)^2}$$

$$\text{Correlation} = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2(y_i-\overline{y})^2}}$$

where y_i is the real solar wind speed, x_i is the predicted solar wind speed, $\overline{y_i}$ is the mean real speed, $\overline{x_i}$ is the mean predicted speed, and n is the total number of data points.

4.3 Other Experimental Details

4.3.1 Missing data

Missing images are substituted with valid observations no more than 30 minutes removed from the missing datum. Missing solar wind speed data are interpolated from available data no more than 30 minutes removed. Time steps with no valid data for filling in the missing observations are discarded.

4.3.2 Hyper-parameter selection

Hyper-parameters are chosen using a Bayesian parameter sweep using the software Weights & Biases (Biewald, 2020) based on the performance of the validation set. For cost reasons, the sweep is conducted at 120 minutes resolution for only 30 epochs.

4.3.3 Training process

The loss function of the network is the default implementation of pytorch's mean squared error (squared L2 norm) (Paszke et al., 2019). The optimizer method to update the weights of the network is the default implementation of the Adam optimizer in pytorch as well (Kingma & Ba, 2014). Batch size is fixed at 64.

4.3.4 Computation

All experiments were run on V100 Nvidia GPU, resulting in a total compute of about 900 GPU hours.

4.4 Year 2018 Evaluation

Solar activity can vary significantly based on position in the solar cycle, so only testing on 2018 only gives the performance of the model in that part of the solar cycle. It therefore cannot be representative of the generalisation of the model to other periods of the solar cycle. However, Raju and Das (2021) provide results for a model trained on solar imaging data with the entire year of 2018 held out for evaluation. As an extra experiment and to compare to their study, a model will be trained with the training and test set schedule shown in Figure 2b. Notably, Figure 2b features a 27 day test buffer before the start of the 2018 test set. This buffer is present because of Raju and Das' concern of 27 day resurgence causing the training and test sets to not be independent. Our view is that since this model is forecasting at a 4 day forecast, any image before that 4 days could be used to train a model in a production system to make that 4 day forecast (especially using the method of online learning). Despite the dependence, this 27-day old image would be one of the most important images you would want to train on. Where the dependence matters for forecasting purposes is crucially when the images are less than the forecast horizon apart. This explains our choice of 4 day buffer otherwise. However, for the point of comparison, this 27 day buffer is kept. Otherwise, all experimental procedures as detailed will remain the same as with the 5-fold split.

5 Results and Discussion

5.1 Comparison to Previous Works

Table 1 shows the comparison of our methodological and modelling pipeline, used with a range of feature extractors, against the most recent state of the art forecasting model in the field and two naive persistence model benchmarks. Notably, all of the models trained under our pipeline improve on the work by Upendran et al. (2020) by at least 8.8% in RMSE and 12.7% in correlation. Indeed, our pipeline with the GoogleNet feature extractor, which is the same feature extractor as was used in the Upendran et al. (2020) model, demonstrated the total improvement our pipeline has delivered. It lowered the RMSE by 9.2% and increased the correlation by 14.6%. Furthermore, our best performing model, based off the Swin Vision Transformer, improves on the state of the art by 11.1% in RMSE and 17.4% in correlation. The model also outperforms at the 1, 2, and 3 day time horizon because the 4 day forecast could also be used for those. Finally, transformer feature extractors outperformed convolutional ones by about 1 to 2% in either metric when used in our model pipeline.

Table 1: Performance of our solar models compared to Upendran et al. (2020) forecasting solar wind speed using the EUV data at a 4 day forecast horizon in the period May 2010 to December 2018. Upendran et al. (2020).

Model	**RMSE**	**% Improvement**	**Correlation**	**% Improvement**
Persistence(4 day)	127.59	-57.1%	0.080	-85.2%
Persistence(27 day)	100.86	-24.2%	0.426	-21.1%
Upendran et al. (2020)	81.21	-	0.54	-
Our models				
Solar InceptionNet v4	74.09	8.8%	0.609	12.7%
Solar DenseNet	73.92	9.0%	0.611	13.1%
Solar GoogleNet	73.71	9.2%	0.619	14.6%
Solar ResNet	73.52	9.5%	0.618	14.4%
Solar TNT	72.70	10.5%	0.629	16.5%
Solar Vision Transformer	72.66	10.5%	0.630	16.7%
Solar Swin Transformer	72.21	11.1%	0.634	17.4%

Table 2 compares the performance of our best performing model, that is the one based on the Swin Transformer feature extractor, and the two persistence benchmarks against the predictions Raju and Das (2021) produced for the year 2018. This setup differs from that of table 1 in that table 1 tests the models on data examples sampled from the whole dataset, and thus across the solar cycle. The present comparison is made solely with respect to the solar cycle conditions present in the year 2018, as chosen by Raju and Das (2021). Our model shows a significant improvement of 8.3% in RMSE and 17.1% in correlation over the performance achieved by Raju and Das (2021).

