ContextNet: Deep learning for Star Galaxy Classification

Noble Kennamer [1] **David Kirkby** [2] **Alex Ihler** [1] **Javier Sánchez** [2]

Abstract

We present a framework to compose artificial neural networks in cases where the data cannot be treated as independent events, our particular motivation is star galaxy classification for ground based optical surveys. Due to a turbulent atmosphere and imperfect instruments, a single image of an astronomical object is not enough to definitively classify it as a star or galaxy. Instead the context of the surrounding objects imaged at the same time need to be considered in order to make an optimal classification. The model we present is divided into three distinct ANNs: one designed to capture local features about each object, the second to compare these features across all objects in an image, and the third to make a final prediction for each object based on the local and compared features. By exploiting the ability to replicate the weights of an ANN, the model can handle an arbitrary and variable number of individual objects embedded in a larger exposure. We train and test our model on simulations of a large up and coming ground based survey, the Large Synoptic Survey Telescope (LSST) and compare to the state of the art approach, showing improved overall performance as well as better performance for a specific class of objects that are important for the LSST.

1. Introduction

The Large Synoptic Survey Telescope (LSST) is a ground based photometric survey that will commence in early 2022 and will observe for 10 years. It will image the entire available sky every three nights and produce over one petabyte of data a year. The science goals of the survey are vast, ranging from the detection of dark energy and dark matter signatures to mapping small objects in the Solar System such as near-Earth asteroids (LSST Science Collaboration et al., 2009). The scale of the data and the types of measurements require sophisticated data analysis methods and present a great opportunity for machine learning scientists to work with domain scientists on fundamental science questions. Conversely, the astronomical data that is collected creates new challenges for the machine learning community and promotes the development of new methods. Collaborations between machine learning scientists and astronomers have been steadily growing and diverse in methods and applications, including a deep learning approach for analyzing strongly lensed systems (Hezaveh et al., 2017), a probabilistic graphical model for processing astronomical images (Regier et al., 2015), an ensemble approach for classification of supernova (Lochner et al., 2016), and many others. These projects have contributed to both fields, which is also the aim for our work. In this paper we present the specific application of star galaxy classification, for which we develop a novel framework for composing neural network models in order to make advances in the field of astronomy.

Star galaxy classification is one of the first processing steps in the data analysis pipeline of any astronomical survey; its foundational nature means that it affects almost every subsequent step of the pipeline (Jurić et al., 2015). The inputs to star galaxy classification are a collection of small, cutout images of detected sources in a single exposure, and the outputs are the predictions for each detected source in that exposure. A single exposure records the photon counts of each pixel of a CCD exposed for a short period of time. For LSST, each exposure is a 16 megapixel grid with an exposure time of 15 seconds. The entire telescope is made up of 189 of these CCD detectors (Kahn et al., 2010). It is expected that approximately 1200 sources will be detected and thus need to be classified for each exposure. Figure 1 shows a single exposure on a log scale with 16 representative cutouts to the right. Note that this exposure contains 1273 detected objects.

In an ideal world for astronomers (one without an atmosphere), this is a rather simple problem. Stars can be thought of as point sources of light, while galaxies have some spatial distribution. Thus in order to make an accurate prediction all one has to do is measure the size of a source and, if it is greater than a certain threshold, it can be safely classi-

[1]Department of Computer Sciences, University of California, Irvine [2]Department of Physics and Astronomy, University of California, Irvine. Correspondence to: Noble Kennamer <nkenname@uci.edu>.

Proceedings of the 35^{th} International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018.

