

SEMANTIC SEGMENTATION OF SOLAR RADIO SPIKES AT LOW FREQUENCIES

PEARSE C. MURPHY*

LESIA, Observatoire de Paris, Université PSL, Sorbonne Université, Université Paris Cité, CNRS, 92190 Meudon, France

STÉPHANE AICARDI

DIO, Observatoire de Paris, Université PSL, CNRS, 75014 Paris, France

BAPTISTE CECCONI

LESIA, Observatoire de Paris, Université PSL, Sorbonne Université, Université Paris Cité, CNRS, 92190 Meudon, France

CARINE BRIAND

LESIA, Observatoire de Paris, Université PSL, Sorbonne Université, Université Paris Cité, CNRS, 92190 Meudon, France

THIBAUT PECCOUX

LESIA, Observatoire de Paris, Université PSL, Sorbonne Université, Université Paris Cité, CNRS, 92190 Meudon, France and
IPSA, 94200 Ivry-sur-Seine, France

Version March 14, 2024

ABSTRACT

Solar radio spikes are short lived, narrow bandwidth features in low frequency solar radio observations. The timing of their occurrence and the number of spikes in a given observation is often unpredictable. The high temporal and frequency of resolution of modern radio telescopes such as NenuFAR mean that manually identifying radio spikes is an arduous task. Machine learning approaches to data exploration in solar radio data is on the rise. Here we describe a convolutional neural network to identify the per pixel location of radio spikes as well as determine some simple characteristics of duration, spectral width and drift rate. The model, which we call SpikeNet, was trained using an Nvidia Tesla T4 16 GB GPU with $\sim 100\,000$ sample spikes in a total time of 2.2 hours. The segmentation performs well with an intersection over union in the test set of ~ 0.85 . The root mean squared error for predicted spike properties is of the order of 23%. Applying the algorithm to unlabelled data successfully generates segmentation masks although the accuracy of the predicted properties is less reliable, particularly when more than one spike is present in the same 64×64 pixel time-frequency range. We have successfully demonstrated that our convolutional neural network can locate and characterise solar radio spikes in a number of seconds compared to the weeks it would take for manual identification.

1. INTRODUCTION

Solar radio bursts occur over a wide frequency range with varying spectral and temporal characteristics. At decametric wavelengths there are five canonical types of radio burst which can have durations ranging from seconds to hours and spectral widths up to tens of MHz. Alongside these, there exist a wide variety of short duration, narrow spectral width bursts. These bursts generally fall into three categories; solar radio spikes (Tarnstrom & Philip 1972), S-bursts (Ellis 1969) and striae (de La Noe & Boisshot 1972), with individual striae and spikes sharing a number of characteristics (Clarkson et al. 2021). These short lived bursts can give an insight into energy transfer that takes place over a number of milliseconds (e.g. Clarkson et al. 2021) as well as the tur-

bulent nature of plasma in the corona (e.g. Kontar et al. 2017; Reid & Kontar 2021). They may also offer a way of remotely measuring the magnetic field in the corona (e.g. Melnik et al. 2010; Clarke et al. 2019). This underscores the importance of understanding the origins of short duration, narrow bandwidth solar radio bursts.

Modern state of the art radio telescopes such as the LOw Frequency ARray (LOFAR; van Haarlem et al. 2013) and its extension, the New Extension in Nançay Upgrading LOFAR (NenuFAR; Zarka et al. 2012) have the temporal and spectral resolution necessary to analyse short duration, narrow bandwidth solar radio bursts in great detail (Morosan et al. 2015; Clarke et al. 2019; Kontar et al. 2017; Sharykin et al. 2018; Clarkson et al. 2021, 2023, Briand et al. *in prep*). However, the sporadic occurrence of spike bursts coupled with the vast data rates of the high resolution observations means that finding them manually can be and arduous and time con-

*pearse.murphy@obspm.fr

suming task. Modern machine learning methods are the ideal solution for this as, once the model is trained, they eliminate the need for human identification of spikes and determining their characteristics. The application of machine learning algorithms to solar radio data has seen increased popularity in recent literature. Already, we have seen the application of the YOLO (you only look once; Redmon et al. 2016) convolutional neural network (CNN) to detecting solar radio spikes in the 1.1 GHz–1.34 GHz range (Lv et al. 2023), while Scully et al. (2021, 2023) used YOLO to detect Type III radio bursts in the 10 MHz–90 MHz range.

