Characterization of exoplanets' light curves with neural networks

Author: Arnau Aguasca Cabot.
Facultat de Física, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain.
Advisor: Francesc Xavier Luri Carrascoso.
Facultat de Física, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain*.

Abstract: A study of the implementation of deep learning using artificial neural networks is undertaken aiming to reduce processing time and human supervision in the characterization of exoplanets' light curves. Firstly, to understand the problem and the techniques involved, a convolutional neural network proposed by Shallue & Vanderburg for the Kepler mission is studied and recreated. Secondly, different alternative neural networks are proposed and compared with the original one, aiming to improve the classification performance.

I. INTRODUCTION

Since the first transiting exoplanet was discovered in 1999 [1], detections of exoplanets through transiting photometry have substantially increased, and the number of exoplanets discovered through photometric observations now represents a total of 76,7% of the discoveries [2].

A transiting exoplanet is seen as a periodic U shape of decreasing luminosity in the light curve of the host star. Nevertheless, there are other sources that can decrease the luminosity of a star and cause identification errors. For instance, variable stars, like δ Scuti variables, which have fluctuations in their luminosity because of variations of their outer layers. Also, binary stars, which represents a 1,3% of the stars observed by Kepler [3] produce dips in the light curves: depending on the distance between the two stars, the transit of one star in front of the other can be seen as a V shape of decreasing luminosity or as a sinusoidal shape [4]. Furthermore, their secondary eclipse can also produce an exoplanet false positive [5]. In addition, instrument noise [6] and stellar variability [7] in the host star can also be sources of error.

In the last decade the number of transit surveys has gone up, and consequently the amount of data to process has significantly increased. In order to speed up the selection of the interesting targets, many algorithms have been developed. Firstly, an initial selection is done through a pipeline, which identifies periodic signal events that may be consistent with an exoplanet; this is known as the threshold crossing event (TCE). Next, a characterization of these TCE is done looking for exoplanet transit signals. One option for this characterization is using deep learning through neural networks.

In this work the use of neural networks in order to characterize exoplanet transits is studied. The aim is, firstly, to reproduce the results of Shallue & Vanderburg [8] with their best convolutional neural network. Secondly, to propose alternative configurations and compare the results aiming to define a more efficient neural network.

In this project, processed data from [8] has been used and can be downloaded here [9]. This data is from Kepler Object of Interest Catalog-Q1-Q17 DR24 [10], but before using these samples, and according to [8], a pre-processing is performed, which consists of phase-fold the observations centring the TCE in the light curve. Two different representations of the light curve are used, one is a global view of the light curve, and the other one is focused on the TCE. Furthermore, the light curves are normalized and the low-frequency variability is removed.

All TCEs in the Kepler Object of Interest Catalog - Q1-Q17 DR24 have been labelled, and used as a training set in *autovetter* (see [11] for more detail); this classification is divided into three different classes: a consistent signal that corresponds to a planet candidate (PC), astrophysical false positives (AFP) and a non-transiting phenomenon (NTP). In this work the same labelled training set as in [11] will be used in order to train and test the neural networks.

The second section of this paper is a brief introduction to what a neural network is and how they can be evaluated. In the third section, the Shallue & Vanderburg's model is described. In the fourth section, the evaluation and comparison of our reproduction of the model made by Shallue & Vanderburg is presented. In the fifth section, alternative models are proposed and evaluated. Finally, the conclusions are presented.

II. NEURAL NETWORKS

An artificial neural network (ANN) is a network composed by nodes (also called neurons), which use weights aiming to recreate the behaviour of biological neurons and the connections between them. ANNs are used in order to solve artificial intelligence problems such as pattern recognition.

Every node is a computational unit that relates the inputs ($\hat{x}$) with the outputs ($\hat{y}$) through an activation function (Φ), as a function of a set of weights ($\hat{w}$) and sometimes a bias value (b), as in equation 1:

$$\hat{y} = \Phi \left\{ \sum_{i=1}^{d} w_i x_i + b \right\} \qquad (1)$$

In this case, the input to the activation function has 2d variables implementing the behaviour of the network.

These activation functions are commonly used to give non-linearity to the neural network. They allow solving non-trivial problems using less nodes [12]. During neural network training, the bias and weights are adjusted in order to minimise the loss function.

