MNRAS **000**, 1–8 (2015) Preprint 16 July 2016 Compiled using MNRAS LATEX style file v3.0

Applying Deep Learning to the Image Photo-z Problem

Samantha E. Thrush[1]★ and Robert J. Brunner[1,2,3,4]
[1]*Department of Astronomy, University of Illinois, Urbana, IL 61801 USA*
[2]*Department of Physics, University of Illinois, Urbana, IL 61801 USA*
[3]*Department of Statistics, University of Illinois, Champaign, IL 61820 USA*
[4]*National Center for Supercomputing Applications, Urbana, IL 61801 USA*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT
When large data sets of galaxies are collected, their associated redshifts are usually collected by finding their photometric redshifts, which takes additional time for a survery to complete. However, if deep learning via deep convolutional neural networks (ConvNets) are utilized, photometric redshifts of galaxies can be predicted without the need for photometry to be completed by the associated survey. In this paper, we present an application of ConvNets on reduced and pixelated galaxy images and redshift data from the Sloan Digital Sky Survey (SDSS). We will show that ConvNets are able to accurately predict the redshift of a sample galaxy image in a way that is competitive with previous works on this subject.

Key words: methods: data analysis – techniques: image processing – methods: statistical – surveys – galaxies:statistics.

1 INTRODUCTION

HERE IS AN EXAMPLE OF NORMAL TEXT. HERE IS AN EXAMPLE OF A CITED PIECE OF TEXT (Ross et al. 2011; Seo et al. 2012).

IN ORDER TO START A NEW PARAGRAPH, THERE MUST BE A BLANK LINE IN BETWEEN TWO PARAGRAPHS.

HERE I REFER TO A SECTION BY NAME IN LATEX, AKA SECTION 2.

2 DATA

2.1 Sloan Digital Sky Survey

TEXT AFTER A FIGURE. HERE IS AN EXAMPLE OF A FOOT NOTE[1].

AN EXAMPLE OF "COMPUTER-LOOKING" TEXT: (`zWarning = 0`)

AN EXAMPLE OF ITALICIZED TEXT: *r*

AN EXAMPLE OF THE "TIMES" OPERATOR BETWEEN TWO NUMBERS: 48×48 pixels with luptitude values in each pixel.

AN EQUATION PRESENTED INLINE WITH THE TEXT: $10.7 < r < 23.1$, and the galaxies in this sample have a mean redshift of $z \sim 0.36$. The *ugriz* color-color space of the sample is plotted in Figure 1.

The first two panels of Figure 11 show the number of objects and the fraction of stars in the test set as functions of *r*-band magnitude. Similarly, Figure 13 shows the number of objects and the fraction of stars in the test set as functions of $g - r$ color.

We do not show the distributions for the training and validation sets in Figures 11 and 13 to avoid cluttering the plots.

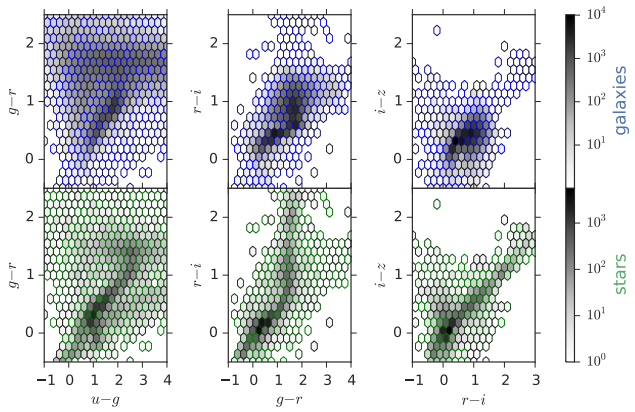


Figure 1. Hexbin plots showing the color-color space distribution of galaxies (blue, top row) and stars (green, bottom row) in the SDSS data set used in this analysis. It is easy to see that there is significant overlap between stars and galaxies in these color-color spaces.

★ thrush2@illinois.edu
[1] http://skyserver.sdss.org/casjobs/

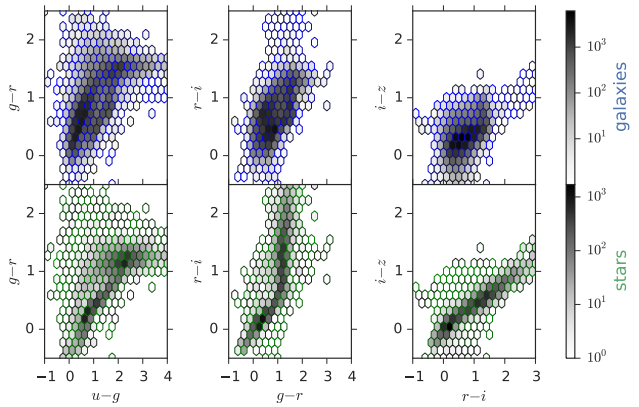


Figure 2. Hexbin plots showing the color-color space distribution of galaxies (blue, top row) and stars (green, bottom row) in the CFHTLenS data set used in this analysis.

2.2 Canada-France-Hawaii Telescope Lensing Survey

We also use photometric data from the Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS[2]; Heymans et al. 2012; Erben et al. 2013; Hildebrandt et al. 2012).

3 DEEP LEARNING

3.1 Neural Networks

An artificial neuron in most artificial neural networks is represented as a mathematical function that models a biological neural structure (Figure 3a). Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ be a vector of inputs to a given neuron, $\boldsymbol{w} = (w_1, w_2, \ldots, w_n)$ be a vector of weights, and b be the bias. Then, the output of the neuron is

$$y = \sigma(\boldsymbol{w} \cdot \boldsymbol{x} + b), \tag{1}$$

where σ is the activation function (or *non-linearity*).

