

# Uncertain Photometric Redshifts with Deep Learning Methods

A. D’Isanto<sup>1</sup> †

<sup>1</sup>Heidelberg Institute for Theoretical Studies (HITS)  
Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg - GERMANY  
email: [antonio.disanto@h-its.org](mailto:antonio.disanto@h-its.org)

**Abstract.** The need for accurate photometric redshifts estimation is a topic that has fundamental importance in Astronomy, due to the necessity of efficiently obtaining redshift information without the need of spectroscopic analysis. We propose a method for determining accurate multimodal photo-z probability density functions (PDFs) using Mixture Density Networks (MDN) and Deep Convolutional Networks (DCN). A comparison with a Random Forest (RF) is performed.

**Keywords.** techniques: galaxies: distances and redshifts, photometric, methods: data analysis, surveys, (galaxies:) quasars: general etc.

---

## 1. Introduction

Determination of distances for astronomical objects through redshift acquired in the recent years an increasingly importance, having a fundamental role in cosmological research. In fact, it is well known that redshift is a fundamental step of the cosmic distance ladder. Redshift is traditionally obtained through spectroscopic analysis but due to long integration times and costly instrumentation requirements, it is not possible to measure it for all objects. Therefore, a convenient alternative is the estimation of photometric redshifts, e.g. based on measurements of pure photometry. However, the uncertainty of such an approach is much higher than the measurement errors obtained from spectroscopy. For this reason, the astronomical community has focused in the uncertainty quantification of redshift estimates through probability density functions (PDFs), instead of using simple point estimates. In this work we propose two neural network models based on Mixture Density Networks (MDN) (Bishop 1994). We use a deep MDN as first architecture, designed to use photometric features as inputs and to generate PDFs. The second architecture is a combination of a Deep Convolutional Network (DCN) (LeCun *et al.* 1998) with a MDN with the purpose to obtain photo-z PDFs based on images as input. We will show that this approach achieves better predictions due to its use of image data that - in contrast to using pre-defined features - allows to capture more details of the objects. We compare the results obtained with a commonly used tool in the related literature, the Random Forest (RF) (Breiman 2001).

## 2. Deep learning algorithms

In the next two subsections we give a description of the deep learning algorithms used for the experiments.

† The author gratefully acknowledges the support of the Klaus Tschira Foundation.

#	Type	Size	Maps	Activ
1	input	28x28	/	/
2	Conv	3x3	256	tanh
3	Pool	2x2	256	tanh
4	Conv	2x2	512	tanh
5	Pool	2x2	512	tanh
6	Conv	3x3	512	ReLU
7	Conv	2x2	1024	ReLU
8	MDN	500	/	tanh
9	MDN	100	/	tanh
10	output	15	/	Eq. 2.1

**Table 1.** DCMDN architecture

### 2.1. Mixture Density Network

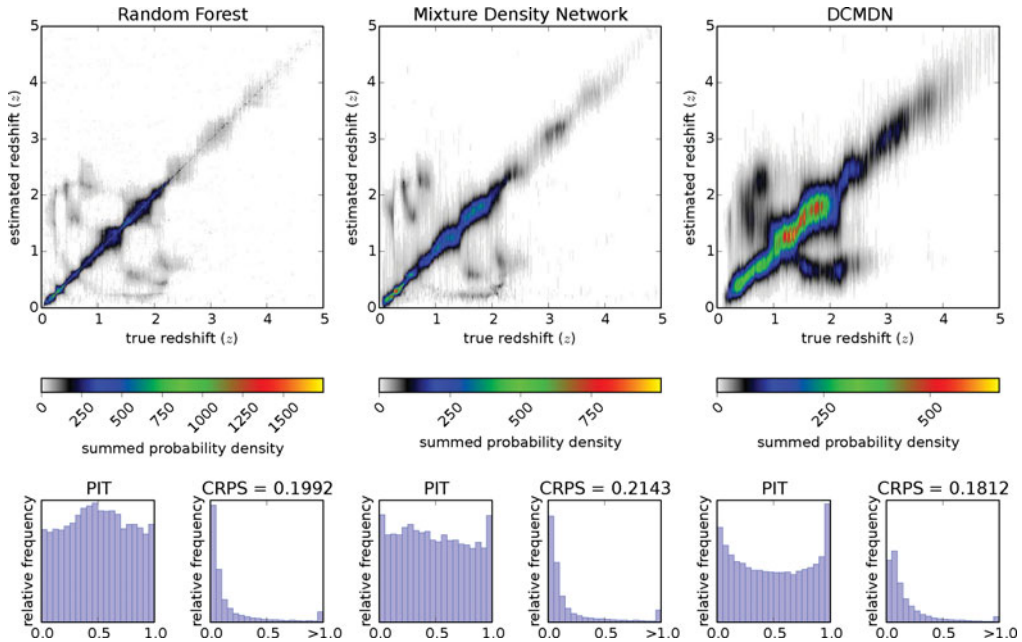
A Mixture Density Network (Bishop 1994) is the combination of a feed-forward neural network and a Gaussian mixture model. The outputs of the network parametrize the Gaussian mixture  $p(\theta|x) = \sum_{j=1}^n \omega_j \mathcal{N}(\mu_j, \sigma_j)$ , i.e. they define the means, variances, and weights. Thus the MDN produces a multi-modal PDF suitable for the case of photo-z, which is flexible enough to represent a multi-modal behavior. The means, variances and weights, are then obtained by the outputs  $z$  of the network:

