EPJ Web of Conferences **265**, 00009 (2022) https://doi.org/10.1051/epjconf/202226500009
Multi-line Diagnostics of the Interstellar Medium

Invertible Neural Networks in Astrophysics

Ralf S. Klessen[1,2,*]

[1]Universität Heidelberg, Zentrum für Astronomie, Institut für Theoretische Astrophysik, Albert-Ueberle-Str. 2, 69120 Heidelberg, Germany
[2]Universität Heidelberg, Interdisziplinäres Zentrum für Wissenschaftliches Rechnen, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany

Abstract. Modern machine learning techniques have become indispensable in many fields of astronomy and astrophysics. Here we introduce a specific class of methods, invertible neural networks, and discuss two specific applications, the prediction of stellar parameters from photometric observations and the study of stellar feedback processes from on emission lines.

1 Introduction

Astronomy and astrophysics have always been highly data-intensive sciences. Large astronomical survey projects, conducted at modern Earth-bound or spaceborne observatories, or massively parallel astrophysical simulations, run at national and European supercomputing facilities, lie at the very forefront of the current "big data" deluge.

Whereas in the past, researchers were able to keep up with the amount of data pouring in by employing manual control and analysis methods, it became clear in recent years that this is no longer feasible, and that more automated and self-guided approaches are needed. Modern large-scale observational surveys or numerical simulation projects, such as those mentioned above, would not be possible without highly automated data-reduction pipelines that convert the raw data from the telescope or the supercomputer into a dimensionally-reduced and more science-ready form. Only this reduction enables an efficient analysis and astrophysical interpretation of the wealth of information available. Key concepts of artificial intelligence, driven by the ever increasing capabilities of modern machine learning techniques, are currently becoming a focal point of these developments. New neural network designs and supervised or unsupervised learning schemes allow for a comprehensive analysis of complex multi-scale astrophysical data with unprecedented accuracy and speed.

Here we introduce and discuss invertible neural networks (INNs) [1–3, 30], which have been successfully applied in the astronomical and astrophysical context to the analysis of star clusters [31], planetary systems [21], stellar feedback processes [25], or the analysis of galaxy mergers [14] in recent proof-of-concept studies.

2 Machine Learning in Astronomy and Astrophysics

Machine learning employs statistical models to predict the characteristics of a dataset using samples of previously collected data without relying on physical models of the system. The

*e-mail: klessen@uni-heidelberg.de

introduction of machine learning for solving regression, classification and clustering problems has revolutionized scientific research, and in particular has provided effective methods for analyzing big astronomical data [17, 24]. In order to construct a model from observed data, many methods rely on human-defined classifiers or 'feature extractors' [22]. However, complex problems require algorithms that automate feature extraction by learning from large amounts of data. Such self-learned feature extraction algorithms are an integral part of the deep learning family, which is based on the construction of artificial neural networks (NNs) [20]. While training NNs requires significant computational power, they achieve far higher levels of accuracy than classic machine learning for many non-linear problems.

There have been several recent studies that employ NN approaches to solve prediction tasks in astronomy and astrophysics. In the context of star cluster research, similar to the focus here, classical convolutional NNs have been employed to study stellar properties either from spectral [15, 41] or photometric [33, 44, 46] data, or they have been trained on data from the European astrometric Gaia satellite to predict properties of stellar clusters in the Milky Way [10, 28]. Classical NN methods have also been used for analyzing and classifying galaxy properties, either based on training data from large observational surveys [45] or from numerical simulations of cosmic structure formation [23, 47]. Other studies in this context have focused on determining the properties of the underlying dark-matter halos [12, 43] and on identifying merging galaxies or merger remnants from images [7, 8, 11, 19].

3 Invertible Neural Networks

A specific type of NN architecture based on the concept of normalizing flows [27] are invertible neural networks (INNs). They have been introduced to address complex and highly ambiguous inverse problems [1]. Unlike classical neural networks, which solve the inverse problem directly, INNs learn the forward process by using additional latent output variables to capture the information otherwise lost. Leveraging their invertible architecture, INNs then derive a solution for the inverse process without additional cost. Conditioned on the observations and the latent variable distribution, INNs can predict full posterior distributions, which is highly advantageous when studying multi-modal or degenerate problems, or when investigating complex correlations between parameters.