A schematic representation is shown in Figure 3b.

The gradients in 1 can be computed using the *backpropagation* procedure (Rumelhart et al. 1988),

$$\frac{\partial L}{\partial y_i} = y_i - \hat{y}_i \tag{2}$$

$$\frac{\partial L}{\partial x_i} = \frac{\partial L}{\partial y_i}\frac{\partial y_i}{\partial x_i}. \tag{3}$$

At each hidden layer, the derivative of the loss function with respect to the output of each unit is a weighted sum of the derivatives with respect to the inputs of the subsequent layer,

$$\frac{\partial L}{\partial y_j} = \sum_i w_{ji}\frac{\partial L}{\partial x_i} \tag{4}$$

$$\frac{\partial L}{\partial x_j} = \frac{\partial L}{\partial y_j}\frac{\partial y_j}{\partial x_j}. \tag{5}$$

Finally, once the gradients $\partial L/\partial x_j$ have been computed, it is straightforward to compute the gradients for the weight,

$$\frac{\partial L}{\partial w_{jk}} = y_k\frac{\partial L}{\partial x_j}. \tag{6}$$

[2] http://www.cfhtlens.org/

3.2 Convolutional Neural Networks

Mathematically, we replace the dot product in Equation 1 with a sum of convolutions. Thus, the k-th feature map is given by

$$y^k = \sigma\left(\sum_m \boldsymbol{w}_m^k * \boldsymbol{x}_m + b^k\right), \tag{7}$$

where we sum over the set of input feature maps, $*$ is the convolution operator, and $\boldsymbol{w}_m^k$ represent the filters.

3.3 Neural Network Architecture

We present the overall architecture of our ConvNet model in Table 1.

The architecture of Krizhevsky et al. (2012) uses relatively large receptive fields (11×11) in the first convolutional layers. Zeiler & Fergus (2014) and Dieleman et al. (2015) also use large receptive fields of 7×7 and 6×6 in the first convolution layer, respectively.

In the interest of scientific reproducibility, we make all our code available at https://github.com/EdwardJKim/dl4astro.

4 REDUCING OVERFITTING

Our convolution neural network has 11×10^6 learnable parameters, while there are only 4×10^4 images in the training set.

4.1 Data Augmentation

One common method to combat overfitting is to artificially increase the number of training data by using label-preserving transformations (Krizhevsky et al. 2012; Dieleman et al. 2015, 2016). Each image is transformed as follows:

- Rotation: Rotating an image does not change whether the object is a star or a galaxy. We exploit this rotational symmetry and randomly rotate each image by a multiple of 90°.
- Reflection: We flip each image horizontally with a probability of 0.5 to exploit mirror symmetry.
- Translation: We also have translational symmetry in the images. Given an image of size 48×48 pixels, we extract a random contiguous crop of size 44×44. Each cropping is equivalent to randomly shifting a 44×44 image by up to 4 pixels vertically and/or horizontally.
- Gaussian noise: We introduce random Gaussian noise to each pixel values by using a similar method to Krizhevsky et al. (2012).

These data augmentation schemes add almost no computational cost, as they are performed on the CPU while the GPU is training the ConvNets on images.

4.2 Dropout

We use a regularization technique called dropout (Hinton et al. 2012) in the fully-connected layers.

4.3 Model Combination

The posterior probability that a source is a galaxy is given by

$$P(G|\boldsymbol{x}, \boldsymbol{D}, \boldsymbol{M}, \boldsymbol{E}) = \sum_{e \in \boldsymbol{E}} P(G|\boldsymbol{x}, \boldsymbol{M}, e)\, P(e|\boldsymbol{D}), \tag{8}$$

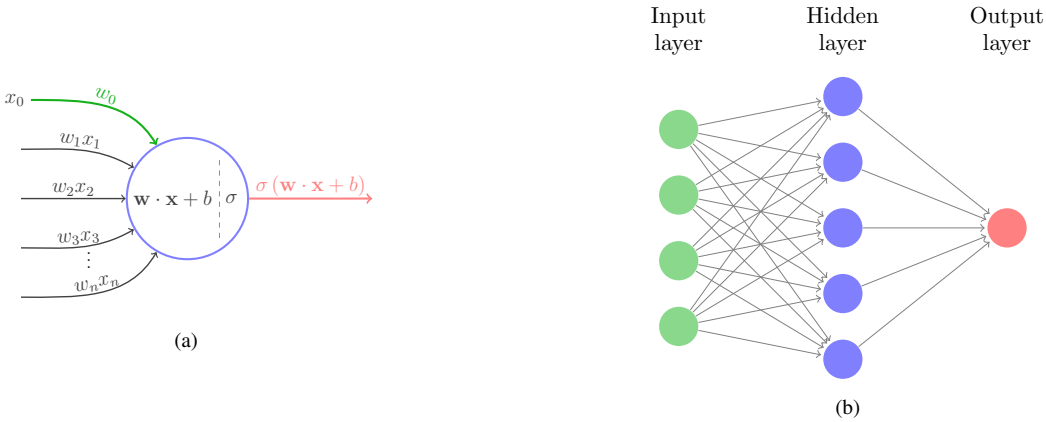


Figure 3. (a) A mathematical model of a biological neuron. (b) A schematic diagram of a neural network with one hidden layer.

