

Classifying complex Faraday spectra with convolutional neural networks

Shea Brown,¹★ Brandon Bergerud,¹ Allison Costa,¹ B. M. Gaensler^{ib},² Jacob Isbell,¹
Daniel LaRocca,¹ Ray Norris^{ib},³ Cormac Purcell,⁴ Lawrence Rudnick^{ib},⁵ and
Xiaohui Sun⁶

¹*Department of Physics and Astronomy, The University of Iowa, Iowa City, IA 52245, USA*

²*Dunlap Institute for Astronomy and Astrophysics, The University of Toronto, Toronto, ON M5S 3H4, Canada*

³*Western Sydney University, Locked Bag 1797, Penrith South, NSW 2751, Australia*

⁴*Research Centre for Astronomy, Astrophysics, and Astrophotonics, Macquarie University, NSW 2109, Australia*

⁵*Minnesota Institute for Astrophysics, University of Minnesota, 116 Church Street SE, Minneapolis, MN 55455, USA*

⁶*Department of Astronomy, Yunnan University, and Key Laboratory of Astroparticle Physics of Yunnan Province, Kunming 650091, China*

Accepted 2018 October 24. Received 2018 October 24; in original form 2017 November 8

ABSTRACT

Advances in radio spectropolarimetry offer the possibility to disentangle complex regions where relativistic and thermal plasmas mix in the interstellar and intergalactic media. Recent work has shown that apparently simple Faraday rotation measure spectra can be generated by complex sources. This is true even when the distribution of rotation measures in the complex source greatly exceeds the errors associated with a single component fit to the peak of the Faraday spectrum. We present a convolutional neural network that can differentiate between simple Faraday thin spectra and those that contain multiple (two) Faraday thin sources. We demonstrate that this network, trained for the upcoming Polarization Sky Survey of the Universe’s Magnetism early science observations, can identify two component sources 99 per cent of the time, provided that the sources are separated in Faraday depth by > 10 per cent of the full width at half-maximum of the Faraday point spread function, the polarized flux ratio of the sources is > 0.1 , and that the signal-to-noise ratio (S/N) of the primary component is > 5 . With this S/N cut-off, the false positive rate (simple sources misclassified as complex) is < 0.3 per cent. Work is ongoing to include Faraday thick sources in the training and testing of the convolutional neural network.

Key words: methods: analytical – Physical data and processes: polarization – methods: data analysis – methods: numerical – methods: statistical.

1 INTRODUCTION

Faraday rotation of linearly polarized radio emission gives unique insight into the properties of the intervening magnetoionic medium. Measurements of rotation measures (RMs) of background polarized radio sources probe astrophysical magnetic fields in a variety of environments like the solar corona (Kooi et al. 2017), H II regions (Harvey-Smith, Madsen & Gaensler 2011), the interstellar medium of the Milky Way (Han et al. 1997; Sun & Reich 2010; Wolleben et al. 2010; Pshirkov et al. 2011; Van Eck et al. 2011; Jansson & Farrar 2012; Oppermann et al. 2012, 2015; Akahori et al. 2013), external galaxies (Han, Beck & Berkhuijsen 1998; Gaensler et al. 2005; Mao et al. 2012; Bernet, Miniati & Lilly 2013), and the intracluster (Enßlin & Vogt 2003; Murgia et al. 2004; Bonafede et al. 2010, 2013) and intergalactic (Akahori & Ryu 2011; Akahori et al. 2014;

Vacca et al. 2016) medium. Traditionally, Faraday rotation has been measured by fitting the change in polarization angle χ as a function of wavelength squared (λ^2), parametrized by the RM defined by

$$\chi(\lambda^2) = \chi_0 + \text{RM}\lambda^2, \quad (1)$$

where χ_0 is the intrinsic polarization angle of the radio emission. This linearity with λ^2 is only valid for the case of a single synchrotron emitting source with an intervening cloud of magnetized thermal plasma, i.e. only when the rotating and emitting medium are completely separated. If there is mixing between the rotating and emitting plasmas, or there is more than one source–screen pair within the same beam, χ will in general not be a linear function of λ^2 . Other limitations that arise with this method include the so-called $n\pi$ ambiguity, where multiple solutions can fit a straight line