The term *deep neural network* is used when the neural network has multiple levels of non-linear operators and when many hidden layers between the input and the output layer are present.

In this paper, a deep convolutional neural network (CNN) has been used. We include here just a concise description of this type of ANN, and we refer the reader to the provided references for details. This kind of neural networks use the convolution in at least one layers. The terms kernel or filter are used for the structural units (like a vector or a matrix) where the weights of the layer are placed. In order to do the convolution, the kernel is placed in every possible position on

---
* Electronic address: xluri@fqa.ub.edu

the input (whether the convolution is done in the input layer) or in the feature map (feature representation in hidden layers) and the scalar product is performed [12].

Shallue & Vanderburg [8] and other works [13][14] proved that the use of convolutional neural networks gives good and efficient results in exoplanet characterization.

Neural networks are trained using cross-validation so as to optimize the network parameters [15]. This means that the dataset is divided into parts: one is used to train the model (training set) while the other is used to validate the results during the process (validation set), and finally additional data is used in order to test the final model (test set). It is common that the size of the training set is larger than the validation set, for instance, 70-30 or 80-20. In this case Shallue & Vanderburg divided their dataset using 80% for training, 10% for validation and finally 10% for testing.

Neural networks make use of randomness in order to explore all the parameter space, reach faster the lowest minimum in this mapping, achieve better generalization, etc. For example, randomness is used in the initialization of weight values.

Due to the importance of randomness in machine learning, one model using the same dataset can give slightly different results in different training runs. In order to avoid the problem and achieve uniform fulfilment in the mapping of the inputs to the outputs, a common solution is to run the same model multiple times and average the different outputs. Jointly, a group of such neural networks reduces the residual generalization error [15].

A. Evaluation of a neural network performance

In order to assess a model, the concepts of accuracy and loss are used, which represent respectively how the model is fitting the training dataset and if the model under study is losing the ability to predict in the general cases. Furthermore, the concepts of precision and recall are commonly used for the same purpose. Both are plotted in the precision-recall curve. Precision is defined as the number of true positives over the number of true positive plus false positive (high precision relates to low false positive rate). On the other hand, recall is the rate between true positives over true positive plus false negative (high recall relates to low false negative rate). In addition, the term of AUC (area under the curve) it is often used, which means, in the case of plotting the true positive vs false positive rate, the probability of a randomly selected sample being classified as positive higher than negative. In this case, as an exoplanet or a false positive, respectively. This kind of plot is called the receiver operating characteristic (ROC) curve.

In order to implement our neural network, the software library TensorFlow [16] has been used. It was chosen because Shallue & Vanderburg developed their neural networks with it and, also, because it is well documented.

III. SHALLUE & VANDERBURG MODEL

The best neural network proposed by [8] is called Astronet, which is a deep convolutional neural network. Shallue & Vanderburg decided to use the light curves of TCE as inputs of the neural network.

The input light curves were obtained after processing the data of Kepler Object of Interest Catalog-Q1-Q17 DR24.

Firstly, for each TCE light curve, any transit that corresponded to any other TCE was extracted. Secondly, the low-frequency variability was also removed. Thirdly. Two different phase-folding of the light curves were performed, centring the event in the middle, aiming to create two different representations of the light curves, the local and the global view:

On the one hand, the phase-folding was done dividing the light curve into 2001 uniformly equally spaced intervals through the TCE period $\left(\lambda = \frac{t}{2001} \right)$, where t is the TCE period. The median fluxes in each interval were calculated and the global view was obtained (Figure 1).

On the other hand, in the case of the local view, the observations through a range of 4 times the transit duration in either side were divided into 201 equally spaced intervals $\left(\lambda = \frac{8d}{201} \right)$, where d is the transit duration. But in this view, so as to make transits more visible and reduce scatter, the median fluxes were calculated using not only values within the intervals, but also the ones from the adjacent intervals. In particular, the width of the range of values used by Shallue & Vanderburg was $\delta = 0.16d$ [8], that represents a total width of 4λ intervals.