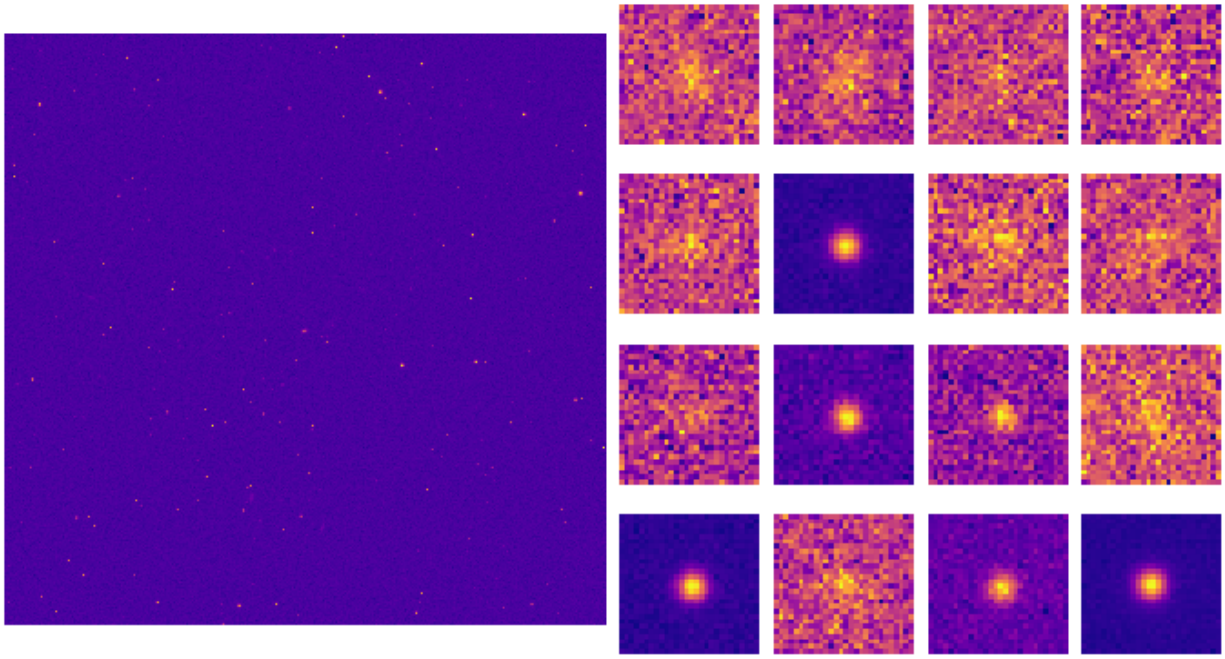

Figure 1. Left: an example of a single exposure with pixel values plotted on a log scale. Right: 16 cutout images of detected sources in this exposure. The entire exposure has 1273 detected sources.

fied as a galaxy. However with a turbulent atmosphere and imperfect optics, light is spread out before it reaches the detector. This spreading function is often called a point spread function and it changes both spatially and temporally. Thus one cannot make predictions assuming that the sources are IID across exposures. Instead one must use a model that is capable of handling the confounding factor of the point spread function (PSF) in order to make accurate predictions. Figure 2 shows an example star and an example galaxy with and without the PSF.

In the last several years there have been tremendous advances in computer vision, especially regarding the use of convolutional networks for classification (Krizhevsky et al., 2012). One of the main motivators of this work is to take advantage of this progress by applying it to the field of astronomy. However this cannot be done without significant modification. Section 2 will show how this problem is currently being solved by astronomers and why current vision models from machine learning need to be extended in order to tackle this problem. Section 3 details the framework we have developed for composing neural network models to predict on non IID instances and to predict simultaneously for a variable number of instances. Our empirical studies are presented in Section 4, showing how our model is able to achieve better results than what is currently used in astronomy. We present our results on simulations of LSST observations using the GalSim image simulation package (Rowe et al., 2015), which was designed and developed by a large group of domain scientists. GalSim is designed to meet the stringent requirements of high precision image analysis applications such as weak gravitational lensing, for current datasets and for future astronomical surveys including LSST. In Section 5 we discuss future work, focusing on how our compositional framework can be further generalized to handle even greater diversity in inputs.

2. Related Work

In this section we describe how astronomers currently solve star galaxy classification, and why machine learning models that assume independence between cutouts are inadequate.

2.1. Measuring Extendedness

When classifying a cutout as a star or galaxy, the most important signal is whether the object is "extended" or not. Objects that are unextended are thought to be stars (point-like sources of light), while objects that are extended are galaxies. Astronomers have several different methods of measuring extendedness, but as shown by Garmilla (2016), they all achieve similar performance and are closely related. Thus we focus on only one of these related techniques. To measure the extendedness astronomers fit two models to each cutout object, one fitting a point spread function, and the second of which also includes a galaxy model. Both of

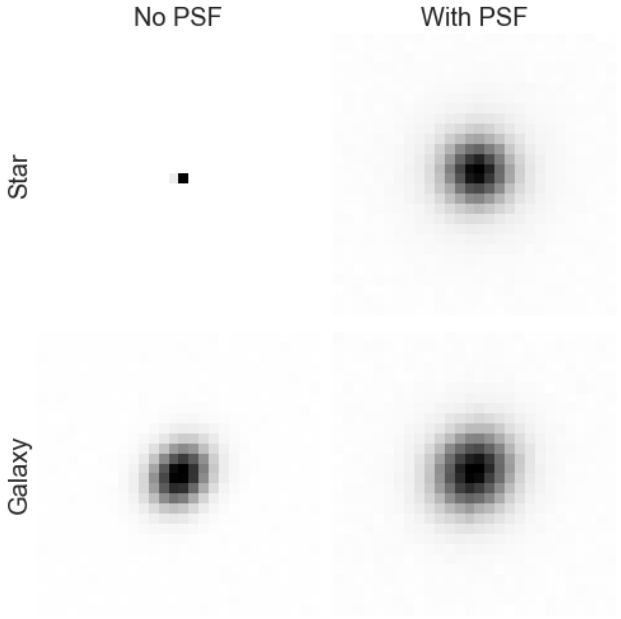

Figure 2. Example images of a star and galaxy with and without the PSF. The first row shows a star and the second row shows a galaxy. The first column is the object without a PSF applied and the second column shows the object with the PSF applied. As we can see without the PSF a star is a point like source of light but becomes spatially distributed when the PSF is applied. Thus making the problem of discriminating between stars and galaxies more challenging. Note the difference in noise level between the star and galaxy has to do with the galaxy being sampled at a dimmer magnitude. Stars at similar magnitude will have a similar noise level.

