Vision transformers and techniques for improving solar wind speed forecasts using solar EUV images

Filip Svoboda*
Department of Computer Science
University of Cambridge
15 JJ Thomson Ave, Cambridge, CB3 0FD
fs437@cam.ac.uk
* *equal contribution*

Edward Brown*
Department of Computer Science
University of Cambridge
15 JJ Thomson Ave, Cambridge, CB3 0FD
ejeb4@cam.ac.uk
* *equal contribution*

Nigel P. Meredith
British Antarctic Survey
High Cross, Madingley Rd, Cambridge CB3 0ET
nmer@bas.ac.uk

Nicholas D. Lane
Department of Computer Science
University of Cambridge
15 JJ Thomson Ave, Cambridge CB3 0FD
ndl32@cam.ac.uk

Abstract

Extreme-ultraviolet images taken by the Atmospheric Imaging Assembly make it possible to use deep vision techniques in the prediction of solar wind speed - a difficult, high-impact, and unsolved problem. This study uses vision transformers and a set of methodological and modelling improvements to deliver an 11.1% lower RMSE error, and a 17.4% higher prediction correlation compared to the previous state of the art models. Furthermore, our analysis shows that vision transformers combined with our pipeline consistently outperform convolutional alternatives. Additionally, the best vision transformer outperforms the best convolutional model by 1.8% in RMSE and 2.6% in correlation with the ground truth solar wind speed.

1 Introduction

Solar wind, a stream of charged particles, is emitted from the sun and arrives 2-7 days later at Earth. Earth itself is largely shielded from moderate solar winds by its magnetosphere, but these winds can cause geomagnetic disturbances - one of the main sources of uncertainty in orbital estimation [1]. Extreme solar winds, can disrupt satellites, impact communication, and even damage power grids. Consequently, accurately forecasting the solar wind speed is very important for our modern society.

This study forecasts the solar wind speed, as published by OmniWeb [10], using the Extreme UV images taken by the Solar Dynamics Observatory (SDO) using the Atmospheric Image Assembly (AIA) [8] and preprocessed in the SDOML dataset [4] available at `https://purl.stanford.edu/vk217bh4910`. Results for forecasting at a four day lag from a single 211 Å image are presented - but this forecast could be used for any lag up to 4 days.

Fourth Workshop on Machine Learning and the Physical Sciences (NeurIPS 2021).

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%
Former state of the art [14]	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 1: Performance of our solar models relative to previous works predicting solar wind speed using the EUV data at a 4 day forecast horizon in the period May 2010 to December 2018. The same results hold for 1, 2, and 3 day forecast horizons and for histories of up to 4 images tested by [14].

The previous state of the art is by Upendran et al [14]. Their model relied on an ImageNet pre-trained GoogleNet [12] feature extractor, one LSTM cell, and a fully connected output layer. At the 4-day time horizon, they report an RMSE of 81.21 and 0.54 prediction correlation to the ground truth.

This study developed a collection of significant methodological and modelling advancements. First, it proposes one technique of data pre-processing, three techniques for dataset organization, and one of data augmentation. Furthermore, it improves on the previous architecture construction methodology by streamlining it and so reducing its over-fitting. Finally, it tests our pipeline on a comprehensive selection of vision feature extractors in order to test its generality and resilience.

Our best performing model was the solar Swin Transformer [9] which improved on the state of the art by 11.1% in RMSE and 17.4% in the correlation. Using the same feature extractor, GoogleNet, our pipeline outperforms the previous state of the art by 9.2% in RMSE and 14.6% in correlation. Of our solar models, all deliver an improvement of at least 8.80% in RMSE and 12.70% in correlation.

2 Proposed Approach

Our methodology differs significantly from that used in previous works in the following aspects:

Image pre-processing: The EUV images are at their current resolution 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 at the edges of the solar disk, and then down sample this, smaller, image to achieve the desired resolution. 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, the relevant, portion of the image less aggressively.

The data split buffer: We train and validate using a 5-fold cross-validation strategy. The dates of the SDOML dataset are split into chunks of 20 days. A pattern is created of 3 chunks of training dates, one of validation and one of test. This pattern is shifted along the full date range of the images five times so that each chunk serves its turn in the test set. This method is similar to Upendran et al, except, in ours a buffer of 4 days is added between the chunks. This is significant, because without the buffers our training validation and test sets will not be independent as the the solar wind speed is highly auto-correlated at up to 4 day periodicity. The 1 day auto-correlation in 2010-2018 was 0.70.

The sampling frequency: We replace the previously used daily sampling resolution with a 30 minute schedule, because solar wind speeds can change significantly at up to hourly resolutions.

The Carrington rotation: The sun rotates on average every 27.28 days, this is one Carrington rotation[11]. As such, the solar features that affected the solar wind speed at a given point come round back 27 days later and produce similar effects. Thus, the solar wind speed is, also, auto-correlated at the Carrington rotation periodicity - 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.

The north-south flip: 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.