YOLO performs what is known as object detection, where an object is located in an image and a bounding box which determines its height and width is drawn around it. The task of locating an object in an image on a per pixel basis is known as semantic segmentation. The varied and irregular morphology of solar radio spikes make it a problem particularly suited for the use of semantic segmentation. Murphy et al. (2023) used a CNN called UNET (Ronneberger et al. 2015) to perform semantic segmentation on NenuFAR dynamic spectra to automatically detect the presence of solar radio emission. Here we adapt the work of Murphy et al. (2023) to perform segmentation for solar radio spikes and determine their location, duration, spectral width and drift rate in NenuFAR dynamic spectra.

The design and training procedure for our model is described in Section 2. We describe the results of the training in Section 3 and give further discussion on the application of our CNN in Section 4. We conclude with a summary in Section 5.

2. METHOD

Our goal is to develop a CNN to automatically generate a mask of where a solar radio spike occurred in the time-frequency domain as well as predict the spike’s central location, duration, spectral width and drift rate. In order to achieve this, we must first perform pre-processing of the NenuFAR dynamic spectra in order to create a dataset suitable for training the CNN.

2.1. Preparing the datasets

The time and frequency location, spectral width, duration and drift rate of 1000 spike bursts were identified manually by summer student Thibault Peccoux, to whom we owe a debt of gratitude for this tedious work. The spikes were located between 10 MHz and 90 MHz in NenuFAR observations on 2022-02-02, 2022-05-29, 2023-05-02, 2023-06-01 and 2023-07-10. These are the same data used by Briand et al. (in prep). Figure 1 shows a dynamic spectrum of a number of radio spikes occurring on 2022-02-02 over ~ 9 s in the frequency range 25 MHz–60 MHz. The figure also shows radio frequency interference (RFI) which is a common feature of low frequency observations and thus our dataset must incorporate this.

These spikes formed the basis for the training, validation and test datasets. We performed a stratified split based on the date of observation so that each split had the same proportion of spikes that were identified on each date. The percentage of spikes in each split were 64% training, 16% validation and 20% test. The inputs to the CNN are 64×64 pixel Stokes I and Stokes V/I dynamic spectra which we hereafter refer to interchangeably as

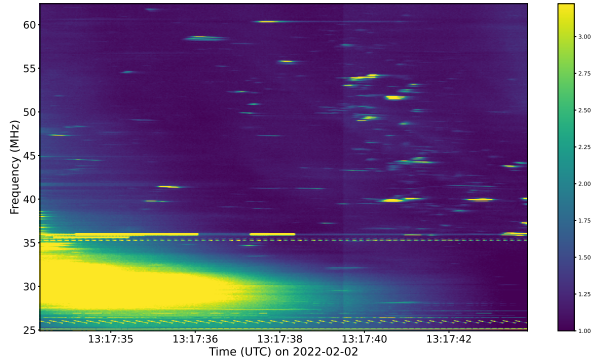


FIG. 1.— A sample of solar radio spikes on 2022-02-02. The colourbar is intensity relative to the background. The large number of spikes in such a short time range indicate the difficulty in manually annotating their every occurrence. Radio frequency interference and the end of a type III solar radio burst in the frequency range 25 MHz–35 MHz also pose a challenge.

tiles. One time pixel corresponds to ~ 21 ms and one frequency pixel is ~ 6.1 kHz. This results in a time duration of 1.34 s and a frequency range of 390.625 kHz for each tile. CNNs often see improved performance when the input values lie within a small range, thus we normalised these to the background and rescale in the range 0–1. The frequency range of the dynamic spectrum in MHz was also included as an input. In order to generate the ground truth output for the model, a segmentation mask was produced for each spike burst using the CUSUM-slope method described by Murphy et al. (2023).

To increase the number of samples available for training, we applied random shifts in time and frequency to the spikes in the training and validation sets. The same shifts were also applied to the segmentation masks. In an effort to improve the robustness of the model when dealing with unseen data, we also located times where no spike occurred to produce samples of the background and added these to the training set. These also included samples in common RFI bands $\lesssim 30$ MHz and at ~ 70 MHz. In total the training set comprised of 90 435 samples and the validation set 16 000. The test set did not undergo any augmentation and contained 200 samples.

For each spike in the dataset we have a list of its characteristics, i.e., its location in time and frequency, its duration, its spectral width and its drift rate. We determine the time and frequency location in terms of pixels while duration, spectral width and drift rate are in physical units of s, MHz and MHz s^{-1} respectively.

