Autoencoding Galaxy Spectra

Peter Melchior [1 2] **ChangHoon Hahn** [1] **Yan Liang** [1]

Abstract

We introduce a generative model for galaxy spectra based on an autoencoder architecture. Our encoder combines convolutional and attentive elements to identify important spectral features. The decoder is a fully-connected network, tasked with generating restframe spectra, which are then explicitly redshifted to the observed redshifts and resampled to match the spectral resolution and coverage of the instrument. The architecture thus reflects the astrophysical dependencies of a data-generating process that exhibits two fundamental degrees of freedom for each galaxy, namely its redshift and the characteristics of its restframe spectrum, and learns a compressed data-driven parameterization of the latter. We train this model on 100,000 optical spectra from SDSS, and find that it generates highly realistic galaxy spectra and excellent representations of the inputs. However, the desired redshift-independent encoding is possible only by augmenting the training spectra with artificially altered redshifts. Doing so establishes redshift invariance at the price of restricting the utilized spectral features to a consensus set that is accessible for any redshift covered by the training data, thereby limiting the information extracted from all spectra.

1. Introduction

Spectroscopy is critical to understand the physical processes that occur in galaxies across cosmic time. It is challenging because the data are noisy and riddled with artifacts, e.g. from telluric contamination, and because of the simultaneous presence of two types of variation: of the physical processes within galaxies and of their redshifts. The latter causes a stretching of the observed spectrum with respect to its restframe shape, which is determined by the former.

The construction of data-driven spectrum models has therefore been limited either to nearby galaxies, where the cosmological redshift can be ignored (Moustakas et al., 2006; Brown et al., 2014), or to a reduced wavelength range that is accessible for all galaxies after they have been artificially transformed to some consensus redshift (Yip et al., 2004; Portillo et al., 2020). Either approach reduces the number of usable galaxies or the wavelength range over which galaxies can be useful. The assumption that any given spectrum can be represented by a linear combination of a small number of basis vectors (Yip et al., 2004) or prototypical templates (Calzetti et al., 1994; Kinney et al., 1996) further curtails the complexity of spectrum models. Despite (or because of) its simplicity, linear models are widely used for inferring redshifts from galaxy spectra (Bolton et al., 2012; Ross et al., 2020) or broadband photometry (Benitez, 2000; Brammer et al., 2008), as well as for the generation of mock spectra from large cosmological simulations (Fagioli et al., 2018; Wechsler et al., 2021). On the other hand, theoretical models cannot reproduce high-quality observed spectra, not even when restricted to specific galaxy subpopulations (e.g. Tojeiro et al., 2011). Many physics-based approaches also treat separately the continuum from stellar emission and emission lines from nebular emission (Baldwin et al., 1981; Kewley et al., 2019). This practice creates a disconnect between a galaxy's stellar population and its gas content, which then affects the capabilities of subsequent analysis efforts (Cappellari, 2017; Leja et al., 2017). In summary, neither data-driven nor theoretical models currently capture the full information content of galaxy spectra.

However, the widespread adoption of template libraries suggests that the space of galaxy spectra is in fact low-dimensional. More explicitly, Portillo et al. (2020) demonstrated that a high-fidelity reconstruction of galaxy spectra can be achieved by an autoencoder architecture (Hinton & Zemel, 1994; Kingma & Welling, 2013) with a latent space of just six dimensions. We extend that work with a specifically designed architecture that combines an attentive convolutional encoder with an explicit redshift transformation after the decoder. It allows the exploitation of the full spectrum of all galaxies in a survey to form a super-model that exceeds the wavelength range and spectral resolution of any individual spectrum.

[*]Equal contribution [1]Department of Astrophysical Sciences, Princeton University, Princeton NJ 08544, USA [2]Center for Statistics and Machine Learning, Princeton University, Princeton, NJ 08544, USA. Correspondence to: Peter Melchior <peter.melchior@princeton.edu>.

ICML 2022 Workshop on Machine Learning for Astrophysics, Baltimore, Maryland, USA, 2022.