The advantage of invertible architectures is that the network automatically learns the inverse process when it is trained to approximate a known forward process. When considering degenerate problems or when taking uncertainties into account, an information loss is unavoidable in the forward process, such that different sets of physical parameters $\mathbf{x}$ are mapped onto identical observations $\mathbf{y}$. Consequently, the degenerate $\mathbf{y}$ cannot uniquely explain the corresponding $\mathbf{x}$. By introducing latent variables $\mathbf{z}$ that capture the information loss during the forward process, we can ensure a bijective mapping that could not be achieved with $\mathbf{x}$ and $\mathbf{y}$ alone. The original INN architecture links $\mathbf{x}$ and a unique pair of [$\mathbf{y}$, $\mathbf{z}$], making a bijective forward mapping $f(\mathbf{x}) = [\mathbf{y}, \mathbf{z}]$ and a inverse mapping $\mathbf{x} = f^{-1}(\mathbf{y}, \mathbf{z}) = g(\mathbf{y}, \mathbf{z})$. The forward process has to be deterministic and there are certain requirements towards the intrinsic dimensionalities of $\mathbf{x}$ and $\mathbf{y}$. Zero padding is necessary if the dimension of $\mathbf{x}$ is smaller than the dimension of [$\mathbf{y}$, $\mathbf{z}$].

The cINN architecture, as illustrated in Figure 1, avoids these problems [2, 3, 29, 30, 37]. It uses a different mapping system by considering the observations $\mathbf{y}$ in both the forward and inverse process as a condition $\mathbf{c}$: $f(\mathbf{x}; \mathbf{c} = \mathbf{y}) = \mathbf{z}$, $\mathbf{x} = g(\mathbf{z}; \mathbf{c} = \mathbf{y})$ [3]. This approach has the advantage that there are no assumptions or restrictions about the intrinsic dimensionalities of $\mathbf{x}$ and $\mathbf{y}$. It has the additional advantage that for very high-dimensional or complex datasets $\mathbf{y}$, we can include a feature extraction network in the conditioning block and fully integrate it in the training process [2, 3]. This allows to employ cINNs in image processing tasks [13, 32].

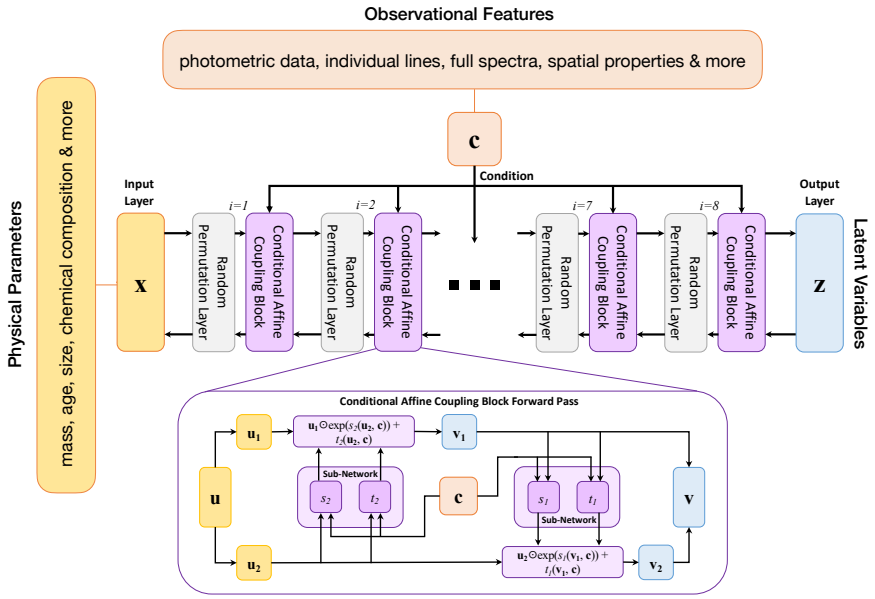

Figure 1. Schematic overview of the cINN architecture with physical input parameters **x** and observational features **y** as a condition **c**. The latent variables **z** capture the information loss during the (forward) training phase. The particular example depicted here consists of eight affine coupling blocks interchanged with permutation layers. It was developed as proof-of-concept for the analysis of diagnostic emission lines from young star clusters discussed by Kang and colleagues [25]. The zoom-in panel of the conditional affine coupling block shows how the information is passed through the block in the forward direction. Further details are provided by Ksoll and collaborators [31].

