

Fast visual inspection of Euclid gravitational lens challenge data

Neal Jackson

Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, School of Natural Sciences, Faculty of Science and Engineering, University of Manchester, Oxford Road, Manchester M13 9PL

31 January 2024

ABSTRACT

Identification of rare objects, such as strong gravitational lenses, in surveys is a serious challenge in an era where surveys can contain billions of objects. Although neural networks can reliably process very large numbers of images and be trained to recognise types of structure, most surveys will inevitably incorporate a final stage in which images will be inspected by eye. Efficient inspection requires attention to the details of image display, in particular for surveys which are conducted in several bands and where the signal-to-noise and resolution may vary between bands. An algorithm in which significant pixels are identified is described; each significant pixel is classified as a blue or red pixel and painted with a colour which compromises between display of the colour and not obscuring subtle differences in pixel brightness. 10^5 cutout images simulated by the Euclid Strong Lensing Working Group were processed in a blind test, resulting in a virtually 100% purity sample with completeness ranges from 6–10%. 6000 images could be processed per hour, corresponding to a potential rate of a million objects per investigator per month.

Key words: gravitational lensing: strong – surveys

1 INTRODUCTION

For most of the history of astronomy, data were inspected by eye. The eye’s capacity for pattern recognition makes it suitable for discovery of rare objects, such as gravitational lenses, in samples of images. The combination of the eye’s ability to recognize particular shapes, together with complementary information about colour, is particularly appropriate for gravitational lens systems, where a background object is imaged into multiple, tangentially distorted images whose colours are generally different from that of the lensing object.

In many early surveys for gravitational lenses, visual inspection was used either as a primary recognition tool or as a major part of the selection process. For example, all 16503 images from the radio-frequency Cosmic Lens All-Sky Survey (CLASS) were visually inspected by multiple observers during the production of the final sample of 22 lenses (Myers et al. 2003; Browne et al. 2003), despite the intrinsically simple radio structure of the flat-spectrum, pointlike sources which formed the parent sample. 4000 sources were inspected by eye as part of the MIT-Greenbank survey, to recognise more complex lens systems involving lensed extended radio sources (Hewitt et al. 1988, 1989).

In large surveys at optical wavebands, simple cuts in morphology or colour have been used in many cases to fil-

ter large samples of potentially lensed sources, before a final stage of optical inspection or confirmatory spectroscopy. Morphology cuts include the identification of quasars which are not pointlike on optical images (e.g. Inada et al. 2003; Lemon et al. 2017, 2018, 2019) and thus may indicate lensed quasar systems, the detection of bright red lensing galaxies (Faure et al. 2008) or the presence of arcs, especially blue arcs (e.g. Lenzen et al. 2004; Seidel & Bartelmann 2007; Cabanac et al. 2007; Gavazzi et al. 2012, 2014) to find lensed galaxies.

More recent surveys for strong lenses have become steadily larger, with parent population sizes increasing from a few tens of thousands to millions, and shortly to billions with projects such as the Vera Rubin Observatory (Large Synoptic Survey Telescope) and Euclid. This means that automatic selection is needed to provide a sample that can be inspected in detail. One option is to expand the human selection using citizen science projects, of which the most advanced is SPACE WARPS (Geach et al. 2015; Marshall et al. 2016; More et al. 2016). The other strategy is to move automatic selection beyond simple algorithmic prescriptions to machine learning. The use of convolutional neural networks and similar architectures is now widely used for selection of lens morphologies (Petrillo et al. 2017; Hartley et al. 2017; Schaefer et al. 2018; Avestruz et al. 2019; Petrillo et al. 2019a,b). In blind tests using simulated data, the best neural

network algorithms often achieve better results than eyeball inspection (Metcalf et al. 2019). However, machine learning algorithms must be provided with training sets. This can be done using simulated data, provided that it can be guaranteed that the simulated data accurately captures all of the features of the real data including the sources of false positives. Alternatively, it can be done using real data, using by-eye inspection of a small sample. In either case it is possible for the eye to detect unexpected or untypical patterns which may be missed by an algorithm whose training data does not contain them.

The aim of this work is to optimize the speed at which by-eye detection of gravitational lenses may be accomplished, while maintaining as far as possible the completeness (detection of real lenses) and purity (freedom from false positives) that would be achieved by more leisurely inspection. Previous investigations have typically involved inspection of a few thousand objects over a period of weeks. A more intensive inspection of the full COSMOS sample, of which the subset with bright red galaxies was published by Faure et al. 2008, was undertaken by Jackson 2008 which covered 250000 objects over a period of four weeks. Subsequent increases in speed resulted in the processing of 100000 objects over 48 hours by two investigators (Metcalf et al. 2019). Here I explore the maximum feasible rate at which sample inspection for gravitational lens identification can be made. The final rate achieved, with careful attention to method, process and image display, is about 6000 objects per hour, corresponding to approximately 1 million objects processed per investigator-month.