Finally, we determined an objectness score for each sample. This is used to determine how certain the model is that a spike exists in the sample at all. An objectness of 0 means the model is certain there is no spike while an objectness of 1 means the model is certain there is a spike. We determine the objectness of each sample as the ratio of total number of pixels in the shifted segmentation mask to that of the segmentation mask before it was shifted. Thus, any spike that is not fully in a tile will have an objectness score less than one.

2.2. Model architecture and training the model

The primary output of our model, affectionately referred to as SpikeNet, is a mask of every pixel where a solar radio spike occurred in time and frequency. As mentioned in Section 1, this type of task is known as semantic

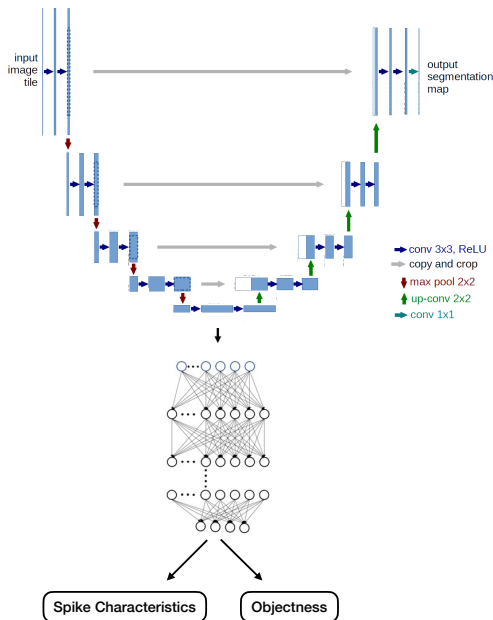


FIG. 2.— Architecture of our model. The top half of the model is the UNET adapted from [Ronneberger et al. \(2015\)](#). We pass the outputs from UNET’s bottleneck layer to a fully connected network. This in turn has two outputs; the spike characteristics and the objectness.

segmentation and a common CNN used in this regard is called UNET. We use the implementation of UNET in the `segmentation_models` Python library ([Iakubovskii 2019](#)) as a starting point.

As well as trying to produce a segmentation mask from these inputs, we require our model to determine the central location in time and frequency, duration, spectral width and drift rate of the spike. Thus, in addition to the decoder part of the UNET, we included a fully connected deep neural network with 5 hidden layers of 30 neurons each to determine these characteristics, as well as the objectness score. The model architecture is depicted in Figure 2. We take the values in the bottleneck layer of the original UNET to use as the input to the fully connected network. The final hidden layer is connect to two separate output layers. This then gives two separate outputs, the 5 spike characteristics and the objectness score.

We apply a different loss for each of the outputs; binary cross entropy for the segmentation performed by UNET, mean squared error for the regression of spike characteristics and binary cross entropy for the objectness. The total loss measured during training was the weighted sum of each of these losses. We give a weight of 10 to the regression loss and a weight of 1 to segmentation and objectness loss. The model was trained using a linearly increasing training rate for the first 5 epochs before implementing a cosine decay. The total validation loss was monitored per epoch and the training was stopped once it failed to decrease for three epochs in a row. The model was trained for a total of 26 epochs which took 2.2 hours on an Nvidia Tesla T4 16 GB GPU.

3. RESULTS

The values of the individual and total loss functions per epoch are shown in Figure 3. The loss for the train-

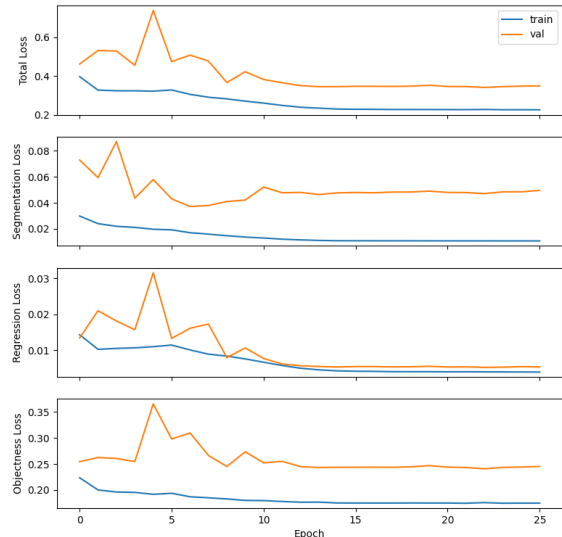


FIG. 3.— Evolution of loss with respect to training epoch. The training (blue) and validation (orange) loss is shown in each panel for the total loss, segmentation loss, regression loss and objectness loss.

ing set (blue) decreased rapidly in all cases while the validation loss (orange) decreased slowly before forming a plateau. The gap between the training and validation loss could indicate under fitting or that the data samples in the training set were not representative of those in the validation set.

