Quasar microlensing light curve analysis using deep machine learning

Georgios Vernardos,1⋆ and Grigorios Tsagkatakis2

1 Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700AV Groningen, the Netherlands
2 Institute of Computer Science - Foundation for Research and Technology (FORTH), GR-71110, Heraklion, Greece

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

We introduce a deep machine learning approach to studying quasar microlensing light curves for the first time by analyzing hundreds of thousands of simulated light curves with respect to the accretion disc size and temperature profile. Our results indicate that it is possible to successfully classify very large numbers of diverse light curve data and measure the accretion disc structure. The detailed shape of the accretion disc brightness profile is found to play a negligible role, in agreement with Mortonson et al. (2005). The speed and efficiency of our deep machine learning approach is ideal for quantifying physical properties in a ‘big-data’ problem setup. This proposed approach looks promising for analyzing decade-long light curves for thousands of microlensed quasars, expected to be provided by the Large Synoptic Survey Telescope.

Key words: accretion, accretion discs – gravitational lensing: micro – quasars: general.

1 INTRODUCTION

Quasar microlensing is a unique way of measuring the structure of AGN at the very small scales inaccessible by other techniques. The magnification effect of the microlenses - stellar mass compact objects lying within lensing galaxies - on small angular diameter sources, like an accretion disc surrounding a supermassive black hole, can be very pronounced, allowing one to study their structure on 10−6 arcsec scales, corresponding to a few tens of AU on the source plane (see the review by Schmidt & Wambsganss 2010). Such measurements are putting classic accretion disc models, particularly the standard thin disc model (Shakura & Sunyaev 1973), to the test (Abolmasov & Shakura 2012): quasar accretion discs are found to be larger than expected (Morgan et al. 2010; Blackburne et al. 2011; Jimenez-Vicente et al. 2014; Chartas et al. 2016) while their temperature profiles range from shallower (Bate et al. 2008; Rojas et al. 2014; MacLeod et al. 2015; Bate et al. 2018) to steeper (Eigenbrod et al. 2008; Floyd et al. 2009; Blackburne et al. 2011; Jimenez-Vicente et al. 2014).

Microlensing snapshot and monitoring observations can produce single-epoch or time series of observations (light curves). Light curve analysis techniques allow one to extract physical information from year-long observations, at the cost of higher model complexity and much more demanding computations. The method of Kochanek (2004) provides a powerful Bayesian analysis approach to studying quasar structure by generating O(106) simulated light curves for each combination of model parameters and fitting them to the data (Poindexter et al. 2008). This computationally demanding technique requires careful use of priors (e.g. on the source effective velocity components), and may have to resort to independent modelling of light curve segments to keep the computations feasible (Poindexter & Kochanek 2010).

Microlensing analyses are based on some knowledge of the mass-density distribution of the lensing galaxy and its environment (the macromodel), which depends on the quality of the imaging data, the presence of satellite galaxies, possible existence of dark matter substructures, etc. This model determines the values of the convergence, κ, and shear, γ, at the locations of the multiple quasar images, which in turn set the microlensing properties via the magnification map (Paczynski 1986; Kayser et al. 1986). If these parameters, or the further partition of the convergence into smooth (dark) and compact (stellar) matter components (e.g. Schechter & Wambsganss 2002), are not accurately known, the resulting microlensing effect can vary significantly (Vernardos & Fluke 2014). Factoring in such uncertainties in current microlensing light curve analysis techniques quickly leads to intractable computations.

The number of known quasar lenses is set to increase from a few hundreds to thousands by all-sky survey facilities currently under construction. The Large Synoptic Survey Telescope (LSST Science Collaborations 2009) will provide...
multi-wavelength light curves for all these lenses over a period of a decade or more. Analyzing this large volume of data will require new, more efficient methods than the current state-of-the-art techniques, which have been designed to perform optimally for single systems with high quality observations.

Machine learning algorithms have provided promising solutions to ‘big-data’ problems in astronomy (e.g. Héa 2014; Huertas-Company et al. 2015; Tuccillo et al. 2018; Sti-vaktakis et al. 2018). Convolutional Neural Networks (CNNs Lecun et al. 2015) are a particular class of such algorithms that can model complicated non-linear relationships by repeated abstract processing layers. CNNs have been successfully used in gravitational lensing to find lenses (Lanusse et al. 2018; Petrillo et al. 2019) and model them (Hezaveh et al. 2017).

In this paper, we use CNNs to model quasar microlensing light curves and extract the size and temperature profile of the accretion disc. Our training set, based on simulated light curves, and the neural network architecture are presented in Section 2. Section 3 consists of the training results and a series of more advanced tests on different types of light curves. Conclusions and implications for future work are discussed in Section 4.

Throughout this paper a cosmological model with $\Omega_m = 0.26$, $\Omega_\Lambda = 0.74$, and $H_0 = 72$ km s$^{-1}$ Mpc$^{-1}$ is used.

2 METHOD

There are two main factors affecting the performance of CNN algorithms: the training set and the design of the network itself. Here we present the steps in generating mock quasar microlensing light curves, used to create the training set and a series of additional test sets, and the main features of the adopted CNN architecture.