Table 1. Summary of ConvNet architecture and hyperparameters. Note that pooling layers have no learnable parameters.

type	filters	filter size	padding	non-linearity	initial weights	initial biases
convolutional	32	5×5	-	leaky ReLU	orthogonal	0.1
convolutional	32	3×3	1	leaky ReLU	orthogonal	0.1
pooling	-	2×2	-	-	-	-
convolutional	64	3×3	1	leaky ReLU	orthogonal	0.1
convolutional	64	3×3	1	leaky ReLU	orthogonal	0.1
convolutional	64	3×3	1	leaky ReLU	orthogonal	0.1
pooling	-	2×2	-	-	-	-
convolutional	128	3×3	1	leaky ReLU	orthogonal	0.1
convolutional	128	3×3	1	leaky ReLU	orthogonal	0.1
convolutional	128	3×3	1	leaky ReLU	orthogonal	0.1
pooling	-	2×2	-	-	-	-
fully-connected	2048	-	-	leaky ReLU	orthogonal	0.01
fully-connected	2048	-	-	leaky ReLU	orthogonal	0.01
fully-connected	2	-	-	softmax	orthogonal	0.01

where $\boldsymbol{x}$ is the input, $\boldsymbol{D}$ is the data set, $\boldsymbol{M}$ is the set of models (i.e., different transformations of input images), and e is an element in the ensemble space $\boldsymbol{E}$ of possible model combinations. By Bayes' Theorem, the posterior probability of e given $\boldsymbol{D}$ is given by

$$P(e|\boldsymbol{D}) = \frac{P(e)}{P(\boldsymbol{D})}\prod_{d\in\boldsymbol{D}} P(d|e) \propto P(e)\prod_{d\in\boldsymbol{D}} P(d|e)\,. \tag{9}$$

For probabilistic classifiers, we can directly use the probabilistic predictions and write Equation 9 as

$$P(e|\boldsymbol{D}) \propto P(e)\prod_{i=1}^{N} \hat{y}_i y_i + (1-\hat{y}_i)(1-y_i)\,. \tag{10}$$

5 TREES FOR PROBABILISTIC CLASSIFICATIONS

TPC is a part of a publicly available software package called MLZ[3] (Machine Learning for Photo-z).

$$I_G(D_{\rm node}, X) = I_d(D_{\rm node}) - \sum_{x\in{\rm values}(X)} \frac{|D_{{\rm node},x}|}{|D_{\rm node}|} I_d\left(D_{{\rm node},x}\right), \tag{11}$$

where $D_{\rm node}$ is the training data in a given node, X is one of the possible dimensions (e.g., magnitudes or colors)

[3] http://pythonhosted.org/MLZ/

Table 2. The definition of the classification performance metrics.

Metric	Meaning
AUC	Area under the Receiver Operating Curve
MSE	Mean squared error
c_g	Galaxy completeness
p_g	Galaxy purity
c_s	Star completeness
p_s	Star purity
$p_g(c_g = x)$	Galaxy purity at x galaxy completeness
$c_s(p_s = x)$	Star completeness at x star purity
CAL	Calibration error with overlapping binning
$\mid \Delta N_g \mid /N_g$	Absolute error in number of galaxies

6 RESULTS AND DISCUSSION

In this section, we first describe the performance metrics that were used for evaluating the models.

6.1 Classification Metrics

Probabilistic classification models can be considered as functions that output a probability estimate of each source to be in one of the classes (e.g., a star or a galaxy).

6.1.1 Receiver Operating Characteristic Curve

When we have no information about the operating condition when evaluating the performance of classifiers, there are effective tools such as the Receiver Operating Characteristic (ROC) curve (Swets, Dawes & Monahan 2000).

6.1.2 Completeness and Purity

We define the galaxy *completeness* c_g (also known as recall or sensitivity) as the fraction of the number of true galaxies classified as galaxies out of the total number of true galaxies,

6.1.3 Mean Squared Error

We also use the mean squared error (MSE; also known as the Brier score (Brier 1950)) as a performance metric. We define MSE as

$$\mathrm{MSE} = \frac{1}{N}\sum_{j=1}^{N}\left(y_j - \hat{y}_j\right)^2, \tag{12}$$

The MSE can be considered as both a score function that quantifies how well a set of probabilistic predictions is calibrated, or a loss function.

6.1.4 Calibration Error

A fully probabilistic classifier predicts not only the class label, but also its confidence level on the prediction.

6.1.5 Number of galaxies

Ideally, the probabilistic output of a classifier would be used in subsequent scientific analyses.

6.2 CFHTLenS

As described in Section 3.2, we train our ConvNet model by monitoring its performance on the validation set.

Figure 7 shows the calibration curve that compares P_{gal}, the fraction of objects that are galaxies (as determined from their spectra), to P_{conv}, the probabilistic output that our ConvNet model produces.

6.3 SDSS

We have also trained and tested our ConvNet model on the SDSS data set, and we present in Table 4 the same six metrics for ConvNet, $\mathrm{TPC}_{\mathrm{morph}}$, and $\mathrm{TPC}_{\mathrm{phot}}$.

In Figure 11, we compare the galaxy purity and star completeness values for ConvNet, $\mathrm{TPC}_{\mathrm{morph}}$, and $\mathrm{TPC}_{\mathrm{phot}}$ as a function of r-band magnitude for the differential counts in the SDSS data.

7 CONCLUSIONS

We have presented a convolutional neural network for classifying stars and galaxies in the SDSS and CFHTLenS photometric images.

