MagNet: Deep Machine Vision for the Cosmic Dawn

Category: Computer Vision

Warren Morningstar (SUID: wmorning) & Dylan Rueter (SUID: tdr38)

Abstract

We present a method for reconstructing the background sources of gravitational lenses using deep convolutional recurrent neural networks. We limit this work to images where the only form of image distortion is the non-linear lensing transformation. We then implement a Recurrent Inference Machine to learn an iterative update method to solve for the background source. We train this machine using a suite of 10^5 simulated images of gravitational lenses, and test on a subset of an additional test set of 3×10^4 images. We find that the network is able to efficiently reconstruct the morphology of the background sources in a qualitatively accurate way. Furthermore, the method recovers the flux of the background source accurately. We discuss applications of the model and future improvements to the training procedure that will need to be made in order to realize these applications.

1 Introduction

We currently understand relatively little about how galaxies form. Much of our lack of understanding is driven by the difficulties in observing galaxies during their infancy. Because primordial galaxies are far away, their luminosity has diminished substantially by the time their light reaches us. Furthermore, because they are typically compact only the largest telescopes are capable of discerning any structure in these galaxies due to the size of the typical instrumental point spread function. As a result, the fine details of these incredibly interesting objects have remained mostly unstudied.

Many of the advances in our understanding of galaxy formation will come from the next generation of massive extragalactic imaging surveys. In particular, these surveys will discover several thousand new gravitational lenses. Gravitational lenses act as natural telescopes, revealing the internal structures of distant galaxies that would be too small and too faint to study on their own. However, the lensing boost also produces heavy distortions on the images, causing significant warping and increasing the multiplicity of observed sources. In order to study these sources, this distortion must be corrected.

Current "cutting-edge" techniques to remove the distortion and infer the structure of these sources rely on maximum likelihood (or *a posteriori*) estimates of the source, given a model for the lens as well as a number of assumptions about the internal structures of galaxies. Many of these assumptions are poorly motivated, and are not consistent with existing observations of galaxies that have not been distorted by lensing. Furthermore, these techniques require significant computational cost, taking hundreds to thousands of CPU hours per lens, and requiring constant human supervision.

Mathematically, the transformation from the true image (the source image) to the observed lensed image (the observed image) may be thought of as a matrix multiplication [1], and the addition of noise

$$I = \mathbf{F}\mathbf{L}S + \mathcal{N}(0, \sigma) \tag{1}$$

where the matrix $\mathbf{L} \in \mathbb{R}^{n_i \times n_S}$ maps each pixel in the image of the background source to the corresponding pixels in the image, and the blurring matrix $\mathbf{F} \in \mathbb{R}^{n_I \times n_i}$ introduces the effect of the instrumental point spread function (or a Fourier Transform in the case of interferometric observations). Generically $n_I \neq n_S$, and due to noise introduced to the image the problem of finding the left inverse becomes ill-posed without specification of an *ad hoc* regularization. This regularization neglects our physical understanding of the structures of galaxies (which is not easily expressed mathematically), and thus may cause biases in the inferred structure of the background source. This is currently a limiting factor in using source reconstruction techniques to study the physics of galaxies.

However, due to the magnification and multiplicity of observed images of the background source, it is reasonable to believe that the observed image should contain enough of the necessary information to produce an accurate reconstruction given a reasonable prior. To this end our objective is to use deep learning techniques to simultaneously represent the mapping from lensed image to background galaxy image, while encoding (via training images) a physically motivated prior on the structure of galaxies to aid in the reconstruction process. We describe our training, dev, and test sets in Section 2. In Section 3, we detail our network architecture. In Section 4 we report the results of application of our architecture to a subset of the test images. In Section 5 we discuss our results, along with future extensions and applications of our network.