2 IMAGE DISPLAY

2.1 Algorithm: purpose and execution

Euclid images are taken in four bands: the VIS band in the visible range, and in three bands (Y, H and J) in the infrared. Simulations of realistic cutout images in each of the bands were provided by the Bologna lens group as part of the Euclid Data Challenge II (Metcalf et al. 2020, in preparation). The IR images are provided as 66×66 pixel cutouts with $0''.3$ pixel scale, and the VIS image as 200×200 pixel cutouts with $0''.1$ pixel scale. The IR images have worse spatial resolution by factors of 2–4. A sample dataset is illustrated in Fig. 1.

The challenge of data visualisation is to preserve, as far as possible, the resolution of the VIS image; to incorporate the colour information provided by the multiple bands and ensure that objects of different colours can clearly be differentiated; to make sure that gradients in intensity and gradients in colour are clearly differentiable; and to suppress noise, which distracts the eye, while preserving any systematic features in the image and painting them in the correct colour. The painting algorithm has to preserve both the colour differences, and also not obscure the subtle differences in brightness which are needed to pick out, for example, faint arcs and rings in the presence of background emission. No simple algorithm achieves all of these objectives, which are all necessary in order to carry out fast visual inspection of large numbers of images. We first describe the adopted algorithm, and then discuss the undesirable consequences of less complicated algorithms.

The overall philosophy of the algorithm is to identify pixels which are significantly above the noise. Of those pixels, we then identify pixels which are bluer than the rest of the image, and pixels which are redder than the rest; these two classes of pixels are painted separately. The colourscale for each pixel is chosen according to Fig. 2, as a compromise between the use of the bluest colours (which obscures the intensity information, and corresponds to $\theta = 90^\circ$ on the figure) and use of the full intensity range (which corresponds to $\theta = 0^\circ$, and extinguishes the colour information). The steps used are as follows:

- (i) The Y, H and J images are shifted and interpolated on to the 200×200 grid of the VIS images.
- (ii) Two smoothed images are produced from the VIS image, one (henceforth the V image) smoothed by 0.5 pixels, and one (henceforth B) smoothed by 0.95 pixels. This gives five images (B, V, Y, H, J) for further processing.
- (iii) For all five images, a mask is made to isolate source pixels. A pixel is considered part of a source if it is 1.2σ above the background, and 5 connected pixels are required for a source to be identified as above the background. Pixels not identified as part of sources, using this mask are then examined and used to determine the median background level¹. The median of each image's background is then subtracted from all its pixel values, and the resulting images are normalised by dividing by the sum of all pixel values.
- (iv) An IR image is derived by averaging the Y, H and J images, leaving three images (V , B , $I = (Y+H+J)/3$) for further processing. The previous step is used to calculate the standard deviation of pixels in each of these three images (σ_V , σ_B , σ_I), in regions without flux from sources. This step implies that the final image will be a two-colour rather than three-colour image.
- (v) An array b is formed, which is 1 wherever $B > 3\sigma_B$, and either $I > 3\sigma_I$ and $B > 1.05I$, or $B > 3\sigma_I$ and $I < 3\sigma_I$; otherwise it is 0. Italic font represents pixel values in the corresponding array. This array b isolates pixels which are significantly blue.
- (vi) An array r is formed, which is 1 wherever $B > 3\sigma_B$, and either $I > 3\sigma_I$ and $I > 1.05B$, or $B < 3\sigma_I$ and $I > 3\sigma_I$; otherwise it is 0. This array r isolates pixels which are significantly red.
- (vii) The V image is clipped with a user-specified minimum level V_{\min} , and at a maximum level which corresponds to the V_{\max} percentile of sorted pixel values in the image, and is then scaled with square-root scaling between those two values. For the Euclid images we used $V_{\min} = 5.8 \times 10^{-13}$ and $V_{\max} = 99.9$. The pixel values of this scaled V image are defined as V_s .
- (viii) Very bright images are subject to a slightly modified procedure, involving modification of V_{\max} by subtraction of $p_c(\log(\Sigma V) - \log(p_h))$, where p_c and p_h are parameters modifiable by the user. For the Euclid images, values of $p_h = 5.0 \times 10^{-8}$ and $p_c = 1.3$ were used.
- (ix) An RGBI final image is formed, with $G=0$. The R plane pixel values are given by $r(1-V_s)\sin\theta$ and the B plane values by $b(1-V_s)\sin\theta$.

¹ These procedures are carried out with the Python `photutils.make_source_mask` and `astropy.stats.sigma_clipped_stats` functions respectively.

