Recognizing Strong Gravitational Lenses

Chris Davis
cpd@stanford.edu

Andrew McLeod
ajmcleod@stanford.edu

Abstract

The detection of a large, representative set of strong gravitational lenses could greatly aid in our understanding of cosmology. Unfortunately they are quite rare, and the best techniques now revolve around squads of scientists manually scanning through images. This is presently borderline unsustainable and will be laughably inefficient with the advent of the Large Synoptic Survey Telescope. Here we examine the effectiveness of convolution neural networks and transfer learning for automated detection algorithms of strong gravitational lenses. We use images from the SPACE WARPS *project, a citizen science initiative to examine tens of thousands of fields of galaxies for the presence of strong gravitational lenses. We find that using a convolution neural network trained on Galaxy morphologies as a feature extractor performs admirably but markedly worse than the citizen-scientists. Scripts used in the analysis of this paper are freely available at* `https://github.com/cpadavis/strongcnn`. *The images are currently only available to those who contact the author, but will be available to the public in the near future.*

1. Introduction

A consequence of Einstein's Theory of General Relativity is that mass bends the path of light. [4] Most of the time the deflections are very small; the original 'gravitational lens' that tested the veracity of Einstein's theory in the years after the first world war was the sun, which deflected light from stars behind it only a few seconds of arc. However, when light passes through a particularly deep gravitational potential (say, near the center of the dark matter halo of a galaxy cluster), the deflections can be particularly large, resulting in brilliant arcs and multiple images. These strong deflections due to light passing through a deep gravitational potential are termed strong gravitational lenses. The gravitational lensing signal can heuristically be thought of as a trade off between a couple of factors: larger gradients in the gravitational potential create larger distortions (and larger gradients in the gravitational potential tend to reside near to the center of the foreground galaxy or galaxy cluster), while the more separated the foreground and background objects are, the bigger the proportional distortion on the background object by the foreground (conversely, the farther away the background object is, the smaller it appears[1]). [9]

The very existence of these potentials acts as a verification of the Theory of General Relativity, but they can also be used for much more. Strong gravitational lenses are one of the few ways to directly probe the distribution of dark matter, a particle (or possibly family of particles) that does not emit electromagnetic radiation but does have mass and hence interacts gravitationally with normal baryonic matter. [15] This allows us to tally the mass of the largest gravitationally-bound structures in the universe, galaxy clusters, which can give us insight into the formation history of these massive objects. [13] They can also find extremely distant objects. [6] In this way, strong gravitational lenses can then tell us something about the expansion history of the universe, by setting limits on how massive the most massive objects in the universe can be. The properties of the bent light itself can also say much about that expansion history. [14, 11, 5, 2, 7, 17] When an object is strongly-lensed into multiple images, each image travels a different span of space and time. When an object does not vary much with time, these different path lengths have no practical import. However, if the object varies appreciably quickly (say it is a distant quasar, a supermassive black hole at the center of a galaxy whose accretion disk emits high-energy radiation at varying rates) then these different path lengths can be used to pin down the rate of expansion of the universe. Finally, strong gravitational lenses also have the potential to rule out or validate alternatives to General Relativity. [18]

Unfortunately, for how useful strong gravitational lenses are, they are also extremely rare. A next generation optical survey like the Large Synoptic Survey Telescope or the Euclid space telescope can expect to find only ten thousand lenses in the whole sky, while it will find ten billion galaxies. [16] Currently in astronomy there are only order hundreds of strong gravitational lenses known, mostly

[1] Note that this is only true in the "low redshift" universe: when objects are farther away than a cosmological redshift of $z \approx 2.5$ or about 2.5 Gyr after the birth of the Universe, they will actually *grow* in angular extent.