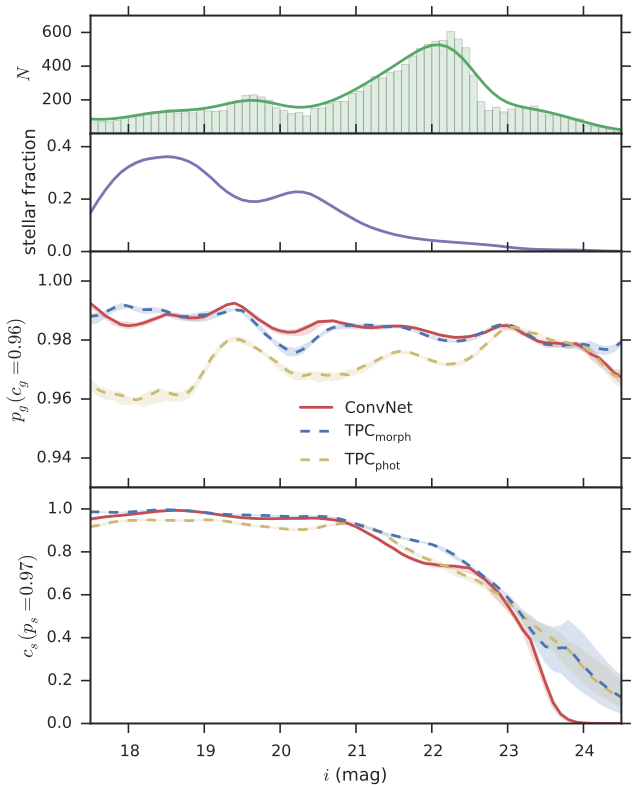


Figure 4. Galaxy purity and star completeness values as functions of the i-band magnitude (differential counts) as estimated by kernel density estimation (KDE) in the CFHTLenS data set. The top panel shows the histogram with a bin size of 0.1 mag and the KDE for objects in the test set. The second panel shows the fraction of stars estimated by KDE as a function of magnitude. The bottom two panels compare the galaxy purity and star completeness values for ConvNet (red solid line), $\mathrm{TPC}_{\mathrm{morph}}$ (blue dashed line), and $\mathrm{TPC}_{\mathrm{phot}}$ (yellow dashed line) as functions of magnitude. The 1σ confidence bands are estimated by bootstrap sampling.

ACKNOWLEDGEMENTS

We acknowledge support from the National Science Foundation Grant No. AST-1313415. RJB acknowledges support as an Associate within the Center for Advanced Study at the University of Illinois.

This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575.

This work is based on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/DAPNIA, at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council (NRC) of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This research used the facilities of the Canadian Astronomy Data Centre operated by the National Research Council of Canada with the support of the Canadian Space Agency. CFHTLenS data processing was made possible thanks to significant computing support from the NSERC Research Tools and Instruments grant program.

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Sci-

Table 3. A summary of the classification performance metrics as applied to the CFHTLenS data. The definition of the metrics is summarized in Table 2. The bold entries highlight the best performance values within each column.

classifier	AUC	MSE	$p_g(c_g = 0.96)$	$c_s(p_s = 0.97)$	CAL	$\lvert\Delta N_g\rvert/N_g$
ConvNet	**0.9948**	0.0112	**0.9972**	0.8971	**0.0197**	**0.0029**
$\text{TPC}_{\text{morph}}$	0.9924	**0.0109**	0.9963	**0.9268**	0.0245	0.0056
TPC_{phot}	0.9876	0.0189	0.9927	0.8044	0.0266	0.0101

Table 4. A summary of the classification performance metrics as applied to the SDSS data.

classifier	AUC	MSE	$p_g(c_g = 0.96)$	$c_s(p_s = 0.97)$	CAL	$\lvert\Delta N_g\rvert/N_g$
ConvNet	0.9952	0.0182	0.9915	0.9500	**0.0243**	0.0157
$\text{TPC}_{\text{morph}}$	**0.9967**	**0.0099**	**0.9977**	**0.9810**	0.0254	**0.0044**
TPC_{phot}	0.9886	0.0283	0.9819	0.8879	0.0316	0.0160

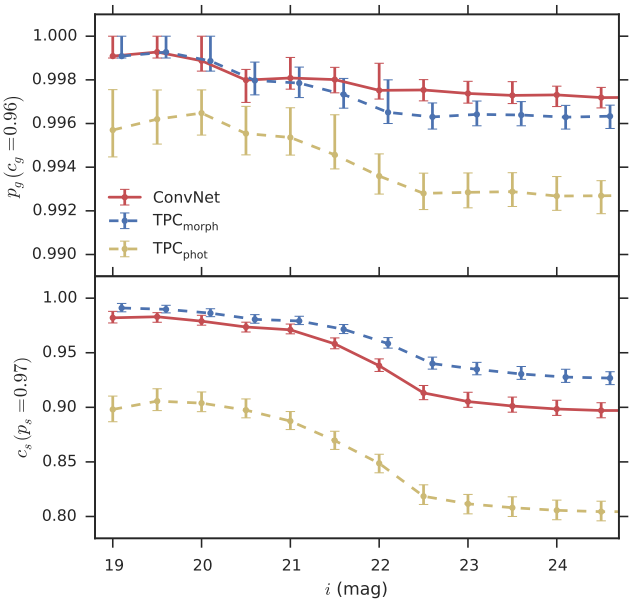


Figure 5. Galaxy purity and star completeness as functions of the i-band magnitude (integrated counts) in the CFHTLenS data set. The upper panel compares the galaxy purity values for ConvNet (red solid line), $\text{TPC}_{\text{morph}}$ (blue dashed line), and TPC_{phot} (yellow dashed line). The lower panel compares the star completeness values. The 1σ error bars are computed following the method of Paterno (2003) to avoid the unphysical errors of binomial or Poisson statistics.