2.1 Generating mock light curves

Generating light curves follows the same approach used in previous work (Vernardos et al. 2015) and elsewhere in the literature (e.g. Kochanek 2004). The necessary components for this procedure are: the accretion disc brightness profile, the Einstein radius of the microlenses, the magnification map, and a velocity model for the accretion disc moving across it. Our assumptions on the size and shape of the accretion disc profile, which are the main focus of this work, and the rest of the parameters involved are described below.

The size of the accretion disc as a function of wavelength is assumed to have the following parametric form:

$$r = r_0 \left( \frac{\lambda}{\lambda_0} \right)^\nu,$$  \hspace{1cm} (1)

where $\lambda = \lambda_{obs}/(1+z_S)$, with $\lambda_{obs}$ being the observing wavelength and $z_S$ the quasar redshift, $\nu$ the power law index, and $r_0$ the size of the disc observed at the rest wavelength $\lambda_0 = 102.68$ nm. This is a frequently used parametric model (e.g. Jimenez-Vicente et al. 2014; Bate et al. 2018) that captures the accretion disc size dependence on physical parameters like the Eddington luminosity, the accretion efficiency, and the black hole mass, in $r_0$. Temperature is inversely proportional to wavelength for accretion disc emission that is a superposition of black body spectra. This is the case for the thin disc model that has $\nu = 4/3$.

The half-light radius, $r_{1/2}$, which contains half the luminosity of the accretion disc, is set to the size obtained from equation (1). Although there is evidence that the shape of the disc profile does not play a significant role in microlensing observations (Mortonson et al. 2005), this has not been examined thoroughly with respect to light curves, and not at all as a function of $\kappa, \gamma$. Here, we assume different fiducial parametric shapes aiming to cover as wide a range of brightness variations as possible. The different brightness profiles have the following shapes: gaussian, gaussian with an inner edge (‘gaussian-hole’), smooth with an inner edge (‘hole’), sinusoidal dropping with radius (‘wavy’), uniform, and exponential. These are all symmetric two-dimensional profiles so that the brightness per unit surface, $I(r)$, is only a function of the radius (see Table 1). Their one-dimensional intersections are shown in Fig. 1. Apart from the size, each profile may have a different number of free parameters, which are listed in Table 1. All the profiles are truncated at $2 \times r_{1/2}$ (the full profiles will only induce a small magnification offset due to their flatness beyond $2 \times r_{1/2}$) and normalized to a luminosity of unity.

The scale length of microlensing is set by the Einstein radius of the microlenses on the source plane:

$$R_{Ein} = \sqrt{\frac{4G\langle M \rangle D_S D_{LS}}{c^2 D_L}},$$ \hspace{1cm} (2)

where $D_S$, $D_L$, and $D_{LS}$ are the angular diameter distances to the source, the lens, and between the lens and the source respectively, $\langle M \rangle$ is the average microlens mass, $G$ the gravitational constant, and $c$ the speed of light. Here we use microlenses with $1 \, M_\odot$ and the fiducial lens and source redshifts of 0.5 and 1.2, to get $R_{Ein} = 4.65 \times 10^{16}$ cm. However, as it turns out only the relative values of $r_{1/2}$ with respect...
The smooth matter fraction, $s$, which is the fraction of smoothly distributed matter (as opposed to compact microlenses) over the total matter, is another important parameter for microlensing studies (e.g. Vernardos et al. 2014). Here we always set $s = 0$ for simplicity. A varying $s$ has a similar effect to the magnification maps as changing the $\kappa, \gamma$, which is out of the scope of this work.

Table 1. Shapes of the accretion disc surface brightness profile, $I(r)$. The normalization factor corresponds to the luminosity, calculated as the two-dimensional integral of the surface brightness. Every shape has a parameter associated to the profile size, which depends on the half-light radius, $r_{1/2}$, the radius containing half the luminosity. The ‘gaussian+hole’ profile has an inner edge, determined by $R_{\text{in}}$ that is equal to the radius of the innermost stable circular orbit of a black hole with mass $M$ ($G$ is the gravitational constant and $c$ the speed of light). The ‘wavy’ profile has an additional parameter $n$, which is the number of nodes of the profile within two half-light radii. The ‘exponential’ profile was not used in the training set (Section 2.2.1) but it was used to test the performance of our classification method on data that it was not trained on (Section 3.3.1).

| name          | analytic form of $I(r)$                                    | normalization factor         | size parameter (function of $r_{1/2}$) | other parameters |
|---------------|-----------------------------------------------------------|------------------------------|---------------------------------------|------------------|
| gaussian      | $\exp\left(-\frac{r^2}{2\sigma^2}\right)$               | $\frac{1}{2\pi\sigma^2}$    | $\sigma = \frac{r_{1/2}}{1 + \frac{r_{1/2}}{R_{\text{in}}}}$ | -                |
| gaussian+hole | $\exp\left[-\frac{r^2}{2\sigma^2}\right]$ for $r \leq R_{\text{in}}$  $\exp\left[\frac{R_{\text{in}}}{2\sigma^2}\right]$ for $r > R_{\text{in}}$ | $\frac{1}{2\pi\sigma^2}$    | $\sigma = \sqrt{\frac{r_{1/2}^2 - R_{\text{in}}^2}{2\text{log}^2}}$ | $R_{\text{in}} = \frac{6GM}{c^2}$ |
| hole          | $\frac{1}{r}(1 - \frac{R_{\text{in}}}{r})$               | $\frac{3R_{\text{in}}}{2\pi}$ | $R_{\text{in}} = \frac{r_{1/2}}{3}$   | -                |
| wavy          | $\frac{\sin^2(\beta r)}{\beta}$                         | $\frac{\beta}{\pi n^2}$     | $\beta = \frac{n\pi}{2r_{1/2}}$      | $n$              |
| uniform       | $1$                                                       | $\frac{1}{\pi R_{\text{in}}^2}$ | $R = \sqrt{\frac{1}{2}r_{1/2}}$      | -                |
| exponential   | $(\exp(\frac{r^2}{\mu^2}) - 1)^{-1}$                     | $0.062 \frac{r_{1/2}}{\mu^2}$ | $\sigma = 0.4106 r_{1/2}$            | -                |