* E-mail: shea-brown@uiowa.edu

in λ^2 space,¹ and bandpass depolarization, where the polarization angle of the emission rotates so significantly across the bandwidth that partial cancellation occurs. The wide-band spectral capability of modern radio telescopes has allowed the use of RM synthesis (Burn 1966; Brentjens & de Bruyn 2005; Sun et al. 2015), which can address problems of bandwidth depolarization and multiple emitting/rotating regions along the line of sight (or within the same beam). RM synthesis inverts the complex polarization spectrum $P(\lambda^2) = Q(\lambda^2) + iU(\lambda^2)$ into a Faraday spectrum:

$$F(\phi) = K \int_{-\infty}^{+\infty} P(\lambda^2) e^{-2i\phi(\lambda^2 - \lambda_0^2)} d\lambda^2, \quad (2)$$

where K is a constant, and ϕ is the ‘Faraday depth’ of the emission, given by

$$\phi(r) = K' \int_r^0 n_e \mathbf{B} \cdot d\mathbf{l}, \quad (3)$$

where K' is a constant, n_e is the electron density, \mathbf{B} is the magnetic field vector, and $d\mathbf{l}$ is an infinitesimal distance along the line of sight from the synchrotron source located at r . In the simple case described in equation (1), $\phi = \text{RM}$. The development of deconvolution algorithms for the Faraday spectra (e.g. Heald 2009; Bell et al. 2013) has further improved the ability of RM synthesis to reveal multiple sources along the line of sight.

Farnsworth, Rudnick & Brown (2011) described an ambiguity in RM derived from $\chi(\lambda^2)$ fitting and RM synthesis where two Faraday components model can produce a consistent single component solution that is neither of the input components nor their mean. This ambiguity can lead to an error in ϕ derived from these methods that is greater than what one would naively calculate from the uncertainty in fitting the peak of $F(\phi)$. In the era of large radio surveys meant to produce grids of background RMs for the archival science, there is a need to distinguish between ‘simple’ foreground screens and more complex sources.

One such survey is the Polarization Sky Survey of the Universe’s Magnetism (POSSUM; Gaensler et al. 2010), which will be conducted with the Australian Square Kilometre Array Pathfinder (ASKAP; Johnston et al. 2008) and will measure more than one million polarized sources in the frequency range of 1130–1430 MHz over 75 per cent of the sky. An Early Science survey, which is being conducted as part of ASKAP’s science commissioning observations, will make use of only 12 antennae of ASKAP, but the frequency coverage will be extended to 700–1800 MHz. The extended frequency coverage of the Early Science survey is ideal for identifying and investigating complex Faraday spectra, provided that these spectra can be identified in an automated way. There are a variety of ways in which a Faraday spectrum can deviate from a single source with a foreground Faraday screen (called Faraday thin; Brentjens & de Bruyn 2005). It can be Faraday thick (caused by significant mixing of Faraday rotating and emitting plasma), have multiple Faraday thin components, or there can be external/internal Faraday dispersion (modulation due to rapidly changing Faraday rotating cells along the line of sight, or within a single beam).

The initial data release of the POSSUM survey will be a catalogue of sources with *simple* Faraday spectra and their associated properties. Simple spectra come from sources with a single Faraday rotating screen in front of them, i.e. their polarization angle

would obey equation (1) for all λ^2 values. This POSSUM Polarization Catalogue (PPC) will not include complex sources, which are any sources that are not simple as defined above, so the pipeline producing the catalogue must have an efficient and effective way of determining/testing the complexity of a source. This test needs to be able to (1) determine whether a Faraday spectrum is complex in general, including cases where there are more than one well-separated Faraday sources (peaks) in the spectrum; (2) determine whether a given peak is Faraday thin or not; and (3) provide some way of assessing how sure we are of the resulting classification. Initial work has shown that the second moment of the clean components resulting from a Faraday cleaning procedure can provide some discriminating power (e.g. Brown 2011; Anderson et al. 2015), while more recent work has focused on the model fitting the polarization spectrum and examining the statistically more likely model in a Bayesian sense (e.g. O’Sullivan et al. 2012, 2017; Sun et al. 2015; Purcell & West 2017).

To this end, we present the construction of a convolutional neural network (CNN) that can automatically classify a Faraday spectrum as either simple or complex. While it does not eliminate the ambiguity presented in Farnsworth et al. (2011), the network does quantify the range of parameters over which we have a high certainty that the classification is correct. In Section 2 we describe the construction and training of the CNN, in Section 3 we outline the testing of the network on simulated data, and in Section 4 we summarize the limitations of the metric and discuss future improvements.

2 CONVOLUTIONAL NEURAL NETWORK

A few of the major difficulties for developing a test for complexity are (a) the ad hoc nature of the initial choice of a metric, (b) the significant work required to find a metric threshold appropriate for the science goals, and (c) estimating the uncertainty in the accuracy of the metric/threshold combination. For these reasons, we have explored CNNs (LeCun & Bengio 1995; Krizhevsky, Sutskever & Hinton 2012) as potential classifiers.

2.1 Background

Traditional neural networks (Cybenko 1989), when used as classifiers, are functions that map data in some space (like an image, a spectrum, or just a list of features) into a probability that the data represent some class of object. Typically the function is approximated by a layered ‘network’ of matrix multiplications interspersed with non-linear scalar functions. The essence of machine learning is that, as we do not know what values these multidimensional matrices should have ab initio, we must train the network by feeding it example data where the classification is known (often called the ‘ground truth’ in literature) and adjusting the elements of the matrices to get closer to the correct answer. This is usually accomplished through *gradient descent*, where the values are adjusted to minimize a loss function (e.g. least-squares difference between the true and predicted answers).