Other considerations: Missing images are substituted with valid observations no more than 30 minutes removed from the missing datum. Missing speed data is interpolated from available data no more than 30 minutes removed. Time steps with no valid data for filling in the missing observations are discarded. Hyper-parameters are chosen using a Bayesian parameter sweep using the software Weights & Biases [2]. For cost reasons, the sweep is conducted at 120 minutes resolution, on one quarter of the data. The metrics used to analyze to overall performance of the algorithm are the root mean squared error (RMSE) and Pearson correlation of the ground truth to the predicted values.

Model building in previous work relied on a convolutional feature extractor pre-trained on ImageNet in combination with an LSTM cell and a fully connected layer [14]. 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. We put forward three fundamental amendments to the aforementioned methodology.

First, we suggest to re-train the feature extractor, rather than to use its ImageNet weights. This we believe to be strictly necessary due to the wide gap between the EUV and the ImageNet datasets.

Second, we propose extracting a single, rather than multiple layers of activations from the feature extractor. Later layers, are by definition non-linear projections of the earlier ones and so can be expected to be highly correlated with the earlier layer features. This correlation will likely significantly worsen the model's multi-collinearity and thus degrade its performance.

Third, we propose using a single image rather than a history of multiple consecutive images as inputs to the models. 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.

Finally, we account for the wind speed from one Carrington rotation ago by adding it as an input feature along with the image-extracted features to the earlier of the two final fully connected layers.

As feature extractors, we chose GoogleNet [12] and its close relative the InceptionNet v4 [13] as it is at the core of the current state of the art [14]. We added ResNet [6] and DenseNet [7] to represent simpler convolutional architectures. Finally, we included the Vision Transformer [3], and its two hierarchical extensions, the Transformer-in-Transformer (TNT) [5] and the Swin Transformer [9]. All experiments were ran on V100 Nvidia GPU, resulting in a total compute of about 900 GPU hours.

3 Results

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 model in the field and two naive persistence model benchmarks. Notably, all of the models trained under our pipeline improve on the work by [14] by at least 8.8% in RMSE and 12.7% in correlation. Indeed, GoogleNet, the same feature extractor used in Upendran et al's model, demonstrated the absolute improvement our pipeline has delivered. Specifically, it lowered the RMSE by 9.2% and increased the correlation by 14.6%. Furthermore, our best performing model, based off Swin Vision Transformer, improves on the state of the art by 11.1% in RMSE and 17.4% in correlation. Similarly, transformers as a model family outperformed convolutions by about 1 to 2% in either metric. The same pattern of results holds true for the full range of experiments in Upendran et al [14], that is for 1, 2, and 3 day forecast horizons as well as for adding histories of up to 4 days long.

Ablation study To demonstrate the 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, those between the validation and training sets become part of the validation set, thus making the two dependent and hurting the model's generalization to the test set.

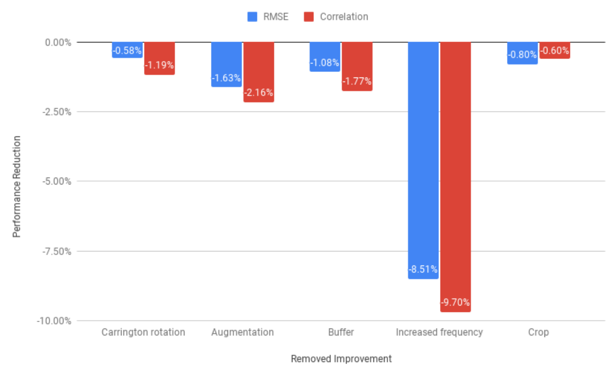
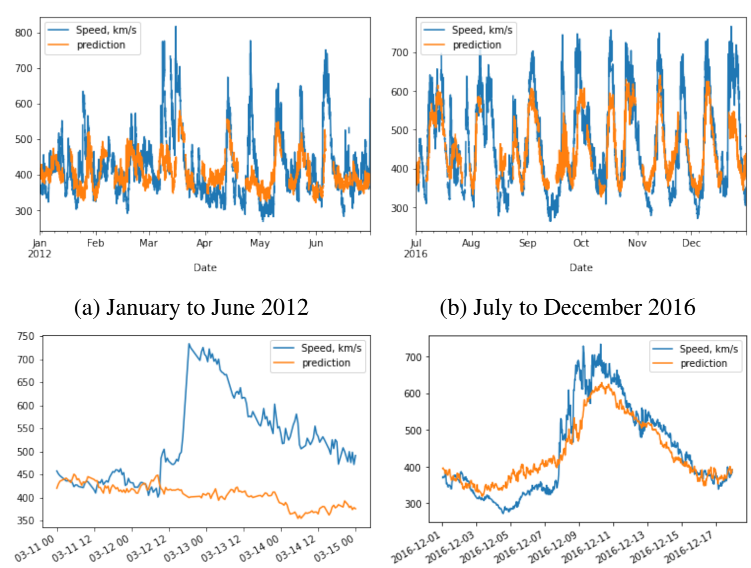


Figure 1: Performance decreases resulting from removing one improvement at a time.