Table 2: Performance of our solar models relative to (Raju & Das, 2021) predicting solar wind speed using EUV data at a 4 day forecast horizon in for the year 2018.

Model	**RMSE**	**% Improvement**	**Correlation**	**% Improvement**
Persistence(4 day)	118.76	-52.3%	-0.027	-104.9%
Persistence(27 day)	85.16	-9.2%	0.464	-15.6%
Raju and Das (2021)	78	-	0.55	-
Our model				
Solar Swin Transformer	71.65	8.3%	0.644	17.1%

5.2 Ablation Study

To demonstrate the stand-alone effect of our suggested techniques on the results, we conducted a study whereby each improvement is removed one at a time and the performance reduction reported. In the case of dropping the buffers, the no-buffer condition was implemented by making those buffers between the validation and training sets become part of the validation set, thus removing the separation between the two sets whilst adhering to a test-validation-train split that is comparable to that of the original condition.

Figure 4 shows that the dominant improvement has been the adjustment of the sampling frequency, excluding it causes 8.51% deterioration in RMSE and 9.70% in correlation.

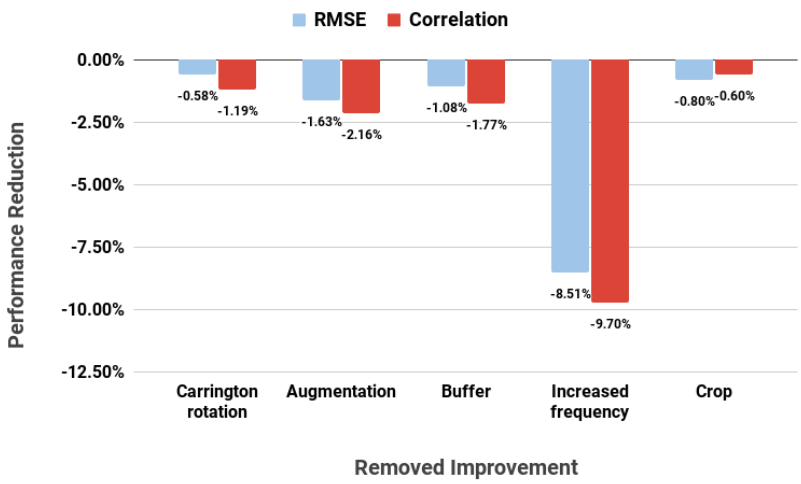


Figure 4: Performance reduction resulting from removing one improvement at a time.

Excluding the other 4 methodological, improvements delivers between 0.58% and 1.63% RMSE deterioration, and between 0.6% and 2.16% fall in correlation. While these figures are modest in magnitude, it ought to be pointed out that the benefits appear uncorrelated between the methods, and when they are all combined they deliver a significant improvement over the previous works.

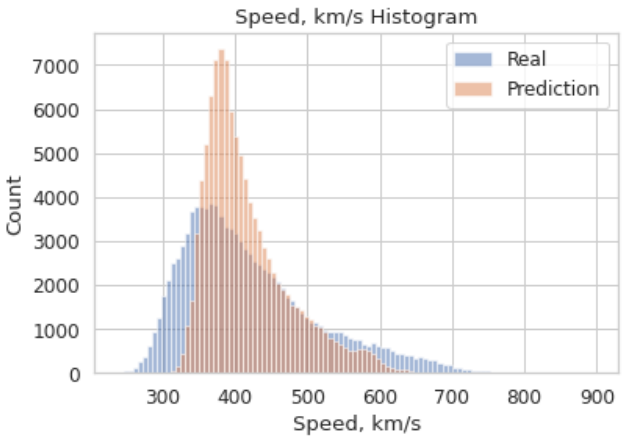
5.3 Prediction Analysis

Next, we analyze the predictions made by the best performing Swin Transformer model to get better understanding what aspects of the solar wind speed prediction task it gets right, and where, if at all, lie its systematic biases.