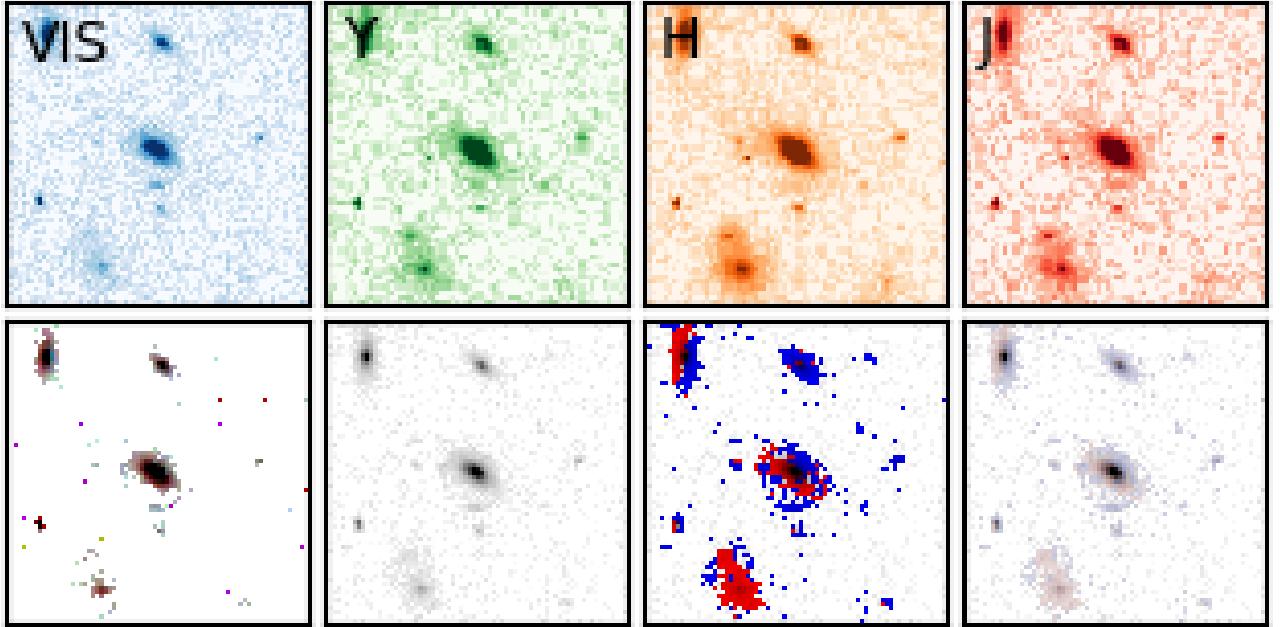


Figure 1. Top row: original data in VIS, Y, H, J. Bottom row: Left: simple algorithm, with an upper and lower cutoff and square-root scaling in between, with the VIS, Y and H images used as blue, green and red colours. The other three panels show the algorithm described in the text, with $\theta = 0^\circ, 90^\circ$ and 25° respectively.

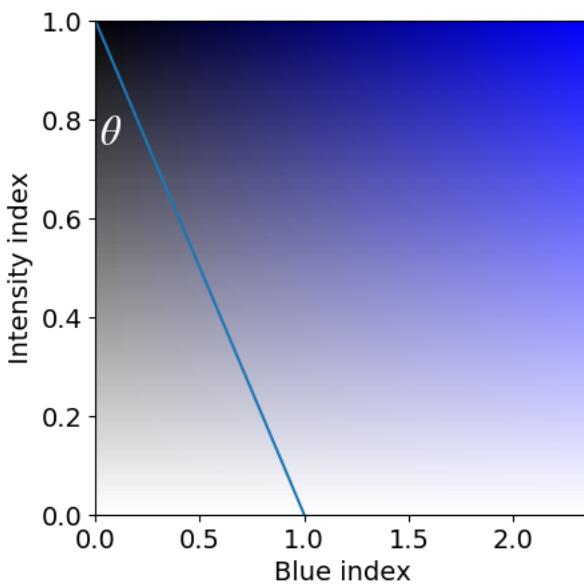


Figure 2. Colour locus for “blue” points in the image. The picture is defined to be square, for which choice the angle θ refers to the angle between the y -axis and the line from the point $(0,1)$ to the point $(1,0)$.

(x) Finally, the I plane pixel values are given by $1 - (1 - V_s) \cos \theta$ for pixels where $b = 1$ or $r = 1$. The overall arrangement amounts to painting each pixel with a colour along the

line in Fig. 2, where the colours are chosen from the red or blue version of the Figure.

Fig. 1 also shows the effect of other algorithms. Fairly extensive experiments were undertaken with simpler arrangements, notably variations of using simple three-colour combinations using three of the bands, together with upper and lower cutoffs and square-root scaling of the pixel values in between. It was found not to be possible to achieve as good results as the more complicated algorithm described in the rest of this subsection.