The key metric to determine the accuracy of our segmentation masks to the ground truth is the intersection over union (IoU) or Jaccard index. It is defined, as the name suggests, as

$$\text{IoU}(A, B) = \frac{|A \cap B|}{|A \cup B|}, \quad (1)$$

where A is the ground truth and B is the prediction. The top panel of Figure 4 shows the IoU for the validation set reaching $\sim 85\%$. The bottom panel of Figure 4 shows that the root mean squared (RMS) error of the regression to predict spike characteristics decreases with each epoch. Comparing the mean RMS for all characteristics of 0.0736 to the mean of the true characteristics gives an approximate error of 23% for predicted characteristics.

It is more interesting, and perhaps more informative, to take a detailed look at the model outputs in Figure 5. Here we show a batch of 32 samples from the validation set and overlay the centre of the spike (red circle/orange triangle), a bounding box described by the total duration and spectral width, and the drift rate. We also overlay the predicted segmentation masks in white. As we expected from the histograms in Figure 6, the predicted characteristics mostly match well with the ground truth but not for every sample. In particular, any time there is more than one spike in the tile, the predicted location is often somewhere between the two spikes. This has likely occurred because the ground truth segmentation masks and spike characteristics assume only one spike is present in each tile thus the model hasn’t learned how to handle multiple spikes. We discuss possible methods to correct

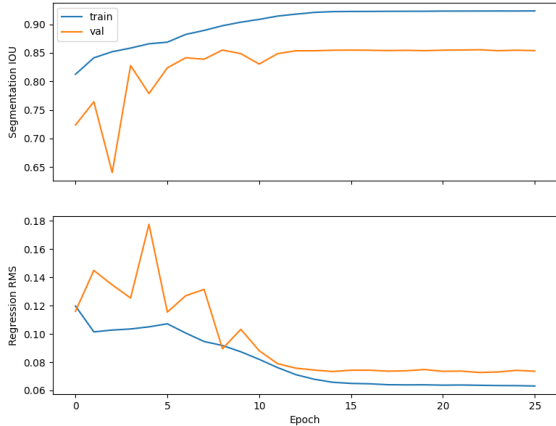


FIG. 4.— The evolution of model metrics with increasing training epoch. Panel a shows the intersection over union for the training (blue) and validation (orange) sets. Similarly, the root mean squared error is shown in the bottom panel.

for this in the next section.

The predicted segmentation masks are output as polygons in the TFCat (Cecconi et al. 2023) format. A selection of these are currently available at <https://doi.org/10.25935/m5cq-f460>. This can be queried using the Astronomical Data Query Language (ADQL, Yasuda et al. 2004). The scripts to prepare the data and train the model are available at <https://gitlab.obspm.fr/pmurphy/spikenet>.

4. DISCUSSION

Machine learning methods are being applied to solar radio data in a growing number of use cases, as outlined in Section 1. We have found success with our CNN in segmenting and characterising solar radio spikes. The high IoU scores obtained during training indicate an accurate recreation of the ground truth segmentation masks. Although Figure 5 shows a good overall agreement between the ground truth and predicted and characteristics the model still gives a mean RMS error of $\sim 23\%$. In Figure 6 we investigate the effect of this error on the model’s predictions by comparing the histogram of each spike characteristics for the ground truth (blue) and predicted value (orange) of the validation set. The distribution of predicted central times and frequencies are similar to the ground truths and the predicted durations follow the same overall shape as the ground truths. The distributions for spectral width and drift rate however, do not fully agree with those of the ground truth. From this, we infer that the dominant source of error is related to the predicted spectral width.

The histograms of the ground truth characteristics can also guide us to where the dataset selection for our model may need improvement. For example, we see peaks in the ground truth central time and frequencies at 0, 0.5 and 1. The peaks at 0 and 1 are likely a result of samples where the central point lies outside of the tile while the central peak is due to the validation set including a sample of every spike in the centre of the tile as well as any samples that have undergone a very small shift during augmentation described in Section 2. The shape of the central time distribution can be attributed to the fact that the random shift applied during augmentation was based on the duration of the spike burst. The frequency

shifts, on the other hand, were a result of a uniform random distribution. Perhaps the accuracy of the regression would improve if the time shifts were drawn from a uniform distribution as well as it will be more used to seeing spikes near the edges of a tile.

