DESA - a Multi-Modal approach for Stellar Astrophysics

Ilay Kamai [1] **Alex M. Bronstein** [1] **Hagai B. Perets** [1]

We introduce Dual Embedding for Stellar Astronomy (DESA) model, a novel multi-modal model designed for representational learning of stellar light curves and spectra. DESA operates in two main stages: first, it trains individual encoders for each modality using a hybrid approach that combines supervised and self-supervised learning; second, it integrates the individual embeddings through a unique module, DualFormer. DualFormer is motivated by the orthogonality of light curve and spectra and incorporates three key innovations: (1) a transformer-based module that merges both inter-modality and intra-modality information, (2) a specialized loss function that ensures alignment across modalities while preventing collapse, and (3) a linear projection layer that extracts meaningful information from the joint embedding space. This projection generates physically relevant features, useful for tasks like similarity search and anomaly detection. We evaluate DESA on several downstream tasks, including zero-shot classification of Color-Magnitude Diagram (CMD) classes, few-shot regression of stellar magnitude and color values, and fine-tuning for binary detection and stellar age prediction. In all cases, DESA outperforms state-of-the-art self-supervised and single-modality models, demonstrating its superior performance on astrophysical data. DESA marks a significant step forward in multi-modal, data-driven research in stellar astrophysics.

1. Introduction

Understanding the fundamental properties of stars is key to astrophysics, providing insights into stellar evolution, galactic structure, and planet formation. Traditionally, this has been done through analysis of stellar measurements, such as light curves (photometry) and spectra (spectroscopy). While spectra analysis predicts stellar parameters like $T_{eff}, logg, vsini$ and metallicity (FeH) (García Pérez et al., 2016; Wu et al., 2014), light curve analysis often uses spot modulation to detect periodicity and magnetic activity (Reinhold et al., 2013; McQuillan et al., 2014; Santos et al., 2019; Lu et al., 2020; Santos et al., 2021; Reinhold et al., 2023; Hattori et al., 2025). The rise of deep learning has further impacted stellar astrophysics, with models predicting or classifying stellar parameters from observations and simulations. For example, (Leung & Bovy, 2019), (Bai et al., 2020), (Olney et al., 2020), (Leung & Bovy, 2023), (Li & Lin, 2023), and (Koblischke & Bovy, 2024) used spectra-based deep learning models, and (Blancato et al., 2020), (Claytor et al., 2024),(Kamai & Perets, 2025), and (Claytor & Tayar, 2025) used light curve based models. Most of these approaches, however, rely on a single modality, referred to as unimodal models.

In contrast, multi-modality models combine data from different modalities of the same object, showing great success in NLP and vision, as seen in CLIP (Radford et al., 2021) and its variants. AstroCLIP (Parker et al., 2024) and Maven (Zhang et al., 2024) are examples of such models for astrophysical data. AstroCLIP combines galaxy images and spectra, and Maven combines light curves and spectra of supernova for classification and redshift estimation.

Here, we present DESA, the first multi-modal model for stars. While similar in some ways to previous approaches (e.g., using a two-step model), DESA offers a new perspective on multimodality in astronomical data. A diagram of DESA is shown in the upper panel of Figure 1.

2. Related Work

2.1. Contrastive Learning

Contrastive self-supervised methods use 'positive' and 'negative' pairs to create an embedding space where positive pairs are close and negative pairs are distant. These methods have been highly successful in the vision domain, with works like SimCLR (Chen et al., 2020). However, classical contrastive methods have drawbacks, such as the need for large batch sizes to adequately represent negative samples and the simplified assumption of 'positive' and 'negative' pairs, which can be problematic in domains with continuous transitions, like stars. Another issue is collapsing, where the model creates trivial features. Several approaches have addressed these challenges, including SimSiam (Chen & He, 2020), MoCo (He et al., 2019), and BYOL (Grill et al., 2020). Despite these issues, contrastive methods remain popular. For example, both (Parker et al., 2024) and (Zhang et al., 2024) used contrastive methods.