2.2 Production of by-eye identifications for multiple images

Once the image panels are made for each of the individual cutout images, they are pasted together in groups of 8×4 into larger panels, an example of which is shown in Fig. 2.1. Each larger panel is presented to the user sequentially, and the panel’s brightness stretch can be adjusted by a slider at the bottom of the panel. Identifications are made by clicking numbers from 1 (possibly a lens) to 4 (certain lens) within each of the subpanels, and a marker appears on each subpanel which has been identified as a potential lensed system. Judgements can be revisited while the panel is displayed, and the list of identifications is stored in a file. With practice, each group of 32 images can be inspected in approximately 20 seconds, leading to a rate of about 6000 objects per hour. This is approximately double the rate achieved in the previous Euclid data challenge Metcalf et al. (2019), and the improvement can be attributed to the improvement in

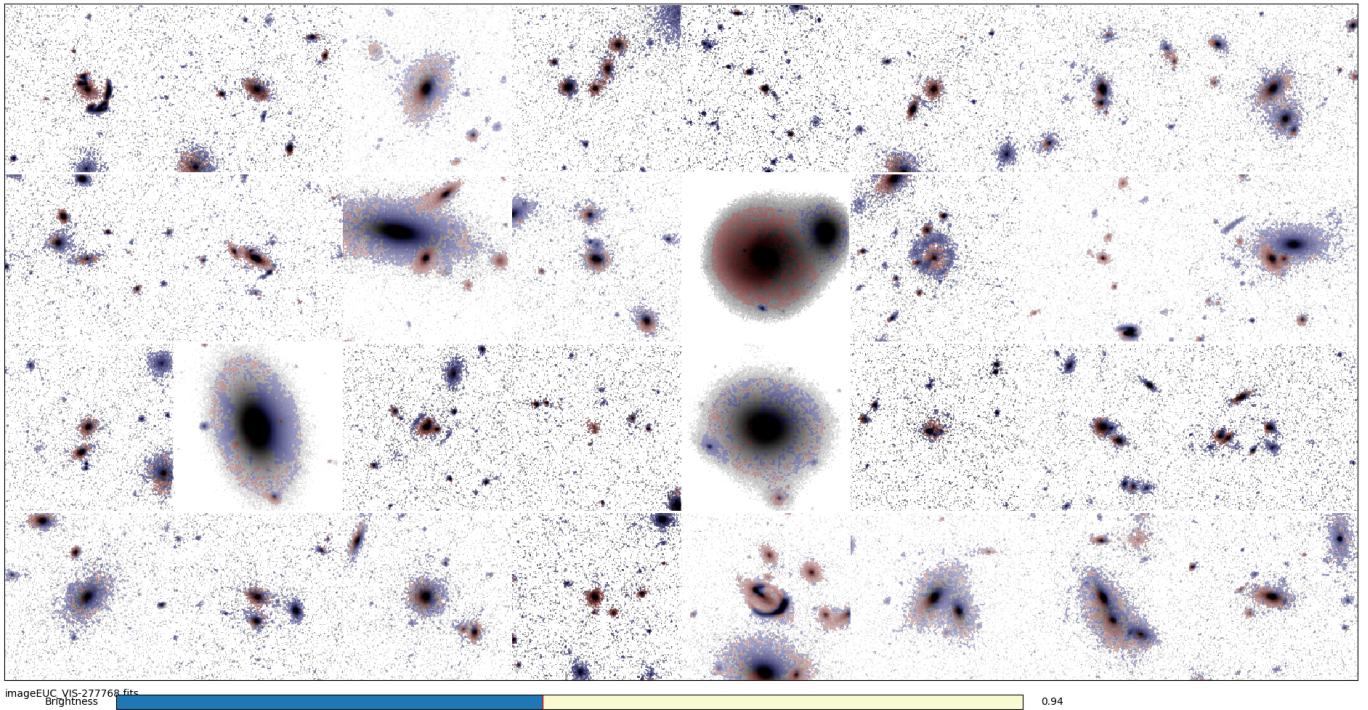


Figure 3. An example screen of 32 sources. With practice, lenses in such a screen can be identified manually in about 20 seconds.

the colour scaling which reduces clutter in the images and makes instant judgements much easier. We consider that this rate is likely to be the highest achievable, however; at this point it is probably limited by the human capacity for ingestive and decision-making. Nevertheless, this rate implies that about 1 million objects could in principle be inspected by an observer in about 1 month.

3 RESULTS FROM THE EUCLID DATA CHALLENGE II

The Euclid Consortium’s working group on strong lensing ran a data challenge in 2018 (Metcalf et al. 2019) to test strong lens-finding codes. Participants classified a test sample of 100000 objects, within a short time period (48 hours). A number of machine learning methods were represented, including a variety of convolutional neural network (CNN) algorithms, a Support Vector Machine (SVM) and an earlier version of the human-inspection algorithm discussed in this paper. In general, deep-learning CNNs performed very well, with human inspection ranked approximately in the middle of the ranking. A second challenge, on which this analysis is based, was held early in 2020 (Metcalf et al. 2020, in preparation) which included more realistic background and source galaxies, with realistic false positives from intrinsic structure such as polar ring galaxies.