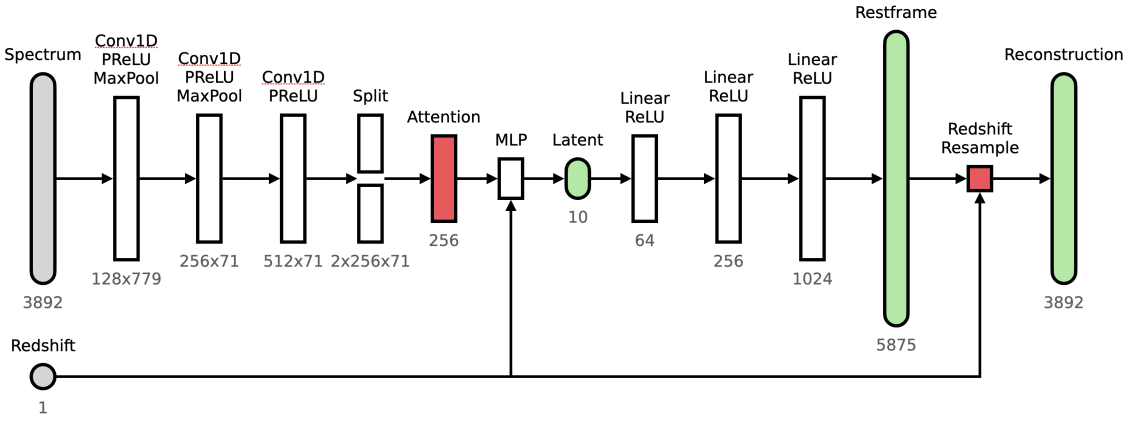

Figure 1. AE architecture with an attentive convolutional encoder and an explicit redshifting transformation after the decoder.

2. Method

We set up a standard autoencoder (AE), and use as loss a weighted MSE

$$L(\theta) = \sum w \odot (y - m_\theta(y))^2 / \sum w, \qquad (1)$$

where $y \in \mathbb{R}^{N \times L}$ denotes a batch of spectra, $m_\theta(y)$ the reconstructed spectra from the autoencoder model, $w_{n,l} = 1/\sigma^2_{n,l}$ the inverse variance weight of the l-th spectral element for the n-th galaxy, and $\odot$ the element-wise multiplication. This loss corresponds to a log-likelihood for normal-distributed data, which is appropriate for optical spectra.

The architecture is shown in Figure 1. We base the encoder on the work by Serrà et al. (2018) as a generic model for 1D data sets. The architecture starts with 3 convolutional layers with progressively wider kernel sizes (5, 11, 21) and max-pooling, which translates $L = 3892$ spectral elements into 512 channels for 71 wavelength segments. It then applies attention in wavelength direction to these channels, i.e. it splits the channels into two parts, h and k ($\in \mathbb{R}^{256\times71}$), and combines them as $e = h \cdot \mathrm{softmax}(k)$, where the dot product and the softmax operate on the last, i.e. the wavelength dimension. In essence, the encoder scans for the presence of features across the spectrum, and then reports their amplitude regardless of their location. This is important for galaxy spectra because increasing the redshift translates all features towards larger wavelengths. The final embedding of e to S latent variables is achieved by a small multi-layer perceptron that is conditioned on the redshift. The step is meant to compensate for the presence or absence of some spectral features e, depending on the redshift of the galaxy.

The decoder, a simple 3-layer MLP with (64, 256, 1024) dimensions and leaky ReLU activation functions, generates a representation of the spectrum at redshift $z = 0$, the so-called "restframe". This internal model is then analytically redshifted to the known galaxy redshift and resampled to the observed spectrum resolution. We perform the last step as a linear interpolation because the spectral resolution of SDSS is sufficient to resolve strong emission lines.

Applying the redshifting transformation explicitly in the generator part of the autoencoder removes the ambiguity between the location of spectral features in restframe and in the observed frame. It is thus much easier to train the critical aspect, namely the spectral encoding of restframe features, without also having to learn the analytically known effects of redshift. A similar approach was employed by (Lanusse et al., 2021) for galaxy images altered by a variable point-spread function.

To enable the redshifting transformation, the decoder's internal model must encompass the maximum range of restframe wavelengths accessible for SDSS spectra of galaxies up to $z_{\max} = 0.5$, i.e $\lambda_{\min} = 3784/(1 + z_{\max})$ Å to $\lambda_{\max} = 9270$ Å. The model has $R = L(1 + z_{\max}) = 5785$ linearly spaced spectral elements, which prevents that any observed spectrum is internally undersampled.

3. Data

We retrieve 100,000 galaxy spectra from SDSS Main Galaxy Sample (Strauss et al., 2002). SDSS MGS is a highly complete sample of galaxies out to moderate redshifts. From the full sample of $\sim$650,000 spectra, we use spectra classified as galaxies, with accurate redshifts, and with no quality flags. In particular, we impose standard quality flags and only use galaxies with low redshift errors $\sigma_z < 10^{-4}$. Each spectrum has $L = 3892$ spectral elements and covers the wavelength range $\lambda = 3784 \ldots 9270$ Å. We use calibrated spectra, inverse variance weights, and masks from the spectroscopic reduction of SDSS Data Release 16 (Ahumada et al., 2020). We normalize the spectra by dividing out the median flux. In masked areas, we set the weights to zero.