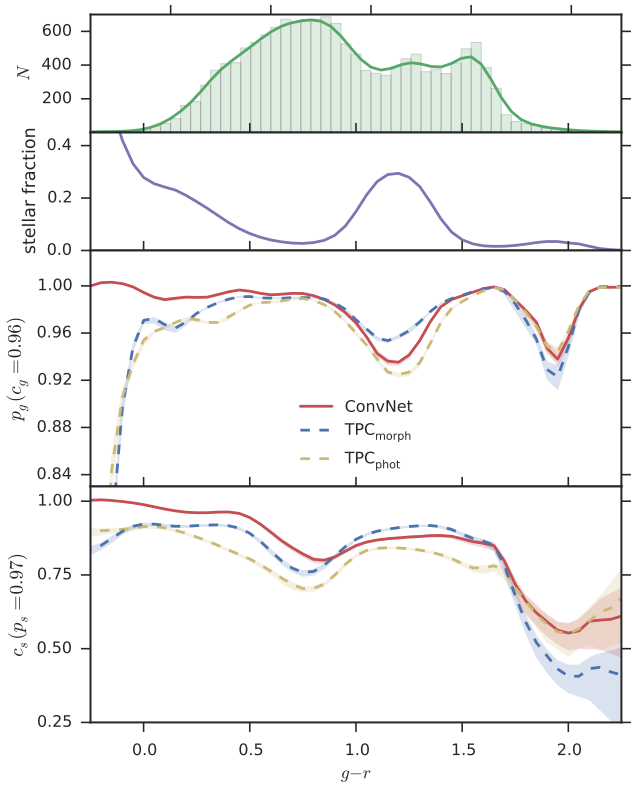


Figure 6. Similar to Figure 4 but as a function of $g-r$ color. The bin size of histogram in the top panel is 0.05.

ence Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is http://www.sdss3.org/.

SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

REFERENCES

Alam S., et al., 2015, ApJS, 219, 12
Ball N. M., Brunner R. J., Myers A. D., Tcheng D., 2006, ApJ, 650, 497
Ball N. M., Brunner R. J., Myers A. D., Strand N. E., Alberts S. L., Tcheng D., Llorà X., 2007, ApJ, 663, 774
Ball N. M., Brunner R. J., Myers A. D., Strand N. E., Alberts S. L., Tcheng D., 2008, ApJ, 683, 12
Banerji M., et al., 2010, MNRAS, 406, 342

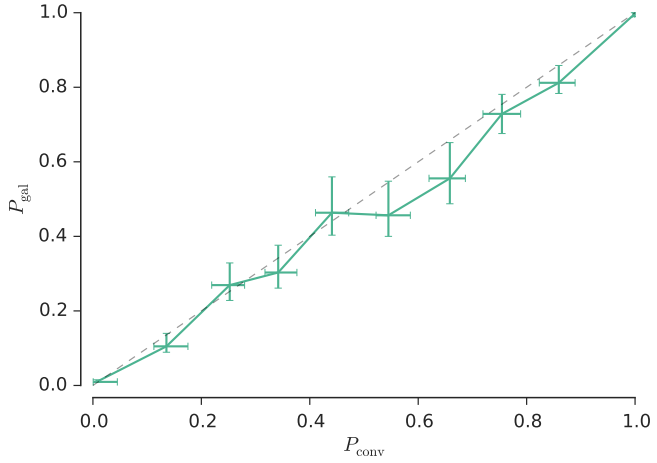


Figure 7. The calibration curve for ConvNet as applied to the CFHTLenS data set. $P_{\rm gal}$ is the fraction of objects that are galaxies, and $P_{\rm conv}$ is the probabilistic output generated by ConvNet. The dashed line displays the relationship $P_{\rm gal} = P_{\rm conv}$. The 1σ error bars are computed following the method of Paterno (2003).