Finally, both representations were normalized with median value set to 0 and the minimum value to -1.
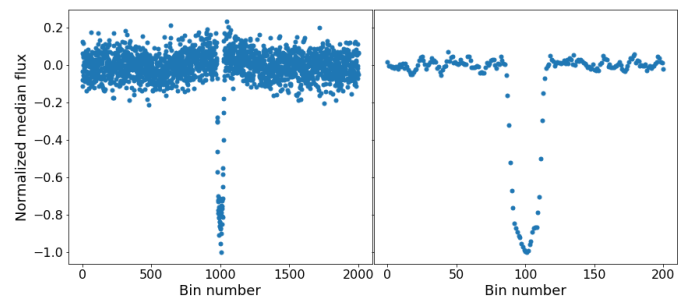

FIG. 1: Light curve in the global view (left) and local view (right) of Kepler-134 b.

The structure of Astronet is shown in Figure 2. It can be seen that the neural network has two different inputs, the left one is the global view of the TCE, and the right one is the local view. After some convolution, Max-Pool and Fully connected layers, an output is returned using a sigmoid activation function in the last layer. The output has values between 0 and 1, which represent the probability of the TCE to be an exoplanet. It uses a sigmoid function because this function usually fits well when the dataset is imbalanced.

The parameters like the number of units in a layer, the kernel size, the activation function, etc. are called hyperparameters.

Depending on the hyperparameters that are chosen in the different layers, the model can achieve better or worse results. The hyperparameters used by Shallue & Vanderburg are the ones that optimize their model. They are shown in Figure 2 inside the parentheses in each layer.

Furthermore, Astronet is trained with 50 epoch and 64 batch size. The first one is the number of times the model will work through the training dataset, and the other one is the number of samples that are used in each training step in order to change the weight of the model. Also, it uses an optimization algorithm called Adam algorithm with the following parameters: $\alpha = 10^{-5}$, $\beta_1 = 0.9, \beta_2 = 0.999$ and $\epsilon = 10^{-8}$.
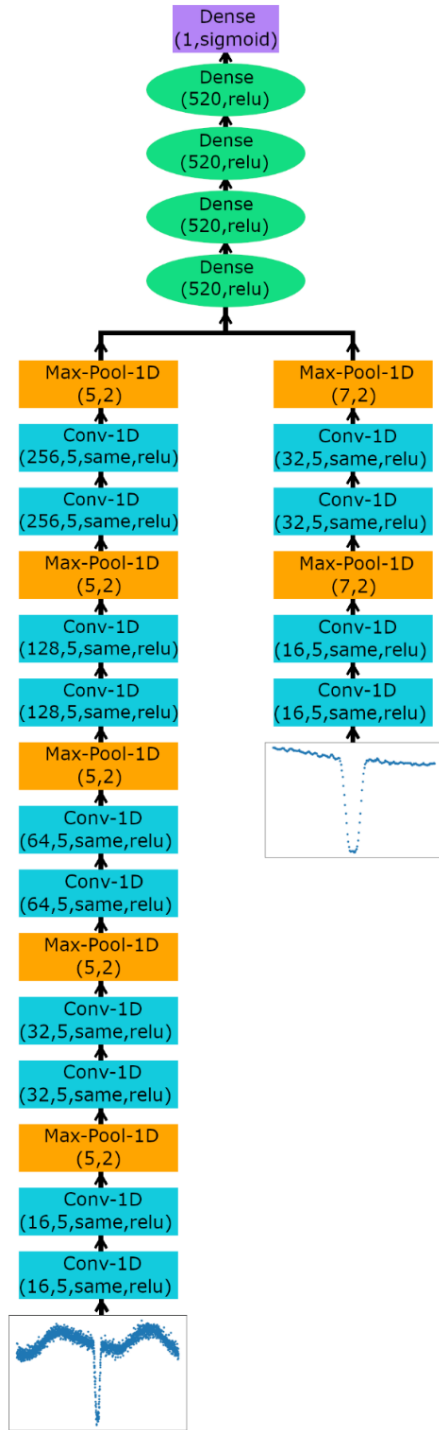
FIG. 2: Diagram of Astronet. Blue rectangles are 1-dimension convolution layers, in parentheses from left to right, the filters, kernel size, padding and activation function. Orange rectangles are 1-dimension Max-Pool layers, in parentheses form left to right, the pool size and strides. Green ellipses are fully connected layers (in Tensorflow they are called dense), in parentheses, from left to right, the number of units and the activation function. Finally, the purple rectangle is the output layer which is a fully connected layer with a single unit using the function sigmoid as activation function.