As mentioned above, the characterisation of spikes is noticeably less accurate when multiple spikes are present in a tile. Performing semantic segmentation for distinct objects is known as instance segmentation. One network that performs well in these kinds of tasks is called Mask R-CNN (Region based CNN; He et al. 2018). Mask R-CNN utilises a region proposal network to locate areas in each image which likely includes the desired object. We borrow the idea of objectness from Mask R-CNN and implement it in our network as a measure of how sure the model is that it “sees” a spike. Hou et al. (2020) have adapted the region proposal network of Faster R-CNN (Ren et al. 2016) to locate solar radio spikes in the 150 MHz–500 MHz range. It would be a significant undertaking for the authors of this work to implement a similar network in the 10 MHz–90 MHz range, particularly because there is no reference in Hou et al. (2020) as to where the code to train their network exists.

4.1. Application to unseen, unlabelled data

The ultimate test for our machine learning model is to apply it to a dynamic spectrum that does not exist in the training, validation or test dataset and thus does not have any ground truth masks or spike characteristics. This is instructive to qualitatively determine the robustness of the model to unseen data and inform how the model can be further developed and put into practical use. Figure 7 shows a Stokes I dynamic spectrum between 11:03:59 and 11:04:05 on 2022-02-02 in the frequency range 47.5 MHz–50 MHz over 4 seconds. Overlaid in transparent white are the segmentation masks as well as the central locations, marked by orange triangles, duration and spectral width, denoted by orange bounding boxes, and drift rate, orange lines.

We see that the core of the spikes are well captured by the predicted segmentation masks but the model fails to predict most of the spikes duration. Similarly, the central point and bounding boxes do not cover the full spike. We attribute this to the spikes not falling fully into the tiles that the original dynamic spectrum was divided into. Similar to the problem of multiple spikes appearing in one tile, this behaviour could be correct by using a region proposal network. Nevertheless, the performance on unseen data is promising and, if used with appropriate caution, could allow for the analysis of solar radio spikes without having to manually find them.

5. CONCLUSION

We have trained a machine learning model on $\sim 100\,000$ samples of solar radio spikes using an Nvidia Tesla T4 16 GB GPU. The training took 2.2 hours in total and resulted in a high $\sim 80\%$ IoU and relatively low < 0.1 RMS error. The model successfully produced segmentation masks for radio spikes in the validation set and predicted the spike characteristics, i.e. location in time and frequency, duration, spectral width and drift rate. The model does have some shortcomings when applied to unseen data. For example, a certain number of spike