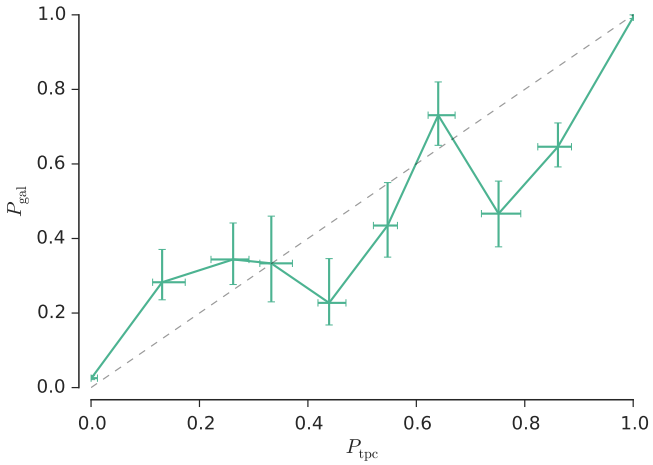
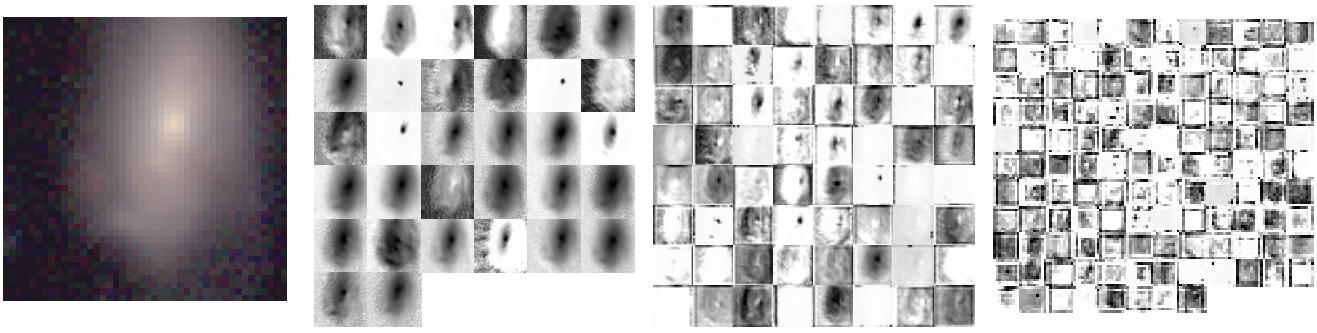
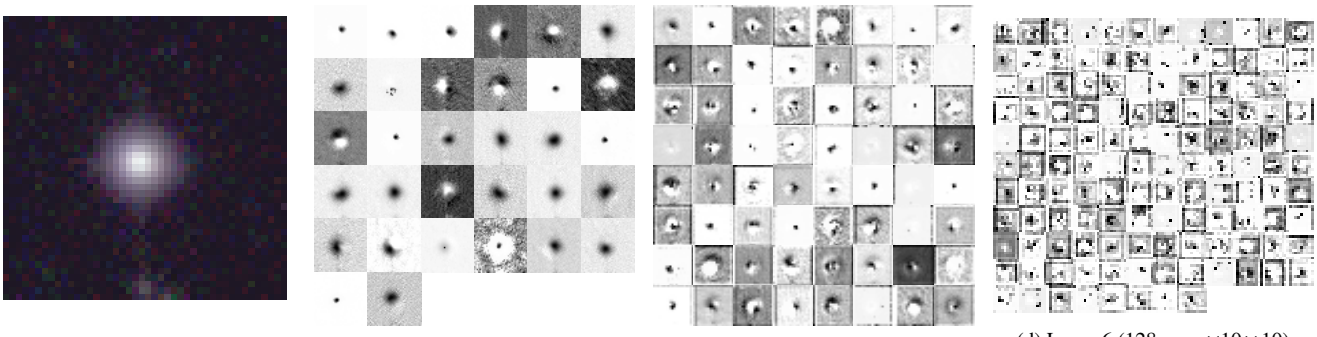


Figure 8. Similar to Figure 7 but for $\mathrm{TPC}_{\rm morph}$. $P_{\rm tpc}$ is the probabilistic output generated by $\mathrm{TPC}_{\rm morph}$.



(a) Input (5 bands×44×44) (b) Layer 1 (32 maps×40×40) (c) Layer 3 (64 maps×20×20) (d) Layer 6 (128 maps×10×10)

Figure 9. (a) A sample 44×44 RGB image of a galaxy in the CFHTLenS data set. The RGB image is created by mapping R → *i* band magnitude, G → *r* band magnitude, and B → *g* band magnitude. (b) Activations on the first convolutional layer when a 5×44×44 image is fed into the network. (c) Activations on the third convolutional layer. (d) Activations on the sixth convolutional layer. Each image in (a), (b), and (c) is a feature map corresponding to the output for one of the learned features.



(a) Input (5 bands×44×44) (b) Layer 1 (32 maps×40×40) (c) Layer 3 (64 maps×20×20) (d) Layer 6 (128 maps×10×10)

Figure 10. Similar to Figure 9 but for a star in the CFHTLenS data set.

Bengio Y., Boulanger-Lewandowski N., Pascanu R., 2013, in Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. pp 8624–8628
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Breiman L., 2001, Machine learning, 45, 5
Breiman L., Friedman J., Stone C. J., Olshen R. A., 1984, Classification and regression trees. CRC press
Brier G. W., 1950, Monthly weather review, 78, 1
Carrasco Kind M., Brunner R. J., 2013, MNRAS, 432, 1483
Carrasco Kind M., Brunner R. J., 2014a, MNRAS, 438, 3409
Carrasco Kind M., Brunner R. J., 2014b, MNRAS, 442, 3380
Caruana R., Niculescu-Mizil A., 2004, in Proceedings of the tenth ACM

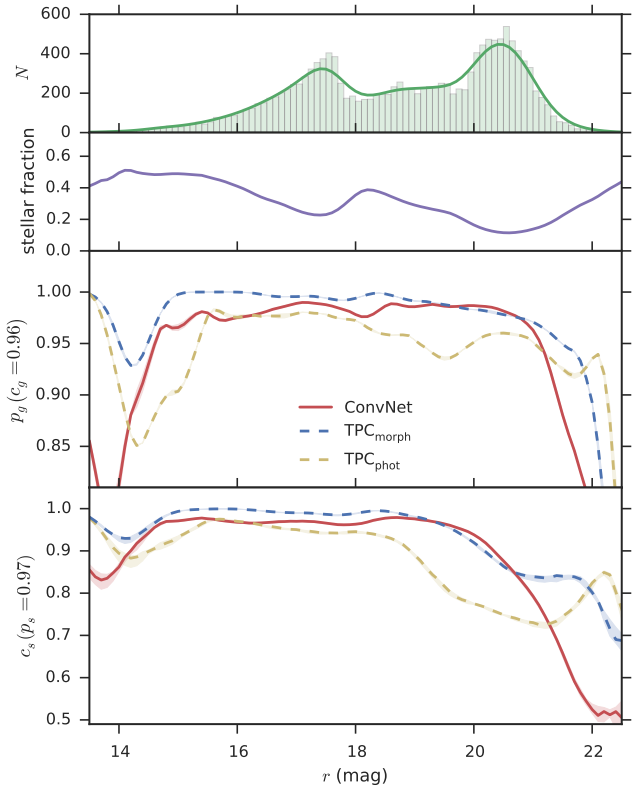


Figure 11. Galaxy purity and star completeness as function of the r-band magnitude for the differential counts in the SDSS data set.