2.2. Regularized Methods

A different line of work focuses on feature-level discrimination rather than instance-level discrimination, as in contrastive methods. This idea is motivated by canonical cor-

relation analysis and was suggested as a self-supervised method by (Zhang et al., 2021) and (Zbontar et al., 2021). The latter was the motivation of the Variance-Invariance-Covariance Regularization (VicReg) architecture (Bardes et al., 2021). VicReg applies three losses to prevent collapse and maintain alignment: ensuring embedding variance is large, forcing the covariance between features to be the identity matrix, and minimizing the L_2 distance between embeddings. This method does not require negative pairs and has been shown to outperform contrastive methods without requiring large batch sizes.

3. Multi-Modal Neural Network for Stellar Astrophysics

3.1. Hybrid Training of Individual Modalities

We begin by training individual modalities separately, as in (Parker et al., 2024) and (Zhang et al., 2024), but with a hybrid framework. This framework adds a supervised head to the self-supervised model and trains with the following loss function:

$$\mathcal{L}_{hybrid} = (1-\lambda)\mathcal{L}_{ssl} + \lambda\mathcal{L}_{supervised}, \quad (1)$$

where λ controls the balance between self-supervision and supervision. This idea, used by (Walmsley et al., 2022) for unimodal galaxy models, ensures the embeddings are physically meaningful. in Appendix B, we show a training example and plot $\mathcal{L}_{ssl}$, $\mathcal{L}_{supervised}$, and $\mathcal{L}_{hybrid}$. It can be seen that the addition of $\mathcal{L}_{supervised}$ indeed contributes to the final loss. The spectra encoder uses a CNN followed by a Conformer module (Gulati et al., 2020), modified with Rotary Position Embedding (RoPE) (Su et al., 2021). The self-supervised approach for spectra is Masked-Filling, where 15% of the spectrum is masked, with 80% replaced by zero and 20% by a random value. The model uses a CNN decoder to reconstruct the spectrum and a Conformer-MLP branch to predict stellar parameters. $\mathcal{L}_{ssl}$ is the Mean Squared Error (MSE) between the masked and filled spectra, while $\mathcal{L}_{supervised}$ is Conformalized Quantile Regression (CQR) (Romano et al., 2019), which creates guaranteed confidence intervals. We also incorporated signal-to-noise ratio (SNR) as a weight for the final loss.

For the light curve encoder, we use a contrastive-hybrid method. Light curves are augmented into two views via cropping, and both the Autocorrelation Function (ACF) and Fast Fourier Transform (FFT) are added as channels. The views are processed by a CNN encoder and Conformer, with embeddings combined via SimSiam framework (Chen & He, 2020), and a 2-layer MLP that predicts the rotational period. Here, $\mathcal{L}_{ssl}$ is the cosine similarity loss from SimSiam, and $\mathcal{L}_{supervised}$ is again CQR.

3.2. DualFormer

The next step is to combine the embeddings from the pre-trained individual encoders. Here we are using a novel approach specifically tailored for light curve and spectra multi-modality. This is motivated by the observation that the information relationships between light curves and spectra of stars are very different from those found in NLP and vision modalities. While text describe its corresponding image, light curve and spectra show different relationships. They both partially describe the star in complementary ways. Intuitively, light curves and spectra can be seen as orthogonal views of the star. Of course, the measurements are not mathematically orthogonal, since the measurements come from different surveys, with potentially different bands and sensitivities, and can be taken at very different times, which makes the relationships more complicated. Moreover, in both text and images (as well as spectra and images like in AstroCLIP), the dynamics of the system are not manifested in the data. Contrary to that, light curves measure time-dependent phenomena by design. This creates a time-dependent information relationship in the case of light curve and spectra alignment. As such, the shared information between light curves and spectra is more complicated and may be degenerate. Another uniqueness of astronomical data is the importance of prior knowledge. In astrophysics, we usually have some extra information about objects. This can be, for example, stellar parameters that are known with good accuracy. As mentioned in 3.1, this information can be used to train individual encoders, but it can also be crucial during the alignment process, since this information is modality-invariant. These differences suggest that standard multi-modal approaches might not be sufficient in our scenario, and that a specific model is needed. The lower panel of Figure 1 shows a diagram of DualFormer. The inputs are the final features from the light curve and spectra encoders. They are first processed in a transformer-like module with a modified MHSA; instead of self-attention, we use both self-attention and cross-attention, where the former focuses on in-modality relationships, and the latter focuses on cross-modality relationships. Next, we aggregate the information using average pooling, add conditional prior information, and project both features through the same linear layer, A. This layer is the effective bottleneck of the network and should store the important shared information. Specifically, we use A for the projection of one feature branch and A^T for the projection of another branch. To align the features while preventing collapse, we are motivated by (Zhang et al., 2021) and (Bardes et al., 2021) but with some modifications. We use the same covariance loss that decorrelates features:

$$\mathcal{L}_{cov}(x_i, x_j) = \frac{1}{d}\sum_{k\neq l}[Cov(x_i, x_j)]^2_{kl}, \quad (2)$$

Model	T_{eff} *MAE (K)*	*logg MAE (dex)*	FeH *MAE (dex)*
DESA (ours)	**91.56**	0.168	**0.069**
StarGRUNet	93.77	**0.162**	0.070

Table 1. Result of spectra encoder. Ground truth labels are from APOGEE. See text for details.

But we use it both inside each branch and between branches; specifically, we chose to decorrelate the features after projection, p_1, p_2 in Figure 1, and the full covariance loss is:

$$\mathcal{L}_{cov} = \mathcal{L}_{cov}(p_1, p_1) + \mathcal{L}_{cov}(p_2, p_2) + \mathcal{L}_{cov}(p_1, p_2) \quad (3)$$

In addition, instead of a point-wise MSE loss between features, we use the following loss term:

$$\mathcal{L}_{duality} = MSE(\langle z_1, p_1 \rangle, \langle z_2, p_2 \rangle), \quad (4)$$

where z_1, z_2 are the features before the projection by A, p_1, p_2 are the projected features, and $\langle \cdot \rangle$ is the standard inner product. $\mathcal{L}_{daulity}$ can be seen as a less constrained version of the invariance term from (Bardes et al., 2021): writing $p_1 = Az_1$ and $p_2 = A^T z_2$ we see that $L_{duality}$ requires equality of the following quadratic forms:

$$z_1^T A z_1 = z_2^T A^T z_2 \quad (5)$$

We see that while the standard invariance term requires z_1 and z_2 to be identical vectors, $\mathcal{L}_{duality}$ does not even require them to lie on the same hyper-surface (since in general $A \neq A^T$), but does require the same amount of information to be extracted from the features. This gives much more freedom for z_1 and z_2 to be different, but constrains the projections, namely A, and A^T. Since A is not necessarily hermitian, we expect the meaningful information to be stored in a shared vector space of A and A^T. Ideally, this would be the eigenspace of A. Therefore, our final feature space is the projection of the feature vectors (z_1, z_2 in Figure 1) unto the eigenspace of A. This can be written as:

$$f = (z_1 + z_2)^T V, \quad (6)$$

where z_1, z_2 are the pre-projection feature vectors and V are the eigenvectors of A after training.

Although this motivation might sound appealing, it is not guaranteed to work well. To test the architectural choices of DualFormer, we conducted an ablation study and tested different attention mechanisms and different uses of A (with and without transpose). The ablation study results are shown in Appendix B. We see that the suggested architecture outperforms the alternatives.

4. Results

We train the full model using low-resolution spectra from LAMOST (Zhao et al., 2012; Wang et al., 2022) and light curves from Kepler (Mathur et al., 2017). The implementation details of our model and all baselines are detailed in Appendix A. First, we present the results of the hybrid pre-training of individual modalities. The spectra encoder performs similar to StarGRUNet (Li & Lin, 2023), with better MAE for T_{eff} and FeH, and competitive results for $logg$ as can be seen in Table 1. While the results are similar, the main improvement over previous works is the fact that we used the entire SNR range, while StarGRUNet and similar works ((Li et al., 2022) for example) used some range of SNR values. This is because we integrated the SNR information into the training by weighting the final loss according to the SNR. In Appendix B, we show the MAE as a function of SNR for APOGEE test set and LAMOST test set. We see that the SNR sensitivity of DESA is better than that of StarGRUNet (see figure 11 in their paper). For the light curve encoder, our model achieves an RMSE of 2.61 days, outperforming (Blancato et al., 2020), the only work that used only real light curves (and no simulations) for training, by a factor of 2.