IV.   EVALUATION & COMPARISON

In order to compare the re-implementation made in this paper and Astronet, the terms explained in section II.A are used. The final results obtained and the ones of Shallue & Vanderburg's model are compared and shown in Table I. It can

be seen that the test accuracy for both of them is very close. In addition, both AUC values are close too, indicating that we have successfully reproduced the network.

	Test Accuracy	AUC
Astronet	0.960	0.988
Reproduction	0.957	0.990

TABLE I: Test accuracy and AUC (in the true positive vs. false positive) for Shallue & Vanderburg's model (Astronet) and the reproduction.

A reason for having slightly different results could be the random initialization of the weights and the serendipity in the training process.

Also, the precision-recall curve of both models is shown in Figure 3. It can be seen that the Shallue's precision-recall curve is a bit higher than the one achieved in this work; this means that they can choose a threshold where the precision and recall are higher, in other words, lower false positive and false negative rate.
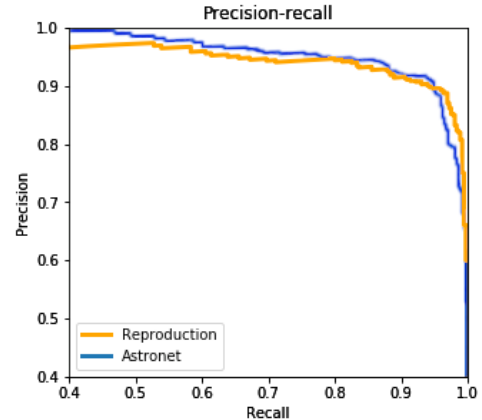

FIG. 3. Precision vs. recall for Astronet and the reproduction. The Astronet's curve is extracted from [8].

Note that despite the area under the curve in the Precision-Recall graphic for Astronet being higher than the achieved in this work, the AUC in the reproduction is greater than Shallue & Vanderburg's model. One must note that the AUC and the area under the curve in the Precision-Recall are not the same. Probably this difference is due to the fact that the AUC overestimates the model skills when the dataset is imbalanced [17], as in this case, where the number of light curves labelled as exoplanet are 3600 in front of 12137 light curves labelled as not exoplanet (AFP + NTP).

In order to understand how the neural network works, we have experimented by setting parts of the light curve to zero and see how the returned probability changes. Using this process, Shallue & Vanderburg proved that the neural network learns to identify the primary and secondary eclipses in the global view [8].

Now, so as to comprehend in detail the role of each part of both inputs, and also find any weaknesses in this model, we did something similar to the method by Shallue & Vanderburg. In this case, nine different configurations were tested setting to zero different parts of the local and global view. For example, while the global view was unchanged, the local view was set all zero. The most illustrative configuration was the one where we "cleansed" the light curve except the part where the luminosity of the star decreased, thus simulating a noise-free

reference level. Figure 3 shows the light curve after applying this change. It can be easily compared with the original light curve in Figure 1.
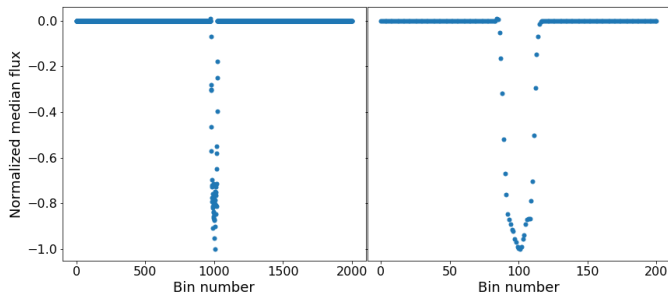

FIG. 3: Light curve in the Global view (left) and local view (right) of Kepler-134 b. All values are set to zero except the TCE.

When an exoplanet was processed using this configuration, the probability returned of being an exoplanet dropped off from 91% to only 10%. A reasonable explanation of a decline of 80% is that possibly the neural network has learned to identify an exoplanet if in the light curve, the luminosity decreases more than the standard deviation of the noise. As the noise was set all to zero except at the transit, the neural network could not compare its magnitude with the noise, then it could not identify the transit as due to an exoplanet, and the probability significantly decreased.