The aim of lens classification is to maximise the number of true positive (TP) identifications, while minimising the number of false positives (FP) where non-lenses are erroneously identified as lenses. If we write TN for the number of non-lenses correctly identified as non-lenses, and FN as the number of lenses incorrectly identified as non-

lenses, we can define two standard quantities: the precision, $P \equiv TP/(TP + FP)$, which describes the purity of the lens sample, and the recall, $R \equiv TP/(TP + FN)$, which describes the completeness of the sample, i.e. what fraction of lenses in the examined sample are actually detected.

In the actual blind challenge, 10^5 fields were simulated, as described in Section 2. 76920 of these fields contained lenses and the rest did not; however, many of the lens fields contained lens system whose lensed structure was not visible. Objects were considered to be lenses if the effective magnification was $\mu > 1.6$ and if the number of source pixels exceeded 20. These criteria reduced the 76920 lens-containing fields to 49596 regarded as actually containing lenses; that is, about 1/3 of the fields containing lenses were not classified within the challenge as such due to low magnification. As described in section 2, these fields were observed at a rate of just over 6000 per hour, with 16 hours of observer time spread over two days; results were submitted 46 hours after downloading of the data. The truth data was supplied after the close of the challenge.

Objects were classified on a scale from 0 (no evidence for lensing) to 4 (certain or nearly certain lens). Table 1 shows the raw outcome, considering the $\mu > 1.6$ criterion to define a lens.

It is useful to examine the false positives further. If we examine the 24 class 3 and 4 identifications which are recorded as false positives, we find that all of these (Fig. 4) belong to the 1/3 of images with lensed structure which were not officially classed as lenses. Nevertheless, the lensed structure is clearly visible in all 24 objects. Hence, the objects which have a by-eye classification as highly probable or certain achieve 100% purity in the identification of 8% of lensed systems.