In our case the data are the complex Faraday spectrum from equation (2), and we wish to find a mapping from any given spectrum to the probability that it is complex. As the sources can show up anywhere within the spectrum, the function should be translationally invariant, which can be introduced through the use of convolutions. Instead of having each layer be a large matrix of weights, a CNN is made up of layers of convolutions, where the weights are a convolutional kernel that is shared across the spectrum during the convolution. Instead of having just one kernel per layer, one

¹This can be addressed with spatial continuity arguments in the image plane (Dolag, Vogt & Enßlin 2005).

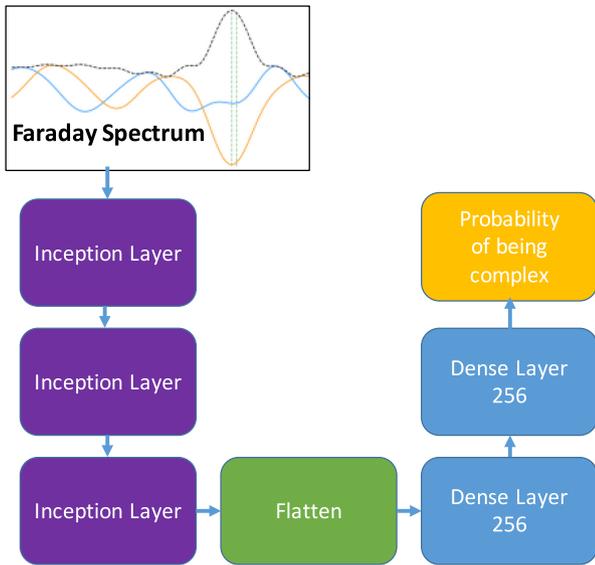


Figure 1. Three inception layers convolutional neural network (CNN) classifier. Each dense layer has an additional dropout (0.5) and activation (ReLU) layer within it.

can apply many kernels to the spectrum and concatenate the results at the end. The single spectrum is now transformed into multiple representations, some of which may be better suited to detect the signal of interest.

A CNN is thus a function that applies multiple convolutions to the spectrum in each ‘layer’, with a non-linear ‘activation function’ applied between layers. A simple, effective, and popular choice is the rectified linear unit (or ReLU; Nair & Hinton 2010), which functionally is just $\text{ReLU}(x) = \max(0, x)$, effectively clipping negative values. The stacking of many such layers is where the ‘deep’ in deep machine learning comes from.

CNNs are ideal for addressing the problems associated with complexity, as the training process will automatically find the series of convolutional kernels that are optimal for classifying the spectra. While model fitting requires large searches over parameter space each time a spectrum is analysed, a CNN samples the space only during training, and requires only a straightforward and efficient ‘feed-forward’ network of matrix operations to classify the sources.

One issue that can arise in using a CNN on our polarization spectra is that sources (or clusters of sources) may have varying widths in Faraday space, so fixing a convolutional kernel size may bias the classifier. To address this, we used an ‘inception’ network initially developed by the GoogLeNet team (Szegedy et al. 2014), which applies a series of convolutions in parallel, each with a different kernel size, and concatenates the results to form the next layer. This further allows the network to search for features in the data that distinguish complex sources. In essence, the network will find the best metric to use during the training process, eliminating the need to choose one at the outset.

2.2 Network architecture

The basic building block for our network architecture (Fig. 1) is an ‘inception layer’ (Fig. 2). Each inception layer will take an input spectrum and apply five parallel channels of convolutions, using four different kernel widths, with kernel sizes of 1, 3, 5, and 23 channels (in Faraday space each channel is 1 rad m^{-2}). The one-

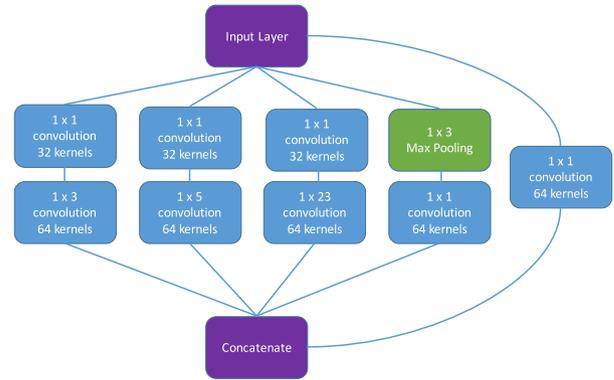


Figure 2. A zoom-in of an inception layer. Each of the convolutional layers (in blue) has additional batch normalization and activation (ReLU) layers within it.