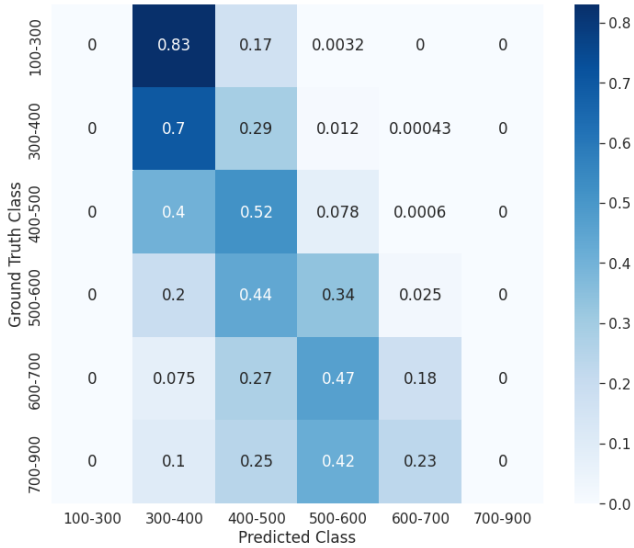
5.3.1 Distribution

Figure 5a shows the distributions of the solar wind speeds predicted by the top model and the underlying ground truth. Both distributions are roughly centered around the same mean with a positive skewness, i.e. they have long right-hand tails. The distributions differ significantly in their kurtosis. The real data has lower kurtosis, that is it has more observations in both its right and left tails. The model's predictions have notably higher kurtosis, as it has a much more pronounced peak at around its mean and much fewer observations in its tails. This is to be expected as the L2 loss function chosen, which all models in this domain use, is known to prioritize the average fit of the model over fitting the extremities. The distributions by themselves, however, do not tell the full story. For that we need to look at figure 5b, which shows the confusion matrix of binned speeds. Both predicted and actual solar wind speeds are split into 4 distinct class bins incremented

by 100 km/s and 2 catch all classes one at each extreme of the distributions. Each block of the confusion matrix corresponds to one combination of a predicted class and a ground truth, i.e. real, class. The value in the block represents the fraction of that real class that were classified as the predicted class. Under a perfect prediction, the blocks would read 1.0 along the diagonal and 0 everywhere else. This would mean that all speeds were correctly predicted in their class. As it is however, our model shows a tendency to over-predict the lower real speeds while under-predicting the higher speeds. Indeed, no speeds that were in the 700-900 km/s range were correctly predicted as such. Similarly, no speeds in the 100-300 range were correctly predicted. This confirms our suspicion that it is the tail observations that are being regressed towards the mean that is driving both the error in the confusion matrix and the difference in the prediction and ground truth distributions.



(a) Distribution of predicted and real speeds



(b) Confusion matrix of binned speeds (km/s)

Figure 5: Distribution and confusion matrix of predicted speeds

5.3.2 Coronal hole area

It has been empirically established that there is a linear relationship between coronal hole area at low latitudes and peak solar wind speed (Nolte et al., 1976; Hofmeister et al., 2018). In order to test whether our model has learned this relationship we need to devise a way of obtaining images with specified coronal hole sizes at the desired latitude. We chose to generate our images using a background of enlarged uneventful solar region and a patch extracted from a coronal hole that can be sized as desired. Each patch size is moved across

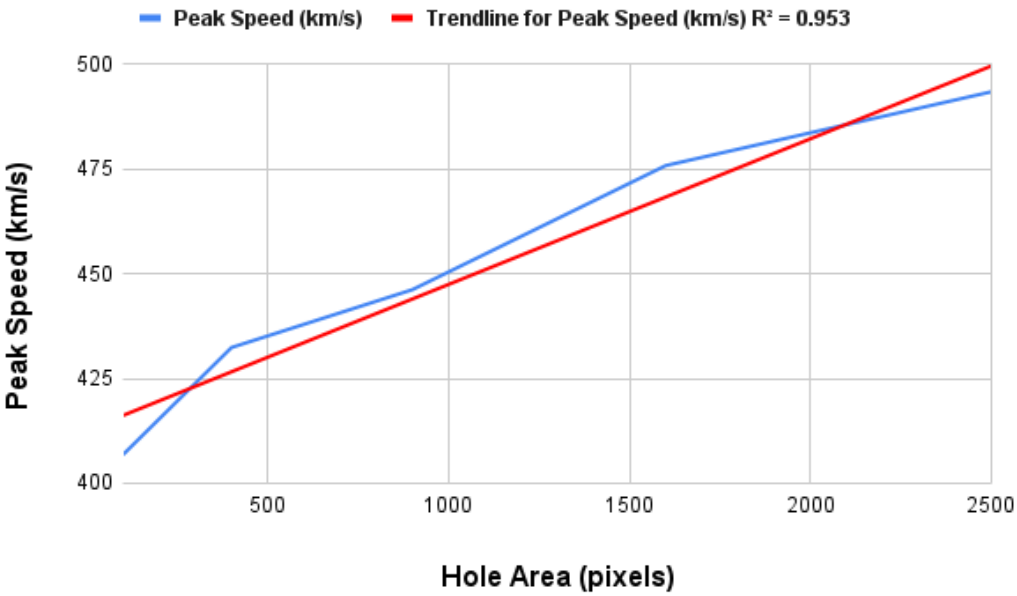


Figure 6: Peak speed of coronal holes at solar equator versus coronal hole area

the image, and the model's peak prediction for that size is recorded. Figure 6 plots the predicted solar wind speeds against the patch sizes. It shows that our model, indeed, succeeded to learn the relationship established by Nolte et al. (1976) and Hofmeister et al. (2018) since its predictions follow very closely the empirically observed linear relationship with a high degree of correlation.