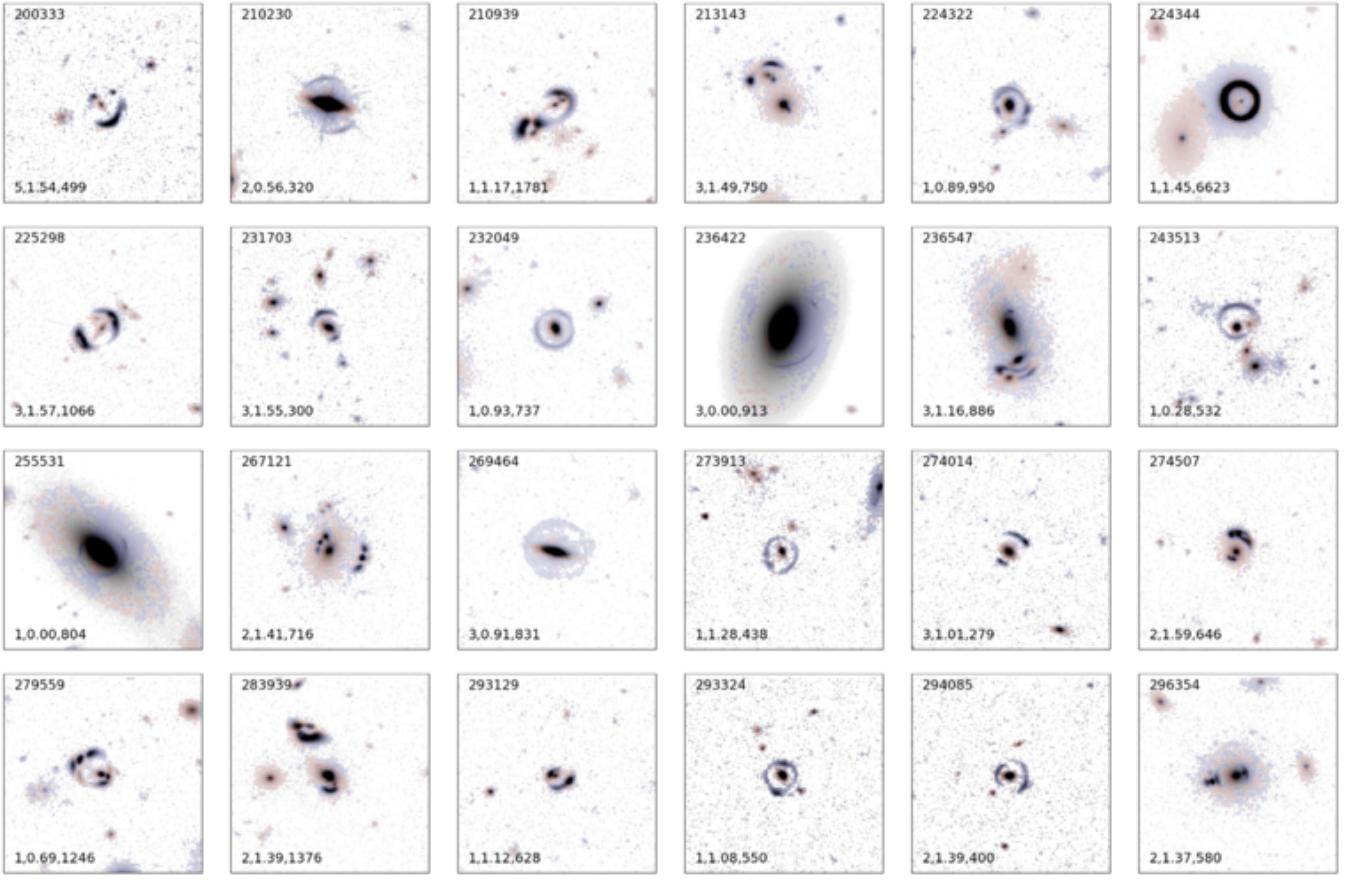


Figure 4. Objects identified as class 4 (certain) lenses by eye which were identified as false positives by the challenge classification. However, all of these images do have a lensed source, albeit with relatively low magnification. The numbers in the top of the plots are the index number within the challenge. The three numbers at the bottom are the number of lensed sources, the effective magnification, and the number of lensed pixels, respectively.

Confidence level	<i>TP</i>	<i>FP</i>	<i>TN</i>	<i>FN</i>
1 (possible)	9671	295	39925	50108
2 (probable)	6602	94	42994	50309
3 (highly probable)	4294	24	45302	50379
4 (certain)	2508	9	47088	50394

Table 1. Numbers of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) for the four different thresholds identified by eye (possible, probable, highly probable, certain). Each row contains all of the detections at confidence levels at or greater than that level. In the highly probable/certain categories, 4294 lenses are correctly detected together with 24 false positives, all of which actually contain lensed structure (see text).

Examining the further 70 objects from class 2 which are recorded as false positives, we find that only 4 actually contain no lensed structure (Fig. 5). Because of the ratio in the parent sample between non-lenses and lensed systems where the lensed structure is too faint to be visible, it is likely that of the order 2–4 more will also be misidentifications as lenses of systems with unobservably faint lensed structure.

In class 2, therefore, we have a sample of 99.9% purity, which contains about 12% of the lensed systems.

Finally, we examine the false negatives: Fig. 6 shows a random sample of 24 simulated lens systems which were not identified as possible lenses by eye. Careful examination of many of these objects shows no likely lens structure, and in other cases lens structure is visible, but could not be confidently distinguished from chance alignments of nearby galaxies without very much higher-quality images.

4 CONCLUSIONS

It is possible to carry out highly efficient inspection of large numbers of astronomical images to detect unusual features consistent with strong gravitational lensing. This requires attention to the algorithm for image display in order to achieve maximum efficiency. The algorithm presented here involves three stages: identification of pixels which are significantly above noise, identification of the colour (red or blue) to be assigned to each of these pixels, and painting of each pixel with an algorithm which compromises between showing the colour difference between pixels while preserving subtle details in pixel brightness.