2 The Dataset

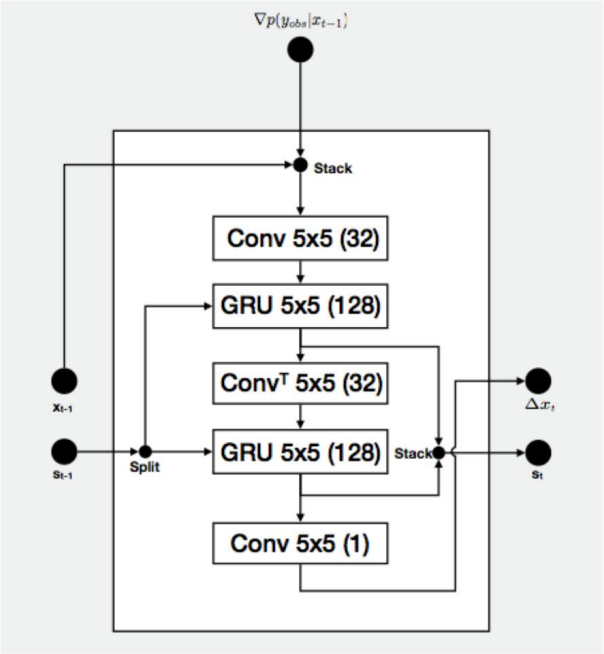

Figure 1: A single RIM cell. This cell takes state from the last iteration, the image from the last iteration, and the gradients of the likelihood with respect to the image from the last iteration as inputs. It outputs its current memory state along with the proposed update to the image.

Our dataset consists of 192×192 single channel images generated using an existing pipeline designed to study gravitational lensing effects using Neural Networks [2]. For our background source, we draw sample galaxy images from the Galaxy Zoo machine learning challenge [3]. We then generate a random realization of a lens model, and place the background source image in a randomly shifted position behind the lens, with a random scaling of the source size. Because the relative position and size of the background source is our target image, and in order to make our objective more uniformly defined, we resample our source image on a uniform grid of size 1.5×1.5 arcseconds centered at the origin. These images will serve as our ground truth to which comparisons will be made. We also require the image of the background source to be magnified by at least a factor of 3 (Indicating that the background source is strongly lensed). We performed this process 130,000 times, giving us a training set of about 100,000 images of lensed galaxies, with up to 15,000 images each for our dev and test sets. We note that this simulation process took a trivial amount of time (a few hours on a single cpu), and thus we could easily simulate more images if needed.

To simulate the effects of the gravitational lens, we use the Singular Isothermal Ellipsoid lens model [4]. One important aspect of the data generation process is that it fully sample the lens parameter space in order for the network to eventually be usable on real lensed images. The observable effects of lensing only depend on the relative displacement between the source and the lens, and so we can center the lens in the image without loss of generality, since we randomly place the source. The angle of rotation within the lensing plane is randomly sampled from a uniform distribution $0 - 2\pi$. The ellipticity of the lens we randomly sample from a uniform $0 - 1$ distribution, reflecting an uninformative prior. We also include an external shear on the observed image. This shear incorporates the effects of mass near the line of sight but external to the gravitational lens itself. Shear has both a magnitude and angle of effect associated with it, which we randomly sample from uniform distributions of range $0 - 0.2$ and $0 - 2\pi$, respectively. The remaining lens parameter is the radius of the Einstein ring, an angular quantity which depends on various distances in the source-lens-observer system and the mass of the lens. Small gravitational lenses, which have small Einstein radii, are very difficult to observe, so we uniformly sample our Einstein radius from $0.1 - 3$ arcseconds since we are primarily interested in the lenses with $\theta_E > 0.5$. As shown in the left column of Figure 2, the resulting distribution of images has significant diversity.