5.3.3 Coronal hole intensity

Obridko et al. (2009) found that the darker the coronal hole, the larger is the peak of the associated high speed stream. We test whether our model learned this empirical relationship by incrementally increasing the minimum brightness of a coronal hole. At each step, any pixel value below the minimum threshold is increased to the minimum value. Figure 7 shows the predicted speed for a large coronal hole visible on the day of 2016-12-06 at 00:00:00 am at various minimum intensities. As we increase the brightness of the coronal hole, the model starts to forecast lower solar wind speeds. This suggests that the model has learned the Obridko et al. (2009) empirical relationship that the darker the hole, the stronger the solar wind.

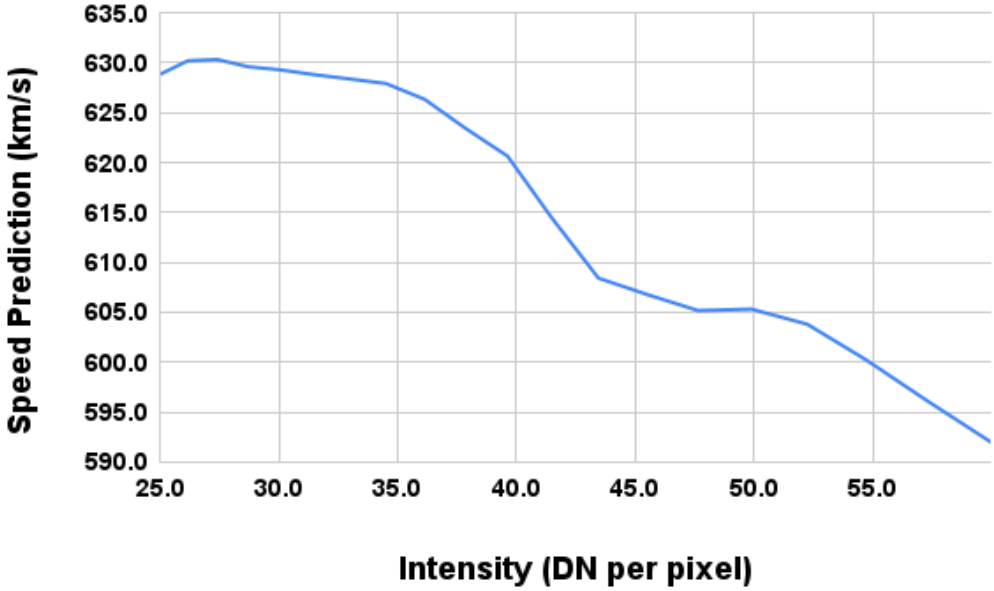
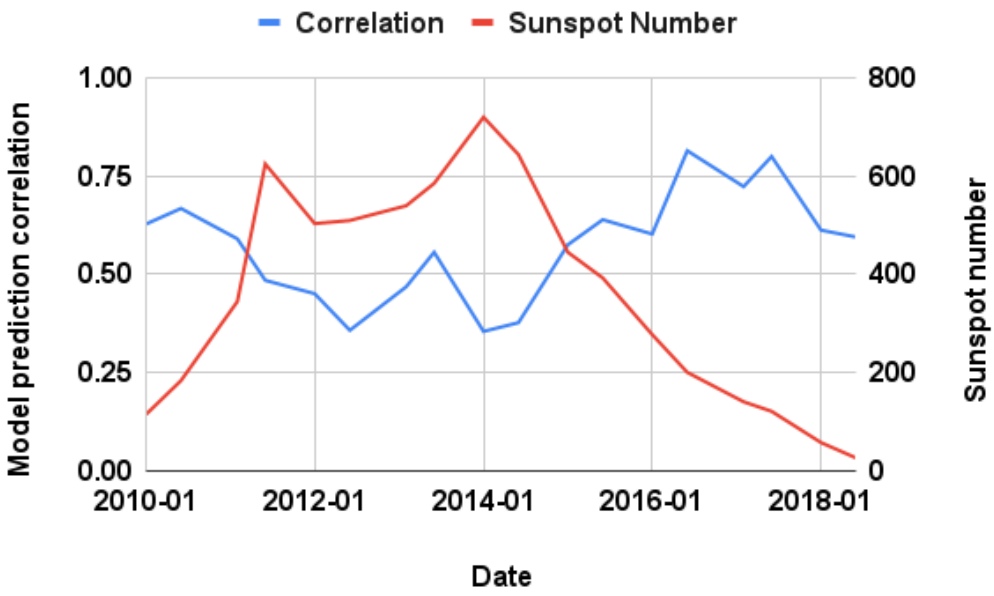