The posterior distribution of physical parameters, $p(\mathbf{x}|\mathbf{y})$, is estimated on the basis of the inverse mapping $f^{-1} = g$. During training, we prescribe the latent variables to have a standard normal probability distribution $p(\mathbf{z}) = N(\mathbf{z}, 0, \mathbf{I})$ with zero mean and unit standard deviation, where $\mathbf{I}$ is the identity matrix with a dimension of $\dim(\mathbf{z}) \times \dim(\mathbf{z})$. Following the inverse process $\mathbf{x} = g(\mathbf{z}; \mathbf{c})$, the posterior distribution is a transformation of the known distribution $p(\mathbf{z})$ to **x**-space, conditioned on the observation.

4 Example 1: INN for Stellar Parameters

In the first example we employ the cINN approach to the task of predicting physical parameters of individual stars based on photometric observations of spatially resolved clusters [31]. In this pilot study we train and test the neural network on synthetic data from the PARSEC stellar evolutionary models [9] and perform a benchmark analysis on real observational data obtained by the Hubble Space Telescope for the young cluster Westerlund 2 [40] and the old globular cluster NGC 6397 [36]. These clusters are chosen to cover the extremes of the cluster range, i.e. very young and very old, in order to gain first insights into the systematics of our approach. We construct the synthetic training sets by adopting isochrone model tables of the correct metallicity for Westerlund 2 and NGC 6397, respectively. The prediction of stellar mass, luminosity, effective temperature and surface gravity works extraordinarily well with posterior distributions that are narrowly constrained around the true values. Determining the

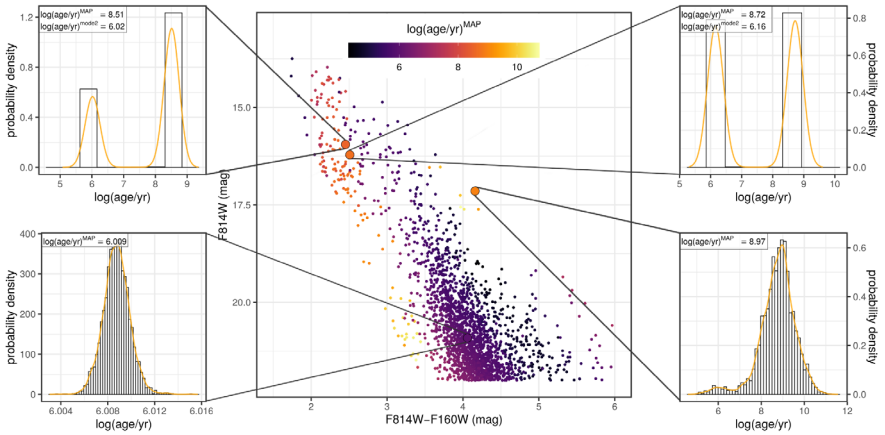

Figure 2. Illustration of the ability of the cINN to capture degeneracies in the physical model and properly cope with multi-modal posterior distribution functions, adopted from Ksoll and collaborators [31]. The middle panel shows a zoom-in of the optical color-magnitude diagram (CMD) of the young massive star cluster Westerlund 2 [40], with its stars color-coded according to the maximum a posteriori (MAP) estimates of log(age). The four smaller panels show the predicted age posterior distributions of highlighted stars. Note that stellar age is one of the most difficult stellar parameters to predict from photometric and spectroscopic observations. The bottom left panel is an example pre-main-sequence star for which our approach provides excellent results, returning a very narrow age distribution at the proposed cluster age. The remaining three cases are taken from stars likely on the turn-on of the main-sequence for which the MAP age estimate is significantly above the suggested age of Westerlund 2.

stellar age is a more difficult task, as illustrated in Figure 2. The predicted posteriors tend to be broader and often exhibit multi-modalities, revealing ample degeneracies in the age prediction. While we can confirm that the true value is part of the predicted distribution in more than 99% of the cases, there are several instances where it does not coincide with the most likely outcome of the posterior, falling into a second peak instead.