discovered by 'eyeball squads' of graduate students. The small number means that target criteria must be somewhat broad in order to maintain a relatively high completeness. Using reasonable target criteria to find strong lenses such as looking only at massive galaxies still means that nearly ten million objects will need to be inspected in the next generation in order to find those ten thousand lenses. A team of ten graduate students could expect to spend about 14 years looking at these objects. Computer algorithms are not much better: most current machine learning algorithms are woefully-underpowered for this task, and generally have poor completeness or poor purity – and often poor both. Additionally, some algorithms are better at finding some types of lenses than others; some perform well on the brilliant arcs, but poorly on the multiply-imaged objects, or vice versa. For example, [10] attempt to fit arc-like features in images in order to find strong gravitational lenses, but this means that multiply-imaged quasars are completely ignored. [1] and [3] in contrast develop an algorithm for finding gravitationally lensed quasars based on catalog-level colors and magnitudes, precluding their algorithm finding strong gravitational lens arcs. New algorithms need to be developed to find more strong gravitational lenses, and more strong lenses need to be found to power these algorithms. These algorithms need to not only identify potential lenses accurately, but be able to make strong statements about their contamination rates, as spectroscopic follow-up can be an expensive endeavor. [12] performed spectroscopic follow-up on 9768 galaxies, finding 28 new strong gravitational lens systems, but taking 40 nights of telescope time on expensive telescopes.

SPACE WARPS (Marshall et. al, in prep.) is a citizen-science initiative designed to overcome these two problems. The program has users examine images from the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS) and vote on where they see lenses. Users are also assessed and trained with simulated lenses and known empty fields. By having thousands of users analyze a survey for short amounts of time each, it is hoped that a more complete sample of lenses can be discovered, which can then be fed into lens-finding algorithms to further improve their performance.

2. Problem Statement

In this project we will use images collected by the Canada-France-Hawaii Telescope Legacy Survey to analyze how Convolution Neural Networks can improve automated detection of strong lens systems. We will also assess the performance of citizen-scientists by comparing our results to them. From other graduate work (but not coursework), we have the locations and categories of around one hundred and twenty known strong lenses, three thousand large fields verified to contain no strong lenses, six thousand simulated strong lenses, and several thousand classifications by citizen-scientists of other potential strong lens systems. These will form the core of our training and testing datasets; our metric will be how well a CNN correctly identifies known and simulated lenses and non-lens systems.

We would like to examine the following questions:

- Do we have enough data to reasonably train and test a CNN? Can we get around this by artificially inflating the data, e.g. by adding rotated images?
- How do citizen-scientists do compared with this automated system?
- Can we use the results of citizen-scientists to train the CNN?
- How well does using features extracted from a convolution neural network trained on galaxy morphology perform when determining the presence of strong gravitational lenses?

3. Technical Approach

From approximately 12000 fields of $440 \times 440 \times 3$ fields, we have constructed approximately 30000 cutouts sized $96 \times 96 \times 3$. These cutouts are selected based on where citizen-scientists clicked, on the theory that both 'correct' and 'incorrect' selections provide useful information about the characteristics of gravitational lenses. In general, we have access to two broad classes of images: 'training' and 'test' images. The 'training' images include fields that were verified in advance to not contain any lenses as well as simulated lensed galaxies, quasars, and clusters. Many of the simulated objects are over-exaggerated and extremely obvious, but we also have access to a second 'refinement stage' of the project, where much harder simulations were given to users. The 'test' images are the fields that citizen-scientists viewed, assessing whether a lens was in the field or not. In these 'test' images are 120 known strong gravitational lens systems, which are also included in this set. (The project confirms roughly half of these known lenses for reasonable definitions of completeness and purity.) For all the images we also have an associated probability that the project would evaluate that system as containing a lens.

It is clear that we do not have enough data. Luckily, we also know that our lens objects must obey certain symmetry properties, so it is quite easy to augment our data. For example, we know that strong lens systems should be independent of rotations as well as small amounts of stretching and translation, so our data can be augmented by applying those transformations to our images.

We train a classifier on this data using two different methods. First, we code our own convolutional net in