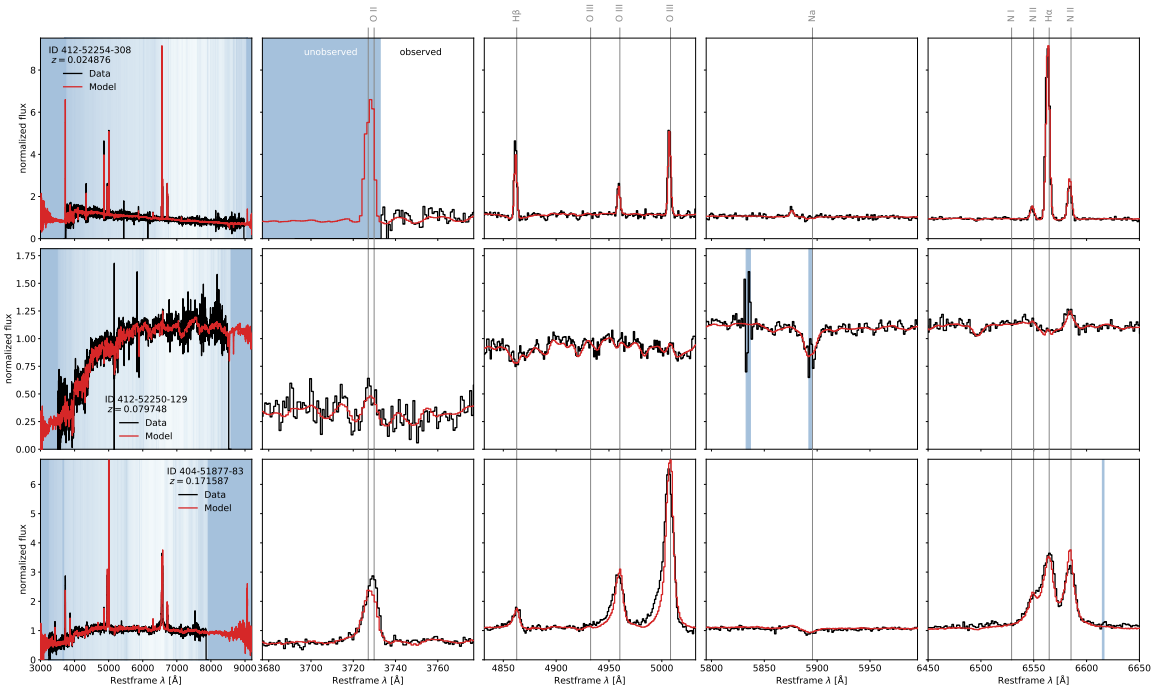

Figure 2. Example restframe spectra and their reconstruction by the AE-10 model, ordered by redshift (*top to bottom*). The left column shows the entire spectrum, with blue shading denoting the statistical weights (brighter is higher). Other columns zoom in to specific emission and absorption lines. Wavelengths that are unobserved or masked by the processing pipeline are shown in dark blue.

4. Results

We implement this architecture in `pytorch` (Paszke et al., 2017) and train it with 70% of the parent sample for 300 epochs with the Adam optimizer (Kingma & Ba, 2015) on a NVIDIA V100 GPU. Figure 2 shows the original and reconstructed spectra of the AE model with $S = 10$ latent variables for several galaxies from the withheld test sample, spanning $z = 0 \ldots 0.2$. The AE-10 model provides an excellent fit to spectra from different types of galaxies at different redshifts, with losses indicating a fit with residuals smaller than the noise level. More qualitatively, good fidelity is evident from the recovery of the highly variable behavior of the NII-Hα complex (last column) for a starforming (first row) to a quiescent galaxy (second row) or with substantial line broadening (last row).

We want to emphasize that the AE spectrum model exceeds the spectral resolution and observed wavelength range for every galaxy, inferring the missing parts from similar galaxies at different redshifts. The establishment of an extended restframe model is most obviously confirmed by the prediction of the expected but unobserved OII doublet (first row, second column).