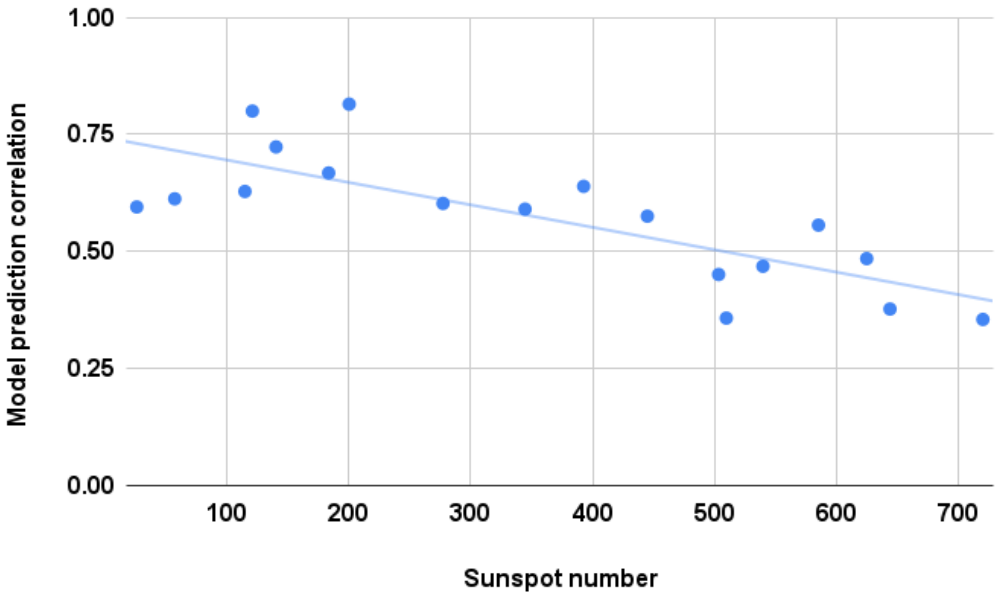
Figure 7: Speed prediction vs minimum pixel intensity for a coronal hole pictured on 2016-12-10

5.3.4 Solar cycle variability

The performance of the model is highly dependent on position in the solar cycle. Figure 8a plots the correlation of the model prediction with the ground truth (blue) at 6 month intervals against the number of sunspots (red) in that period. The sunspot number represents the solar cycle. Notably, the model's prediction correlation to ground truth is much better in the declining phase of the solar cycle, that is in the 2016 to 2018. At the same time, it performs much worse around the peak of the solar cycle in 2014. This relationship is confirmed when we view the data as correlation-sunspot number couples and visualize them in a scatter plot. This is shown in the figure 8b. We observe a strong, 0.78, negative correlation of the number of sunspots and the model prediction correlation to the ground truth. Since sunspot number is used to measure the solar cycle, this suggests that the model performance is highly dependent on the solar cycle and more specifically on the prevalent type of solar activity in a given period.



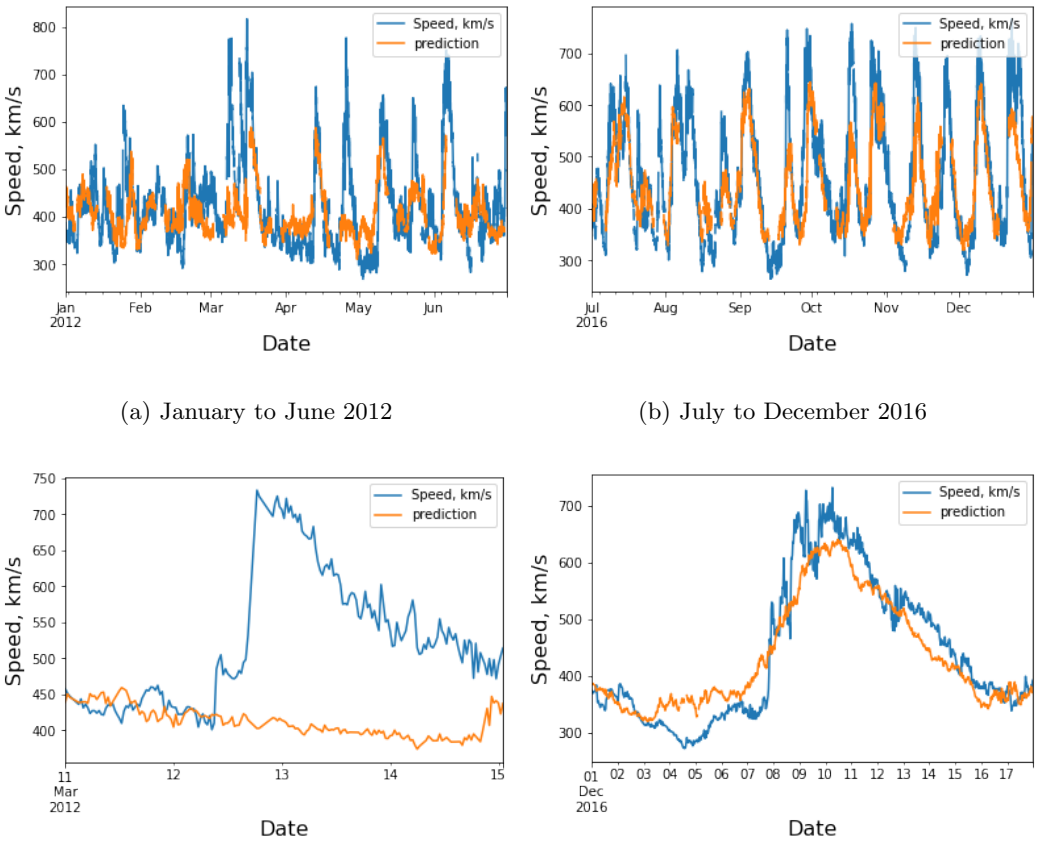
(a) Model prediction correlation and sunspot number vs date