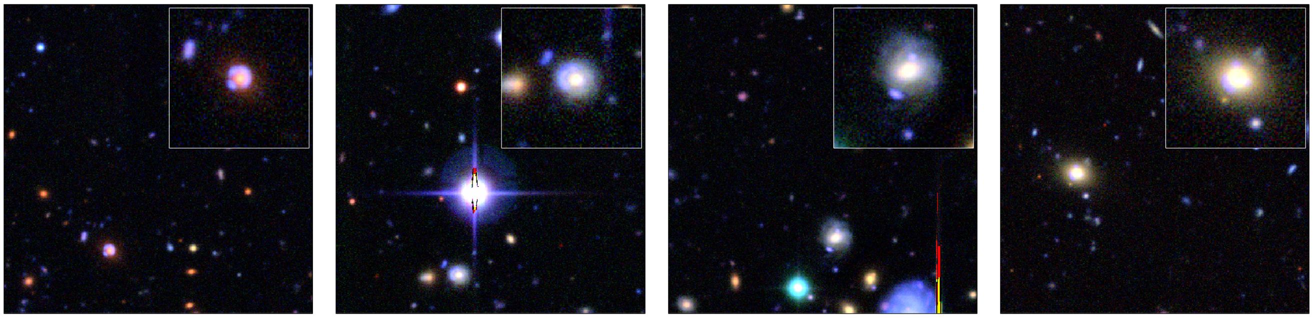

Figure 1. Typical sims. Insets indicate the location of the lens in the image. These insets are fed into our training system.

Figure 2. Typical Space Warps duds. Insets indicate regions where volunteers typically clicked. These insets are then fed into our training system.

python using THEANO. Second, we apply transfer learning techniques to train on a convolutional net galaxy morphology classifier, which has graciously been made available to us by Ryan Keisler and which achieved 7th place in the 2014 Galaxy Zoo Kaggle competition. This classifier runs 96×96 images through 3 convolutional layers and 2 fully-connected layers and predicts a galaxy to have one of 37 enumerated morphologies. (See Figure 4.) We train classifiers on top of the first fully-connected layer, which has 500 neurons.

3.1. The SPACE WARPS Catalog

SPACE WARPS is a web-based service that enables the discovery of strong gravitational lenses in wide-field imaging surveys by large numbers of people. Carefully produced color composite images are displayed to volunteers via a classification interface which records their estimates of the positions of candidate lensed features. Simulated lenses, and expert-classified non-lenses, are inserted into the image stream at random intervals; this training set is used to give the volunteers feedback on their performance, and to estimate a dynamically-updated probability for any given image to contain a lens. Low probability systems are retired from the site periodically, concentrating the sample towards a set of candidates; this “stage 1” set is then re-classified by the volunteers in a second refinement stage. This “stage 2” has a different set of training images, ones that are generally considered ‘harder’. Most stage 1 simulated lenses are very obvious[2], while simulated lenses in stage 2 are often much more subtle. [3] Figures 2 and 1 show example stage 2 fields with cutouts inlaid. Notice that while the first three images in Figure 1 are very clearly strong gravitational lenses [4], the fourth is very difficult to find. Unfortunately, we would very much like to find these, because there are many such systems and they contain important information about the mass structures at the centers of galaxies.[5] Figure 2 highlights the difficulties of this task. Each ‘dud’ has features that conceivably look like strong gravitational lensing, but are in actuality some other confounding effect: color gradi-

[2] Very bright and blue quasars multiply-imaged around a small red galaxy, very bright, separated, and full Einstein rings.

[3] Dim multiply-imaged quasars of varying magnitude, dim and incomplete Einstein rings located close to a galaxy.

[4] For the neophyte: the first is a broken blue arc around a central red galaxy; the second is a blue arc around a central yellow galaxy; the third is a multiply-imaged blue quasar (images appear above and below the central galaxy), the fourth is a broken dim arc located behind a very bright foreground galaxy.

[5] The trade off is this: the rate of alignment between foreground and background objects decreases as one decreases the area around a foreground object, but the strength of strong gravitational distortions – and the signal we can pull out from identifying such systems – increases as one gets closer to the center of the foreground object. A yet further complication to this is that background objects are naturally fainter than foreground objects, *and* the foreground objects with the highest gravitational potential (and hence the largest distortions of background images) tend to also be the brightest objects. Both these complications render the task even more difficult.

ents from variations in the Point Spread Function between the different color bands, dust surrounding a galaxy, galaxies that are actually in the same cluster, and chance alignments of background galaxies and foreground stars.

The fields users observe are 440×440 size images, containing multiple potential locations for strong gravitational lenses, although it is unlikely that a field contains more than one strong gravitational lens. In order to generate 96×96 images of lenses and non-lenses, we use the recorded estimates of the positions of candidate lensed feature. More specifically, we apply the DBSCAN clustering algorithm, which agglomeratively grows clusters such that that are within a minimum distance and contain a minimum number of samples. DBSCAN is a convenient choice of a clustering algorithm because it has a well-defined way of rejecting outliers, which we generically interpret as genuine "mis-clicks" on the part of users. We are very generous in the definition of a cluster and only require two members within 100 pixels of each other to form a cluster.