Notice that one can use different ways to estimate how the neural network works by varying the input and seeing how the output changes. Nevertheless, its internal functioning is opaque, as the weights fitted during the training do not provide any clue about why these values were chosen, and how they affect in achieving the desired performance. That is why neural networks are usually described as black-boxes.

V. PROPOSED MODELS

After a successful reproduction of the work of Shallue & Vanderburg, different models can be proposed, putting into practice the knowledge acquired through the previous work.

Firstly, the introduction of auxiliary parameters to the neural network was tested. For example, as the light curve in the local and global view are normalized, part of the lost data in the process can be useful to characterise an exoplanet. For example, in a distant eclipsing binary, the eclipses make sharp transits.

Secondly, as it has been suggested in the section IV, if the neural network takes into account the relationship between the depth of the signal and the standard deviation of the noise, maybe introducing the SNR of the light curve as an additional parameter could help the neural network to identify exoplanets more efficiently.

The data of the depth and the SNR used in these models are also from the database of Kepler Object of Interest Catalog - Q1-Q17 DR24 [10].

Although Shallue & Vanderburg proved in [8] that using both local and global view in their net achieved better results, the use of only one view with parameters is also tested as one of the proposed alternative models.

The different models tested are shown in Table II.

The first test was to check whether a simpler or a complex structure could achieve better results. These were only the cases of the models using the local view with and without

auxiliary parameters (see Table II). Their performances were improved when two convolution layers and two (Local view and Local view + D) or five (Local view + P) dense layers were used. Also, the use of a Max-Pool layer (with pool size equal to 5) between both convolution layers seemed to improve the performance. Furthermore, the number of epochs required for training the neural network could be reduced because the neural network started overfitting.

The other models achieved better results using the same structure as Astronet.

Secondly, the number of units in the fully connected layers was changed. For instance, the units tested were 128, 256, 512. The results of the validation loss as a function of the epochs showed that 128 units were the best choice for models with both auxiliary parameters. But, when only one auxiliary parameter was used, 256 units were a better choice. Nevertheless, no sign of improvement was found using these number of units in the other model's performance. Finally, the last hyperparameter tested was the use of padding. In all cases the use of padding improves the performance, as Shallue & Vanderburg proved.

The other hyperparameters were not changed because the proposed models differ only in the part of adding one or two auxiliary parameters. Then, using part of the hyperparameters of Astronet was likely a good choice.

In Table II, the best results of the test accuracy and AUC obtained by different models are shown.

	Test Accuracy	AUC
Astronet	0.960	0.988
Global & Local view + P	0.963	0.988
Global & Local view + D	0.960	0.988
Local view	0.934	0.976
Local view + D + SNR	0.929	0.972
Local view + D	0.928	0.977
Global view	0.957	0.987
Global view + D + SNR	0.955	0.987
Global view + D	0.957	0.986

TABLE II: Test accuracy and AUC calculated by different neural networks. Where D represents de auxiliary parameter depth, SNR the auxiliary parameter of the SNR and P means both depth and SNR.

It can be seen that only two models have the same or very slightly higher performance than Astronet. Furthermore, the models Local view and Global view achieve better results than the ones in [8].

In order to comprehend the relevancy of the values of the depth and SNR in the model, something similar to the last part of section IV is done. If the depth was set to zero, almost all probabilities returned by the neural network did not change. But when the value introduced was very high, for instance, 10^5 ppm, all probabilities becomes zero. This suggests that introducing the depth as an auxiliary parameter does not have an important role when transits' depths are slight, but it could be useful in cases of false positives with sharp depths.

On the other hand, in the case of varying the value of the SNR, the probability of being an exoplanet was barely unaffected when it was set to zero. Nevertheless, for high values of SNR, the probability decreased. These behaviours probably mean that the dataset used is biased in the way that a significant part of the light curves labelled as exoplanet have

low SNR. This hypothesis is proposed as there is no reason why a star with an exoplanet must have low SNR.

However, the results in Table II seem to support the fact that using these auxiliary parameters reduce the performance of the neural network in cases where only one view is used.