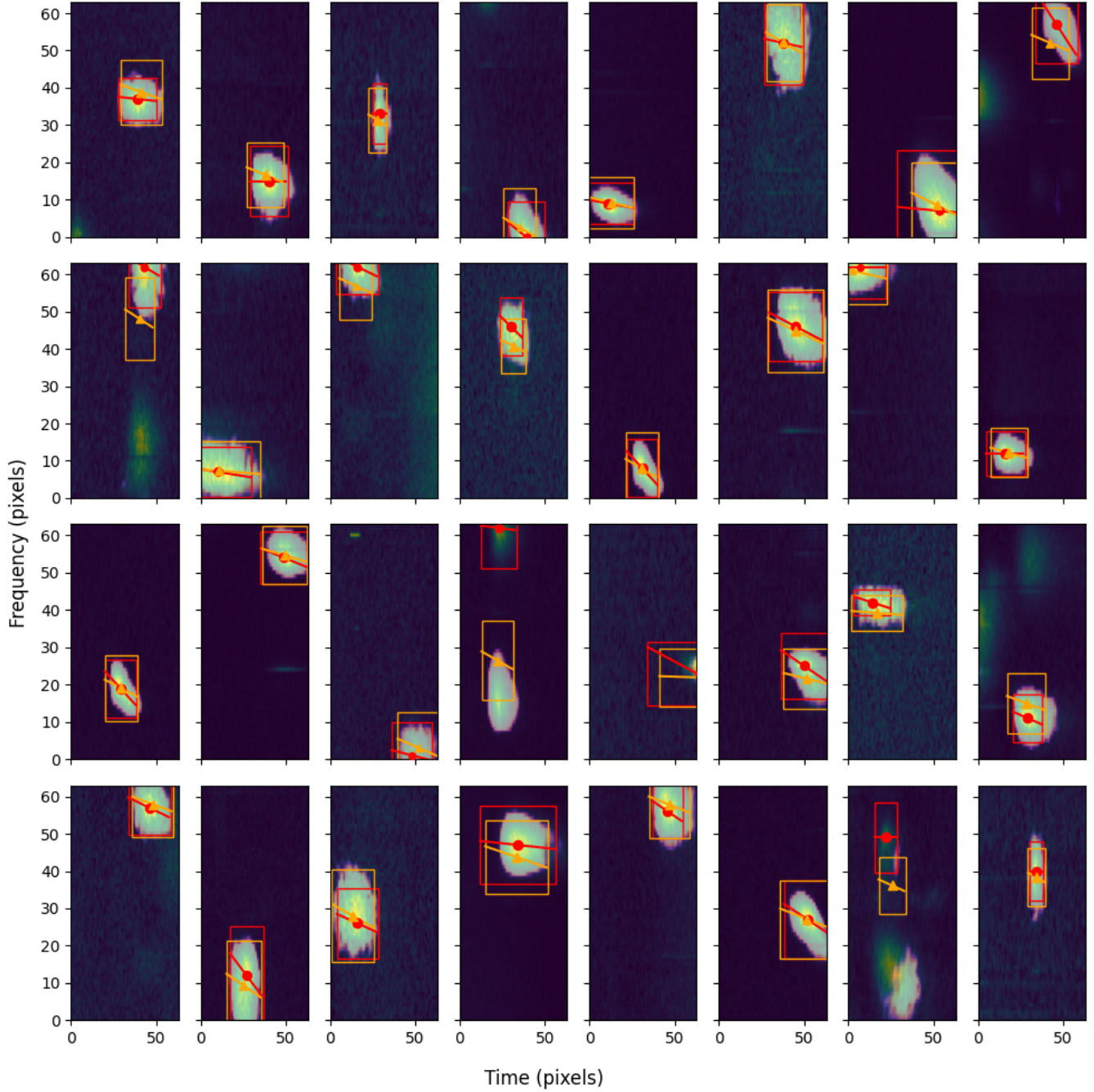


FIG. 5.— A comparison of the model performance compared to the ground truth. Each panel represents a sample from the validation set and depicts the dynamic spectrum of the spike burst with an overlay of the predicted segmentation mask. The ground truth spike position is given by a red circle while an orange triangle denotes the predicted position. The bounding box, determined by the spike duration and spectral width, as well as its drift rate is also plotted in red for ground truth and orange for model prediction. The axes of the panels denote time (x axis) and frequency (y axis) in pixels.