The fear of noisy clusters is this: the real task of our techniques is to distinguish strong gravitational lens systems from other configurations of galaxies (for example, random alignments of galaxies). Noisy clicks end up creating random cutouts of the field, slightly changing the task of our classifier to distinguishing strong gravitational lens systems from random cutouts from the field of a galaxy survey. Inclusion of noise however ended up not being an issue for stage 2: non-lenses, where the correct action on the part of the user is to leave no marker, have a median of 28 markers in stage 2, while simulated lenses (where the correct answer is to click at a specific location) have a median of 180 markers. In stage 1, the simulated lenses have a median of 80 markers, while the duds have a median of 3 markers (and a mean of 9.6). It may be the case that noise is injected in the stage 1 non-lens sample.

Overall, our base dataset has 24,177 images from stage 1, of which 5159 are of simulated lenses, and 1876 images from stage 2, of which 151 are simulated lenses. We also have 9030 classifications that stage 2 users made of images in the CFHTLS survey where it is unknown whether they contain a lens or not. From these classifications, a list of approximately 40 candidate strong gravitational lensing objects have been found which will soon receive spectroscopic follow-up to confirm whether they are strong gravitational lensing systems or not.[6] A future project with this work would be to link the probabilities from the SPACE WARPS system with the probabilities obtained by a detection algorithm.

[6]Spectroscopy can yield precise redshifts of different objects in a field. This way, if different parts of a strong gravitational lens arc are at the same redshift, or if the multiply-imaged quasars are, then one can confirm that we are really seeing such a system.

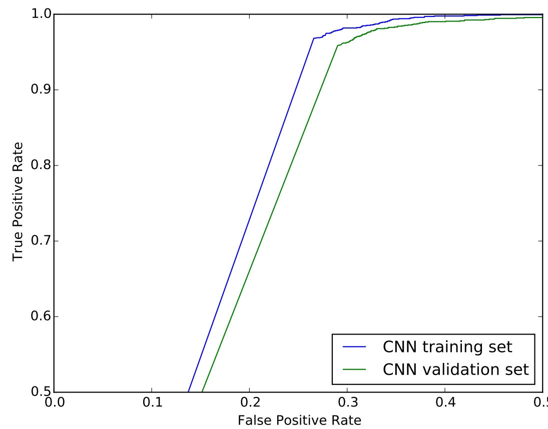

Figure 3. Receiver Operating Curves extracted from the training and validation sets used with our Convolutional Neural Network after training for 35 epochs. These results should be compared with the SPACE WARPS Stage 1 Receiver Operating Curve in Figure 5, which outperforms our network for false positive rates below ~ 0.3. However, above this point our Convolutional Net achieves higher true positive rates than any of the other (machine or human) methods herein considered.

4. Results

4.1. Convolution Neural Network

In order to retain our ability to experiment with non-traditional training methods and architectures, we chose to implement our Convolutional Neural Network (CNN) from the ground up using THEANO. We settled on a five-layer layer architecture consisting of two convolution/max dropout layers, a fully connected layer, and a softmax layer. Training was carried out using RMSprop with a decay rate of 0.9, and the model was regularized using L_2 normalization with a regularization strength of 10^{-4} and dropout with probability 0.5 at each layer . We chose our convolution layers to have stride 4 and depth 10 (with padding to maintain the input image size), and the subsequent max pooling layers to have stride 2. Assigning our fully connected layer 50 neurons (and including the paramters in the subsequent softmax layer) gave our model a total of 288,492 trainable parameters.