This result seems to suggest that, by exploiting the known redshift transformation in the decoder path, the autoencoder indeed has established a redshift-independent extraction of spectral features to define its latents. However, closer inspection reveals that this is not the case. By artificially altering the redshift of observed spectra between $0 \leq z \leq 0.5$, and removing all parts of the spectra that are shifted out of the observable wavelength range (see an example in Figure 3), we find that while the reconstructions remain accurate, the corresponding positions in latent space are not stable. That complicates the interpretation of the latent space location as a unique descriptor of a galaxy's restframe spectrum. More concerning, spectra of identical galaxies observed at sufficiently different redshift do not end up at the same latent space location and thus fail to contribute to a robust training of the restframe model. In other words, when trained on a range of redshifts, the model is pieced together from many local solutions that combine a latent location and a redshift to yield a good fit of an observed spectrum. This behavior is enabled by a flexible decoder that can turn different latent locations into essentially indistinguishable spectra. These local solutions could ultimately be defined by a very small set of training spectra, defying the intention of constructing a common model from all galaxies.

A possible remedy is consistency regularization (Sinha & Dieng, 2021), where a penalty term is added to the loss that discourages different latent positions for training data that differ only by a desired invariance transformation. Another common practice in comparable situations is to train with

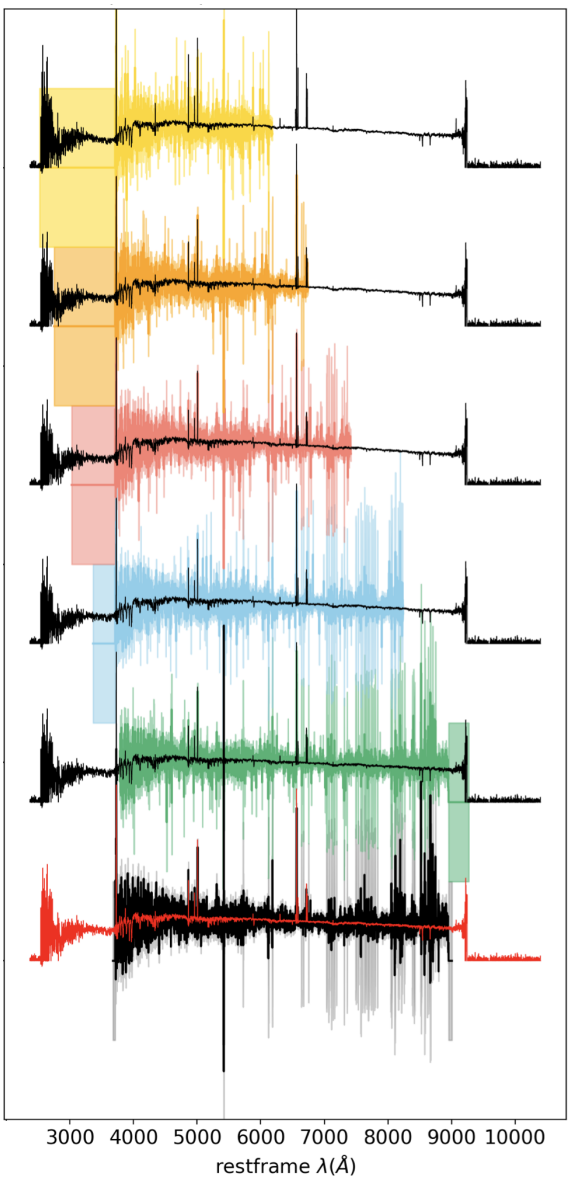

Figure 3. Artificially redshifted spectra (from bottom to top: $z = 0, 0.12, 0.25, 0.38, 0.5$, in color) produce visually good fits (black lines), but differ in latent space location, unless the training is augmented with such artificial redshift examples.

data augmentation (e.g. Mittal et al., 2020). We therefore repeat the training with artificially altered redshifts, and we indeed find redshift-invariant encoding positions from either consistency penalty or data augmentation. Both procedures, while effective, come at a price that is, to our knowledge, not broadly acknowledged. As is evident from Figure 3, to achieve robustness under artificial redshift changes, the encoder learns to only utilize spectral features that are available at all redshifts encountered during training. Doing otherwise, i.e. using one set of features at redshift z_1 and another one at z_2, would inevitably lead to different latent locations. With increasing redshifts these commonly available consensus features get increasingly relegated to the bluest side of the spectrum, utilizing ever fewer information from the observed spectra. If we were to span a redshift range up to $z \approx 1$ during training, we would not find any consensus features at all, forcing the autoencoder to trade off consistency against reconstruction fidelity.

Our findings are further supported by preliminary experiments with parameter inference from either full or encoded spectra, which show weakened constraining power after encoder compression, even though the autoencoded model has average errors below the noise level of the observation.