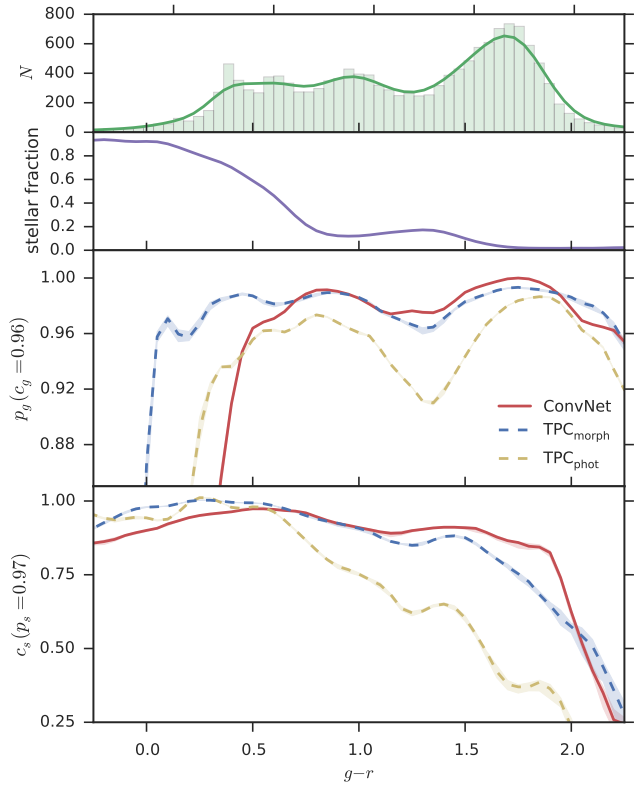


Figure 13. Similar to Figure 11 but as a function of $g-r$ color. The bin size of histogram in the top panel is 0.05.

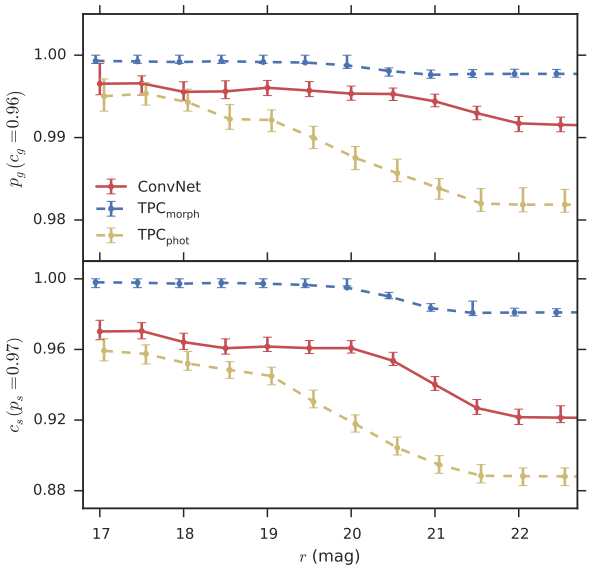


Figure 12. Galaxy purity and star completeness as functions of the r-band magnitude for the integrated counts in the SDSS data set.

SIGKDD international conference on Knowledge discovery and data mining. pp 69–78
Crocce M., et al., 2016, MNRAS, 455, 4301
Davis M., et al., 2003, Astronomical Telescopes and Instrumentation, pp 161–172
DeGroot M. H., Fienberg S. E., 1983, The statistician, pp 12–22
Desai S., et al., 2012, ApJ, 757, 83
Dieleman S., et al. 2015, Lasagne: First release., doi:10.5281/zenodo.27878
Dieleman S., Willett K. W., Dambre J., 2015, MNRAS, 450, 1441
Dieleman S., De Fauw J., Kavukcuoglu K., 2016, arXiv preprint arXiv:1602.02660
Eisenstein D. J., et al., 2011, AJ, 142, 72
Erben T., et al., 2013, MNRAS, p. stt928
Fadely R., Hogg D. W., Willman B., 2012, ApJ, 760, 15
Fukushima K., 1980, Biological cybernetics, 36, 193
Garilli B., et al., 2008, A&A, 486, 683
Garilli B., et al., 2014, A&A, 562, A23
Gwyn S. D., 2012, AJ, 143, 38
He K., Zhang X., Ren S., Sun J., 2015, arXiv preprint arXiv:1512.03385
Henrion M., Mortlock D. J., Hand D. J., Gandy A., 2011, MNRAS, 412, 2286
Heymans C., et al., 2012, MNRAS, 427, 146
Hildebrandt H., et al., 2012, MNRAS, 421, 2355
Hinton G. E., Srivastava N., Krizhevsky A., Sutskever I., Salakhutdinov R. R., 2012, arXiv preprint arXiv:1207.0580
Hoyle B., 2015, arXiv preprint arXiv:1504.07255
Huertas-Company M., et al., 2015, ApJS, 221, 8
Ivezić Ž., Connolly A. J., VanderPlas J. T., Gray A., 2014, Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data. Princeton University Press
Kamdar H. M., Turk M. J., Brunner R. J., 2016a, MNRAS, 455, 642
Kamdar H. M., Turk M. J., Brunner R. J., 2016b, MNRAS, 457, 1162
Kim E. J., Brunner R. J., Kind M. C., 2015, MNRAS, 453, 507