Due to memory and time constraints, we have not yet trained this CNN on all our of data. The results of training it on 8,000 randomly selected images from the SPACE WARPS Stage 1 data set for 35 epochs is shown in Figure 3, where we have plotted the Receiver Operating Curve obtained from our model's predictions on the training set and a test set composed of another (non-overlapping) randomly selected 8,000 images from the SPACE WARPS Stage 1 data set. This result should be easily improved on by training on more of our images, by adding data augmentation, and by

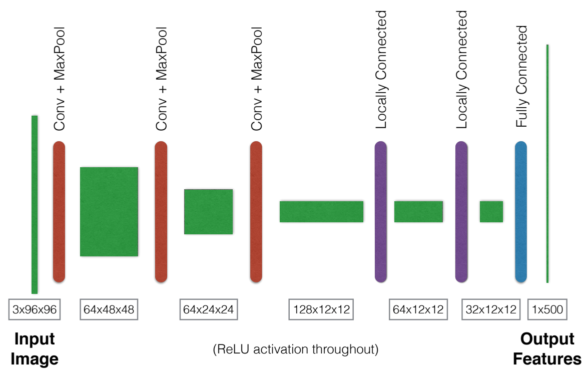

Figure 4. Architecture of the convolution neural network trained on the galaxy zoo morphologies. The feature vector we use comes from the fully connected layer.

increasing the complexity of our CNN architecture. However, it already shows that training a CNN solely on the SPACE WARPS data set is a viable option, given enough CPU power and time.

4.2. Transfer Learning

Transfer Learning relies on the idea that convolution neural networks that perform similar tasks pick up similar features, as well as the observation that lower levels in neural networks tend to be quite generic in the features they pick out. They provide an answer to the scenario when there are too few data to effectively train a complex system like deep convolution neural networks: start the training of your new system from the results of training a similar system. For the scope of this project, we chose to examine the effects of transfer learning from galaxy morphology to strong gravitational lens identification. In both cases, an input image of a galaxy is fed into the network, and some classification is read out. Additionally, both networks need to differentiate shapes in the central regions of the galaxy image (for example to find bars in spiral galaxies) from outer regions (spiral arms, arcs, multiply-lensed systems). The Galaxy Zoo competition provides an ideal candidate for transfer learning because of these facts, and also because the image quality is comparable between the Sloan Digital Sky Survey (the telescope survey on which the Galaxy Zoo images were based) and the CFHTLS survey. We have on hand a convolution neural network trained to classify galaxy morphology provided by Ryan Keisler [7]. The architecture of that network can be seen in Figure 4. We do the simplest thing possible: we run the convolution neural network as a feature extractor, and take images from the fully connected layer. Thus we transform a $96 \times 96 \times 3$ image into a 500 feature vector. We then train these feature vectors on various classifiers (Random Forest, Support Vector Machine, Softmax) and evaluate results against a test set. We distinguish between stage 1 and stage 2 data because the simulated lenses changed between the two sets. The code that produces the feature vectors also augments the data by automatically producing feature vectors of flips and rotations of the input images. This allows us to increase the size of our input dataset nearly 20-fold.

We train these datasets on three classifiers: Random Forests (which are an ensemble of decision trees trained on the data), Softmax and linear Support Vector Machines. We use stochastic gradient descent for the latter two classifiers. We create a test dataset by randomly extracting 20 percent of the dataset and setting it aside. We also ensure that any data augmentation stays in the training or test sets. Our goal with all the above systems is not to find the maximal accuracy, but to find some reasonable trade-off between the true positive rate and the false positive rate: we want to find as many lenses as we can, but we also know that confirmation of these lenses by spectroscopic follow-up is an expensive endeavor such that we want to minimize the number of non-lenses that make it into our candidate list. Because of this, any potential candidate list we would make from any of our classifiers has a relatively hard threshold at a false positive rate of 0.2.

The resultant Receiver Operating Curves can be observed in Figure 5. In general we find that support vector machines perform the best as a classifier on the feature vectors, but that the feature vectors perform more poorly than the SPACE WARPS users. We must caveat though that SPACE WARPS does not create a validation dataset against which to test the performance of the system. Even if we don't perform quite as well as the citizen-scientists, we consider this a promising baseline for future performance by transfer learning: we have not even begun to consider potential performance gains by retraining the convolution neural network on the SPACE WARPS data.

5. Discussion

We briefly attempt to answer the questions we posed at the beginning of this study:

Do we have enough data to reasonably train and test a CNN?

We do, and we have easy ways to augment the data through rotations, flips, and (in the future) other transformations that leave the identification of a strong gravitational lens invariant.