5. Conclusion

We present a novel architecture to encode spectra of galaxies that explicitly acknowledges the simultaneous variation of two parameters of the data generating process: the intrinsic spectrum of a galaxy and its redshift. We find good reconstruction fidelity for a wide range of galaxy types and redshifts. Our approach enables many applications that depend on generative models of galaxies, e.g. to generate realistic spectra, detect outliers, better infer properties such as redshift, and cross-calibrate different surveys.

On the other hand, we find evidence that, despite the redshift transformation being explicitly performed in the decoder path, the resulting latent space locations are not redshift-invariant. We attribute this finding to a preference for exploiting a broad range of spectral features, whose presence or absence at different redshift alters the encoding result. The lack of an agreement on how to encode spectra as a function of redshift complicates the interpretation of the latent space. For instance, it means that one cannot directly leverage information learned from spectra at one redshift to guide the reconstruction of spectra at another redshift, or more generally distinguish between the apparent effects of redshift and actual physical differences in the spectra of galaxies at different redshifts.

Modified training with either a consistency regularization or data augmentation substantially improves the redshift-invariant encoding, but is achieved by limiting the wavelength range over which spectral features are considered to a consensus that is accessible for all galaxies in the training data. This reduces the information that is effectively extracted from all galaxies, leading e.g. to weakened constraints from parameter inference based on encoded spectra. We welcome recommendations for establishing invariant encodings that leverage all available information even from samples that lack consensus in the observed data space.