to $R_{\text{Ein}}$ are important and the absolute value of $R_{\text{Ein}}$ is irrelevant for the results presented here.

The effects of microlensing are simulated by magnification maps; pixellated representations of the source plane caustics due to the microlenses. The main parameters\(^1\) of these maps are the convergence, $\kappa$, and shear, $\gamma$, which determine the density and clustering of the caustic network (see Fig. 2). We use high quality magnification maps from the GERLUMPH\(^2\) parameter survey (Vernardos et al. 2014) that cover a large range of $\kappa, \gamma$ parameters. These square maps are 25 $R_{\text{Ein}}$ wide, enough to allow for many independent and long light curves to be extracted, and have a 10,000 pixel resolution on a side, enough to resolve the details of the different disc profile shapes (e.g. the inner edge of the ‘gaussian+hole’ profile, at $9 \times 10^{14}$ cm, corresponds to 8 pixels for $R_{\text{Ein}} = 4.65 \times 10^{16}$ cm).

In order to extract microlensing observables for accretion discs that are not point-sources ($r_{1/2} > 1$ pixel, or $> 0.0025 R_{\text{Ein}}$, which is always the case here), the magnification map and the disc profile need to be convolved first. This leads to a reduced available map area, the effective map, due to the convolution edge effects. The largest profiles used here are $\sim 0.8 R_{\text{Ein}}$ wide, leaving an effective map of 24.2 $R_{\text{Ein}}$ (9,680 pixels) to extract light curves from.

The combination of the velocity vectors of the observer, the lens, and the source, results in an effective velocity of the source across the magnification map (see Kochanek 2004, for a description of such a velocity model). The length of a light curve on the magnification map is set by the magnitude of this velocity multiplied by the time range spanned by the observations (usually up to years). The orientation of the map with respect to the effective velocity depends on the direction of the external shear, which can be derived from macromodels of the lensing galaxy. However, these model components are more relevant to individual object studies and are outside the scope of this work that focuses on the size of the accretion disc with respect to $R_{\text{Ein}}$. Therefore, the following broad assumptions are made: the light curve trajectories have a random direction, all light curves have the same length, and the light curve data are continuous. An example of mock light curves is shown in Fig. 2.

2.2 Analysis

The dramatic increase in the volume of astronomical observations has led to a situation where the manual analysis of highly dynamic observations by human experts is creating a bottleneck in the observation-to-knowledge pipeline. To address this challenge, cutting-edge automated data analysis systems based on machine learning are gaining significant traction. To that end, we consider deep machine learning approaches based on one-dimensional CNNs for supervised modeling of light curve data, where annotated examples are utilized for training multi-layer deep neural networks that can then characterize new observations. We discuss the methodology for generating the training data set first and the architecture of the proposed CNN next.

2.2.1 Training set

Using the procedure described above, a set of mock light curves is generated for training the deep machine learning analysis algorithm. The Einstein radius is set to $4.65 \times 10^{16}$ cm, corresponding to a lens and a source at redshifts 0.5 and 1.2 respectively, and 1 $M_\odot$ microlenses. The half-light radius that sets the size of all the different accretion disc profiles varies in the range 0.02 – 0.2 $R_{\text{Ein}}$. The additional parameters for the ‘gaussian+hole’ and ‘wavy’ profiles are

---

1 The smooth matter fraction, $s$, which is the fraction of smoothly distributed matter (as opposed to compact microlenses) over the total matter, is another important parameter for microlensing studies (e.g. Vernardos et al. 2014). Here we always set $s = 0$ for simplicity. A varying $s$ has a similar effect to the magnification maps as changing the $\kappa, \gamma$, which is out of the scope of this work.

2 http://gerlumph.swin.edu.au/
Figure 2. Mock light curves from magnification maps with \( \kappa = 0.09, \gamma = 0.55 \) (top) and \( \kappa = 0.8, \gamma = 0.1 \) (bottom). The light curve trajectories on the maps (red arrows) have the same length (5× \( R_{\text{Ein}} \)) and a random orientation. Mock light curves for all the profile shapes listed in Table 1 are shown for two different accretion disc profile sizes (\( r_{1/2} \) equals to \( 0.02 R_{\text{Ein}} \) and \( 0.1 R_{\text{Ein}} \) respectively, for the top and bottom panels next to each map). The magnification, \( \mu \), is with respect to the macromodel induced magnification, \( \mu_{\text{macro}} = \left(1 - \kappa\right)^2 - \gamma^2 \), which is different for each map. Although the light curves may appear indistinguishable on large scales (\( \sim R_{\text{Ein}} \)), the effect of small scale profile structure is visible during a caustic-crossing event, as indicated in the inset plot in the third panel on the right-hand side.