Next, we evaluate fine-tuning on various tasks: zero-shot, few-shot, and full fine-tuning. For zero-shot Color Magnitude Diagram (CMD) classification, we used labels from (D. et al., 2025) and a Gaussian Mixture Model (GMM) on UMAP-reduced features. For few-shot, we used linear regression on a small subsample (20% of the test set) to

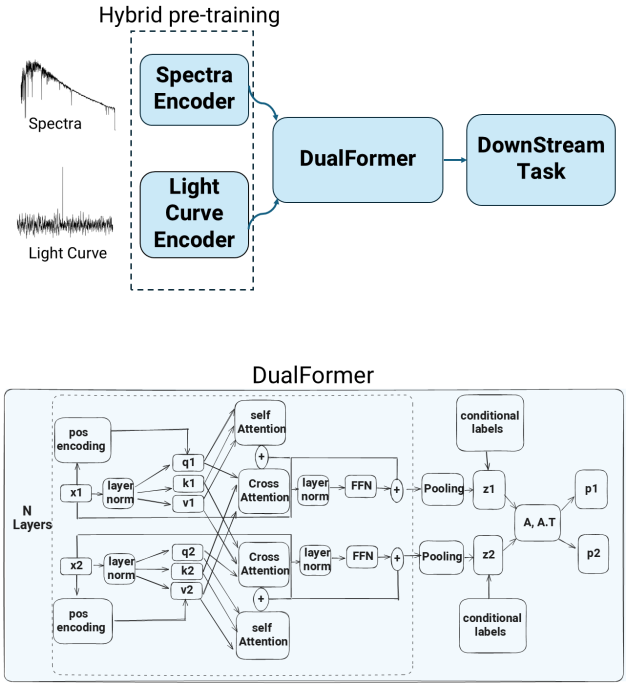

Figure 1. Left panel - High-level diagram of the entire model. Right panel - Detailed diagram of the DualFormer module.

Model	*GMM zero-shot accuracy*	*Linear Regression R^2*	*$BP-RP$ Accuracy*	*G_{mag} Accuracy*
DESA (ours)	**0.40**	**0.920**	**0.677**	**0.711**
MoCo	0.18	-0.001	0.208	0.159
MoCo Clean	0.11	-0.0004	0.202	0.160
SimSiam	0.15	-0.0003	0.202	0.158
VICReg	0.25	-0.0003	0.202	0.158

Table 2. Result of zero-shot CMD clustering (first column), and few-shot regression. See text for details.

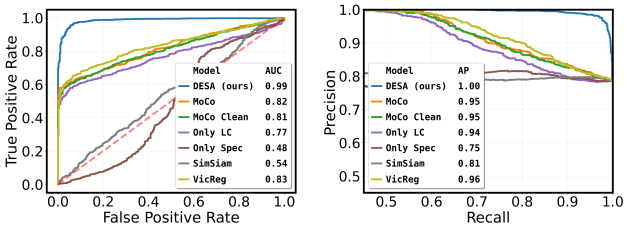

Figure 2. Experimental results of binary detection for different models. The left panel shows ROC curves. The right panel shows precision-recall curves.

Model	*Age MAE (Gyr)*	*Age RMSE (Gyr)*
DESA (ours)	**0.61**	**0.94**
MoCo	0.81	1.28
MoCo Clean	0.78	1.23
SimSiam	1.24	1.81
VICReg	0.78	1.23
Only Spectra	1.30	1.70
Only Light Curve	0.80	1.25

Table 3. Experimental result of stellar age prediction for different models.

predict de-reddened BP-RP and absolute G-Magnitude. In both cases, our model outperform alternative models with a very large margin (Table 2).