5 Example 2: INN for HII Region Diagnostics

Another application of the INN architecture is the study of physical properties of extragalactic star clusters and star-forming clouds from individual emission lines of HII regions. We present a cINN that predicts the posterior distribution of seven physical parameters (cloud mass, star formation efficiency, cloud density, cloud age as in the age of the first generation stars, age of the youngest cluster, the number of clusters, and the evolutionary phase of the cloud) from the luminosity of 12 optical emission lines, and test our network with synthetic models that are not used during training. The training database is constructed by using the WARPFIELD Emission Predictor [35], which allows us to collect both cloud properties and corresponding observable quantities (i.e. line luminosity). WARPFIELD-EMP describes the evolution of a cluster, expanding bubble, and the surrounding cloud using the 1D stellar feedback code WARPFIELD [38] and calculates detailed emission predictions based on the output from WARPFIELD with the help of CLOUDY [18] and the radiative transfer code POLARIS [39]. WARPFIELD takes into account several feedback mechanisms (i.e., stellar winds, radiation pressure, thermal gas pressure, supernovae, and gravity) self-consistently.

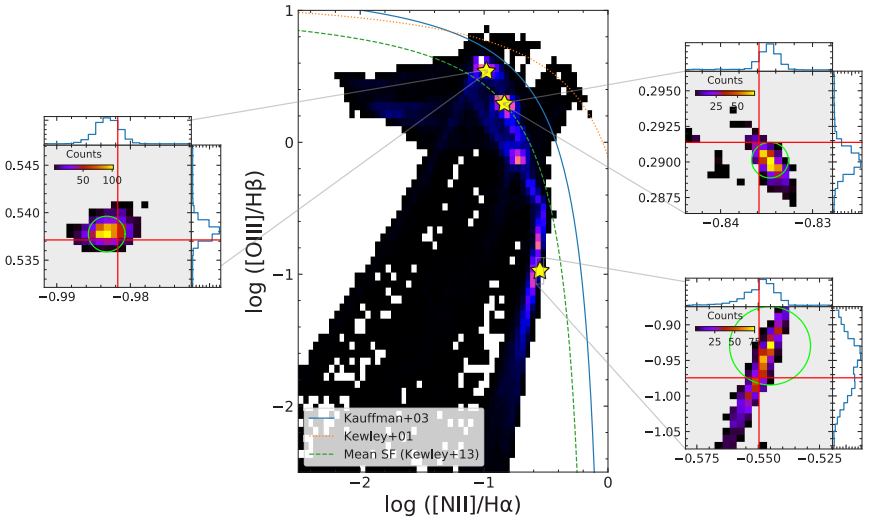

Figure 3. Two-dimensional histogram showing the ratios of diagnostic emission lines of singly ionized nitrogen, [NII], doubly ionized oxygen, [OIII], and atomic hydrogen, Hα and Hβ, for all of the models in the test set, where brighter color indicates a higher number of models. Overlaid as yellow stars are the corresponding values of three representative cases. Zoom-in panels show the distribution of the line ratios that we recover if we sample the posterior distribution for each example model 1024 times and use the resulting values as input for new WARPFIELD-EMP calculations. The true line ratio values for each example model are represented by red lines in these zoom-in panels. Green circles in each zoom-in panel indicate the area in which 68% of the models are included from the centre of the distribution. The image is adopted from Kang et al. [25].

6 Summary

The proof-of-concept studies mentioned here have successfully demonstrated the large potential and versatility of invertible neural networks for astronomical and astrophysical applications. Despite addressing a wide range of different scales and physical systems, our approach has in common that the ground-truth sample used to train the neural network is based on synthetic data, generated either from large-scale numerical simulations, Markov-chain Monte Carlo (MCMC) methods, or a database of physical models to generate the physical feature space **x**, combined with advanced post-processing and radiative transfer methods to produce the corresponding space of synthetic observables **y**. Training on synthetic data is needed in many astrophysical applications, because it is often not possible to build a ground-truth sample based on observational data only, and even in those cases for which sufficiently well understood observations exist, the numbers are usually too low to adequately train a neural network. Furthermore, training on synthetic data gives us a high degree of control over the problem and allows us to better understand the flow of information through the network, so that we can validate the network performance with high precision and accuracy. We can address the important question of measurement errors and internal degeneracies in the astrophysical system, i.e. in the mapping from **x** to **y**, and we can quantitatively assess how they influence the posterior distribution function. Once the INN is fully tested and characterized, its application to real astronomical observation allows us then also to assess the fidelity and accuracy of the underlying physical model that was used to train the network.