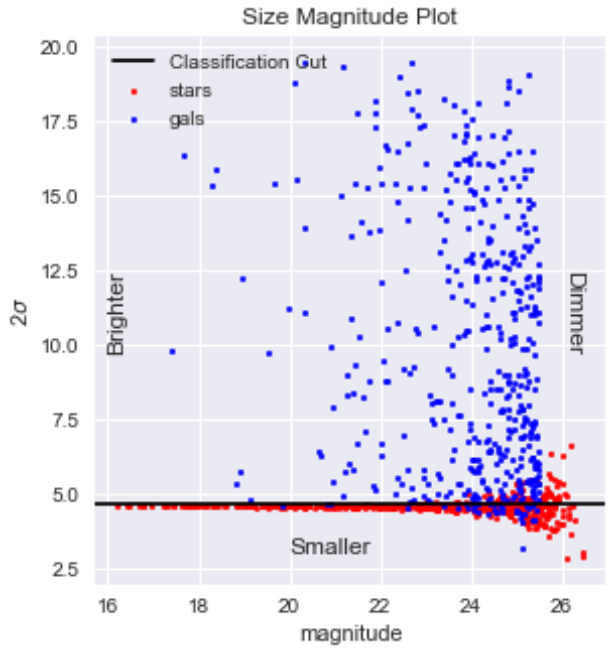

Figure 3. A size magnitude plot for a single exposure. In this figure we illustrate how star galaxy classification is typically solved by measuring the size and magnitude of each object in an exposure and making a classification cut on the size. The blue dots indicate galaxies and the red dots stars. We use a popular method used by the Dark Energy Survey (Jarvis et al., 2016) to measure size, originally developed by Hirata & Seljak (2003) and improved by Mandelbaum et al. (2012). A representative classification cut is shown by the black horizontal line. As we can see, this method achieves strong performance for bright objects, but the performance degrades for dim objects.

these fits are based on selecting from a discrete set of templates and parameterized profiles, using a chi-squared test to identify the best-fitting template for each model. From the two models, we then separately measure Mag_{psf}, the magnitude of the object weighted by the fitted model of the point spread function, and $Mag_{g_{model}}$, the magnitude measured from the galaxy fitted model. Note that the magnitude of an object is a log measure of the brightness of the object: smaller magnitudes mean brighter objects and larger magnitudes mean dimmer objects (inverse scale). We can then measure the extendedness of a single object as $Mag_{psf} - Mag_{g_{model}}$. Typical numbers for magnitudes in modern sky surveys are 15-25. Intuitively, a star being a point-like source of light, the only process responsible for spreading out its light as it travels through the atmosphere is the PSF. Thus the PSF will be the best fitted model and $Mag_{psf} - Mag_{g_{model}}$ should measure zero for a star. On the other hand, for galaxies, $Mag_{psf} - Mag_{g_{model}}$ should differ from zero, since galaxies have an inherent size and the best fitted model will not be Mag_{psf} (Garmilla, 2016).

An example of a typical size magnitude plot for a single exposure can be seen in Figure 3. As this example shows, this method does very well for bright objects (low magnitude), but the star predictions are highly contaminated by galaxies for lower magnitudes.

Astronomers have used the preceding argument to create classifiers based on thresholds of extendedness to determine if an object is a star or galaxy. This method was used for the Sloan Digital Sky Survey (SDSS) (Lupton et al., 2001) and has also been adopted by subsequent surveys including Hyper Suprime-CAM (HSC) (Bosch et al., 2018), Dark Energy Survey (DES) (Jarvis et al., 2016) and is intended to be used for LSST (Jurić et al., 2015). However as sky surveys push deeper into space, gathering data on dimmer objects, this method becomes less effective and is unable to distinguish dimmer objects (Garmilla, 2016). Thus there is a significant need for new methods that are not just better overall, but well suited to dimmer objects. The method we will discuss in Section 3 achieves better performance

both overall and specifically on dimmer objects, as shown in Section 4.

2.2. Non-IID Data and the Role of Context

Deep learning has prompted rapid progress in computer vision especially in regards to object classification and object detection (Krizhevsky et al., 2012; Girshick et al., 2014). This progress provides ample motivation to apply deep learning methods to star galaxy classification. However, it turns out that these methods cannot be applied effectively without significant modification.

In particular, the point spread function acts as a confounding factor on the input data, creating a significant and correlated source of uncertainty. The high variation of the PSF makes it entirely possible, even common, for a star in one exposure with a large PSF to appear bigger (be more extended) than a galaxy in another exposure with a smaller PSF; this situation can be seen in the example in Figure 2. The PSF can vary both across exposures and within different spatial regions of the same exposure. However, the only available evidence of the PSF is its effect on this and nearby objects in the exposure (particularly stars).

Thus, the nearby objects in an exposure provide "context" that is necessary to the prediction task: here, information about the possible magnitude and shape of the PSF. However, this information is entangled with the information of interest, i.e., whether the objects are stars or galaxies. We contend that extracting and sharing this contextual information is critical to effective star galaxy classification.