(b) Model prediction correlation and sunspot number

Figure 8: Model performance compared to sunspot number

Indeed, a key component of the model's performance across the solar cycle is the type of encountered solar features. The top two panels of the figure 9 show the model's performance in early 2012, with 80.81 RMSE and 0.45 correlation, and in late 2016, with 73.32 RMSE and 0.81 correlation. The solar wind behaviour in the later half of 2016, was driven by coronal holes and the high speed solar wind streams associated with them. Whereas, 2012 had a much higher sunspot number and had far more Earth-directed CMEs.



(a) January to June 2012

(b) July to December 2016

(c) Coronal Mass ejection, March 2012

(d) Coronal hole, December 2016

Figure 9: Solar Swin Transformer performance in different parts of the solar cycle and on different solar phenomena

We observe a marked difference in performance between predictions driven by CMEs and coronal holes. Figures 9d and 9c show how the model captures the longer lasting, speed profile of a coronal hole quite well, while missing the speed profile of the sudden CME. This offers an explanation to the pronounced variability in the model's prediction quality. The solar activity in the declining phase is driven by coronal holes. These are more easily picked up by the models. Since the Sun in the later half of 2016 was in the declin-

ing phase, the models' performance was much better. In 2012, a year with far more CMEs, the model performance was reduced, as the models struggled to catch the CMEs.

The failure to fit on the more sudden coronal mass ejections is a chief limitation of the models developed in this space. It can be ascribed to the lack of significant and persistent CME-related features in the EUV images, preventing them from being captured by the models. We note that ML models using solar EUV images alone to forecast other space weather related parameters such as geomagnetic activity as measured by the AE or Kp indices or suprathermal electrons at geostationary orbit would most likely suffer from the same limitation resulting in a similar pattern of behaviour with the best correlations during the declining phase of the solar cycle and the worst correlations around solar maximum.

6 Conclusions

This study uses attention-based machine vision models and a set of methodological and modelling improvements to forecast the solar wind speed at L1 using solar images at 211Å wavelength. These improvements result in 11.1% lower RMSE and 17.4% higher prediction correlation with the ground truth when compared to previous works. Additionally, this study observed that attention-based architectures in general have about 2-3% performance edge in both RMSE and correlation over the previously-used convolutional alternatives. The model's performance is highly dependent on the position in the solar cycle. The model performance is strongly negatively correlated with the sunspot number, as the model performance is better in the declining phase of the solar cycle when the solar wind behaviour is dominated by coronal hole activity. Finally, the model has independently learned two empirical relationships between coronal established by previous publications. First, it complies with the observed linear relationship between coronal hole area and the peak speed associated with it. Second, it learned that the darker the coronal hole, the stronger the solar wind speed associated with it.

Open Research

The SDOML 211 Å image data is available here: `https://purl.stanford.edu/vk217bh4910`

The OmniWeb solar wind data is available here: `https://omniweb.gsfc.nasa.gov/form/omni_min.html`

Code for analysing model output is supplied here: `https://github.com/eddbrown/solar-swin-transformer-output-data`

Acknowledgements

For this study, we acknowledge extensive use of the SDOML (Galvez et al., 2019) dataset. These images are processed versions of images taken by the AIA (Atmospheric Imaging Assembly) instrument aboard the Solar Dynamic Observatory.