channel convolution (often called a ‘ 1×1 ’ convolution; Szegedy et al. 2014) is not a channel convolution at all, but a matrix multiplication in the depth direction, allowing mixing between the real and imaginary parts of the spectrum in the first application, and relationships between different kernels in subsequent layers. We chose the 1, 3, and 5 channel width convolutions because they appear in the original network of Szegedy et al. (2014), and the 23 channel convolution so as to fully sample the full width at half-maximum (FWHM) of a Faraday thin source given our frequency coverage, which is $\text{FWHM} = 22 \text{ rad m}^{-2}$. In Fig. 2, each of these convolutional layers are labelled as blue boxes, and come with two other operations packaged inside. After the convolution is computed, the outputs are normalized (called ‘batch normalization’, which stabilizes the gradient descent during training), and then a ReLU activation function is applied. The layer in the green box, called a ‘maxpooling’ layer (Boureau, Ponce & LeCun 2010), reduces the number of channels in the spectrum by taking the maximum value in a three-channel wide kernel, then removing half the channels. The full network is shown in Fig. 1, and the individual inception layers are shown in Fig. 2. The ‘flatten’ layer will take the deep network of features constructed by the inception layers and project it into a single vector of features to be passed to the dense layers. These dense layers are traditional artificial neural networks (Cybenko 1989), where the weights are simple matrix multiplications. To avoid overfitting during the training process, half of the weights (selected at random) in the dense layers are set to zero during each training iteration (called ‘dropout’; Srivastava et al. 2014). Source code can be found on GITHUB.²

Using this inception layer as our base, we considered and tested several CNNs with different number of layers. By monitoring our loss function as computed on an unseen validation set of data (see Section 3 below), we found that the best network has three convolutional inception layers, along with two fully connected (dense) layers (Fig. 2).

3 TRAINING

In order to train the proposed complexity classifier, we simulated both simple and complex spectra using a realistic observational model. Since our current purpose is to develop a metric for the

²https://github.com/sheabrown/faraday_complexity/blob/master/final/How2Guide.ipynb

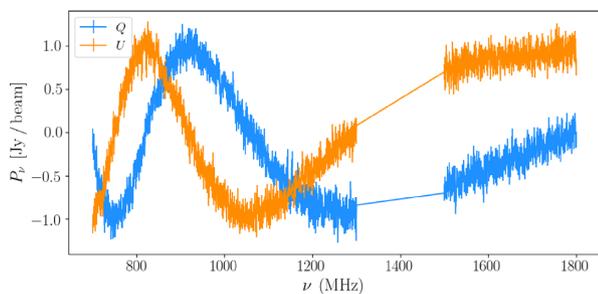


Figure 3. An example of a polarized spectrum sampled with the POSSUM Early Science frequency coverage used in the training data.

Table 1. Two-component parameter space.

Parameter	Symbol	Range
Amplitude 1	P_1	1
Amplitude 2	P_2	[0, 1]
Faraday depth {1, 2}	$\phi_{\{1, 2\}}$	[-50, +50]
Polarization angle {1, 2}	$\chi_{\{1, 2\}}$	[0, $+\pi$]
Noise/channel	σ	[0, 0.333]

POSSUM Early Science observations, we have used the proposed frequency coverage of the ASKAP 12 Early Science survey (700–1300 and 1500–1800 MHz), with a total of 900, 1-MHz channels (see Fig. 3). The current work focuses only on two-component models (with both components being Faraday thin), which are believed to be the dominant source of complex spectra (e.g. Anderson et al. 2015; O’Sullivan et al. 2017). The two emitting regions can have different polarized amplitudes (P_1 and P_2), as well as different intrinsic polarization angles (χ_1 and χ_2) and foreground Faraday depths (ϕ_1 and ϕ_2). The combined polarized spectrum is given by

$$P(\lambda^2) = P_1 e^{2i(\chi_1 + \phi_1 \lambda^2)} + P_2 e^{2i(\chi_2 + \phi_2 \lambda^2)}. \quad (4)$$

We also simulate simple sources using equation (4), but with $P_2 = 0$. To train the network, we generated 130 000 sources (100 000 training set and 30 000 for validation), roughly half of which were complex (two-component) sources, and the other half were simple.³ Table 1 shows the parameter space that was sampled at random from a uniform distribution. We used the simplifying assumption that both the noise and total intensity have no spectral dependence. As a first step, we chose to train the network on the Faraday spectra $F(\phi)$ only, as it will be computed as part of the POSSUM pipeline, but in theory the network can be trained using the polarization spectrum directly as well. For each source, a polarization spectrum was created first, and then a Faraday spectrum was created using the standard inversion formula of Brentjens & de Bruyn (2005):

$$F(\phi) \approx K \sum_{i=1}^N P_i e^{-2i\phi(\lambda_i^2 - \lambda_0^2)}, \quad (5)$$

where $K = 1/N_{\text{channels}}$, λ_0^2 is the average λ^2 of the channels, and P_i is the measured complex polarization in channel i . Fig. 3 shows an example of a complex polarized $P(\nu)$, and Fig. 4 shows a selection of Faraday $F(\phi)$ spectra for complex sources in the training set.