In Appendix B, we show the UMAP reduced features, colored by the CMD classes used for clustering. We see that our model shows a much more informative UMAP, with natural separation, while other models mix classes (Dwarfs and Subgiants, for example). In Appendix B we show the results of few-shot learning. The left panel shows the results of color and magnitude predictions (which are compared to baselines and summarized in Table 2) and the right panel shows predictions of T_eff and $log(\frac{L}{L_\odot})$. In both cases we see that the model not only learn the individual labels, but also the correct relationships. This suggests that we can recover stellar diagrams and even create new diagrams.

For fine-tuning tasks, we add a transformer prediction head and fine-tune on binary detection and stellar age estimation. Both tasks are challenging and require both photometric and spectroscopic information. For baselines, we used the same models as in the zero-shot and few-shot experiments, with the addition of the individual pretrained encoders, to test unimodal models. For binary detection, we use a curated sample from (D. et al., 2025), which consists of binaries and single stars. We achieved 96% accuracy, F1, and AUC, outperforming all alternatives as can be seen in Figure 2. We also outperformed the recent supervised work by (Jing et al., 2025), which trained a supervised model and reported AUC of 95%. It is worth mentioning that one of the advantages of our model is the fact that it learns the task of binary detection from different detection methods, while other models (like (Jing et al., 2025)) usually use a specific method. This is a result of the fact that it combines photometric and spectroscopic information and implies that it has the potential for better generalization.

For age prediction, a particularly challenging task, we train on gyrochronology ages from (Bouma et al., 2024) and (Lu et al., 2024) and achieve an RMSE lower than 1 Gyr, outperforming all other models (Table 3). Similar to the binary detection task, the fact that our model uses multi-modal information enables potential age estimation from a combination of different methods.

5. Conclusions

We presented DESA, a new multi-modality model for stellar astrophysics. DESA backbone consists of pre-trained uni-modality encoders that show state-of-the-art perforemence, and an alignment module, DualFormer, which is motivated by the observation that astrophysical data is unique and different compared to common multi-modality domains such as vision and NLP. We demonstrate the effectiveness of DESA in various ways, using zero-shot, few-shot, and fine-tuning experiments for challenging tasks like binary detection and stellar age inference that require information from both modalities. DESA consistently outperforms all baselines on all experiments with large margins, proving its superiority in the astronomical domain. These results demonstrate that DESA is not merely a predictive model, but a foundation model capable of extracting physically meaningful structure from heterogeneous data. We anticipate that DESA will serve as a powerful framework for future data-driven discovery in large stellar surveys, facilitating population studies, anomaly detection, and improved parameter estimation across the HR diagram.

References

Bai, Y., Liu, J., Wang, Y., and Wang, S. Machine-learning regression of extinction in the second gaia data release. *The Astronomical Journal*, 159(3):84, feb 2020. doi: 10.3847/1538-3881/ab63d5. URL https://dx.doi.org/10.3847/1538-3881/ab63d5.

Bardes, A., Ponce, J., and LeCun, Y. VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning. *arXiv e-prints*, art. arXiv:2105.04906, May 2021. doi: 10.48550/arXiv.2105.04906.

Blancato, K., Ness, M., Huber, D., Lu, Y., and Angus, R. Data-driven derivation of stellar properties from photometric time series data using convolutional neural networks. *arXiv e-prints*, art. arXiv:2005.09682, May 2020. doi: 10.48550/arXiv.2005.09682.

Bouma, L. G., Hillenbrand, L. A., Howard, A. W., Isaacson, H., Masuda, K., and Palumbo, E. K. Ages of Stars and Planets in the Kepler Field Younger than Four Billion Years. *apj*, 976(2):234, December 2024. doi: 10.3847/1538-4357/ad855f.

Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. A Simple Framework for Contrastive Learning of Visual Representations. *arXiv e-prints*, art. arXiv:2002.05709, February 2020. doi: 10.48550/arXiv.2002.05709.