[7] rkeisler@stanford.edu

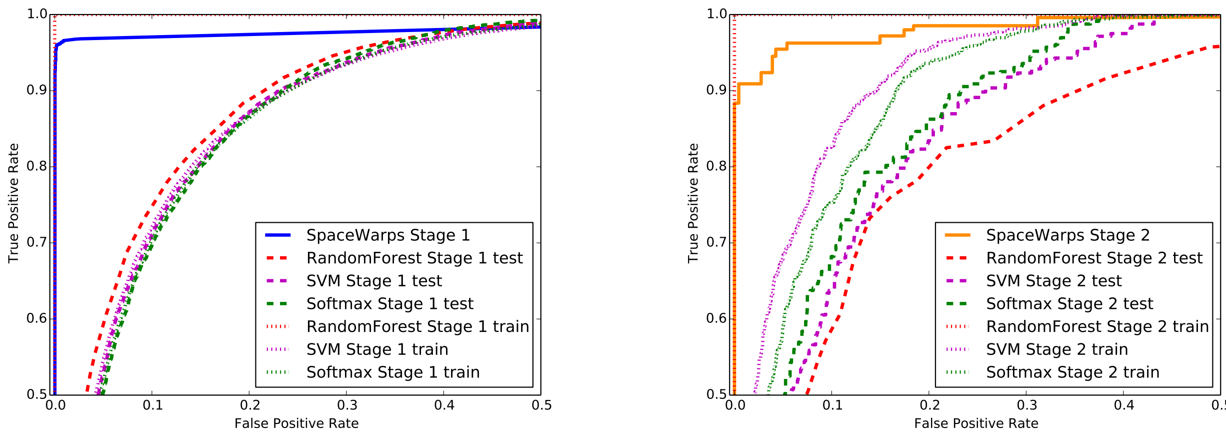

Figure 5. Receiver Operating Curves the SPACE WARPS system and different linear classifiers trained on feature vectors extracted from a convolution neural network originally used to determine galaxy morphologies. We find that of the linear classifiers on the feature vectors, Softmax classifications perform best on the test dataset, however all the feature vectors perform worse than the users themselves. Note that the x-axis stops at a false positive rate of 0.5, and the y-axis begins at a true positive rate of 0.5. Truly random guessing (which results in a 1:1 relationship between the true positive rate and the false positive rate) would not show up on this graph.

How do citizen-scientists do compared with this automated system?

Currently the citizen-scientists outperform our automated systems.

Can we use the results of citizen-scientists to train the CNN?

In the sense that the citizen-scientists are used in creating the cutouts for our system, we find their results quite useful. We were unable to use them for much more than that, however. Future work could involve calibrating our classifications against the probabilities from the SPACE WARPS citizen-scientist classification system.

How well does using features extracted from a convolution neural network trained on galaxy morphology perform when determining the presence of strong gravitational lenses?

We appear to do admirably well in that our receiver operating curves obtain remarkably high true positive rates for modestly low false positive rates. However, the extracted features do worse than the citizen-scientists from SPACE WARPS.

6. Conclusions

The need for new automated detection algorithms for finding strong gravitational lenses will only become more pressing in the next decade, as it becomes infeasible for scientists to scan images by eye for such systems. Using the SPACE WARPS dataset, we have examined how convolution neural networks trained both on this particular dataset and on other datasets can perform at the detection task.

While our Convolutional Neural Net trained on just the SPACE WARPS dataset has not yet outperformed citizen-scientists at sufficiently low false positive rates, there is ample reason to believe marked improvements can be made with more time and CPU power. In addition to training on more images (and augmented images) for longer, we can expect to gain a significant reduction in training time by taking advantage of THEANO's GPU capabilities and by implementing batch normalization. [8] It may also be beneficial to explore larger and more expressive CNN architectures.

In our application of Transfer Learning, we find that features extracted from a convolution neural network trained on the classification of galaxy morphology (with a linear support vector classifier on top for converting the feature vector to a binary "lens" and "not-lens") performs admirably but markedly worse than the citizen-scientists trained on the dataset. Further work examining improvements by retraining the whole neural network could lead to a generic classification machine that takes images from *any* galaxy survey and states whether the image contains a strong lens or not. Additionally there is much potential in both direct convolution neural networks and transfer learning from other networks in linking the classification outputs of the networks with the probability estimates of the SPACE WARPS system, which also examined several thousands more "unknown" systems and could lead to more gravitational lenses being identified.