³Each time the data simulator created a spectrum, there was a 50 per cent probability that it would be complex. The probability is an adjustable parameter in the source code.

The network was trained using batch stochastic gradient descent (SGD; Duda Hart & Stork 2012) on the training set of 100 000 sources, with 30 000 sources withheld for cross-validation during the training. The SGD minimizes a loss function, and we used the binary cross-entropy (Hinton & Salakhutdinov 2006) for this purpose. One epoch of training consists of adjusting the network weights with SGD using 100 000 sources, then checking how well it does on the 30 000 validation sources. The training lasted for 100 epochs (taking roughly 3 h on a 3.30 GHz CPU with 32 GB of RAM), though no improvement in the loss function of the validation set was found after 55 epochs. The weights found on epoch 55 were saved and used for testing below.

4 RESULTS AND DISCUSSION

The trained network was then applied to 100 000 new sources randomly generated using the same parameter space as the training set. The output for each source is a value p between 0 and 1 that can be thought of as the probability that the source is complex. A perfect network would return a 1 for every complex source and a 0 for every simple source. Fig. 5 shows a histogram of p for the 100 000 test sources. The distribution is bimodal, indicating that the network was confident about the classification of most sources. The larger spread in the $p = 0$ peak may indicate that the network is picking up on the ambiguity of Farnsworth et al. (2011), something that we investigate below. If we take $p > 0.5$ as the threshold for complexity, we can construct the confusion matrix as shown in Table 2.

The network produces 7.2 per cent false negatives and 3.0 per cent false positives. In order to hunt down the complex sources that are misclassified as simple, we can plot the p for all the complex sources as a function of both the second component’s amplitude P_2 and the absolute separation in the two components’ Faraday depths $\Delta\phi = |\phi_1 - \phi_2|$ (Fig. 6). The majority of false negatives happen at small P_2 and $\Delta\phi$. This is consistent with the results of Farnsworth et al. (2011) and Sun et al. (2015) that point to the difficulty in identifying two component sources when $\Delta\phi < \text{FWHM}$ of the Faraday point spread function. Fig. 7 shows an example of one of the false negatives. The false positive rate is low, and visual inspection of p as a function of σ reveals that these occur mostly at low S/N.

For the purposes of constructing a classifier for large-scale polarization surveys like POSSUM, we would like to exclude the phase space of sources that would likely not make it into the catalogue due to low S/N, as well as sources where the RMs of the two components are close enough to allow probing of a foreground Faraday screen. We therefore searched for the region of phase space in which we can detect >99 per cent of the complex sources, allowing for the false positive rate to adjust appropriately based on the cut-off values. We found that if the minimum S/N of the primary component is 5.0, and restrict our sample to $P_2 > 0.1$, and $\Delta\phi > 2.3 \text{ rad m}^{-2}$ (which is about 10 per cent of the 23 rad m^{-2} FWHM of the Faraday point spread function), the false negatives are reduced to <1 per cent, while the false positives are reduced to <0.3 per cent. Table 3 shows the new confusion matrix with the cut-offs applied to the same simulated data. What these cut-offs mean for the initial POSSUM catalogue (the PPC) is that the network is 99 per cent confident that the Faraday spectrum is simple, with the understanding that a secondary component can be hiding in the above phase space. The probability returned by the network can be recorded for each source, allowing one to flag sources where p is close to the nominal cut-off of $p < 0.5$ for the PPC. We should note that one can trade a higher S/N cut-off