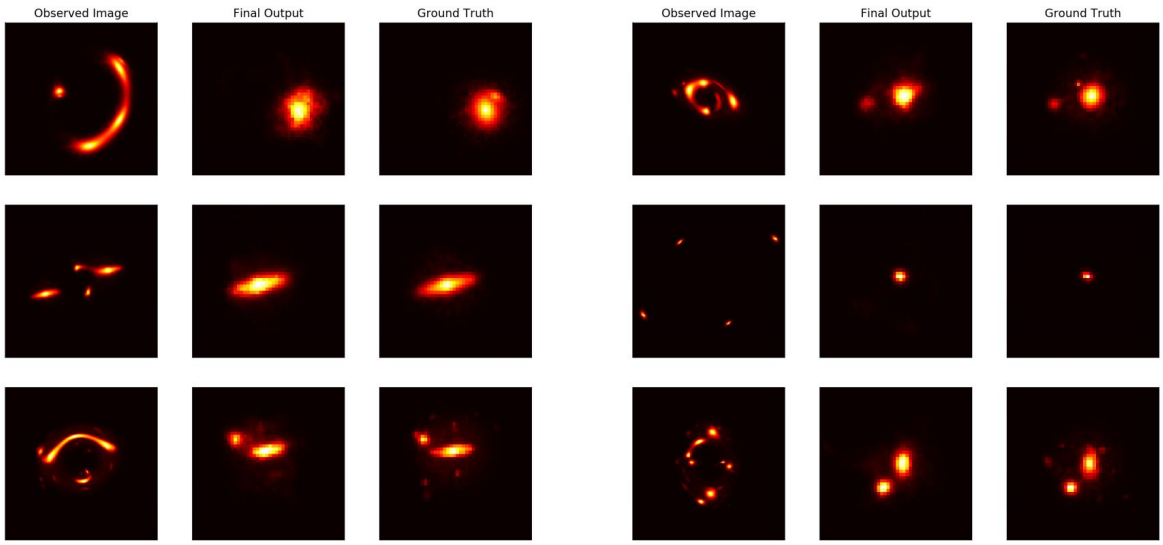

Figure 2: Six example reconstructions of images in our test set using our network. In all cases, the left column shows the observed source, the middle column shows the output reconstruction, and the right column shows the ground truth. It is apparent that the network is able to reconstruct the morphology of the background source regardless of its size, ellipticity or number of components.

3 Network Architecture

To learn the optimization algorithm for reconstructing the background source, we use a Recurrent Inference Machine (RIM) [5]. This network was developed specifically for the purposes of iteratively solving inverse problems. The architecture of an RIM cell is shown in Figure 1 and a description of a single time step in the network is as follows:

1. The image of the attempted source reconstruction from the previous time step is fed to a raytracing simulation to predict the observed image (the forward model). The log-likelihood of the model is computed (for the purposes of this project, we use a mean-squared error as the log-likelihood).

2. The gradients of the log-likelihood with respect to each pixel in the attempted source reconstruction image is computed. The current trial image, and the gradients image are stacked together.

3. The stacked images are sent to a 5×5 convolutional layer, with 32 output channels and a linear activation.

4. The resulting feature cube is sent to a convolutional Gated Recurrent Unit (GRU), with a 5×5 convolution, and a tanh activation function. Other than the mathematical operations being convolutions (and thus the inputs, gates, states, and outputs being image cubes), this cell is otherwise identical to a traditional GRU. The update and forget gates are each computed with their own 5×5 convolutions, and receive a sigmoid activation. The number of output channels is 128.

5. The output of the GRU is sent to a Transpose-convolutional layer, and is downsampled to 32 output channels. The activation of this layer is linear.

6. The output of the Transpose-convolutional layer is sent to a second convolutional GRU, again with 5×5 filters and a tanh activation function.

7. The output of the second GRU is sent to a final 5×5 convolutional layer, with a single output channel to produce the element-wise update rule to the previous source reconstruction image.

8. The update image is added to the previous source reconstruction to get the current predicted image. This image is passed to future iterations of the network.

This procedure is then unrolled over time steps, where the input to each step is the previous image, and the output of each step is the updated image. By using a GRU as the recurrent cell, the network is able

to pass state information to the next step. The update and forget gates determine which information should be added and removed from the memory state.