Furthermore we acknowledge the use of the OMNIWeb service at their provided measurements of the solar wind speed at L1.

Regarding software, we acknowledge extensive use of python and the python packages numpy (Harris et al., 2020) and pytorch (Paszke et al., 2019).

For experiment tracking and general project organisation, we acknowledge extensive use of the software provided by Weights and Biases (Biewald, 2020).

N. P. Meredith and R. B. Horne would like to acknowledge support from the NERC grants NE/V00249X/1 (Sat-Risk) and NE/R016038/1.

References

Bailey, R. L., Reiss, M. A., Arge, C. N., Möstl, C., Henney, C. J., Owens, M. J., ... Hinterreiter, J. (2021). Using gradient boosting regression to improve ambient solar wind model predictions. *Space Weather*, *19*(5), e2020SW002673. (e2020SW002673 2020SW002673) doi: https://doi.org/10.1029/2020SW002673

Biewald, L. (2020). *Experiment tracking with weights and biases.* (Software available from wandb.com)

Borovsky, J. E., & Denton, M. H. (2006). Differences between cme-driven storms and cir-driven storms. *Journal of Geophysical Research: Space Physics*, *111*(A7). doi: https://doi.org/10.1029/2005JA011447

Burlaga, L., & Lepping, R. (1977). The causes of recurrent geomagnetic storms. *Planetary and Space Science*, *25*(12), 1151-1160. doi: https://doi.org/10.1016/0032-0633(77)90090-3

Chandorkar, M., Furtlehner, C., Poduval, B., Camporeale, E., & Sebag, M. (2019). Dynamic time lag regression: predicting what & when. In *International conference on learning representations.*

Chen, G.-m., Xu, J., Wang, W., Lei, J., & Burns, A. G. (2012). A comparison of the effects of cir- and cme-induced geomagnetic activity on thermospheric densities and spacecraft orbits: Case studies. *Journal of Geophysical Research: Space Physics*, *117*(A8). doi: https://doi.org/10.1029/2012JA017782

Cranmer, S. R. (2009). Coronal holes. *Living Reviews in Solar Physics*, *6*, 3,66. doi: https://doi.org/10.12942/lrsp-2009-3

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 ieee conference on computer vision and pattern recognition* (p. 248-255). doi: 10.1109/CVPR.2009.5206848

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *CoRR*, *abs/2010.11929*.

Galvez, R., Fouhey, D. F., Jin, M., Szenicer, A., Muñoz-Jaramillo, A., Cheung, M. C. M., ... Thomas, R. (2019, may). A machine-learning data set prepared from the NASA solar dynamics observatory mission. *The Astrophysical Journal Supplement Series*, *242*(1), 7. doi: 10.3847/1538-4365/ab1005

Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., & Brendel, W. (2018). Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv preprint arXiv:1811.12231*.

Han, K., Xiao, A., Wu, E., Guo, J., Xu, C., & Wang, Y. (2021). *Transformer in*

transformer.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... Oliphant, T. E. (2020, September). Array programming with NumPy. *Nature*, *585*(7825), 357–362. doi: 10.1038/s41586-020-2649-2

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *2016 ieee conference on computer vision and pattern recognition (cvpr)* (p. 770-778). doi: 10.1109/CVPR.2016.90

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Hofmeister, S. J., Veronig, A. M., Temmer, M., Vennerstrom, S., Heber, B., & Vrsnak, B. (2018). The dependence of the peak velocity of high-speed solar wind streams as measured in the ecliptic by ace and the stereo satellites on the area and co-latitude of their solar source coronal holes. *Journal of Geophysical Research. Space Physics*, *123*, 1738 - 1753.

Horne, R. B., Phillips, M. W., Glauert, S. A., Meredith, N. P., Hands, A. D. P., Ryden, K. A., & Li, W. (2018). Realistic worst case for a severe space weather event driven by a fast solar wind stream. *Space Weather*, *16*(9), 1202-1215. doi: https://doi.org/10.1029/2018SW001948

Huang, G., Liu, Z., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261-2269.

Kilpua, E., Balogh, A., Von Steiger, R., & Liu, Y. (2017, 01). Geoeffective properties of solar transients and stream interaction regions. In (p. 295-338). doi: 10.1007/978-94-024-1588-9_9

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kozyra, J. U., Crowley, G., Emery, B. A., Fang, X., Maris, G., Mlynczak, M. G., ... Yee, J.-H. (2006). Response of the upper/middle atmosphere to coronal holes and powerful high-speed solar wind streams in 2003. In *Recurrent magnetic storms: Corotating solar wind streams* (p. 319-340). American Geophysical Union (AGU). doi: https://doi.org/10.1029/167GM24

Lemen, J. R., Akin, D. J., Boerner, P. F., Chou, C., Drake, J. F., Duncan, D. W., ... others (2011). The atmospheric imaging assembly (aia) on the solar dynamics observatory (sdo). In *The solar dynamics observatory* (pp. 17–40). Springer.