Complex Sources

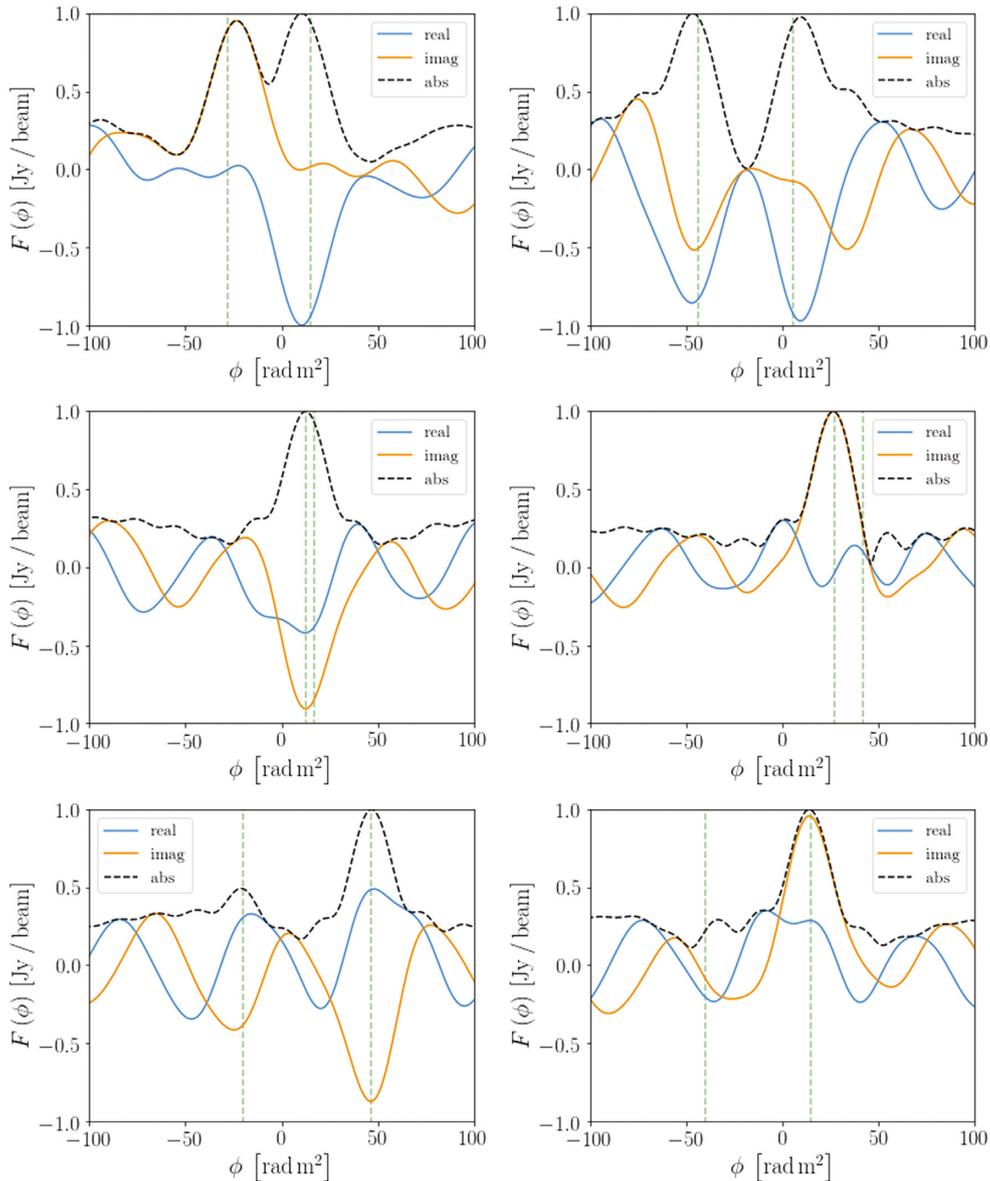


Figure 4. Complex Faraday spectra from the training data (with ϕ_1 and ϕ_2 given by the vertical green dashed lines). No deconvolution (RM clean) was performed.

to allow a narrower $\Delta\phi$ and still reach the 99 per cent, something that might be advantageous depending on the science goal.

5 CONCLUSION AND FUTURE WORK

We have constructed a CNN that is able to distinguish between simple Faraday sources and those that contain two Faraday thin components, demonstrating on simulated POSSUM Early Science data that it can detect 99 per cent of complex sources with <0.3 per cent false positive rate in a realistic and useful region of the source parameter phase space. The training and application of this network for other observational parameters is straightforward, needing only the frequency coverage to be changed. The most obvious future development of the network would include (1) lifting the simplification on the flat spectral index and channel independent noise, (2)

allowing for modified RM synthesis that includes channel weights in equation (2), and (3) the inclusion of Faraday thick and three component sources during training. Including complexity beyond this may prove impractical, as O’Sullivan et al. (2017) were able to fit just about any source using a combination of three Faraday thin components. Given the power of CNNs used in commercial applications, the inclusion of point (3) into the training would also allow for multiple classes beyond a binary simple/complex classification.

The CNN classifier presented here is not specific to the ASKAP telescope, and in theory could be trained on simulated data from any radio telescope that can produce a Faraday spectrum. This method may prove particularly useful for low-frequency dipole antennas, where one of the major problems is to evaluate and remove instrumental polarization. Given sufficiently realistic simulated data, the

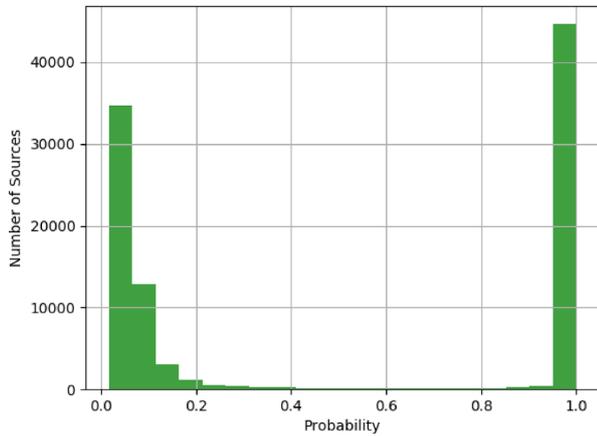


Figure 5. Histogram of probability values for the 100 000 test sources.