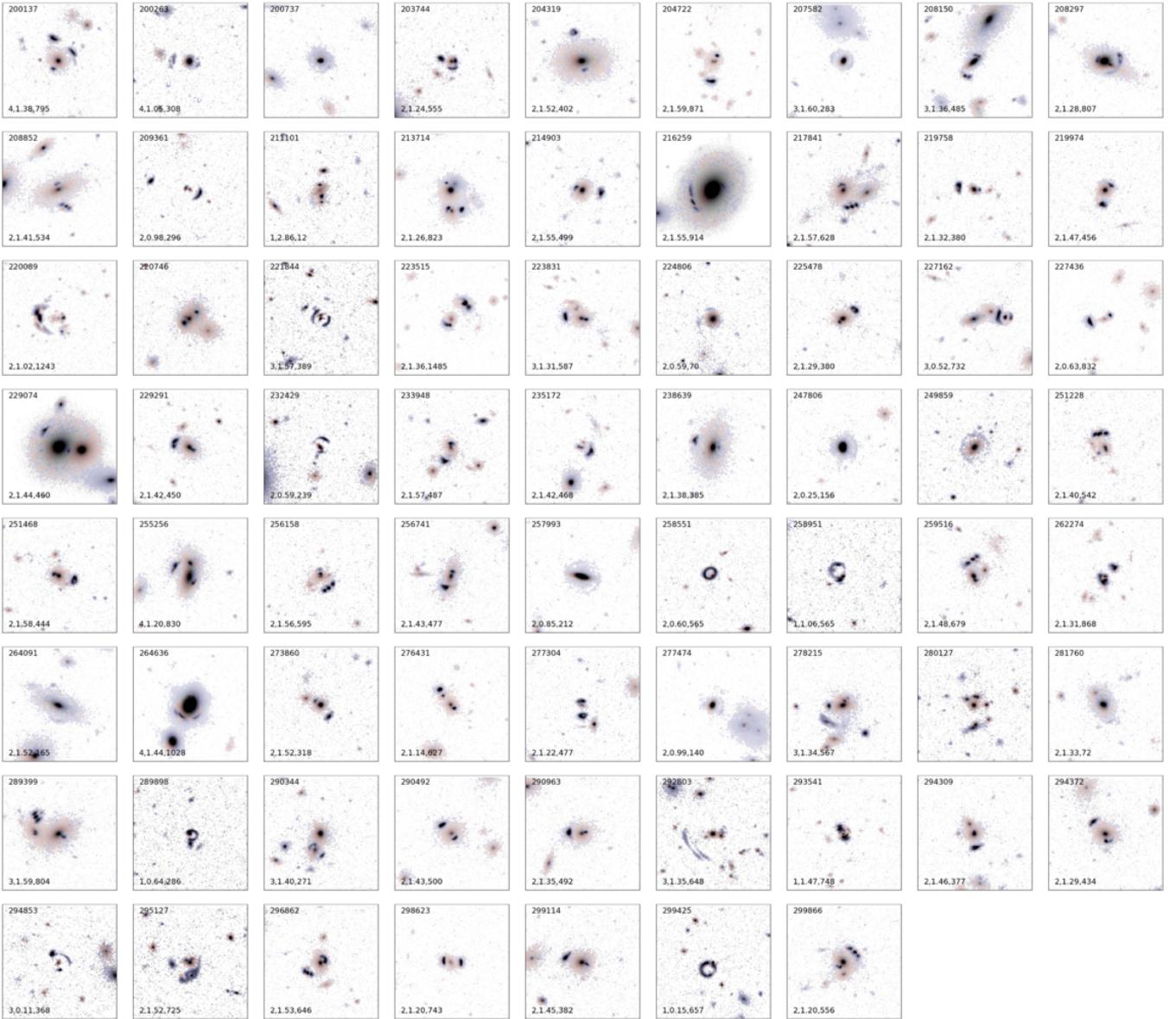


Figure 5. Objects identified as class 2 (probable) or class 3 (highly probable) lenses by eye which were identified as false positives by the challenge classification. Four of these (marked 200737, 220746, 249859, 280127) have no lensed source; the other 66 have lensed sources, but with lower magnification.

Testing of this procedure on a blind sample of 10^5 simulated multicolour images, with different colour bands having different resolution and signal-to-noise, has shown that reliable identifications are possible at a rate of 6000 per hour for a single inspector. With a realistic lens set, of the order of 6–8% of lens systems can be identified with close to 100% accuracy.

In the near future, instruments such as Euclid and the Vera Rubin Observatory (formerly the Large Synoptic Survey Telescope) will yield images of billions of astronomical objects. It is obviously unfeasible to classify such large numbers manually. However, one would ideally wish to apply automated classifiers to such large samples which passed as many potentially interesting objects as possible for subsequent human inspection, not only to discover strong lenses

but also to find unexpected or exotic lenses. In this case, the rate of 6000 classifications per hour gives an upper limit to the number of objects in a post-automatic-classification sample of a few million, which can be inspected by a number of observers in a few weeks.

REFERENCES

- Avestruz C., Li N., Zhu H., Lightman M., Collett T. E., Luo W., 2019, *ApJ*, **877**, 58
- Browne I. W. A., et al., 2003, *MNRAS*, **341**, 13
- Cabanac R. A., et al., 2007, *A&A*, **461**, 813
- Faure C., et al., 2008, *ApJS*, **176**, 19
- Gavazzi R., Treu T., Marshall P. J., Brault F., Ruff A., 2012, *ApJ*, **761**, 170