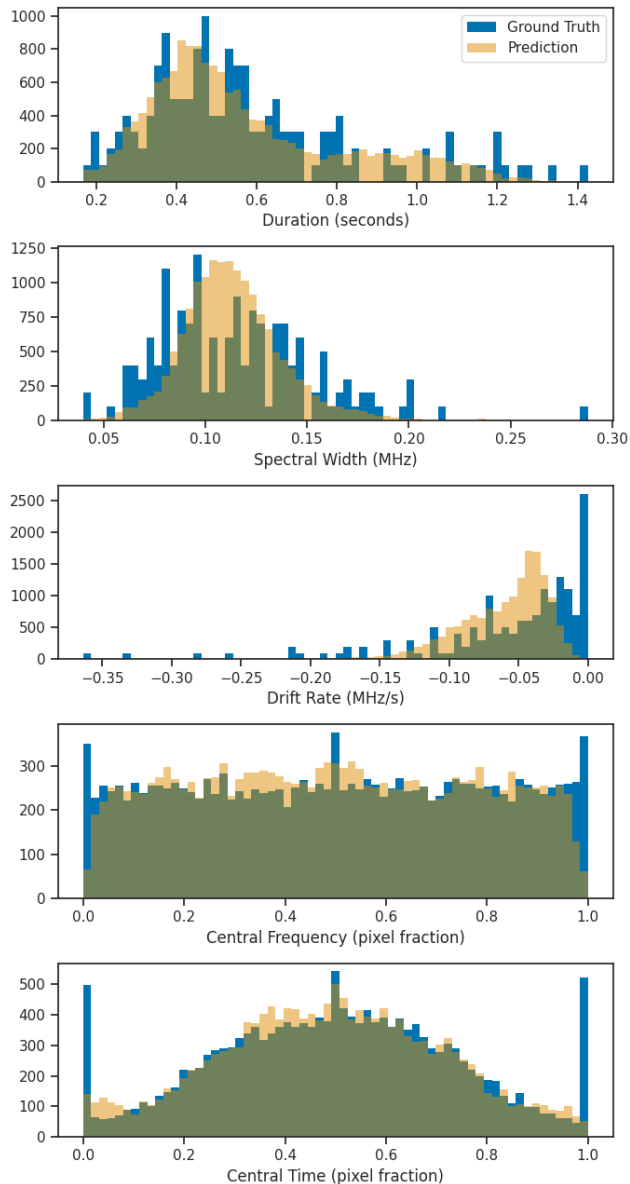


FIG. 6.— A comparison of the histograms for the ground truth and predicted spike characteristics. The time and frequency position are determined as a fraction of the width (height) of the dynamic spectrum in terms of pixels. We see an abundance of spikes in the validation set have centres outside the width of the tile, resulting in the peaks at 0 and 1. There is also always 1 sample with the spike directly centred in the tile which may explain the peak at 0.5 in time and frequency location.

masks are abruptly cut off and the model was less successful in determining the position of spikes when more than one spike appeared in the 64×64 pixel dynamic spectrum sample. This is due to the original dynamic spectrum being divided into 64×64 tiles thus, any burst not fully in one tile is unlikely to be fully recovered in the mask. A region proposal network to first determine where spikes are likely to be is the best solution for this.

With the current data rate generation of LOFAR and NenuFAR, not to mention the ongoing development of next generation low frequency arrays such as the Square Kilometre Array (SKA) which are set to produce petabytes of data, the ability to automatically de-

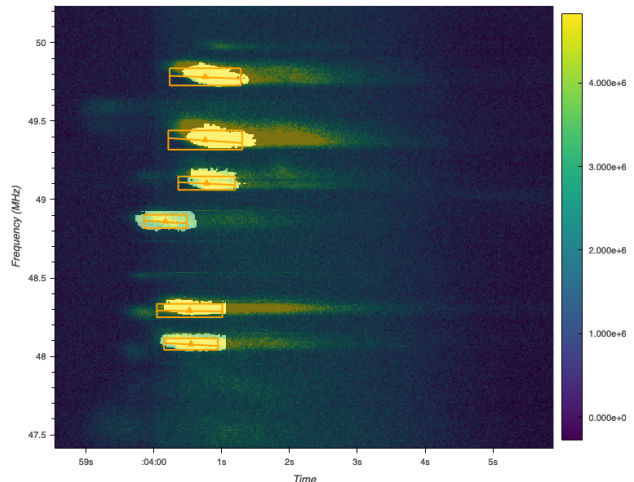


FIG. 7.— Performance of the model on unseen data. The panel shows a Stokes I dynamic spectrum at the same frequency and temporal resolution as Figure 1. Overlaid on the dynamic spectrum is the segmentation mask detected by the CNN. Orange triangles denote the predicted centre position of the spikes while their duration, spectral width, and drift rate are shown as the orange bounding boxes and lines respectively. The segmentation masks capture the core of each spike well but fail when a spike extends across a tile.

tect fine scale features in solar radio observation has become increasingly important. This task is ideally suited to machine learning methods, the latest of which we have presented here. Our CNN drastically reduces the time to locate and identify solar radio spikes from minutes to less than a second per spike. While the accuracy of the predicted spike characteristics could be improved, the use of our CNN trivialises finding spikes hidden in solar radio data and allows for more time to be dedicated to their analysis using whichever methods a particular researcher desires. Future development of SpikeNet could see it used in near real-time on NenuFAR solar observations to dynamically determine an appropriate time resolution. For example, should SpikeNet detect a certain number of spikes over a given time period, keep this data at its native resolution. Otherwise, the time resolution should be rebinned. This will alleviate pressure on the fixed storage space allocated to solar observations with NenuFAR. The expected data rates of the SKA (Scaife 2020) mean that dynamically adapting the recording resolution could drastically reduce the storage required for solar observations in the future.

ACKNOWLEDGEMENTS

The authors would like to thank Thibault Peccoux for his work in identifying and characterising solar radio spikes. This work has been supported by the Europlanet-2024 Research Infrastructure project, which received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 871149 and a grant from Paris Observatory science council under the MINERVA project. We acknowledge the use of the Nançay Data Center (CDN – Centre de Données de Nançay) facility. The CDN is hosted by the Observatoire Radioastronomique de Nançay (ORN) in partnership with the Observatoire de Paris, the Université d’Orléans, the Observatoire des Sciences de l’Université d’Orléans (OSUC) and the French Centre National de la Recherche Scientifique (CNRS). The CDN is sup-

ported by the Région Centre-Val de Loire (département du Cher). The ORN is operated by the Observatoire de Paris, associated with the CNRS. The research lead-

ing to these results has received funding from the European Union's Horizon Europe Programme under the EXTRACT Project, grant agreement n° 101093110.