Table 2. Confusion matrix: before cut-offs.

Predicted \rightarrow	Simple	Complex
True simple	48 318	1481
True complex	3618	46 583

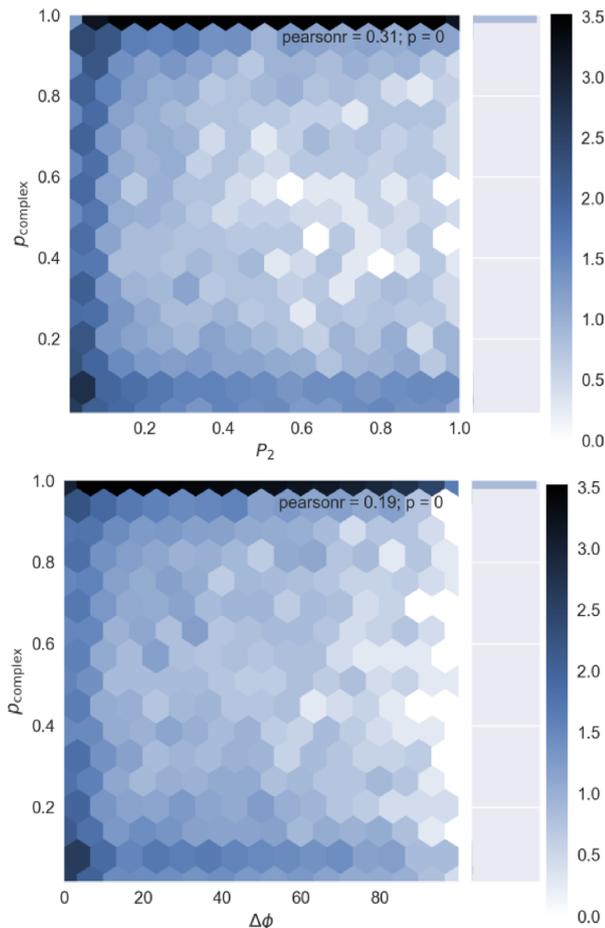


Figure 6. Top: probability output of the network versus the relative flux of the second component for simulated complex sources in the test data set. Colour scale is the log of the number of sources. The black triangles at the top are the high density of p values around 0.99, and the right-hand subplot is a histogram showing that the vast majority of sources were classified as complex. Bottom: like above, but plotted versus the absolute difference in the Faraday depth ($\Delta\phi$) between the two components.

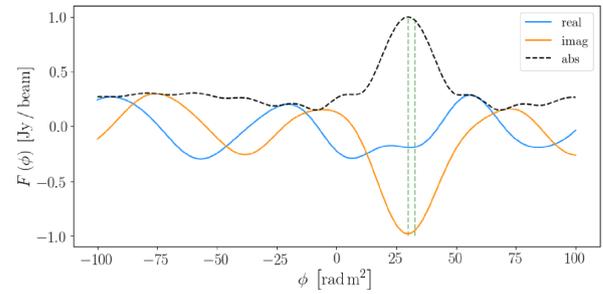


Figure 7. Top: Faraday spectrum of a complex source misidentified as simple by the classifier. The two Faraday depths are labelled with vertical green dashed lines.

Table 3. Confusion matrix: after cut-offs.

Predicted \rightarrow	Simple	Complex
True simple	29 281	69
True complex	247	25 337

algorithm can be trained to distinguish among instrumental and real features.

ACKNOWLEDGEMENTS

The Dunlap Institute is funded through an endowment established by the David Dunlap family and the University of Toronto. BMG acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) through grant RGPIN-2015-05948, and of the Canada Research Chairs program. Partial support for LR comes from NSF grant 17-14205 to the University of Minnesota. XS is supported by the National Natural Science Foundation of China under grant No. 11763008.