The training set is extracted from a single magnification map having \( \kappa = 0.09, \gamma = 0.55 \) (top left panel of Fig. 2). The length of the light curves is arbitrarily set to 5 \( R_{\text{Ein}} \) (\( \approx2,000 \) pixels\(^3\)), which is long enough to permit several caustic crossings to occur, yet short enough to allow for a large number of independent light curves to be extracted from the 24.2 \( R_{\text{Ein}} \)-wide effective maps. Thousands of light curves are generated for each profile and filtered to keep the ones having a maximum magnification of at least 1.5. This is to avoid flat light curves without any meaningful features, which have a high chance to occur due to the large regions of the magnification map that are not covered by caustics (see top left panel of Fig. 2). For each \( r_{1/2} \) and profile shape, 900 light curves are selected for the training set, which is extracted from a single magnification map having \( \kappa = 0.09, \gamma = 0.55 \) (\( s = 0 \)), shown in the top left panel of Fig. 2. The length of the light curves is arbitrarily set to 5 \( R_{\text{Ein}} \) (\( \approx2,000 \) pixels\(^3\)), which is long enough to permit several caustic crossings to occur, yet short enough to allow for a large number of independent light curves to be extracted from the 24.2 \( R_{\text{Ein}} \)-wide effective maps. Thousands of light curves are generated for each profile and filtered to keep the ones having a maximum magnification of at least 1.5. This is to avoid flat light curves without any meaningful features, which have a high chance to occur due to the large regions of the magnification map that are not covered by caustics (see top left panel of Fig. 2). For each \( r_{1/2} \) and profile shape, 900 light curves are selected for the training set, which is extracted from a single magnification map having \( \kappa = 0.09, \gamma = 0.55 \) (\( s = 0 \)), shown in the top left panel of Fig. 2. The length of the light curves is arbitrarily set to 5

\[ R_{\text{Ein}} \approx 2,000 \text{ pixels}^3, \]
Table 2. Different sizes for the accretion disc profiles, parameterized by the half-light radius, $r_{1/2}$ (see also Table 1). Each $r_{1/2}$ value is assigned to a class that is used to train the machine learning algorithm described in Section 2.2.2. A value of $R_{\text{Ein}}=4.64 \times 10^{14}$ cm is used and the corresponding observed wavelengths are calculated from equation (1) with $r_0=2.59 \times 10^{15}$ cm and $\nu=4/3$, for a source at redshift 1.2. However, as discussed in Sections 3 and 4, the absolute values of $R_{\text{Ein}}$, $r_{1/2}$, and the corresponding $\lambda_{\text{obs}}$ do not affect the results of the analysis and are only listed for illustrative purposes.

| class | $r_{1/2}$ [$R_{\text{Ein}}$] | $r_{1/2}$ [×10$^{14}$ cm] | $\lambda_{\text{obs}}$ [nm] |
|-------|-----------------|-----------------|-----------------|
| 0     | 0.02            | 9.3             | 104.78          |
| 1     | 0.04            | 18.6            | 176.22          |
| 2     | 0.06            | 27.9            | 238.86          |
| 3     | 0.08            | 37.2            | 296.38          |
| 4     | 0.10            | 46.5            | 350.37          |
| 5     | 0.12            | 55.8            | 401.71          |
| 6     | 0.14            | 65.1            | 450.95          |
| 7     | 0.16            | 74.4            | 498.45          |
| 8     | 0.18            | 83.7            | 544.49          |
| 9     | 0.20            | 93.0            | 589.26          |

Table 3. Architecture of a layers-block component of our CNN.

| Type          | Shape | Parameters |
|---------------|-------|------------|
| Convolution   | (32,10) | 10272      |
| ReLU          | -     | -          |
| Convolution   | (32,50) | 51232      |
| ReLU          | -     | -          |
| Add           | -     | 0          |
| Pooling       | (3,2) | 0          |
| Batch Normalization | -     | 0          |

In total, the proposed CNN is parameterized by 673,098 parameters that are tuned during the training stage to minimize the categorical cross entropy between predicted ($\hat{y}$) and actual labels ($y$). For a single example classified in one out of 9 classes, this is given by:

$$\mathcal{L}(y, \hat{y}) = -\sum_{c=1}^{C} \mathbb{1}_{y_i \in C_c} \log p(\hat{y}_i),$$  \hspace{1cm} (3)

where $\mathbb{1}_{y_i \in C_c}$ is the indicator function taking a value of one only for the correct class $c$, and $p(\hat{y}_i)$ is the predicted probability associated with each class. Minimization of the loss function is achieved by employing the Adam optimizer (Kingma & Ba 2014, with learning rate $10^{-3}$, decay $10^{-5}$). To quantify the performance of the CNN architecture, we report the prediction accuracy as the percentage of correctly predicted $r_{1/2}$ classes, averaged over all examples (either training or validation, see Fig. 3). To maximize the accuracy, the training stage involves the repetitive tuning of parameters for 1000 epochs, where each epoch corresponds to a single pass over the entire training set (45,000 light curves).