References

[1] L. Ardizzone, et al., arXiv 1808.04730 (2018)
[2] L. Ardizzone, et al., arXiv 1907.02392 (2019)
[3] L. Ardizzone, et al., arXiv 2105.02104 (2021)
[4] M. Bellagente, et al., SciPost Physics **9**, 074 (2020)
[5] S. Bieringer, et al., SciPost Physics **10**, 126 (2021)
[6] T. Bister, M. Erdmann, U. Köthe, J. Schulte, arXiv 2110.09493 (2021)
[7] C. Bottrell, et al., MNRAS **490**, 5390 (2019)
[8] C. Bottrell, et al., MNRAS **511**, 100 (2022)
[9] A. Bressan, et al., MNRAS **427**, 127 (2012)
[10] T. Cantat-Gaudin, et al., A&A **640**, A1 (2020)
[11] A. Ćiprijanović, et al., MNRAS **506**, 677 (2021)
[12] M.E. de los Rios, et al., arXiv 2111.08725 (2021)
[13] A. Denker, M. Schmidt, J. Leuschner, P. Maass, Journal of Imaging **7** (2021)
[14] L. Eisert, et al., arXiv 2202.06967 (2022)
[15] S. Fabbro, et al., MNRAS **475**, 2978 (2018)
[16] C. Federrath, et al., Nature Astronomy **5**, 365 (2021)
[17] E.D. Feigelson, G.J. Babu, Modern Statistical Methods for Astronomy (2012)
[18] G.J. Ferland, et al., Revista Mexicana de Astronomia y Astrofisica **53**, 385 (2017)
[19] L. Ferreira, et al., ApJ **895**, 115 (2020)
[20] I.J. Goodfellow, Y. Bengio, A. Courville, Deep Learning (2016)
[21] J. Haldemann, et al., arXiv 2202.00027 (2022)
[22] T. Hastie, et al., The elements of statistical learning (2009)
[23] M. Huertas-Company, et al., MNRAS **499**, 814 (2020)
[24] Z. Ivezic, et al., Statistics, Data Mining, and Machine Learning in Astronomy (2014)
[25] D.E. Kang, et al., MNRAS **512**, 617 (2022)
[26] D.P. Kingma, P. Dhariwal, Advances in Neural Information Proc. Sys., 10215 (2018)
[27] I. Kobyzev, et al., IEEE T. on Pattern Analysis and Machine Intelligence **43**, 3964 (2021)
[28] M. Kounkel, K. Covey, K.G. Stassun, AJ **160**, 279 (2020)
[29] J. Kruse, G. Detommaso, U. Köthe, R. Scheichl, arXiv 1905.10687 (2019)
[30] J. Kruse, et al., arXiv 2101.10763 (2021)
[31] V.F. Ksoll, et al., MNRAS **499**, 5447 (2020)
[32] J.H. Nölke, et al. arXiv 2011.05110 (2020)
[33] R. Olney, et al., AJ **159**, 182 (2020)
[34] A. Paszke, et al., Advances in Neural Information Processing Systems **32**, 8024 (2019)
[35] E.W. Pellegrini, et al., MNRAS **496**, 339 (2020)
[36] G. Piotto, et al., AJ **149**, 91 (2015)
[37] S.T. Radev, U.K. Mertens, A. Voss, L. Ardizzone, U. Köthe, arXiv 2003.06281 (2020)
[38] D. Rahner, E.W. Pellegrini, S.C.O. Glover, R.S. Klessen, MNRAS **483**, 2547 (2019)
[39] S. Reissl, R. Brauer, R.S. Klessen, E.W. Pellegrini, ApJ **885**, 15 (2019)
[40] E. Sabbi, et al., ApJ **891**, 182 (2020)
[41] K. Sharma, et al., MNRAS **491**, 2280 (2020)
[42] D. Trofimova, et al., arXiv 2012.08195 (2020)
[43] R. von Marttens, et al., arXiv 2111.01185 (2021)
[44] W. Wei, et al., MNRAS **493**, 3178 (2020)
[45] C. Wu, et al., MNRAS **482**, 1211 (2019)
[46] L. Yang, et al., arXiv 2112.07304 (2021)
[47] L. Zanisi, et al., MNRAS **501**, 4359 (2021)