Figure 1 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. This shows that, while consecutive images are too highly correlated to meaningfully add to each other's wind speed predictions, they are distinct enough to permit useful training when taken as separate dataset examples.

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 that their great effective importance is ultimately determined by the overwhelming need for highly accurate solar wind speed predictions for systems such as satellite orbit determination.

Critical evaluation of model performance A key component of model's performance is the position in the solar cycle and the type of encountered solar features. Figure 2 demonstrates this on the representative vision transformer example. The top two panels show its performance in early 2012, with 78.64 RMSE and 0.49 correlation, and in late 2016, with 71.25 RMSE and 0.81 correlation. The much higher correlation in 2016 suggests that the solar winds in that 6 months were fitted significantly better than in 2012 despite the moderate difference in the average RMSE over the two periods.



(a) January to June 2012 (b) July to December 2016

(c) Coronal Mass ejection, March 2012
1 tick = 12 hours

(d) Coronal hole, December 2016
1 tick = 48 hours

Figure 2: Solar Vision Transformer performance across the solar cycle, and solar phenomena.

Indeed, we observe a marked difference in performance between predictions driven by specific types of solar phenomena. The lower two panels of figure 2 demonstrate how the longer lasting coronal holes are being captured by the vision transformer and thus correctly reflected in its predictions, while the much more sudden coronal mass ejections fail to be reflected in the model's predictions. This seems to offer an explanation to the pronounced variability in the model's prediction quality. It is likely that the solar wind in 2012 has been less often caused by coronal holes, than by coronal mass ejections and other phenomena which the current models are not geared towards capturing.

The failure to fit on the finer solar features, such as the coronal mass ejections, is the chief limitation of the models developed in this space. Likely, it can be ascribed to the relative sparsity of coronal mass ejections in the data, which limits the model's exposure to them, and to their shorter time frame as they usually occur at significantly shorter timescale than coronal holes do. Combined, this makes them harder to capture and learn from even at our higher sampling frequency.

4 Conclusions

This study uses vision transformers and a set of methodological and modelling improvements to forecast solar OmbiWeb's wind speed using the SDOML EUV data at L1. It proposes improved data pre-processing and augmentation, training set construction and increased sampling frequency. Furthermore, it proposes a streamlined architecture construction that is less prone to multicollinearity and over-fitting. These improvements result in 11.1% lower RMSE and 17.4% higher prediction correlation with the ground truth. Finally, it observes that vision transformer-based architectures have about 2-3% performance edge in both RMSE and correlation over the convolutional alternatives.

References

[1] T. Berger, M. Holzinger, E. Sutton, and J. Thayer. Flying Through Uncertainty. *Space Weather*, 18(1), 2020.

[2] L. Biewald. Experiment tracking with weights and biases, 2020. Software available from wandb.com.

[3] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *CoRR*, abs/2010.11929, 2020.

[4] R. Galvez, D. F. Fouhey, M. Jin, A. Szenicer, A. Muñoz-Jaramillo, M. C. M. Cheung, P. J. Wright, M. G. Bobra, Y. Liu, J. Mason, and R. Thomas. A machine-learning data set prepared from the NASA solar dynamics observatory mission. *The Astrophysical Journal Supplement Series*, 242(1):7, may 2019.

[5] K. Han, A. Xiao, E. Wu, J. Guo, C. Xu, and Y. Wang. Transformer in transformer, 2021.

[6] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

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

[8] J. R. Lemen, 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, et al. The atmospheric imaging assembly (aia) on the solar dynamics observatory (sdo). In *The solar dynamics observatory*, pages 17–40. Springer, 2011.

[9] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo. Swin transformer: Hierarchical vision transformer using shifted windows. *CoRR*, abs/2103.14030, 2021.

[10] OmniWeb. Omniweb data archive. `https://omniweb.gsfc.nasa.gov/form/dx1.html`. Accessed: 2021.

[11] I. Ridpath. *A dictionary of astronomy*. Oxford University Press, 2012.

[12] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. *CoRR*, abs/1409.4842, 2014.

[13] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826, 2016.

[14] V. Upendran, M. C. M. Cheung, S. Hanasoge, and G. Krishnamurthi. Solar wind prediction using deep learning. *Space Weather*, 18(9):e2020SW002478, 2020. e2020SW002478 10.1029/2020SW002478.

Checklist

The checklist follows the references. Please read the checklist guidelines carefully for information on how to answer these questions. For each question, change the default **[TODO]** to [Yes] , [No] , or [N/A] . You are strongly encouraged to include a **justification to your answer**, either by referencing the appropriate section of your paper or providing a brief inline description. For example:

- Did you include the license to the code and datasets? [Yes] See Section **??**.
- Did you include the license to the code and datasets? [No] The code and the data are proprietary.
- Did you include the license to the code and datasets? [N/A]

Please do not modify the questions and only use the provided macros for your answers. Note that the Checklist section does not count towards the page limit. In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [N/A] . There negative societal impacts for forecasting solar wind speed.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No] . *The work once further extended and analyzed will be submitted to the Space Weather Journal. Our code will be made publicly available then. We are more than happy to discuss this point.*
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]