To optimize the network, we use as a loss function the total squared error on the predicted image at *every* time step. By using every reconstruction step in computing the loss, we attempt to push the network toward a faster reconstruction of the source, since taking more steps to converge to the optimal estimate will produce a larger loss. It also penalizes the network for overshooting the true image and making the reconstruction worse given additional time steps. The cost function is just the sum of the loss functions over training examples in a mini-batch.

To minimize the cost function, we employ the adam optimizer using an exponentially decaying learning rate with a timescale of 5000 training steps, a decay rate of 0.96, and an initial rate of 1.0×10^{-6}. Because recurrent neural networks are susceptible to exploding gradients (particularly early in training), we employed gradient clipping in our optimization. We optimized the cost function for 140,000 training steps, with two training examples per batch. This constitutes nearly three epochs of training. We did not use a larger mini-batch size because we were limited by memory requirements. The networks were trained on an NVIDIA Tesla K80 GPU.

4 Results

Outputs of the network for several test examples are shown in Figure 2. The network does a good job of reconstructing the morphology of the main component of the background sources, and as the training progresses further, many of the satellite galaxies present in images of some of the background sources are reconstructed as well. As a quantitative metric of the success of our network, the root-mean-squared error per pixel on our sett of test images is 0.01, which implies a 1% error given that our images have pixel intensities between 0 and 1. We find that while the network is currently able to identify and reconstruct all bright and extended sources in the image, it currently has modest difficulty with compact satellites that have sizes on the order of a single pixel. This will likely improve as the network continues to learn.

In addition to accurately reconstructing the morphology of the background sources, this network is quite accurately reconstructing the total flux, which we define as the sum of the pixel values of the background source. This is very important physically, as measures of brightness in various wavelength bands are important handles for the temperatures and ages of the observed stars. A plot of the network output flux against the flux of the ground truth is shown in figure 3.

5 Discussion

The RIM architecture provides a good method of solving the inverse problem when the forward "corruption" process is understood, and as a result has been quite successful in reconstructing the background sources of gravitational lenses. However, these experiments are somewhat idealized due to the fact that they do not include noise in the observed images. Future applications of this network to images of gravitational lenses will include the presence of noise with significant variations in SNR. This is relatively trivial to include, as the noise in many telescopes is spatially uncorrelated and gaussian distributed with a known noise RMS. Additionally, most astronomical telescopes have additional observational systematics that must be included in the forward model. These effects include blurring by the point spread function (which we discuss in Section 1), masking of pixels in the image (either due to bad pixel columns in the CCD, or cosmic rays incident upon the detector during exposure), and removal of light produced by the foreground lens. Again, many of these corruption processes are well understood and can be included in the forward model. The last of these (lens light removal) is slightly more difficult, but may be performed using Independent Component Analysis (ICA) assuming images of the lens are taken in multiple photometric filters [2].

Because our original network implementation produced excellent results, and because training was relatively slow, we did not experiment significantly with hyperparameter tuning. In particular, the number of channels in the embedding layers and GRU layers of the RIM cell were relatively modest (32 and 128 respectively). It is possible that allowing for more hidden channels would provide additional flexibility that could be useful when the model is extended to deal with more complicated forms of corruption. However, we defer this experimentation to future work.

Additionally, in this work, we have assumed that the lens model is known *a priori*. In general, this assumption may not be perfect. In particular, the parameters of the lens model must also be inferred from the input image. This is mostly a solved problem [2], but we will need to examine the effects that