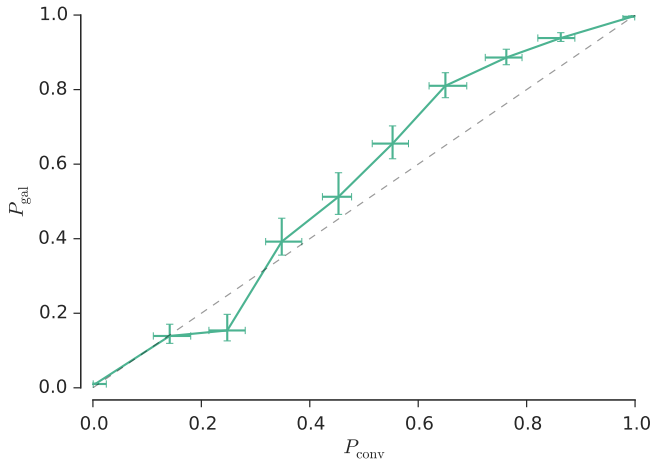


Figure 14. Calibration curve for ConvNet as applied to the SDSS data set.

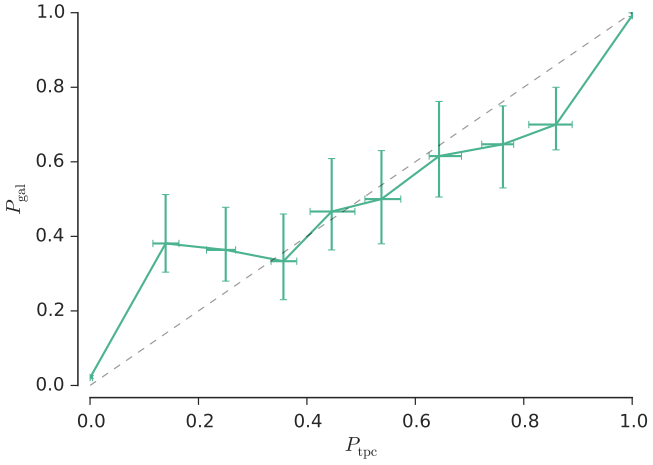


Figure 15. Calibration curve for $\mathrm{TPC}_{\rm morph}$ as applied to the SDSS data set.

Krizhevsky A., Sutskever I., Hinton G. E., 2012, in Advances in neural information processing systems. pp 1097–1105
Le Fèvre O., et al., 2005, A&A, 439, 845
LeCun Y., Bottou L., Bengio Y., Haffner P., 1998, Proceedings of the IEEE, 86, 2278
LeCun Y., Bengio Y., Hinton G., 2015, Nature, 521, 436
Li N., Thakar A. R., 2008, Computing in Science & Engineering, 10, 18
Lupton R. H., Gunn J. E., Szalay A. S., 1999, AJ, 118, 1406
Monteith K., Carroll J. L., Seppi K., Martinez T., 2011, in Neural Networks (IJCNN), The 2011 International Joint Conference on. pp 2657–2663
Nair V., Hinton G. E., 2010, in Proceedings of the 27th International Conference on Machine Learning (ICML-10). pp 807–814
Newman J. A., et al., 2013, ApJS, 208, 5
Odewahn S. C., Stockwell E. B., Pennington R. L., Humphreys R. M., Zumach W. A., 1992, AJ, 103, 318
Paterno M., 2003, Calculating Efficiencies and Their Uncertainties, http://home.fnal.gov/~paterno/images/effic.pdf
Ross A. J., et al., 2011, MNRAS, 417, 1350
Rumelhart D. E., Hinton G. E., Williams R. J., 1988, Cognitive modeling, 5, 1
Russakovsky O., et al., 2015, International Journal of Computer Vision, 115, 211
Saxe A. M., McClelland J. L., Ganguli S., 2013, arXiv preprint arXiv:1312.6120
Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525
Seo H.-J., et al., 2012, ApJ, 761, 13
Sevilla-Noarbe I., Etayo-Sotos P., 2015, Astronomy and Computing, 11, 64
Silverman B. W., 1986, CRC press, 26
Simonyan K., Zisserman A., 2014, arXiv preprint arXiv:1409.1556
Soumagnac M. T., et al., 2015, MNRAS, 450, 666
Suchkov A. A., Hanisch R. J., Margon B., 2005, AJ, 130, 2439
Sutskever I., Martens J., Dahl G., Hinton G., 2013, in Proceedings of the 30th international conference on machine learning (ICML-13). pp 1139–1147
Swets J. A., Dawes R. M., Monahan J., 2000, Scientific American, p. 83
Szegedy C., et al., 2015, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp 1–9
Theano Development Team 2016, arXiv e-prints, abs/1605.02688
Weir N., Fayyad U. M., Djorgovski S., 1995, AJ, 109, 2401
Wolpert D. H., 1996, Neural computation, 8, 1341
York D. G., et al., 2000, AJ, 120, 1579
Zadrozny B., Elkan C., 2001, in ICML. pp 609–616
Zeiler M. D., Fergus R., 2014, in , Computer vision–ECCV 2014. Springer, pp 818–833

This paper has been typeset from a TEX/LATEX file prepared by the author.