References

Ahumada, R., Prieto, C. A., Almeida, A., Anders, F., Anderson, S. F., Andrews, B. H., Anguiano, B., Arcodia, R., Armengaud, E., Aubert, M., Avila, S., Avila-Reese, V., Badenes, C., Balland, C., Barger, K., Barrera-Ballesteros, J. K., Basu, S., Bautista, J., Beaton, R. L., Beers, T. C., Benavides, B. I. T., Bender, C. F., Bernardi, M., Bershady, M., Beutler, F., Bidin, C. M., Bird, J., Bizyaev, D., Blanc, G. A., Blanton, M. R., Boquien, M., Borissova, J., Bovy, J., Brandt, W. N., Brinkmann, J., Brownstein, J. R., Bundy, K., Bureau, M., Burgasser, A., Burtin, E., Cano-Díaz, M., Capasso, R., Cappellari, M., Carrera, R., Chabanier, S., Chaplin, W., Chapman, M., Cherinka, B., Chiappini, C., Doohyun Choi, P., Chojnowski, S. D., Chung, H., Clerc, N., Coffey, D., Comerford, J. M., Comparat, J., da Costa, L., Cousinou, M.-C., Covey, K., Crane, J. D., Cunha, K., Ilha, G. d. S., Dai, Y. S., Damsted, S. B., Darling, J., Davidson, James W., J., Davies, R., Dawson, K., De, N., de la Macorra, A., De Lee, N., Queiroz, A. B. d. A., Deconto Machado, A., de la Torre, S., Dell'Agli, F., du Mas des Bourboux, H., Diamond-Stanic, A. M., Dillon, S., Donor, J., Drory, N., Duckworth, C., Dwelly, T., Ebelke, G., Eftekharzadeh, S., Davis Eigenbrot, A., Elsworth, Y. P., Eracleous, M., Erfanianfar, G., Escoffier, S., Fan, X., Farr, E., Fernández-Trincado, J. G., Feuillet, D., Finoguenov, A., Fofie, P., Fraser-McKelvie, A., Frinchaboy, P. M., Fromenteau, S., Fu, H., Galbany, L., Garcia, R. A., García-Hernández, D. A., Oehmichen, L. A. G., Ge, J., Maia, M. A. G., Geisler, D., Gelfand, J., Goddy, J., Gonzalez-Perez, V., Grabowski, K., Green, P., Grier, C. J., Guo, H., Guy, J., Harding, P., Hasselquist, S., Hawken, A. J., Hayes, C. R., Hearty, F., Hekker, S., Hogg, D. W., Holtzman, J. A., Horta, D., Hou, J., Hsieh, B.-C., Huber, D., Hunt, J. A. S., Chitham, J. I., Imig, J., Jaber, M., Angel, C. E. J., Johnson, J. A., Jones, A. M., Jönsson, H., Jullo, E., Kim, Y., Kinemuchi, K., Kirkpatrick, Charles C., I., Kite, G. W., Klaene, M., Kneib, J.-P., Kollmeier, J. A., Kong, H., Kounkel, M., Krishnarao, D., Lacerna, I., Lan, T.-W., Lane, R. R., Law, D. R., Le Goff, J.-M., Leung, H. W., Lewis, H., Li, C., Lian, J., Lin, L., Long, D., Longa-Peña, P., Lundgren, B., Lyke, B. W., Ted Mackereth, J., MacLeod, C. L., Majewski, S. R., Manchado, A., Maraston, C., Martini, P., Masseron, T., Masters, K. L., Mathur, S., McDermid, R. M., Merloni, A., Merrifield, M., Mészáros, S., Miglio, A., Minniti, D., Minsley, R., Miyaji, T., Mohammad, F. G., Mosser, B., Mueller, E.-M., Muna, D., Muñoz-Gutiérrez, A., Myers, A. D., Nadathur, S., Nair, P., Nandra, K., do Nascimento, J. C., Nevin, R. J., Newman, J. A., Nidever, D. L., Nitschelm, C., Noterdaeme, P., O'Connell, J. E., Olmstead, M. D., Oravetz, D., Oravetz, A., Osorio, Y., Pace, Z. J., Padilla, N., Palanque-Delabrouille, N., Palicio, P. A., Pan, H.-A., Pan, K., Parker, J., Paviot, R., Peirani, S., Ramírez, K. P., Penny, S., Percival, W. J., Perez-Fournon, I., Pérez-Ràfols, I., Petitjean, P., Pieri, M. M., Pinsonneault, M., Poovelil, V. J., Povick, J. T., Prakash, A., Price-Whelan, A. M., Raddick, M. J., Raichoor, A., Ray, A., Rembold, S. B., Rezaie, M., Riffel, R. A., Riffel, R., Rix, H.-W., Robin, A. C., Roman-Lopes, A., Román-Zúñiga, C., Rose, B., Ross, A. J., Rossi, G., Rowlands, K., Rubin, K. H. R., Salvato, M., Sánchez, A. G., Sánchez-Menguiano, L., Sánchez-Gallego, J. R., Sayres, C., Schaefer, A., Schiavon, R. P., Schimoia, J. S., Schlafly, E., Schlegel, D., Schneider, D. P., Schultheis, M., Schwope, A., Seo, H.-J., Serenelli, A., Shafieloo, A., Shamsi, S. J., Shao, Z., Shen, S., Shetrone, M., Shirley, R., Aguirre, V. S., Simon, J. D., Skrutskie, M. F., Slosar, A., Smethurst, R., Sobeck, J., Sodi, B. C., Souto, D., Stark, D. V., Stassun, K. G., Steinmetz, M., Stello, D., Stermer, J., Storchi-Bergmann, T., Streblyanska, A., Stringfellow, G. S., Stutz, A., Suárez, G., Sun, J., Taghizadeh-Popp, M., Talbot, M. S., Tayar, J., Thakar, A. R., Theriault, R., Thomas, D., Thomas, Z. C., Tinker, J., Tojeiro, R., Toledo, H. H., Tremonti, C. A., Troup, N. W., Tuttle, S., Unda-Sanzana, E., Valentini, M., Vargas-González, J., Vargas-Magaña, M., Vázquez-Mata, J. A., Vivek, M., Wake, D., Wang, Y., Weaver, B. A., Weijmans, A.-M., Wild, V., Wilson, J. C., Wilson, R. F., Wolthuis, N., Wood-Vasey, W. M., Yan, R., Yang, M., Yèche, C., Zamora, O., Zarrouk, P., Zasowski, G., Zhang, K., Zhao, C., Zhao, G., Zheng, Z., Zheng, Z., Zhu, G., and Zou, H. The 16th Data Release of the Sloan Digital Sky Surveys: First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra. , 249(1):3, July 2020. doi: 10.3847/1538-4365/ab929e.

Baldwin, J. A., Phillips, M. M., and Terlevich, R. CLASSIFICATION PARAMETERS FOR THE EMISSION-LINE SPECTRA OF EXTRAGALACTIC OBJECTS. *Publications of the Astronomical Society of the Pacific*, 93(551):5, February 1981. ISSN 1538-3873. doi: 10.1086/130766. URL https://iopscience.iop.org/article/10.1086/130766/meta.

Benitez, N. Bayesian photometric redshift estimation. , 536:571–583, 2000. ISSN 0004-637X. URL http://iopscience.iop.org/article/10.1086/308947/meta.