3 RESULTS

Here we present the results of the network training and its application. First, we explore how modifying the neural network architecture can achieve better results with the training set. Predictions can be only as good as the data used to train the algorithm, while a well-trained network is not guaranteed to give good predictions for data of a different kind. This effect of the training/validation data on the predictions is explored in a number of tests. Finally, the ability to correctly recover a fiducial size-wavelength dependence of the accretion disc profiles is measured. We point out that the network output is always one of the classes of $r_{1/2}$ listed in Table 2.

3.1 Convergence

To demonstrate the need for multi-layered CNN architectures, Table 4 reports the achieved accuracy for a limited set of training and validation examples (4500 and 500 respectively) at different epochs and for different architectures.
The consequence of this situation is a known problem affecting CNNs called overfitting (Srivastava et al. 2014). The large gap between training and validation performance when no drop-out regularization is introduced, reaches 90% when no drop-out regularization is introduced, and the corresponding accuracy in the validation set is only 69%. Observe that although the performance on the training set is very high, the performance on the validation set is much lower. For example, the accuracy on the training set is 95%, while the accuracy on the validation set is only 70%. This is a clear indication of overfitting.

Specifically, we consider networks with 2, 3, 4 and 5 layers-blocks, while for the case of 5 layers-blocks, we also explore the impact of drop-out (DO) regularization. For each case, we report the accuracy achieved on the validation set (and the training set in parenthesis). The best architecture used in the following is highlighted in bold.

| Epoch | 200 | 500 | 1000 |
|-------|-----|-----|------|
| 2 blocks | 30.8 (39.9) | 44.7 (55.4) | 56.9 (61.4) |
| 3 blocks | 58.3 (57.8) | 65.2 (71.0) | 65.2 (77.3) |
| 4 blocks | 51.3 (54.1) | 65.5 (71.2) | 67.4 (79.2) |
| 5 blocks | 61.7 (96.6) | 70.5 (99.6) | 69.3 (99.8) |
| 5 blocks (DO) | 66.2 (70.6) | 70.3 (81.4) | 75.3 (90.8) |

Table 4. Accuracy (percentage of the correctly predicted $r_{1/2}$ classes over the total) achieved by different CNN architectures over a limited set of training and validation examples (4500 and 500 light curves respectively). The numbers are for the validation set (training set in parenthesis). The difference between the last two architectures is the inclusion of drop-out (DO) regularization. The best architecture used in the following is highlighted in bold.

3.2 Validation

The validation set consists of a part of the training set that was not used during the actual training. For each $r_{1/2}$ class and profile shape, 100 light curves are set aside as part of the validation set. The final version of the trained network is then used to predict the class of each of these 50,000 light curves.

The confusion matrix is a standard way of visualizing the performance of algorithms dealing with classification problems. For each light curve and regardless of the profile shape, the true value of $r_{1/2}$ (the actual class) and the output of the trained network (the predicted class) are stored in a matrix. If all the entries lie on the diagonal, then all the predicted classes coincide with the actual classes and the network predictions are 100 per cent accurate. We obtain a 96 per cent accuracy for the validation set, whose confusion matrix is shown in the left panel of Fig. 4.

3.3 Further tests

The validation set consists of light curves that are generated from the same underlying magnification map as those used in the training. Thus, they are expected to have very similar properties, which explains the 96 per cent accuracy of the predictions - the network performs well on the same kind of data that it was trained on.

To push the limits of the predictive power of our network, a number of ‘unfair’ tests was performed using data sets that were designed to be different to the training. We examined light curves from a profile with a shape unknown to the network - the exponential profile, from different realizations of statistically equivalent maps (having the same $\kappa, \gamma$ as the training but different caustic networks), maps with varying $\kappa, \gamma$ throughout the parameter space, and difference light curves. In all of these tests the same neural network was used, trained only on light curves generated from the magnification map shown in the top left of Fig. 2 and the profile shapes of Table 1. We point out once more that the network output is always the class of $r_{1/2}$ listed in Table 2, regardless of the profile shape.

3.3.1 Classifying an unknown profile

The first test is to classify light curves from a profile whose shape is unknown to the network, i.e. different to the shapes used in the training set. We used the exponential profile, shown in Table 1. Because this profile is unbounded for $r \to 0$, we introduce an artificial cut off at 10 times the brightness value at $r_{1/2}$ (at $r \approx 0.2 r_{1/2}$) and assume this value for the smaller radii. Due to this sharp peak at $r \to 0$, this profile is not expected to smooth out light curve features by much compared to a point source. Also, this profile is quite...
3.3.3 Difference light curves

In order to obtain microlensing light curves from real observations, one has to subtract the time series of the brightness measurements between pairs of multiple quasar images (after correcting for the time delay). This removes the (unknown) intrinsic variability of the quasar itself, which is the same in all the multiple images, and the residual signal is attributed to microlensing. However, this results in a fundamental issue that cannot be addressed: there is no way to know which of the two images is undergoing some microlensing effect and to what extent (or if both are at the same time). Therefore, to model observational light curves one has to consider such difference light curves.