This assertion is borne out empirically: when we evaluate the performance of a standard classifier framework modeled after AlexNet (Krizhevsky et al., 2012), except using only one output neuron for a binary classifier, on the datasets described in Section 4 we find that the PSF has a major and deleterious effect. The results are reported in Table 1. If we apply a fixed PSF to all objects, the deep convolutional model achieves a strong performance; this is degraded slightly if we allow the PSF to vary over the exposure ("spatial variation") but keep it constant across exposures ("temporal variation"). However, if we move to the more realistic setting of a PSF that varies both spatially in a single exposure and across different exposures, we see performance degrade to random guessing, with little improvement even as the data set size grows.

In our initial work we also tested an R-CNN model (Girshick et al., 2014), which solves the slightly more general problem of simultaneously detecting and classifying objects in an image. In star galaxy classification, detection is typically done by an earlier processing step in the pipeline. And star galaxy detection is a much easier problem than general object detection. Unfortunately the R-CNN achieved poor performance on both detection and classification, perhaps because of the low signal to noise ratio in the data, which is characteristic of the problem. In addition, images are quite large and the vast majority of pixels view empty space (background), suggesting there may be issues due to class imbalance.

These impediments of standard classifier frameworks motivate us to develop a framework for composing neural network models to capture non-IID data effects while also allowing for varying number of objects in a given exposure.

3. ContextNet

The question we are trying to solve is thus not a straightforward classification question, but what we call a *contextual classification* question: the classification of one object cannot be determined without taking into account the context of the surrounding objects. We address this problem by dividing up the modeling procedure into three consecutive steps: local modeling, global modeling and predicting. Each of these steps is associated with its own neural network. The local network is designed to capture local features about each individual object in a single exposure, independent from all other objects in the exposure. It is then replicated for however many objects exist in an exposure. The global network is designed to take in all of the local features and produce global features that describe the exposure as a whole. Finally, the prediction network takes in the local and global features to produce a class prediction; like the local network, it is replicated as many times as there are objects and applied independently to each. A pictorial overview of the model can be seen in Figure 4.

We define a cutout of a single object $X = (x_1, x_2, ..., x_n)$ where x_i is the value of a single pixel. We define an exposure $E = \{X_j\}_{j=1}^{m}$ as collection of cutout images where there are m cutouts in each exposure. (In the sequel we show how our architecture can be extended to exposures containing more than m cutouts.) A single input to our model corresponds to one exposure.

The local network LN is a neural network that takes in a single object (cutout image) and outputs a vector of local features:

$$Y_j = (y_1, y_2, ..., y_k) = LN(X_j) \tag{1}$$

Note that, for our specific application, we also include the two coordinates representing the position of the object in the sky as a local feature in Y. The local network is replicated and applied m times to each cutout in the exposure.

The input to the global network, GN, are the concatenated outputs of the local network:

$$G = GN(Y_1, \ldots, Y_m) \tag{2}$$

Table 1. Models like AlexNet achieve excellent performance on standard image classification datasets. However, these models are not well suited to handle non-IID data. This table shows the performance of an AlexNet-like architecture for star galaxy classification. If all objects are blurred with the same PSF, the model does remarkably well, and only degrades slightly when the PSF changes spatially but is held constant across exposures (temporal variation). However, if the PSF varies between exposures the model fails completely, even when the amount of training data is doubled.

PSF Spatial Variation	PSF Temporal Variation	Training Size	Accuracy
Constant	Constant	5000 exposures	0.97
Varying	Constant	5000 exposures	0.92
Varying	Varying	5000 exposures	0.49
Varying	Varying	10000 exposures	0.51

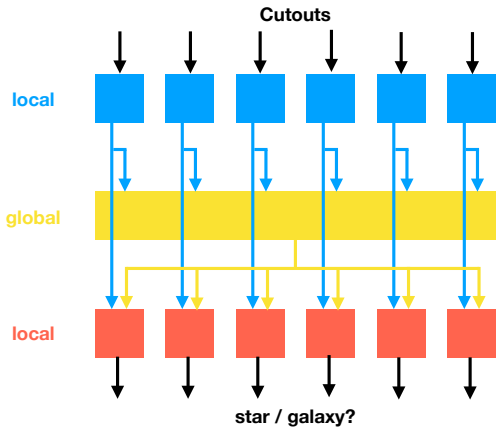

Figure 4. High level overview of the ContextNet architecture. Each color represent a different network and a different step in the modeling process. Blue is the local network that takes in cutouts and produces local features for each object. It is replicated as many times as there are objects in the exposure. Yellow is the global model that takes in all the local features and distills that into global features corresponding to the entire exposure. Red is the prediction network that takes in both the local and global features and makes the final prediction for each object. It is also replicated as many times as there are objects.

The output of GN is a single vector representing the contextual information in the exposure.

Finally, the prediction network, PN, takes in a single object's local feature vector and the global context vector, and is replicated and applied m times for each cutout image in the exposure:

$$P_i = PN(Y_i, G) \tag{3}$$

This results in m predictions, one for each object in the exposure. All three models are trained simultaneously using a binary cross entropy loss for each prediction and propagating the gradients through all three networks. Figure 4 shows a high level layout of the model.