$$\mu_j = z_j^\mu, \quad \sigma_j = \exp(z_j^\sigma), \quad \omega_j = \frac{\exp(z_j^\omega)}{\sum_{i=1}^n \exp(z_i^\omega)}. \quad (2.1)$$

Normally the MDN uses negative log-likelihood as a loss function, but in this work we use the *continuous rank probability score* (CRPS) (Gneiting *et al.* 2005) as loss function. This is to obtain a trained MDN which produces PDFs both well calibrated and sharp as measured by the CRPS, as explained in detail in Polsterer *et al.* (2016).

### 2.2. Deep Convolutional Network

A Deep Convolutional Network is a model in which several convolutional and sub-sampling layers are coupled with a fully-connected network. This architecture is particularly meant to learn from raw image data. In our case, we want to estimate redshifts directly from images, without the need to extract photometric features, so we couple a DCN with a MDN, in order to produce photo-z PDFs directly from SDSS images. We alternate convolutional and pooling layers to generate feature maps and generate a hierarchically compressed representation of the input data. The output of the convolutional network is then taken as input for the MDN which produces a multi-modal predictive density for photo-z. Thereby the extraction of the *feature maps* is automatically done by the network. Those obtained feature maps are then taken as inputs for the fully-connected part. We choose a modified version of the LeNet-5 architecture (LeCun *et al.* 1998), properly coupled with the presented MDN (see Section 2.1), obtaining what we call a Deep Convolutional Mixture Density Network (DCMDN). In Tab. 1 there is the architecture of the DCMDN used for the experiments, designed to run on GPUs, using a cluster equipped with Nvidia Titan X.



**Figure 1.** Results of the predictions obtained with the MDN and the DCMDN, compared with the RF results. For each experiment, three plots are given. The upper plots compare the spectroscopic redshift with the predictive distributions produced by the models, where the color indicates the summed probability density of the distributions. In the two lower plots, the histogram of the PIT values and the histogram of the individual CRPS values, are shown. The mean CRPS value is also given.

### 3. Experiments and Analysis

The data we use for the experiments are taken from the Sloan Digital Sky Survey Quasar Catalog V (Richards *et al.* 2010), based on the 7-th data release of the Sloan Digital Sky Survey (SDSS), consisting in 105,783 spectroscopically confirmed quasars, in a redshift range between 0.065 and 5.46. For the experiments we use a random subsample of 50,000 patterns. For each pattern we take the five *ugriz* magnitudes as input features and the respective images in the same bands. Finally, we compare the performances of MDN and DCMDN with the widely used RF.

The RF, in its original architecture, is not meant to produce PDFs. In order to obtain a distribution, we first collect the predictions  $z_{t,n}$  of each individual decision tree  $t$  in the forest, for every  $n$ -th data item. We take  $T = 256$  number of trees in the forest and define the PDF for the RF by fitting a mixture of 5 Gaussian components to the outputs,  $p(\theta|x) = \sum_{j=1}^5 \omega_j \mathcal{N}(\theta | (\mu_j, \sigma_j))$ , as we described also in Section 2.1 for the MDN.