Difference light curves result by subtracting the brightness in magnitudes (equivalent to dividing the magnification) of two light curves that originated from different quasar images, most likely having quite different $\kappa, \gamma$ parameters. This means that each difference light curve is now associated with two magnification maps rather than just one. To create such pairs of maps, we select 5 maps from the set of statistically equivalent maps used previously and combine them to form 10 unique pairs. It is pointed out that the same accretion disc profile (size and shape) has to be used with both maps before creating the difference light curves. Due to the combinatorial explosion of the size of the data, 10 of the previously generated light curves per profile shape, size, and map are selected, resulting in 50,000 difference light curves in total (5,000 for each class of $r_{1/2}$).

The confusion matrix for this test set is shown in Fig. 4. The matrix is still mostly diagonal, with 42 per cent of the light curves classified correctly. We point out that applying this combination and division operation to the data is expected to introduce new features that the network has not been trained to recognize, in a way, creating a new type of data. Thus, the observed drop in the number of correct predictions is not surprising.

Finally, for this test we used pairs of maps with the same $\kappa, \gamma$ values, although real pairs of quasar images are expected to have quite different ones. As it turns out, the effect of varying $\kappa, \gamma$ is much more prominent even without considering difference light curves, and is examined separately below.

3.3.4 Parameter space

For this test, a set of 108 magnification maps is selected throughout the $\kappa, \gamma$ parameter space, shown in Fig. 5. For each map, profile shape, and size, 200 light curves are selected that have a maximum magnification of at least 1.5, resulting in a total of $\approx 10^6$. The confusion matrices are calculated in each case, but are hard to visualize in the $\kappa, \gamma$ parameter space. Instead, the percentage of the correct predictions is computed for each map (the confusion matrix trace), for all the profile shapes, and shown in Fig. 5.

The highest percentage of correct predictions is between 70 to 80 per cent, and it occurs near $\kappa = 0.09, \gamma = 0.55$, the values that the network was trained on. The network per-
forms well in a wide area around these values, which is not surprising because the caustic networks and the statistical properties of the maps are quite similar there (see fig. 4 of Vernardos & Fluke 2013, for the magnification histograms in the parameter space).

The network performs the worst near the critical line (black line on Fig. 5) for high values of $\kappa$. The critical line is where two macro-images of the quasar merge or annihilate and the macro-magnification $\mu_{\text{macro}} = [(1 - \kappa)^2 - \gamma^2]^{-1}$ diverges. The number of microlenses dramatically increases and caustic networks become denser and superimposed, leading to quite different light curves. The magnification map shown in the bottom panel of Fig. 2 lies much closer to the critical line and the difference with the map used in the training is striking: there are several and constant small variations in the light curve as the source is continuously traversing dense caustic networks, as opposed to simple caustic crossing events displaying the characteristic double peak as the source enters and exits an isolated diamond caustic.

3.4 Recovering accretion disc parameters

Measuring the accretion disc size in a range of wavelengths can constrain the free parameters in equation (1), i.e. the size $r_0$ and slope $\nu$. Fixing $r_0$ and $\nu$ to some fiducial values, in this case $2.59 \times 10^{15}$ cm and 4/3, matches a wavelength to each accretion disc size ($r_{1/2}$). We can then use the sizes predicted in each wavelength by the neural network to infer the underlying $r_0$ and $\nu$ values.

However, we are dealing with what is effectively a pattern recognition problem and any kind of physical scale is irrelevant; stretching or compressing the light curve data in the length or magnification directions does not affect the network predictions, as long as the $r_{1/2}/R_{\text{Ein}}$ and pixels$/R_{\text{Ein}}$ ratios, which determine the relevant size of the profiles and the light curve resolution, remain fixed. Thus, the actual values of $r_0$ and $\nu$ used to match sizes to wavelengths are actually irrelevant.

For each trajectory on the magnification map we extract the light curve data of a given accretion disc profile shape for all values of $r_{1/2}$. Predicting the size of each light curve in such collections leads to a measurement of $r_{1/2}$ as a function of wavelength. Converting these measurements to logarithmic space and performing a simple linear fit leads to a measurement of $r_0$ and $\nu$.

We use the test sets for an exponential profile, statistically equivalent maps, and difference light curves created in Section 3.3, which provide 1,000, 7,000 and 5,000 measurements of $r_{1/2}$ vs wavelength. The resulting values for $\nu$ and $r_0$ are shown in Table 5. The probability density contours and histograms are shown in Fig. 6.

It is pointed out that the shapes of the probability contours in Fig. 6 and the resulting fractional errors on the derived $r_0$ and $\nu$ parameters do not depend on their absolute values and only reflect the extent to which the network fails to predict the correct sizes in each case (reflecting the confusion matrix of each test set). The measurements of $r_0$ and $\nu$ from the validation set (not shown) are therefore expected to be even closer to the true underlying values.