Liu, D. D., Huang, C., Lu, J. Y., & Wang, J. S. (2011, 05). The hourly average solar wind velocity prediction based on support vector regression method. *Monthly*

Notices of the Royal Astronomical Society, *413*(4), 2877-2882. doi: 10.1111/j.1365-2966.2011.18359.x

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. *CoRR*, *abs/2103.14030*.

Meredith, N. P., Horne, R. B., Lam, M. M., Denton, M. H., Borovsky, J. E., & Green, J. C. (2011). Energetic electron precipitation during high-speed solar wind stream driven storms. *Journal of Geophysical Research: Space Physics*, *116*(A5). doi: https://doi.org/10.1029/2010JA016293

Nolte, J., Krieger, A., Timothy, A., Gold, R., Roelof, E., Vaiana, G., ... McIntosh, P. (1976). Coronal holes as sources of solar wind. *Solar Physics*, *46*(2), 303–322.

Obridko, V., Shelting, B., Livshits, I., & Askerov, A. (2009). Relationship between the contrast of coronal holes and parameters of the solar wind streams. *Astronomy reports*, *53*(11), 1050–1058.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in neural information processing systems 32* (pp. 8024–8035). Curran Associates, Inc.

Raju, H., & Das, S. (2021). Cnn-based deep learning model for solar wind forecasting. *Sol Phys*. doi: https://doi.org/10.1007/s11207-021-01874-6

Richardson, I. G., Webb, D. F., Zhang, J., Berdichevsky, D. B., Biesecker, D. A., Kasper, J. C., ... Zhukov, A. N. (2006). Major geomagnetic storms (dst less than or equal to -100 nt) generated by corotating interaction regions. *Journal of Geophysical Research: Space Physics*, *111*(A7). doi: https://doi.org/10.1029/2005JA011476

Ridpath, I. (2012). *A dictionary of astronomy.* Oxford University Press.

St. Cyr, O. C., Howard, R. A., Sheeley Jr., N. R., Plunkett, S. P., Michels, D. J., Paswaters, S. E., ... Lamy, P. L. (2000). Properties of coronal mass ejections: Soho lasco observations from january 1996 to june 1998. *Journal of Geophysical Research: Space Physics*, *105*(A8), 18169-18185. doi: https://doi.org/10.1029/1999JA000381

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S. E., Anguelov, D., ... Rabinovich, A. (2014). Going deeper with convolutions. *CoRR*, *abs/1409.4842*.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. *2016 IEEE Conference on*

Computer Vision and Pattern Recognition (CVPR), 2818-2826.

Tsurutani, B. T., Gonzalez, W. D., Gonzalez, A. L. C., Tang, F., Arballo, J. K., & Okada, M. (1995). Interplanetary origin of geomagnetic activity in the declining phase of the solar cycle. *Journal of Geophysical Research: Space Physics*, *100*(A11), 21717-21733. doi: https://doi.org/10.1029/95JA01476

Turner, N. E., Mitchell, E. J., Knipp, D. J., & Emery, B. A. (2006). Energetics of magnetic storms driven by corotating interaction regions: A study of geoeffectiveness. In *Recurrent magnetic storms: Corotating solar wind streams* (p. 113-124). American Geophysical Union (AGU). doi: https://doi.org/10.1029/167GM11

Upendran, V., Cheung, M. C. M., Hanasoge, S., & Krishnamurthi, G. (2020). Solar wind prediction using deep learning. *Space Weather*, *18*(9), e2020SW002478. (e2020SW002478 10.1029/2020SW002478) doi: https://doi.org/10.1029/2020SW002478

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need..

Wintoft, P., & Lundstedt, H. (1997). Prediction of daily average solar wind velocity from solar magnetic field observations using hybrid intelligent systems. *Physics and Chemistry of the Earth*, *22*(7), 617-622. (Low-Altitude Investigation of Dayside Magnetospheric Boundary Processes) doi: https://doi.org/10.1016/S0079-1946(97)00186-9

Wintoft, P., & Lundstedt, H. (1999). A neural network study of the mapping from solar magnetic fields to the daily average solar wind velocity. *Journal of Geophysical Research: Space Physics*, *104*(A4), 6729-6736. doi: https://doi.org/10.1029/1998JA900183

Yang, Y., Shen, F., Yang, Z., & Feng, X. (2018). Prediction of solar wind speed at 1 au using an artificial neural network. *Space Weather*, *16*(9), 1227-1244. doi: https://doi.org/10.1029/2018SW001955