Chen, X. and He, K. Exploring Simple Siamese Representation Learning. *arXiv e-prints*, art. arXiv:2011.10566, November 2020. doi: 10.48550/arXiv.2011.10566.

Claytor, Z. R. and Tayar, J. New Rotation Periods from the Kepler Bonus Background Light Curves. *arXiv e-prints*, art. arXiv:2506.03248, June 2025. doi: 10.48550/arXiv.2506.03248.

Claytor, Z. R., van Saders, J. L., Cao, L., Pinsonneault, M. H., Teske, J., and Beaton, R. L. Tess stellar rotation up to 80 days in the southern continuous viewing zone. *The Astrophysical Journal*, 962(1):47, feb 2024. doi: 10.3847/1538-4357/ad159a. URL https://dx.doi.org/10.3847/1538-4357/ad159a.

D., G., Mathur, S., García, R. A., Pinsonneault, M. H., Santos, Â. R. G., Beck, P. G., Grossmann, D. H., Schimak, L., Bedell, M., Merc, J., and Escorza, A. Kepler meets Gaia DR3: Homogeneous extinction-corrected color-magnitude diagram and binary classification. *aap*, 696: A243, April 2025. doi: 10.1051/0004-6361/202348735.

García Pérez, A. E., Allende Prieto, C., Holtzman, J. A., Shetrone, M., Mészáros, S., Bizyaev, D., Carrera, R., Cunha, K., García-Hernández, D. A., Johnson, J. A., Majewski, S. R., Nidever, D. L., Schiavon, R. P., Shane, N., Smith, V. V., Sobeck, J., Troup, N., Zamora, O., Weinberg, D. H., Bovy, J., Eisenstein, D. J., Feuillet, D., Frinchaboy, P. M., Hayden, M. R., Hearty, F. R., Nguyen, D. C., O'Connell, R. W., Pinsonneault, M. H., Wilson, J. C., and Zasowski, G. ASPCAP: The APOGEE Stellar Parameter and Chemical Abundances Pipeline. *aj*, 151 (6):144, June 2016. doi: 10.3847/0004-6256/151/6/144.

Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P. H., Buchatskaya, E., Doersch, C., Avila Pires, B., Guo, Z. D., Gheshlaghi Azar, M., Piot, B., Kavukcuoglu, K., Munos, R., and Valko, M. Bootstrap your own latent: A new approach to self-supervised Learning. *arXiv e-prints*, art. arXiv:2006.07733, June 2020. doi: 10.48550/arXiv.2006.07733.

Gulati, A., Qin, J., Chiu, C.-C., Parmar, N., Zhang, Y., Yu, J., Han, W., Wang, S., Zhang, Z., Wu, Y., and Pang, R. Conformer: Convolution-augmented Transformer for Speech Recognition. *arXiv e-prints*, art. arXiv:2005.08100, May 2020. doi: 10.48550/arXiv.2005.08100.

Hattori, S., Angus, R., Foreman-Mackey, D., Yuxi, Lu, and Colman, I. Measuring Long Stellar Rotation Periods (¿10 days) from TESS FFI Light Curves is Possible: An Investigation Using TESS and ZTF. *arXiv e-prints*, art. arXiv:2505.10376, May 2025. doi: 10.48550/arXiv.2505.10376.

He, K., Fan, H., Wu, Y., Xie, S., and Girshick, R. Momentum Contrast for Unsupervised Visual Representation Learning. *arXiv e-prints*, art. arXiv:1911.05722, November 2019. doi: 10.48550/arXiv.1911.05722.

Jing, Y., Mao, T.-X., Wang, J., Liu, C., and Chen, X. Half a Million Binary Stars Identified from the Low-resolution Spectra of LAMOST. *apjs*, 277(1):15, March 2025. doi: 10.3847/1538-4365/ada895.

Kamai, I. and Perets, H. B. Accurate and Robust Stellar Rotation Periods Catalog for 82771 Kepler Stars Using Deep Learning. *aj*, 169(2):59, February 2025. doi: 10.3847/1538-3881/ad99ab.