Varying numbers of objects. In practice, the number of objects (cutout images) varies by exposure. While our basic model assumes a fixed number of objects per exposure, it is easily extended to predict on exposures with $N > m$ objects. Intuitively, we need enough objects to accurately estimate the effect of the PSF in the vicinity, but expect performance to saturate as m increases. In practice, we use a simple ensembeleing technique akin to bagging (Breiman, 1996). First, we define a minimum number of predictions n_p to make on each cutout, We then create $S_n = \lceil n_p * N/m \rceil$ sets of size m, filling each set randomly with cutouts, while ensuring that there are no duplicates in any single set and that each cutout appears in at least n_p sets. Finally, we predict on each set as previously described, and average over the predictions produced for each cutout. We can select m to be large enough to accurately assess the PSF effect, while small enough to be confident that at least m objects will appear in every exposure. We then expect most sets of size m to be sufficient to estimate the context, and averaging over n_p such sets provides a degree of robustness against unlucky set selections.

Another important point is that the architectures of the three different neural networks are quite flexible and can be mixed and matched. For our application, for example, we can use a convolutional network for the local network, and fully connected networks for the global and prediction network. It is also possible to include multiple types of local networks – for example, if the input could consist of differently sized cutouts, we could train a local network for each size, as long as the local features, Y_j, produced had the same dimensions for all local networks. Thus our ContextNet framework is not only capable of handling non-IID data, but also a variable number of inputs, and inputs of different dimensions or types.

4. Experiments

As the start of the LSST survey approaches simulations are being produced by a variety of teams and collaborations in order to test and calibrate all components of the data management (DM) pipeline. This includes the current star

Table 2. Overall comparison of ContextNet with the size magnitude based classifier in the DM stack. ContextNet does significantly better on accuracy, precision and recall. Note that achieving a high precision sample of stars is important for downstream processing tasks.

MODEL	RECALL	PRECISION	ACCURACY
CONTEXTNET	0.96	0.88	0.93
DM CLASSIFIER	0.92	0.82	0.85

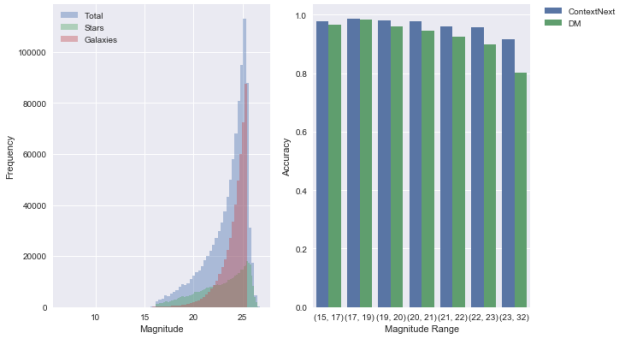

Figure 5. Left: The magnitude distribution for the DESC simulation data set used for testing. Blue indicates the distribution for all objects, red the distribution for galaxies only, and green for stars only. As we can see, galaxies tend to be dimmer than stars. Right: Accuracy of the two models, binned by magnitude, with ContextNet in blue and the DM stack classifier in green. The two models are competitive on brighter objects, but ContextNet gives much better performance for dimmer objects. Performance in this dim region is becoming increasingly important as ground-based astronomical surveys image deeper into the sky.

galaxy classifier within the DM stack. It is this simulated data that we test on and a classifier based on the DM stack that we compare with. With respect to the domain science, this is the most important comparison since the DM based classifier is what will be used if no superior method is produced. The DM classifier has been inherited and modified from the one used in SDSS and is of the form discussed in section 2. The use of simulated data, while not ideal, is necessary since ground truth on real astronomical data is not possible to obtain.

The architecture of our model is as follows: The local network takes in a cutout of dimension (28, 28) and the layers are Conv(Filters=64, kernel=(3, 3)$\rightarrow$ Elu $\rightarrow$ Conv(Filters=128, kernel=(3, 3) $\rightarrow$ Elu $\rightarrow$ Flatten $\rightarrow$ Dense(20) $\rightarrow$ Elu. We chose only 20 local features because cutouts of galaxies are not complex images as seen by figure 1 they are essentially noisy elliptical objects positioned in the center of the image. The global network takes in the concatenation of the local features from 1000 objects in the exposure with layers Dense(1000) $\rightarrow$ Elu $\rightarrow$ Dense(1000) $\rightarrow$ Elu $\rightarrow$ Dense(1000) $\rightarrow$ Elu. We chose 1000 objects as the minimum number per exposure which is approximate for LSST. Each exposure will have anywhere from 1200 - 2200 detected sources and anything less than 1000 means something most likely went catastrophically wrong with the instruments. The prediction network takes in the local features for a single object and the global features and has the following architecture Dense(100) $\rightarrow$ Elu $\rightarrow$ Dense(100) $\rightarrow$ Elu $\rightarrow$ Dense(1) $\rightarrow$ Sigmoid. The final output is the probability that the object is a galaxy and we use binary cross entropy to train.

Our training set consists of 5000 exposures each containing 1000 sources. The test set consists of 1000 exposures and each contain anywhere from 1200-2200 objects[1]. The results are presented in table 2.

As can be seen from Table 2 and Figure 5, ContextNet achieves much better overall performance; perhaps more importantly, most of this boost comes from significantly better performance for dim objects. Dim objects are especially important for LSST and future surveys as they image deeper into the sky, capturing dimmer objects than have previously been measured.