For the RF and the MDN we use as input the 5 magnitudes plus all the possible color combinations, obtaining a 15-dimensional feature vector, respectively. The generated training and test set both contain 25,000 patterns. The DCMDN is trained on the images, that are obtained using the *Hierarchical Progressive Surveys* data partitioning format (Fernique *et al.* 2015) and performing a proper cutout on client side, in order to obtain the desired dimensions (28x28). Each pattern is originally a stack of 5 images in the *ugriz* filters, where every pixel is converted from flux units to *luptitudes* (Lupton *et al.* 1999). As done with the usual features, we additionally form the color images from the *ugriz* images by taking all possible pairwise differences, thus obtaining a stack of 15 images; every object/pattern is then represented by a tensor of dimensions 15x28x28.

In order to have a rotational invariant network, we perform data augmentation, taking rotations of each image at 0, 90, 180, 270 degrees. By doing so, we obtain a training set of 100,000 images, a validation set of 50,000 images and a test set of 50,000 images. Dropout is also applied to limit overfitting.

The results of the experiments are reported in Fig. 1. Following Polsterer *et al.* (2016), we use two statistical tools: the CRPS as a score function, and the *probability integral transform* (PIT) histogram (Gneiting *et al.* 2005), in order to obtain a visual estimation of the quality of the produced PDFs. In the RF experiment, the model reaches a CRPS of 0.20 and the PIT shows a bit of overdispersion. The performance of the MDN is a bit worse than the RF in terms of the CRPS, with a score of 0.21, but exhibits a better calibrated PIT. Using the DCMDN architecture we achieve the best results in terms of the CRPS, with a score of 0.19. The resulting PIT is acceptable, although it is still showing some underdispersion. The reason for the better overall performance of the DCMDN is that the features-based approach use only a fraction of the available information. In fact, in the process of features extraction a lot of information gets lost. Instead, using images, the DCMDN is able to automatically determine thousands of features, leading to a better prediction of the photo-z PDFs.

#### 4. Conclusions

Main purpose of this work is to show a method to produce photo-z PDFs using deep learning architectures. We generate very good probabilistic predictions based on features or images as input, producing a Gaussian mixture model as output. Our proposed architectures show better performances in the comparison carried out with a RF based method. In particular, we show that the proposed DCMDN gives the best performance, as it is able to use the entire information contained in the images. As showed by the PIT analysis, some optimization with respect to calibration can still be done, in order to deal with some dispersion phenomena. We firmly believe that the presented method needs little improvements to become a standard in predicting photo-z PDFs. As regression problems are very common in Astronomy, this approach can easily be applied to many other scientific topics.

#### References

- Christopher M. Bishop. Mixture density networks. Technical report, 1994.
- LeCun, Y., Bottou, L., bengio, Y., & Haffner P., Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278-2324, November 1998
- Leo Breiman. Random forests. *Mach. Learn.*, 45(1):5-32, October 2001.
- Gneiting, T., Raftery, A. E., Westveld, A. H., & Goldman, T. Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation. *Monthly Weather Review*, 133:1098, 2005
- Polsterer, K. L., D'Isanto A., & Gieseke, F. Uncertain photometric redshifts. 2016
- Richards, G. T., Hall, P. B., & Schneider, D. P., *et al.*, VizieR Online Data Catalog: The SDSS-DR7 quasar catalog (Schneider+, 2010). *VizieR Online Data Catalog*, 7260, May 2010
- Fernique, P., Allen, M. G., *et al.* Hierarchical progressive surveys. Multi-resolution HEALPix data structures for astronomical images, catalogues, and 3-dimensional data cubes. *A&A*, 578:A114, June 2015.
- Lupton, R. H., Gunn, J. E., & Szalay, A. S. A Modified Magnitude System that Produces Well-Behaved Magnitudes, Colors, and Errors Even for Low Signal-to-Noise Ratio Measurements. *AJ*, 118:1406-1410, September 1999.