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., and Amodei, D. Scaling Laws for Neural Language Models. *arXiv e-prints*, art. arXiv:2001.08361, January 2020. doi: 10.48550/arXiv.2001.08361.

Koblischke, N. and Bovy, J. SpectraFM: Tuning into Stellar Foundation Models. *arXiv e-prints*, art. arXiv:2411.04750, November 2024. doi: 10.48550/arXiv.2411.04750.

Leung, H. W. and Bovy, J. Deep learning of multi-element abundances from high-resolution spectroscopic

data. *Monthly Notices of the Royal Astronomical Society*, 483(3):3255–3277, March 2019.

Leung, H. W. and Bovy, J. Towards an astronomical foundation model for stars with a transformer-based model. *Monthly Notices of the Royal Astronomical Society*, 527 (1):1494–1520, 10 2023. ISSN 0035-8711.

Li, X. and Lin, B. Estimating stellar parameters from LAMOST low-resolution spectra. *mnras*, 521(4):6354–6367, June 2023.

Li, X., Zeng, S., Wang, Z., Du, B., Kong, X., and Liao, C. Estimating atmospheric parameters from lamost low-resolution spectra with low snr. *Monthly Notices of the Royal Astronomical Society*, 514(3):4588–4600, 06 2022. ISSN 0035-8711.

Loshchilov, I. and Hutter, F. Decoupled Weight Decay Regularization. *arXiv e-prints*, art. arXiv:1711.05101, November 2017. doi: 10.48550/arXiv.1711.05101.

Lu, Y., Angus, R., Agüeros, M. A., Blancato, K., Ness, M., Rowland, D., Curtis, J. L., and Grunblatt, S. Astraea: Predicting long rotation periods with 27 day light curves. *The Astronomical Journal*, 160(4):168, sep 2020. doi: 10.3847/1538-3881/abada4. URL https://dx.doi.org/10.3847/1538-3881/abada4.

Lu, Y., Angus, R., Foreman-Mackey, D., and Hattori, S. In This Day and Age: An Empirical Gyrochronology Relation for Partially and Fully Convective Single Field Stars. *aj*, 167(4):159, April 2024. doi: 10.3847/1538-3881/ad28b9.

Mathur, S., Huber, D., Batalha, N. M., Ciardi, D. R., Bastien, F. A., Bieryla, A., Buchhave, L. A., Cochran, W. D., Endl, M., Esquerdo, G. A., Furlan, E., Howard, A., Howell, S. B., Isaacson, H., Latham, D. W., MacQueen, P. J., and Silva, D. R. Revised stellar properties of kepler targets for the q1-17 (dr25) transit detection run. *The Astrophysical Journal Supplement Series*, 229(2):30, mar 2017. doi: 10.3847/1538-4365/229/2/30. URL https://dx.doi.org/10.3847/1538-4365/229/2/30.

McQuillan, A., Mazeh, T., and Aigrain, S. Rotation Periods of 34,030 Kepler Main-sequence Stars: The Full Autocorrelation Sample. *apjs*, 211(2):24, April 2014. doi: 10.1088/0067-0049/211/2/24.

Olney, R., Kounkel, M., Schillinger, C., Scoggins, M. T., Yin, Y., Howard, E., Covey, K. R., Hutchinson, B., and Stassun, K. G. Apogee net: Improving the derived spectral parameters for young stars through deep learning. *The Astronomical Journal*, 159(4):182, apr 2020. doi: 10.3847/1538-3881/ab7a97. URL https://dx.doi.org/10.3847/1538-3881/ab7a97.

Pan, J.-S., Ting, Y.-S., Huang, Y., Yu, J., and Liu, J.-F. The Scaling Law in Stellar Light Curves. *arXiv e-prints*, art. arXiv:2405.17156, May 2024. doi: 10.48550/arXiv.2405.17156.