Even though it is not ideal to use simulated data, one of its advantages is that we can use the parameters that defined the objects in the simulation to better understand the predictions being made. This is especially important for deep models where interpretability is hard and often not possible. It is also necessary to develop a strong understanding of our model to convince domain scientists to adopt this approach. The main parameters used in the simulation to define stars and galaxies are their brightness or magnitude, size, eccentricity and the amount it is rotated. Stars are modeled as perfectly round objects with negligible size. Galaxies on the other had have distributions over each of these parameters. Figure 6 shows the parameters plotted against the probability of being predicted a galaxy by ContextNet. Intuitively the size of the object is the biggest signal for being predicted a galaxy. This is expected when considering that state of the art models for this problem only use the size of the object to make a prediction. We can also see that the amount the object is rotated or stretched (eccentricity) seems to play a very little role in the prediction. This is somewhat surprising considering that galaxies are the only objects that are rotated and stretched. However this is due to the fact that the PSF can cause stars to appear rotated or stretched making this a much weaker signal. From these plots we can see that the magnitude of the object also appears to be a strong signal for the predictions of ContextNet. This is also a reasonable result given that stars and galaxies have very

[1]Our code and information for data access can be found at https://github.com/NobleKennamer/ContextNet.

Figure 6. In these four plots galaxies are colored purple and stars are pink. The y-axis is the probability of classifying the object as a galaxy. From left to right: the first figure compares the probability of detecting a galaxy with the size of each object, the second with the magnitude, the third with the eccentricity and the last with the amount it was rotated from -90 to 90. Using the simulation parameters can help us interpret the predictions of the model.

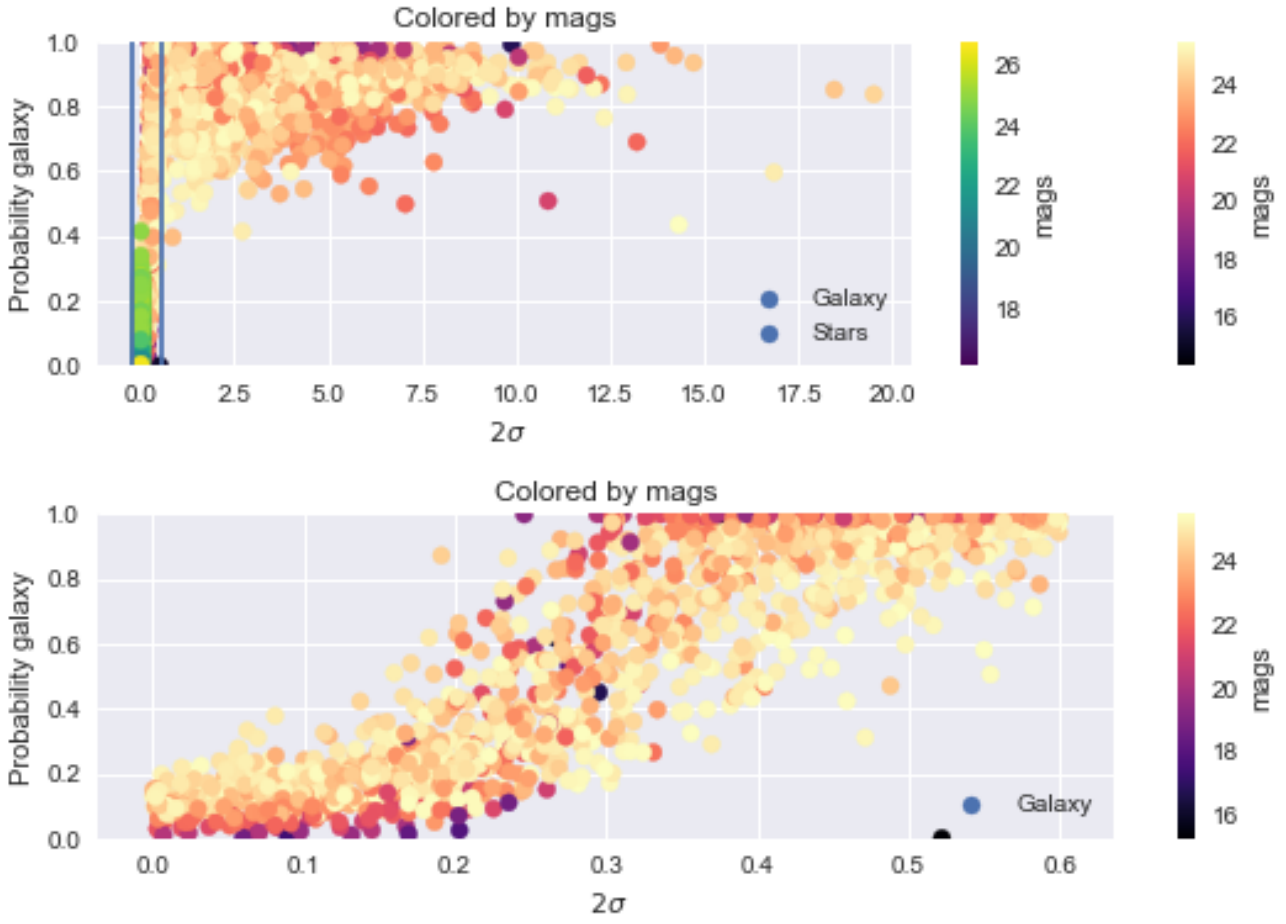

Figure 7. The top figure shows the size on each object on the x axis and the probability of classifying the object as a galaxy on the y-axis and each object is colored by its magnitude. The yellow-red colors are galaxies and the blue-green colors are stars. The bottom figure is a zoomed in figure of the top focusing on smaller objects and with the stars removed.

different magnitude distributions, as seen in Figure 5.