Bolton, A. S., Schlegel, D. J., Aubourg, É., Bailey, S., Bhardwaj, V., Brownstein, J. R., Burles, S., Chen, Y.-M., Dawson, K., Eisenstein, D. J., Gunn, J. E., Knapp, G. R., Loomis, C. P., Lupton, R. H., Maraston, C., Muna, D., Myers, A. D., Olmstead, M. D., Padmanabhan, N., Pâris, I., Percival, W. J., Petitjean, P., Rockosi, C. M., Ross, N. P., Schneider, D. P., Shu, Y., Strauss, M. A., Thomas, D., Tremonti, C. A., Wake, D. A., Weaver, B. A., and Wood-Vasey, W. M. Spectral Classification and Redshift Measurement for the SDSS-III Baryon Oscillation Spec-

troscopic Survey. , 144(5):144, November 2012. doi: 10.1088/0004-6256/144/5/144.

Brammer, G. B., van Dokkum, P. G., and Coppi, P. EAZY: A fast, public photometric redshift code. , 686(2):1503, December 2008. ISSN 0004-637X. doi: 10.1086/591786. URL http://iopscience.iop.org/article/10.1086/591786/meta.

Brown, M. J. I., Moustakas, J., Smith, J.-D. T., da Cunha, E., Jarrett, T. H., Imanishi, M., Armus, L., Brandl, B. R., and Peek, J. E. G. An atlas of galaxy spectral energy distributions from the ultraviolet to the mid-infrared. , 212, June 2014. ISSN 0067-0049. doi: 10.1088/0067-0049/212/2/18. URL https://ui.adsabs.harvard.edu/#abs/2014ApJS..212...18B.

Calzetti, D., Kinney, A. L., and Storchi-Bergmann, T. Dust extinction of the stellar continua in starburst galaxies: The ultraviolet and optical extinction law. , 429:582, July 1994. ISSN 0004-637X. doi: 10.1086/174346. URL https://ui.adsabs.harvard.edu/abs/1994ApJ...429..582C.

Cappellari, M. Improving the full spectrum fitting method: accurate convolution with Gauss-Hermite functions. , 466 (1):798–811, April 2017. doi: 10.1093/mnras/stw3020.

Fagioli, M., Riebartsch, J., Nicola, A., Herbel, J., Amara, A., Refregier, A., Chang, C., Gamper, L., and Tortorelli, L. Forward modeling of spectroscopic galaxy surveys: application to SDSS. , 2018(11):015, November 2018. doi: 10.1088/1475-7516/2018/11/015.

Hinton, G. E. and Zemel, R. Autoencoders, minimum description length and helmholtz free energy. In Cowan, J., Tesauro, G., and Alspector, J. (eds.), *Advances in Neural Information Processing Systems*, volume 6, pp. 3–10. Morgan-Kaufmann, 1994. URL https://proceedings.neurips.cc/paper/1993/file/9e3cfc48eccf81a0d57663e129aef3cb-Paper.pdf.

Kewley, L. J., Nicholls, D. C., and Sutherland, R. S. Understanding galaxy evolution through emission lines. *Annual review of astronomy and astrophysics*, 57(1):511–570, August 2019. ISSN 0066-4146. doi: 10.1146/annurev-astro-081817-051832. URL https://doi.org/10.1146/annurev-astro-081817-051832.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. URL http://arxiv.org/abs/1412.6980.

Kingma, D. P. and Welling, M. Auto-Encoding variational bayes. December 2013. URL http://arxiv.org/abs/1312.6114v10.

Kinney, A. L., Calzetti, D., Bohlin, R. C., McQuade, K., Storchi-Bergmann, T., and Schmitt, H. R. Template ultraviolet to Near-Infrared spectra of star-forming galaxies and their application to K-Corrections. , 467:38, August 1996. ISSN 0004-637X. doi: 10.1086/177583. URL https://ui.adsabs.harvard.edu/abs/1996ApJ...467...38K.

Lanusse, F., Mandelbaum, R., Ravanbakhsh, S., Li, C.-L., Freeman, P., and Póczos, B. Deep generative models for galaxy image simulations. , 504(4):5543–5555, May 2021. ISSN 0035-8711. doi: 10.1093/mnras/stab1214. URL https://academic.oup.com/mnras/article-pdf/504/4/5543/38036124/stab1214.pdf.

Leja, J., Johnson, B. D., Conroy, C., van Dokkum, P. G., and Byler, N. Deriving Physical Properties from Broadband Photometry with Prospector: Description of the Model and a Demonstration of its Accuracy Using 129 Galaxies in the Local Universe. , 837(2):170, March 2017. doi: 10.3847/1538-4357/aa5ffe.

Mittal, A., Soorya, A., Nagrath, P., and Hemanth, D. J. Data augmentation based morphological classification of galaxies using deep convolutional neural network. *Earth Science Informatics*, 13(3):601–617, September 2020. ISSN 1865-0481. doi: 10.1007/s12145-019-00434-8. URL https://doi.org/10.1007/s12145-019-00434-8.