VI.   CONCLUSIONS

In this work the best neural network made by Shallue & Vanderburg has been successfully reproduced, reaching a 0.957 test accuracy compared to the 0.960 reached by Astronet and also almost the same AUC value 0.990 instead of 0.988.

Moreover, models using the global and local view with auxiliary parameters reached the same performance of Astronet, with the one using both depth and SNR achieving very slightly higher test accuracy value (0.963) and the same AUC value (0.988). However, maybe the results could be improved if further computational time is spent trying different configuration and hyperparameters.

In addition, this work has proved that the use of neural networks in order to characterise exoplanets is an efficient option that can reduce the processing time and human supervision. Since to use neural networks a lot of data is needed, further exoplanets discoveries in the following decades will be an important contribution in this area.

In future work, a neural network which uses both views with auxiliary parameters obtained by a parametric modelling of the light curve [18] would be interesting to evaluate.

_____

[1] D. Charbonneau et al, «Detection of Planetary Transits Across a Sun-like Star», *The Astrophysical Journal,* vol. 529, L45–L48, 4 pp., 1999.

[2] NASA Exoplanet Archive. [En línea]. Available: https://exoplanetarchive.ipac.caltech.edu/

[3] B. Kirk et al, «*Kepler* eclipsing binary stars. VII. The catalog of eclipsing binaries found in the entire *Kepler* data set», *The Astronomical Journal*, vol. 151, num. 68, 21 pp., 2016.

[4] A. Prša et al, «*Kepler* eclipsing binary stars. I. Catalog and principal characterizaton of 1879 eclipsing binaries in the first data release», *The Astronomical Journal*, vol. 141, num. 83, 16 pp., 2011.

[5] A. Santerne et al, «The contribution of secondary eclipses as astrophysical false positives to exoplanet transit surveys», *Astronomy & Astrophysics*, vol. 557, num. A139, 8 pp., 2013.

[6] R. L. Gilliland et al, «*Kepler* mission stellar and instrument noise properties», *The Astrophysical Journal,* vol. 197, num. 6, 19 pp., 2011.

[7] T. Kallinger et al, «The connection between stellar granulation and oscillation as seen by the Kepler mission», *Astronomy & Astrophysics*, vol. 570, num. A41, 17 pp., 2014.

[8] C. J. Shallue & Vanderburg, «Identifying Exoplanets with Deep Learning: A Five-planet Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90», *The Astrophysical Journal,* vol. 155, num. 5, 21 pp., 2018.

[9] Dataset processed by Shallue & Vanderburg. [En línea]. Available: https://drive.google.com/file/d/1_MymCJtfdAojVP52hcE6 vtBMDsuziwaj/view?usp=sharing

[10] S. Seader et al, «Detection of potential transit signals in 17 quarters of Kepler mission data», *The Astrophysical Journal*, vol. 217, num.18, 20 pp., 2015.

[11] J. H. Catanzarite, «Autovetter planet candidate catalog for Q1-Q17 Data Release 24 (KSCI-19091-001)», 2015. [En línea]. Available: https://exoplanetarchive.ipac.caltech.edu/docs/KSCI-19091-001.pdf

[12] C. C. Aggarwal, «Neural Networks and Deep Learning», Charm: Springer International Publishing, 2018.

[13] P. Chintarungruangchai & G. Jiang, «Detecting Exoplanet Transits through Machine-learning Techniques with Convolutional Neural Networks», *Publications of the Astronomical Society of the Pacific*, vol. 131, num. 064502, 15 pp., 2019.

[14] A. Dattilo et al, «Identifying Exoplanets with Deep Learning II: Two New Super-Earths Uncovered by a Neural Network in K2 Data», *The Astrophysical Journal,* vol. 157, 21 pp., 2019.

[15] L. K. Hansen & Salamon, «Neural network ensembles», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.12, Issue 10, pp. 993-1001, 1990.

[16] Tensorflow. [En línea]. Available: https://www.tensorflow.org/

[17] J. Davis & M. Goadrich, «The relationship between Precision-Recall and ROC curves», *Association for Computing Machinery*, book title: Proceedings of the 23rd International Conference on Machine Learning, pp. 233-240, 2006.

[18] Z. Mikulášek, « Phenomenological modelling of eclipsing system light curves», *Astronomy & Astrophysics*, vol. 584, num. A8, 2015.