We examine the relationship between predictions and magnitudes further in Figure 7. Here we plot the probability of being classified as a galaxy on the y-axis and the size of each object on the x-axis. We color each object by its magnitude. The top figure shows all objects and the bottom plot zooms in on just the smaller objects. From this we can see an interesting relationship where small bright galaxies can be classified correctly as galaxies or incorrectly as stars. The incorrect classification of these small bright objects makes sense physically given that small bright galaxies do look very much like stars. But the fact that not all of these objects are classified incorrectly tells us that the model is making classifications based on more than just the size and magnitude of the object.

5. Conclusion

In this paper we presented ContextNet a framework for composing neural network models to make predictions for non IID data as well as being able to take in variable number of inputs and possible different types of data. We showed our model achieved better performance for an important problem in astronomy obtaining superior performance to a model that LSST intends to use. The model does particularly better for dim objects, which are becoming increasingly important. We also were able to partially interpret our results by relating parameters of the simulated objects with predictions made by the model. This is a necessary analysis when trying to convince domain scientists to adopt new models. In addition we discussed why standard computer vision techniques failed at this problem and needed to be extended for contextual classification. We intend to extend this model by not just having one type of local network, but several for various cutout sizes. We also intend to extend this work to measure properties about stars and galaxies in addition to classifying them.

Acknowledgements

This work supported in part by NSF under grant IIS-1254071, DARPA under the World Modelers program (W911NF18C0015) and the US Department of Energy award DE-SC0009920,

We would also like to acknowledge the MAPS program at UCI supported by NSF award 1633631.

References

Bosch, J., Armstrong, R., Bickerton, S., Furusawa, H., Ikeda, H., Koike, M., Lupton, R., Mineo, S., Price, P., Takata, T., Tanaka, M., Yasuda, N., AlSayyad, Y., Becker, A. C., Coulton, W., Coupon, J., Garmilla, J., Huang, S., Krughoff, K. S., Lang, D., Leauthaud, A., Lim, K.-T., Lust, N. B., MacArthur, L. A., Mandelbaum, R., Miyatake, H., Miyazaki, S., Murata, R., More, S., Okura, Y., Owen, R., Swinbank, J. D., Strauss, M. A., Yamada, Y., and Yamanoi, H. The Hyper Suprime-Cam software pipeline. *Publications of the Astronomical Society of Japan*, 70:S5, January 2018. doi: 10.1093/pasj/psx080.

Breiman, L. Bagging predictors. *Machine Learning*, 24(2): 123–140, August 1996.

Garmilla, J. A. *Star Galaxy Separation in Hyper Suprime-Cam and Mapping the Milky Way with Star Counts*. PhD thesis, Astrophysical Sciences Department, Princeton University, 2016.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Computer Vision and Pattern Recognition*, 2014.

Hezaveh, Y. D., Levasseur, L. P., and Marshall, P. J. Fast automated analysis of strong gravitational lenses with convolutional neural networks. *Nature*, 548:555–557, August 2017. doi: 10.1038/nature23463.

Hirata, C. and Seljak, U. Shear calibration biases in weak-lensing surveys. *Monthly Notices of the Royal Astronomy*, 343:459–480, August 2003. doi: 10.1046/j.1365-8711.2003.06683.x.

Jarvis, M., Sheldon, E., Zuntz, J., Kacprzak, T., Bridle, S. L., Amara, A., Armstrong, R., Becker, M. R., Bernstein, G. M., Bonnett, C., Chang, C., Das, R., Dietrich, J. P., Drlica-Wagner, A., Eifler, T. F., Gangkofner, C., Gruen, D., Hirsch, M., Huff, E. M., Jain, B., Kent, S., Kirk, D., MacCrann, N., Melchior, P., Plazas, A. A., Refregier, A., Rowe, B., Rykoff, E. S., Samuroff, S., Sánchez, C., Suchyta, E., Troxel, M. A., Vikram, V., Abbott, T., Abdalla, F. B., Allam, S., Annis, J., Benoit-Lévy, A., Bertin, E., Brooks, D., Buckley-Geer, E., Burke, D. L., Capozzi, D., Carnero Rosell, A., Carrasco Kind, M., Carretero, J., Castander, F. J., Clampitt, J., Crocce, M., Cunha, C. E., D'Andrea, C. B., da Costa, L. N., DePoy, D. L., Desai, S., Diehl, H. T., Doel, P., Fausti Neto, A., Flaugher, B., Fosalba, P., Frieman, J., Gaztanaga, E., Gerdes, D. W., Gruendl, R. A., Gutierrez, G., Honscheid, K., James, D. J., Kuehn, K., Kuropatkin, N., Lahav, O., Li, T. S., Lima, M., March, M., Martini, P., Miquel, R., Mohr, J. J., Neilsen, E., Nord, B., Ogando, R., Reil, K., Romer, A. K., Roodman, A., Sako, M., Sanchez, E., Scarpine, V., Schubnell, M., Sevilla-Noarbe, I., Smith, R. C., Soares-Santos, M., Sobreira, F., Swanson, M. E. C., Tarle, G., Thaler, J., Thomas, D., Walker, A. R., and Wechsler, R. H. The DES Science Verification weak lensing shear catalogues. *Monthly Notices of the Royal Astronomy*, 460:2245–2281, August 2016. doi: 10.1093/mnras/stw990.