Moustakas, J., Kennicutt, Jr, R. C., and Tremonti, C. A. Optical star formation rate indicators. , 642(2):775, May 2006. ISSN 0004-637X. doi: 10.1086/500964. URL https://iopscience.iop.org/article/10.1086/500964/meta.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. Automatic differentiation in PyTorch. October 2017. URL https://openreview.net/pdf?id=BJJsrmfCZ.

Portillo, S. K. N., Parejko, J. K., Vergara, J. R., and Connolly, A. J. Dimensionality reduction of SDSS spectra with variational autoencoders. , 160:45, July 2020. ISSN 0004-6256. doi: 10.3847/1538-3881/ab9644. URL https://ui.adsabs.harvard.edu/abs/2020AJ....160...45P.

Ross, A. J., Bautista, J., Tojeiro, R., Alam, S., Bailey, S., Burtin, E., Comparat, J., Dawson, K. S., de Mattia, A., du Mas des Bourboux, H., Gil-Marín, H., Hou, J., Kong, H.,

Lyke, B. W., Mohammad, F. G., Moustakas, J., Mueller, E.-M., Myers, A. D., Percival, W. J., Raichoor, A., Rezaie, M., Seo, H.-J., Smith, A., Tinker, J. L., Zarrouk, P., Zhao, C., Zhao, G.-B., Bizyaev, D., Brinkmann, J., Brownstein, J. R., Rosell, A. C., Chabanier, S., Choi, P. D., Chuang, C.-H., Cruz-Gonzalez, I., de la Macorra, A., de la Torre, S., Escoffier, S., Fromenteau, S., Higley, A., Jullo, E., Kneib, J.-P., McLane, J. N., Muñoz-Gutiérrez, A., Neveux, R., Newman, J. A., Nitschelm, C., Palanque-Delabrouille, N., Paviot, R., Pullen, A. R., Rossi, G., Ruhlmann-Kleider, V., Schneider, D. P., Magaña, M. V., Vivek, M., and Zhang, Y. The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Large-scale structure catalogues for cosmological analysis. , 498(2):2354–2371, October 2020. doi: 10.1093/mnras/staa2416.

Serrà, J., Pascual, S., and Karatzoglou, A. Towards a universal neural network encoder for time series. May 2018. URL http://arxiv.org/abs/1805.03908.

Sinha, S. and Dieng, A. B. Consistency regularization for variational Auto-Encoders. May 2021. URL http://arxiv.org/abs/2105.14859.

Strauss, M. A., Weinberg, D. H., Lupton, R. H., Narayanan, V. K., Annis, J., Bernardi, M., Blanton, M., Burles, S., Connolly, A. J., Dalcanton, J., Doi, M., Eisenstein, D., Frieman, J. A., Fukugita, M., Gunn, J. E., Ivezić, Ž., Kent, S., Kim, R. S. J., Knapp, G. R., Kron, R. G., Munn, J. A., Newberg, H. J., Nichol, R. C., Okamura, S., Quinn, T. R., Richmond, M. W., Schlegel, D. J., Shimasaku, K., SubbaRao, M., Szalay, A. S., Vanden Berk, D., Vogeley, M. S., Yanny, B., Yasuda, N., York, D. G., and Zehavi, I. Spectroscopic Target Selection in the Sloan Digital Sky Survey: The Main Galaxy Sample. , 124(3):1810–1824, September 2002. doi: 10.1086/342343.

Tojeiro, R., Percival, W. J., Heavens, A. F., and Jimenez, R. The stellar evolution of luminous red galaxies, and its dependence on colour, redshift, luminosity and modelling. , 413(1):434–460, May 2011. doi: 10.1111/j.1365-2966.2010.18148.x.

Wechsler, R. H., DeRose, J., Busha, M. T., Becker, M. R., Rykoff, E., and Evrard, A. ADDGALS: Simulated Sky Catalogs for Wide Field Galaxy Surveys. *arXiv e-prints*, art. arXiv:2105.12105, May 2021.

Yip, C. W., Connolly, A. J., Szalay, A. S., Budavári, T., SubbaRao, M., Frieman, J. A., Nichol, R. C., Hopkins, A. M., York, D. G., Okamura, S., Brinkmann, J., Csabai, I., Thakar, A. R., Fukugita, M., and Ivezić, Ž. Distributions of galaxy spectral types in the sloan digital sky survey. *AJS; American journal of sociology*, 128(2):585, August 2004. ISSN 0002-9602, 1538-3881. doi: 10.1086/422429. URL https://iopscience.iop.org/article/10.1086/422429/meta.