4 CONCLUSIONS

We used a machine learning approach based on a Convolutional Neural Network (CNN) to study simulated quasar microlensing light curves for the first time. Our algorithm was trained on and applied to hundreds of thousands of such light curves. We demonstrated that our method is able to recover the correct size of the underlying accretion disc at 96 per cent accuracy for the validation set, and at 82 for a test set of statistically equivalent light curves, generated from magnification maps with the same $\kappa, \gamma$ values but different random microlens positions. When light curves from a profile whose shape is unknown to the network are used (the ‘exponential’ profile), the accuracy drops at 15 per cent but the corresponding confusion matrix shown in Fig. 4 remains almost diagonal (the network underpredicts the sizes by 1-3
The focus of this work is to study light curves in an abstract way, taking into account only the size of the accretion disc with respect to $R_{\text{Ein}}$. Thus, light curve data were assumed to be long ($5\times R_{\text{Ein}}$), continuous, and containing some minimum amount of information (all light curves were selected to have a maximum magnification of $1.5$, guaranteeing at least some variations). The most important caveat is the length of the light curves, which determines the impact of other factors not considered here, like the effective velocity model. However, the applicability of this new technique to real data will ultimately depend on how well observational effects, like sparse data, gaps, irregular sampling, etc, can be taken into account and mitigated.

In this work we demonstrated the feasibility of analyzing hundreds of thousands of microlensing light curves. This method opens up a whole new range of possibilities of quantifying light curve properties systematically and consistently throughout the microlensing parameter space. This is important for the upcoming all-sky surveys like LSST, which will provide data on thousands of microlensed quasars. This new CNN approach was explored along two perpendicular directions: improving the network predictions on a restricted training set by modifying its architecture, and then applying it to a wider range of data. The next step is to couple them, include observational effects, and improve the predictions on all the new architecture/data possibilities (difference light curves, throughout the $\kappa, \gamma$ parameter space, as a function of light curve length, etc). The new technique presented here brings us closer to addressing the pertinent question: “At which point all the available information has been extracted from a quasar microlensing light curve?”. Eventually, the most critical parameter will be the lightcurve length, which is necessary for designing future observations, follow-up campaigns, etc.

ACKNOWLEDGEMENTS

GV is supported through an NWO-VICI grant (project number 639.043.308). The authors would like to thank the referee James H. H. Chan for his comments that improved the final version of this work.

REFERENCES

Abolmasov P., Shakura N. I., 2012, Monthly Notices of the Royal Astronomical Society, 427, 1867

Bate N. F., Floyd D. J. E., Webster R. L., Wyithe J. S. B., 2008, Monthly Notices of the Royal Astronomical Society, 391, 1955

Machine learning and quasar microlensing

Figure 6. Probability density and histograms for the accretion disc parameters $r_0$ and $\nu$ (see equation 1). Measurements are obtained from the exponential profile, equivalent maps, and difference light curve datasets. The shape of the probability density surface depends purely on the ability of the network to correctly predict the accretion disc size, and is related to the confusion matrices of these datasets, shown in Fig. 4. Dashed lines indicate the true underlying values used in this case, i.e. $\nu = 4/3$ and $r_0 = 2.59 \times 10^{14}$ cm. Contours are drawn at the 68, 95, and 99 per cent confidence intervals.

Classifying accretion disc sizes in many wavelengths leads to measurements of the accretion disc temperature profile via equation (1). Fig. 6 shows the probability of such measurements for thousands of light curves from three different test sets. The errors on recovering the correct underlying structure parameters depend purely on the accuracy of the network size predictions. In particular, for the 'exponential' profile shape the absolute size predictions are well-off the truth, as expected from the confusion matrix, but the relative scaling, and hence the power-law index $\nu$, is correctly recovered.

Our network was trained on a single magnification map with $\kappa = 0.09, \gamma = 0.55$ and then applied to the entire $\kappa, \gamma$ parameter space shown in Fig. 5. Predictions near the training map location are systematically higher than elsewhere in the parameter space, indicating that there is a potentially measurable effect of the $\kappa, \gamma$ values on the light curves. This could lead to measurements of the smooth matter fraction - constraining the partition between smooth (dark) and compact (stellar) matter - which are crucial to lifting the degeneracy between dark matter and the initial mass function in the lens (e.g. Oguri et al. 2014).

The focus of this work is to study light curves in an abstract way, taking into account only the size of the accretion disc with respect to $R_{\text{Ein}}$. Thus, light curve data were assumed to be long ($5\times R_{\text{Ein}}$), continuous, and containing some minimum amount of information (all light curves were selected to have a maximum magnification of $1.5$, guaranteeing at least some variations). The most important caveat is the length of the light curves, which determines the impact of other factors not considered here, like the effective velocity model. However, the applicability of this new technique to real data will ultimately depend on how well observational effects, like sparse data, gaps, irregular sampling, etc, can be taken into account and mitigated.

In this work we demonstrated the feasibility of analyzing hundreds of thousands of microlensing light curves. This method opens up a whole new range of possibilities of quantifying light curve properties systematically and consistently throughout the microlensing parameter space. This is important for the upcoming all-sky surveys like LSST, which will provide data on thousands of microlensed quasars. This new CNN approach was explored along two perpendicular directions: improving the network predictions on a restricted training set by modifying its architecture, and then applying it to a wider range of data. The next step is to couple them, include observational effects, and improve the predictions on all the new architecture/data possibilities (difference light curves, throughout the $\kappa, \gamma$ parameter space, as a function of light curve length, etc). The new technique presented here brings us closer to addressing the pertinent question: “At which point all the available information has been extracted from a quasar microlensing light curve?”. Eventually, the most critical parameter will be the light curve length, which is necessary for designing future observations, follow-up campaigns, etc.