Jurić, M., Kantor, J., Lim, K., Lupton, R. H., Dubois-Felsmann, G., Jenness, T., Axelrod, T. S., Aleksić, J., Allsman, R. A., AlSayyad, Y., Alt, J., Armstrong, R., Basney, J., Becker, A. C., Becla, J., Bickerton, S. J., Biswas, R., Bosch, J., Boutigny, D., Carrasco Kind, M., Ciardi, D. R., Connolly, A. J., Daniel, S. F., Daues, G. E., Economou, F., Chiang, H.-F., Fausti, A., Fisher-Levine, M., Freemon, D. M., Gee, P., Gris, P., Hernandez, F., Hoblitt, J., Ivezić, Ž., Jammes, F., Jevremović, D., Jones, R. L., Bryce Kalmbach, J., Kasliwal, V. P., Krughoff, K. S., Lang, D., Lurie, J., Lust, N. B., Mullally, F., MacArthur, L. A., Melchior, P., Moeyens, J., Nidever, D. L., Owen, R., Parejko, J. K., Peterson, J. M., Petravick, D., Pietrowicz, S. R., Price, P. A., Reiss, D. J., Shaw, R. A., Sick, J., Slater, C. T., Strauss, M. A., Sullivan, I. S., Swinbank, J. D., Van Dyk, S., Vujčić, V., Withers, A., Yoachim, P., and LSST Project, f. t. The LSST Data Management System. *ArXiv e-prints*, December 2015.

Kahn, S. M., Kurita, N., Gilmore, K., Nordby, M., O'Connor, P., Schindler, R., Oliver, J., Van Berg, R., Olivier, S., Riot, V., Antilogus, P., Schalk, T., Huffer, M., Bowden, G., Singal, J., and Foss, M. Design and development of the 3.2 gigapixel camera for the Large Synoptic Survey Telescope. In McLean, I. S., Ramsay, S. K., and Takami, H. (eds.), *Ground-based and Airborne Instrumentation for Astronomy III*, volume 7735 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pp. 0, 2010. doi: 10.1117/12.857920.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. In Langley, P. (ed.), *Advances in neural information processing systems*, pp. 1097–1105, 2012.

Lochner, M., McEwen, J. D., Peiris, H. V., Lahav, O., and Winter, M. K. Photometric Supernova Classification with Machine Learning. *The Astrophysical Journal, Supplement*, 225:31, August 2016. doi: 10.3847/0067-0049/225/2/31.

LSST Science Collaboration, Abell, P. A., Allison, J., Anderson, S. F., Andrew, J. R., Angel, J. R. P., Armus, L., Arnett, D., Asztalos, S. J., Axelrod, T. S., and et al. LSST Science Book, Version 2.0. *ArXiv e-prints*, December 2009.

Lupton, R., Gunn, J. E., Ivezić, Z., Knapp, G. R., and Kent, S. The SDSS Imaging Pipelines. In Harnden, Jr., F. R., Primini, F. A., and Payne, H. E. (eds.), *Astronomical Data Analysis Software and Systems X*, volume 238 of *Astronomical Society of the Pacific Conference Series*, pp. 269, 2001.

Mandelbaum, R., Hirata, C. M., Leauthaud, A., Massey, R. J., and Rhodes, J. Precision simulation of ground-based lensing data using observations from space. *Monthly Notices of the Royal Astronomy*, 420:1518–1540, February 2012. doi: 10.1111/j.1365-2966.2011.20138.x.

Regier, J., Miller, A., McAuliffe, J., Adams, R., Hoffman, M., Lang, D., Schlegel, D., and Prabhat, M. Celeste: Variational inference for a generative model of astronomical images. In Bach, F. and Blei, D. (eds.), *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2095–2103, Lille, France, 07–09 Jul 2015. PMLR. URL http://proceedings.mlr.press/v37/regier15.html.

Rowe, B. T. P., Jarvis, M., Mandelbaum, R., Bernstein, G. M., Bosch, J., Simet, M., Meyers, J. E., Kacprzak, T., Nakajima, R., Zuntz, J., Miyatake, H., Dietrich, J. P., Armstrong, R., Melchior, P., and Gill, M. S. S. GALSIM: The modular galaxy image simulation toolkit. *Astronomy and Computing*, 10:121–150, April 2015. doi: 10.1016/j.ascom.2015.02.002.