REFERENCES

- Akahori T., Ryu D., 2011, *ApJ*, 738, 134
 Akahori T., Ryu D., Kim J., Gaensler B. M., 2013, *ApJ*, 767, 150
 Akahori T., Kumazaki K., Takahashi K., Ryu D., 2014, *PASJ*, 66, 65
 Anderson C. S., Gaensler B. M., Feain I. J., Franzen T. M. O., 2015, *ApJ*, 815, 49
 Bell M. R., Oppermann N., Crai A., Enßlin T. A., 2013, *A&A*, 551, L7
 Bernet M. L., Miniati F., Lilly S. J., 2013, *ApJ*, 772, L28
 Bonafede A., Feretti L., Murgia M., Govoni F., Giovannini G., Dallacasa D., Dolag K., Taylor G. B., 2010, *A&A*, 513, A30
 Bonafede A., Vazza F., Brüggén M., Murgia M., Govoni F., Feretti L., Giovannini G., Ogrean G., 2013, *MNRAS*, 433, 3208
 Boureau Y.-L., Ponce J., LeCun Y., 2010, in Furnkranz J., Joachims T., eds, ICML-10 – Proceedings of the 27th International Conference on Machine Learning. International Machine Learning Society, Madison, WI, p. 111
 Brentjens M. A., de Bruyn A. G., 2005, *A&A*, 441, 1217
 Brown S., 2011, Internal POSSUM Report #9: Assess Complexity of RM Synthesis Spectrum. Technical Report, POSSUM Collaboration
 Burn B. J., 1966, *MNRAS*, 133, 67
 Cybenko G., 1989, *Math. Control Signals Syst.*, 2, 303
 Dolag K., Vogt C., Enßlin T. A., 2005, *MNRAS*, 358, 726
 Duda R. O., Hart P. E., Stork D. G., 2012, *Pattern Classification*. Wiley, New York
 Enßlin T. A., Vogt C., 2003, *A&A*, 401, 835
 Farnsworth D., Rudnick L., Brown S., 2011, *AJ*, 141, 191
 Gaensler B. M., Haverkorn M., Staveley-Smith L., Dickey J. M., McClure-Griffiths N. M., Dickel J. R., Wolleben M., 2005, *Science*, 307, 1610

- Gaensler B. M., Landecker T. L., Taylor A. R. POSSUM Collaboration, 2010, *BAAS*, 42, 515
- Han J. L., Manchester R. N., Berkhuijsen E. M., Beck R., 1997, *A&A*, 322, 98
- Han J. L., Beck R., Berkhuijsen E. M., 1998, *A&A*, 335, 1117
- Harvey-Smith L., Madsen G. J., Gaensler B. M., 2011, *ApJ*, 736, 83
- Heald G., 2009, in Strassmeier K. G., Kosovichev A. G., Beckmann J., eds, *Proc. IAU Symp. Vol. 259, Cosmic Magnetic Fields: From Planets, to Stars and Galaxies*. Cambridge Univ. Press, Cambridge, p. 591
- Hinton G. E., Salakhutdinov R. R., 2006, *Science*, 313, 504
- Jansson R., Farrar G. R., 2012, *ApJ*, 757, 14
- Johnston S. et al., 2008, *Exp. Astron.*, 22, 151
- Kooi J. E., Fischer P. D., Buffo J. J., Spangler S. R., 2017, *Sol. Phys.*, 292, 56
- Krizhevsky A., Sutskever I., Hinton G. E., 2012, *Adv. Neural Inf. Processing Syst.*, 25, 1097
- LeCun Y., Bengio Y., 1995, in Arbib M. A., ed., *The Handbook of Brain Theory and Neural Networks*. MIT Press, Cambridge, MA, p. 3361
- Mao S. A. et al., 2012, *ApJ*, 759, 25
- Murgia M., Govoni F., Feretti L., Giovannini G., Dallacasa D., Fanti R., Taylor G. B., Dolag K., 2004, *A&A*, 424, 429
- Nair V., Hinton G. E., 2010, in Furnkranz J., Joachims T., eds, *ICML-10 – Proceedings of the 27th International Conference on Machine Learning*. International Machine Learning Society, Madison, WI, p. 807
- Oppermann N. et al., 2012, *A&A*, 542, A93
- Oppermann N. et al., 2015, *A&A*, 575, A118
- O’Sullivan S. P. et al., 2012, *MNRAS*, 421, 3300
- O’Sullivan S. P., Purcell C. R., Anderson C. S., Farnes J. S., Sun X. H., Gaensler B. M., 2017, *MNRAS*, 469, 4034
- Pshirkov M. S., Tinyakov P. G., Kronberg P. P., Newton-McGee K. J., 2011, *ApJ*, 738, 192
- Purcell C. R., West J., 2017, Internal POSSUM Report #68: Measuring Faraday Complexity. Technical Report, POSSUM Collaboration
- Srivastava N., Hinton G. E., Krizhevsky A., Sutskever I., Salakhutdinov R., 2014, *J. Machine Learning Res.*, 15, 1929
- Sun X.-H., Reich W., 2010, *Res. Astron. Astrophys.*, 10, 1287
- Sun X. H. et al., 2015, *AJ*, 149, 60
- Szegedy C. et al., 2014, preprint ([arXiv:1409.4842](https://arxiv.org/abs/1409.4842))
- Vacca V. et al., 2016, *A&A*, 591, A13
- Van Eck C. L. et al., 2011, *ApJ*, 728, 97
- Wolleben M., Landecker T. L., Hovey G. J., Messing R., Davison O. S., House N. L., Somaratne K. H. M. S., Tashev I., 2010, *AJ*, 139, 1681

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.