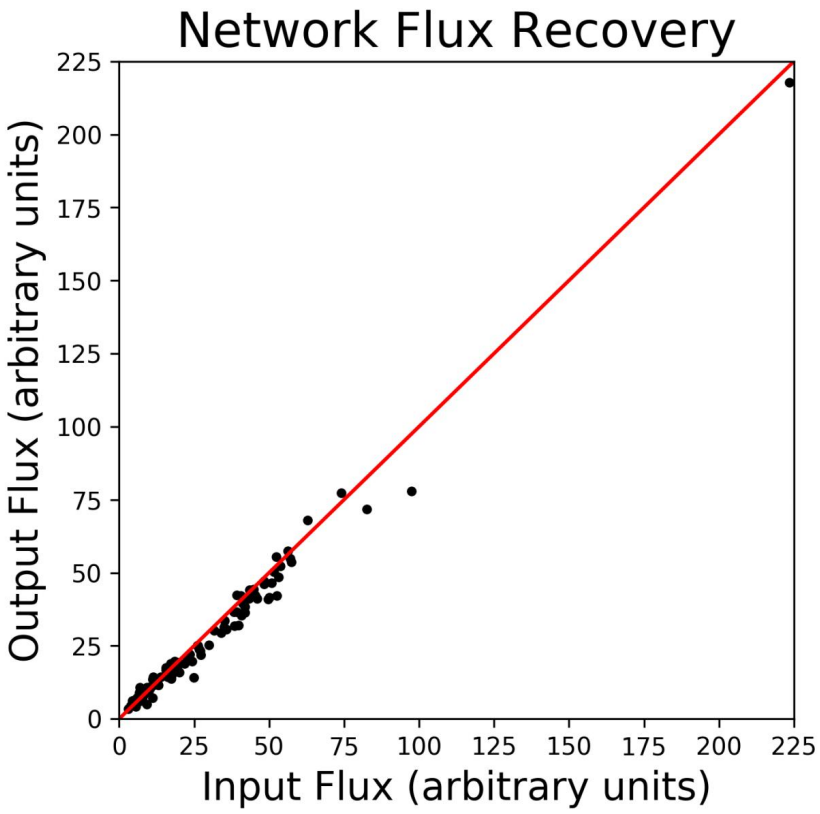

Figure 3: Flux recovery of the images in the test set. The x-axis shows the flux in the image of the ground truth, while the y-axis shows the flux in the image predicted by our network. The red line denotes a 1-to-1 mapping.

small errors in the inferred lens model produce on the reconstructed source. Current uncertainties in lens model predictions from convolutional neural networks are of order 0.01, significantly smaller than 1% errors for certain parameters of the lens. It is, as of now, unclear how (or if) this uncertainty will manifest itself in the predicted images.

Finally, the most potentially interesting extension to this project is to produce source reconstructions on synthesis images produced using interferometers such as ALMA. These observations have a significantly larger data volume than traditional CCD images, and currently are computationally challenging to analyze using alternative techniques. The gains in computational efficiency by using the RIM for source reconstruction here may be up to a factor of 10^8 in evaluation speed, making this an attractive alternative to traditional source reconstruction techniques.

Acknowledgements

The authors would like to thank Patrick Putzky for generously sharing the source code for the RIM module, which made this project possible. This work used the XStream computational resource, supported by the National Science Foundation Major Research Instrumentation program (ACI-1429830)

References

[1] S. J. Warren and S. Dye. Semilinear Gravitational Lens Inversion. *ApJ*, 590:673–682, June 2003.

[2] Yashar D Hezaveh, Laurence Perreault Levasseur, and Philip J Marshall. Fast automated analysis of strong gravitational lenses with convolutional neural networks. *Nature*, 548(7669):555, 2017.

[3] C. J. Lintott, K. Schawinski, A. Slosar, K. Land, S. Bamford, D. Thomas, M. J. Raddick, R. C. Nichol, A. Szalay, D. Andreescu, P. Murray, and J. Vandenberg. Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *MNRAS*, 389:1179–1189, September 2008.

[4] R. Kormann, P. Schneider, and M. Bartelmann. Isothermal elliptical gravitational lens models. *Astronomy & Astrophysics*, 284:285–299, April 1994.

[5] Patrick Putzky and Max Welling. Recurrent inference machines for solving inverse problems. *arXiv preprint arXiv:1706.04008*, 2017.