ACKNOWLEDGEMENTS

GV is supported through an NWO-VICI grant (project number 639.043.308). The authors would like to thank the referee James H. H. Chan for his comments that improved the final version of this work.

REFERENCES

Abolmasov P., Shakura N. I., 2012, Monthly Notices of the Royal Astronomical Society, 427, 1867

Bate N. F., Floyd D. J. E., Webster R. L., Wyithe J. S. B., 2008, Monthly Notices of the Royal Astronomical Society, 391, 1955
Bate N. F., Vernardos G., O’Dowd M. J., Neri-Larios M. D., 2018, Monthly Notices of the Royal Astronomical Society, 479, 4796
Blackburne J. A., Pooley D., Rappaport S., Schechter P. L., 2011, The Astrophysical Journal, 729, 34
Chartas G., et al., 2016, Astronomische Nachrichten, 337, 356
Eigenbrod A., Courbin F., Meylan G., Agol E., Anguita T., Schmidt R. W., Wambsganss J., 2008, Astronomy & Astrophysics, 490, 933
Floyd D. J. E., Bate N. F., Webster R. L., 2009, Monthly Notices of the Royal Astronomical Society, 398, 233
Hála P., 2014, preprint (astro-ph/1412.8341)
He K., Zhang X., Ren S., Sun J., 2016, in Proceedings of the IEEE conference on computer vision and pattern recognition. p. 770
Hezaveh Y. D., Levasseur L. P., Marshall P. J., 2017, Nature, 548, 555
Huertas-Company M., et al., 2015, Astrophysical Journal, Supplement Series, 221, 8
Jimenez-Vicente J., Mediavilla E., Kochanek C. S., Muñoz J. A., Motta V., Falco E., Mosquera A. M., 2014, The Astrophysical Journal, 783, 47
Kayser R.,Refsdal S., Stabell R., 1986, Astronomy & Astrophysics, 166, 36
Kingma D. P., Ba J., 2014, preprint (arXiv:1412.6980)
Kochanek C. S., 2004, The Astrophysical Journal, 605, 58
LSST Science Collaborations 2009, preprint (astro-ph/0912.0201)
Lamisse F., Ma Q., Li N., Collett T. E., Li C. L., Ravanbakhsh S., Mandelbaum R., Péczos B., 2018, Monthly Notices of the Royal Astronomical Society, 473, 3895
Lecun Y., Bengio Y., Hinton G., 2015, Nature, 521, 436
MacLeod C. L., et al., 2015, Astrophysical Journal, 806, 258
Morgan C. W., Kochanek C. S., Morgan N. D., Falco E. E., 2010, The Astrophysical Journal, 712, 1129
Mortonson M. J., Schechter P. L., Wambsganss J., 2005, The Astrophysical Journal, 628, 594
Oguri M., Rusu C. E., Falco E. E., 2014, Monthly Notices of the Royal Astronomical Society, 439, 2494
Paczyński B., 1986, The Astrophysical Journal, 301, 503
Petrillo C. E., et al., 2019, Monthly Notices of the Royal Astronomical Society, 482, 807
Poidexter S., Kochanek C. S., 2010, The Astrophysical Journal, 712, 668
Poidexter S., Morgan N. D., Kochanek C. S., 2008, The Astrophysical Journal, 673, 34
Rojas K., Motta V., Mediavilla E., Falco E., Jiménez-Vicente J., Muñoz J. A., 2014, The Astrophysical Journal, accepted
Schechter P. L., Wambsganss J., 2002, The Astrophysical Journal, 580, 685
Schmidt R. W., Wambsganss J., 2010, Gen. Relativ. and Gravit., 42, 2127
Shakura N. I., Sunyaev R. A., 1973, Astronomy & Astrophysics, 24, 337
Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R., 2014, The Journal of Machine Learning Research, 15, 1929
Stivaktakis R., Tsagkatakis G., Moraes B., Abdalla F., Starck J.-L., Tsakalides P., 2018, preprint (astro-ph/1809.09622)
Tuccillo D., Huertas-Company M., Decenci`ere E., Velasco-Forero S., Domínguez Sánchez H., Dimarco P., 2018, Monthly Notices of the Royal Astronomical Society, 475, 894
Vernardos G., Fluke C. J., 2013, Monthly Notices of the Royal Astronomical Society, 434, 832
Vernardos G., Fluke C. J., 2014, Monthly Notices of the Royal Astronomical Society, 445, 1223
Vernardos G., Fluke C. J., Bate N. F., Croton D. J., 2014, The Astrophysical Journal Supplement Series, 211, 16
Vernardos G., Fluke C. J., Bate N. F., Croton D. J., Vohl D., 2015, The Astrophysical Journal Supplement Series, 217, 23
Figure A1. Confusion matrices (actual versus predicted classes of $r_{1/2}$) per profile shape for the set of statistically equivalent maps. The middle panel of Fig. 4 is the sum of these constituent confusion matrices. Each row of the matrices is normalized to 100 per cent. The higher the percentage on the diagonal the more accurate the network predictions. The different classes are listed in Table 2.