REFERENCES

- Cecconi, B., Louis, C. K., Bonnin, X., Loh, A., & Taylor, M. B. 2023, *Frontiers in Astronomy and Space Sciences*, 9
- Clarke, B. P., Morosan, D. E., Gallagher, P. T., et al. 2019, *Astronomy & Astrophysics*, 622, A204, arXiv: 1901.07424
- Clarkson, D. L., Kontar, E. P., Gordovskyy, M., Chrysaphi, N., & Vilmer, N. 2021, *The Astrophysical Journal*, 917, L32, aDS Bibcode: 2021ApJ...917L..32C
- Clarkson, D. L., Kontar, E. P., Vilmer, N., et al. 2023, *The Astrophysical Journal*, 946, 33, aDS Bibcode: 2023ApJ...946...33C
- de La Noe, J. & Boisshot, A. 1972, *Astronomy and Astrophysics*, 20, 55
- Ellis, G. 1969, *Australian Journal of Physics*, 22, 177, publisher: CSIRO PUBLISHING
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. 2018, *Mask R-CNN*, arXiv:1703.06870 [cs]
- Hou, Y. C., Zhang, Q. M., Feng, S. W., et al. 2020, *Solar Physics*, 295, 1, publisher: Springer Nature
- Iakubovskii, P. 2019, *Segmentation models*
- Kontar, E. P., Yu, S., Kuznetsov, A. A., et al. 2017, *Nature Communications*, 8, 1515, publisher: Nature Publishing Group
- Lv, P. R., Hou, Y. C., Feng, S. W., Du, Q. F., & Tan, C. M. 2023, *Astrophysics and Space Science*, 368, 14
- Melnik, V. N., Konovalenko, A. A., Rucker, H. O., et al. 2010, *Solar Physics*, 264, 103, publisher: Springer Netherlands
- Morosan, D. E., Gallagher, P. T., Zucca, P., et al. 2015, *Astronomy & Astrophysics*, 580, A65, publisher: EDP Sciences
- Murphy, P. C., Cecconi, B., Briand, C., & Aicardi, S. 2023, accepted: 2023-10-23T10:58:13Z
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. 2016, *You Only Look Once: Unified, Real-Time Object Detection*, arXiv:1506.02640 [cs]
- Reid, H. A. S. & Kontar, E. P. 2021, *Nature Astronomy*, 5, 796, arXiv: 2103.08424 Publisher: Nature Research
- Ren, S., He, K., Girshick, R., & Sun, J. 2016, *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*, arXiv:1506.01497 [cs]
- Ronneberger, O., Fischer, P., & Brox, T. 2015, *IEEE Access*, 9, 16591, arXiv: 1505.04597 Publisher: Institute of Electrical and Electronics Engineers Inc.
- Scaife, A. M. M. 2020, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 378, publisher: Royal Society Publishing
- Scully, J., Flynn, R., Carley, E., Gallagher, P., & Daly, M. 2021, in *2021 32nd Irish Signals and Systems Conference (ISSC) (IEEE)*, 1–6, arXiv: 2105.13387
- Scully, J., Flynn, R., Gallagher, P. T., Carley, E. P., & Daly, M. 2023, *Astronomy and Astrophysics*, 674, A218, aDS Bibcode: 2023A&A...674A.218S
- Sharykin, I. N., Kontar, E. P., & Kuznetsov, A. A. 2018, *Solar Physics*, 293, 115, publisher: Springer Netherlands
- Tarnstrom, G. L. & Philip, K. W. 1972, *Astronomy and Astrophysics*, 16, 21, aDS Bibcode: 1972A&A....16...21T
- van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, *Astronomy & Astrophysics*, 556, A2, arXiv: 1305.3550
- Yasuda, N., Mizumoto, Y., Ohishi, M., et al. 2004, 314, 293, conference Name: *Astronomical Data Analysis Software and Systems (ADASS) XIII* aDS Bibcode: 2004ASPC..314..293Y
- Zarka, P., Girard, J. N., Tagger, M., & Denis, L. 2012, *LSS/NenuFAR: The LOFAR Super Station project in Nançay*, conference Name: *SF2A-2012: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics* Pages: 687-694 aDS Bibcode: 2012sf2a.conf..687Z