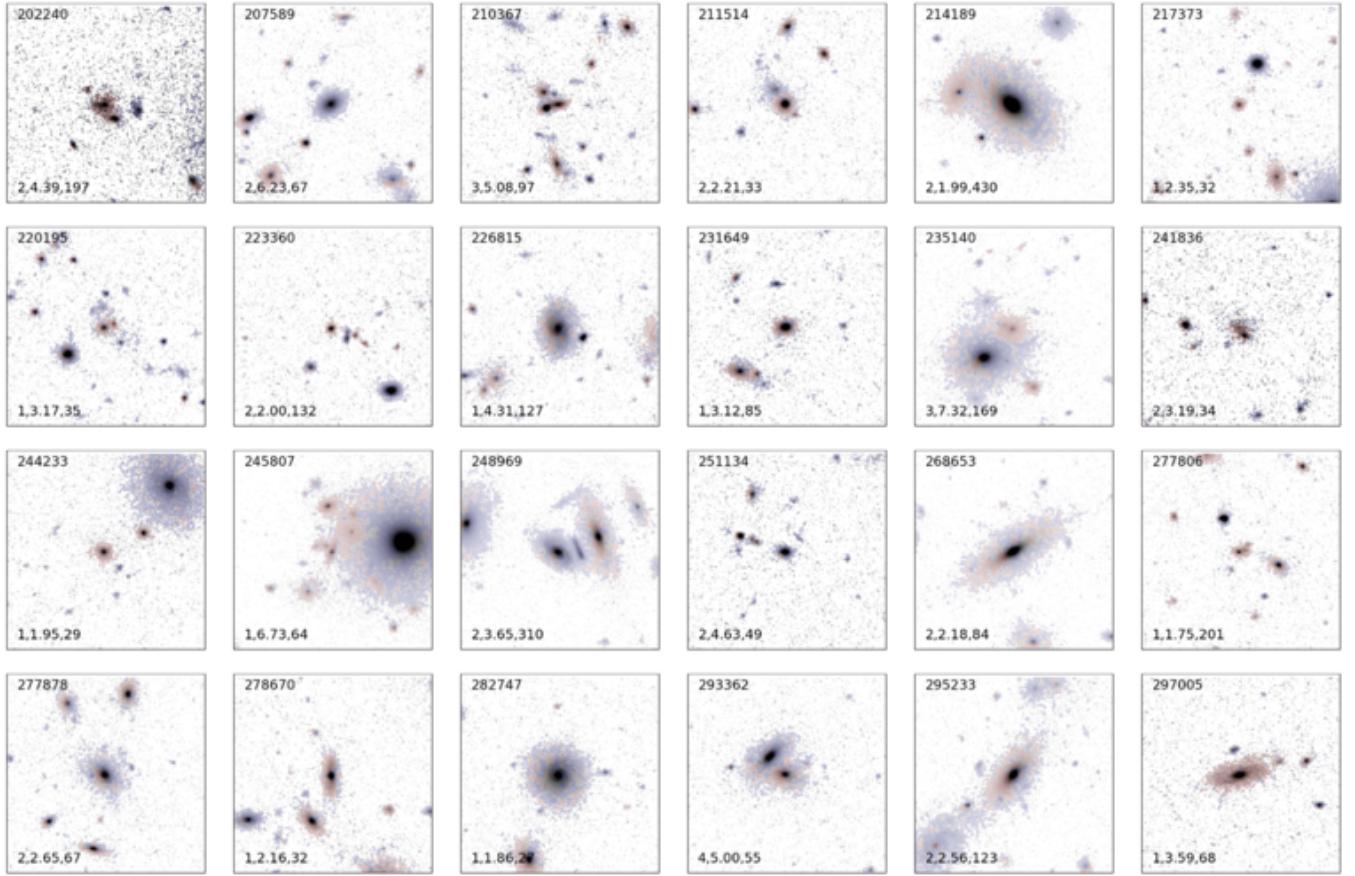


Figure 6. A randomly selected set of false negative sources, which were not selected by eye as potential lenses but which do have lensed structure.

Gavazzi R., Marshall P. J., Treu T., Sonnenfeld A., 2014, [ApJ, 785, 144](#)

Geach J. E., et al., 2015, [MNRAS, 452, 502](#)

Hartley P., Flamary R., Jackson N., Tagore A. S., Metcalf R. B., 2017, [MNRAS, 471, 3378](#)

Hewitt J. N., Turner E. L., Schneider D. P., Burke B. F., Langston G. I., 1988, [Nature, 333, 537](#)

Hewitt J. N., Burke B. F., Turner E. L., Schneider D. P., Lawrence C. R., Langston G. I., Brody J. P., 1989, Results of The VLA Gravitational Lens Survey. p. 147, doi:[10.1007/3-540-51061-3_49](https://doi.org/10.1007/3-540-51061-3_49)

Inada N., et al., 2003, [AJ, 126, 666](#)

Jackson N., 2008, [MNRAS, 389, 1311](#)

Lemon C. A., Auger M. W., McMahon R. G., Koposov S. E., 2017, [MNRAS, 472, 5023](#)

Lemon C. A., Auger M. W., McMahon R. G., Ostrovski F., 2018, [MNRAS, 479, 5060](#)

Lemon C. A., Auger M. W., McMahon R. G., 2019, [MNRAS, 483, 4242](#)

Lenzen F., Schindler S., Scherzer O., 2004, [A&A, 416, 391](#)

Marshall P. J., et al., 2016, [MNRAS, 455, 1171](#)

Metcalf R. B., et al., 2019, [A&A, 625, A119](#)

More A., et al., 2016, [MNRAS, 455, 1191](#)

Myers S. T., et al., 2003, [MNRAS, 341, 1](#)

Petrillo C. E., et al., 2017, [MNRAS, 472, 1129](#)

Petrillo C. E., et al., 2019a, [MNRAS, 482, 807](#)

Petrillo C. E., et al., 2019b, [MNRAS, 484, 3879](#)

Schaefer C., Geiger M., Kuntzer T., Kneib J. P., 2018, [A&A, 611,](#)

A2

Seidel G., Bartelmann M., 2007, [A&A, 472, 341](#)

This paper has been typeset from a [TeX/LaTeX](#) file prepared by the author.