Acknowledgements

We received access to the dataset through Phil Marshall[8], who also graciously explained to us how SPACE WARPS currently works. The convolutional neural network upon which the transfer learning is based was kindly provided to us by Ryan Keisler[9]. All errors in interpretation or otherwise are entirely our own.

References

[1] A. Agnello, B. C. Kelly, T. Treu, and P. J. Marshall. Data mining for gravitationally lensed quasars. *MNRAS*, 448:1446–1462, Apr. 2015.

[2] A. Barnacka, M. Geller, I. P. Dell'antonio, and W. Benbow. Strong lensing, time delays, and the value of H$_0$. *ArXiv e-prints*, Aug. 2014.

[3] J. H. H. Chan, S. H. Suyu, T. Chiueh, A. More, P. J. Marshall, J. Coupon, M. Oguri, and P. Price. Chitah: Strong-gravitational-lens hunter in imaging surveys. *ArXiv e-prints*, Nov. 2014.

[4] A. Einstein. *Relativity: The special and general theory*. Henry Holt and Company, 1920.

[5] A. H. Gonzalez, S. A. Stanford, M. Brodwin, C. Fedeli, A. Dey, P. R. M. Eisenhardt, C. Mancone, D. Stern, and G. Zeimann. IDCS J1426.5+3508: Cosmological Implications of a Massive, Strong Lensing Cluster at z = 1.75. *ApJ*, 753:163, July 2012.

[6] Y. D. Hezaveh. Prospects for Measuring the Mass of Black Holes at High Redshifts with Resolved Kinematics Using Gravitational Lensing. *ArXiv e-prints*, Aug. 2014.

[7] A. Hojjati and E. V. Linder. Next generation strong lensing time delay estimation with gaussian processes. Aug 2014.

[8] S. Ioffe and C. Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *ArXiv e-prints*, Feb. 2015.

[9] C. S. Kochanek. *Part 2: Strong gravitational lensing*, pages 91–268. Springer, 2006.

[10] J. M. Kubo and I. P. Dell'Antonio. A method to search for strong galaxy-galaxy lenses in optical imaging surveys. *MNRAS*, 385:918–928, Apr. 2008.

[11] T. Kundic, E. L. Turner, W. N. Colley, J. R. Gott, III, J. E. Rhoads, Y. Wang, L. E. Bergeron, K. A. Gloria, D. C. Long, S. Malhotra, and J. Wambsganss. A Robust Determination of the Time Delay in 0957+561A, B and a Measurement of the Global Value of Hubble's Constant. *ApJ*, 482:75, June 1997.

[12] I. Momcheva, K. A. Williams, R. J. Cool, C. R. Keeton, and A. I. Zabludoff. A Spectroscopic Survey of the Fields of 28 Strong Gravitational Lenses: The Redshift Catalog. *ArXiv e-prints*, Mar. 2015.

[13] M. Oguri. The Mass Distribution of SDSS J1004+4112 Revisited. *PASJ*, 62:1017–, Aug. 2010.

[14] S. Refsdal. On the possibility of determining Hubble's parameter and the masses of galaxies from the gravitational lens effect. *MNRAS*, 128:307, 1964.

[15] S. Refsdal. The gravitational lens effect. *MNRAS*, 128:295, 1964.

[16] S. Serjeant. Up to 100,000 reliable strong gravitational lenses in future dark energy experiments. Sep 2014.

[17] S. H. Suyu, T. Treu, S. Hilbert, A. Sonnenfeld, M. W. Auger, R. D. Blandford, T. Collett, F. Courbin, C. D. Fassnacht, L. V. E. Koopmans, P. J. Marshall, G. Meylan, C. Spiniello, and M. Tewes. Cosmology from Gravitational Lens Time Delays and Planck Data. *ApJL*, 788:L35, June 2014.

[18] M. A. Troxel, M. Ishak, and A. Peel. The effects of structure anisotropy on lensing observables in an exact general relativistic setting for precision cosmology. *JCAP*, 3:40, Mar. 2014.

[8] `pjm@slac.stanford.edu`

[9] `rkeisler@stanford.edu`