Parker, L., Lanusse, F., Golkar, S., Sarra, L., Cranmer, M., Bietti, A., Eickenberg, M., Krawezik, G., McCabe, M., Morel, R., Ohana, R., Pettee, M., Régaldo-Saint Blancard, B., Cho, K., Ho, S., and Polymathic AI Collaboration. AstroCLIP: a cross-modal foundation model for galaxies. *mnras*, 531(4):4990–5011, July 2024.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., and Sutskever, I. Learning Transferable Visual Models From Natural Language Supervision. *arXiv e-prints*, art. arXiv:2103.00020, February 2021. doi: 10.48550/arXiv.2103.00020.

Reinhold, T., Reiners, A., and Basri, G. Rotation and differential rotation of active Kepler stars. *aap*, 560:A4, December 2013. doi: 10.1051/0004-6361/201321970.

Reinhold, T., Shapiro, A. I., Solanki, S. K., and Basri, G. New rotation period measurements of 67 163 Kepler stars. *aap*, 678:A24, October 2023. doi: 10.1051/0004-6361/202346789.

Romano, Y., Patterson, E., and Candès, E. J. Conformalized Quantile Regression. *arXiv e-prints*, art. arXiv:1905.03222, May 2019. doi: 10.48550/arXiv.1905.03222.

Santos, A. R. G., García, R. A., Mathur, S., Bugnet, L., van Saders, J. L., Metcalfe, T. S., Simonian, G. V. A., and Pinsonneault, M. H. Surface Rotation and Photometric Activity for Kepler Targets. I. M and K Main-sequence Stars. *apjs*, 244(1):21, September 2019. doi: 10.3847/1538-4365/ab3b56.

Santos, A. R. G., Breton, S. N., Mathur, S., and García, R. A. Surface Rotation and Photometric Activity for Kepler Targets. II. G and F Main-sequence Stars and Cool Subgiant Stars. *apjs*, 255(1):17, July 2021. doi: 10.3847/1538-4365/ac033f.

Su, J., Lu, Y., Pan, S., Murtadha, A., Wen, B., and Liu, Y. RoFormer: Enhanced Transformer with Rotary Position Embedding. *arXiv e-prints*, art. arXiv:2104.09864, April 2021. doi: 10.48550/arXiv.2104.09864.

Walmsley, M., Slijepcevic, I., Bowles, M. R., and Scaife, A. Toward Galaxy Foundation Models with Hybrid Contrastive Learning. In *Machine Learning for Astrophysics*, pp. 29, July 2022. doi: 10.48550/arXiv.2206.11927.

Walmsley, M., Bowles, M., Scaife, A. M. M., Shingirai Makechemu, J., Gordon, A. J., Ferguson, A. M. N., Mann, R. G., Pearson, J., Popp, J. J., Bovy, J., Speagle, J., Dickinson, H., Fortson, L., Géron, T., Kruk, S., Lintott, C. J., Mantha, K., Mohan, D., O'Ryan, D., and Slijepevic, I. V. Scaling Laws for Galaxy Images. *arXiv e-prints*, art. arXiv:2404.02973, April 2024. doi: 10.48550/arXiv.2404.02973.

Wang, C., Huang, Y., Yuan, H., Zhang, H., Xiang, M., and Liu, X. The Value-added Catalog for LAMOST DR8 Low-resolution Spectra. *apjs*, 259(2):51, April 2022. doi: 10.3847/1538-4365/ac4df7.

Wu, Y., Du, B., Luo, A., Zhao, Y., and Yuan, H. Automatic stellar spectral parameterization pipeline for LAMOST survey. In Heavens, A., Starck, J.-L., and Krone-Martins, A. (eds.), *Statistical Challenges in 21st Century Cosmology*, volume 306 of *IAU Symposium*, pp. 340–342, May 2014. doi: 10.1017/S1743921314010825.

Zbontar, J., Jing, L., Misra, I., LeCun, Y., and Deny, S. Barlow Twins: Self-Supervised Learning via Redundancy Reduction. *arXiv e-prints*, art. arXiv:2103.03230, March 2021. doi: 10.48550/arXiv.2103.03230.

Zhang, G., Helfer, T., Gagliano, A. T., Mishra-Sharma, S., and Ashley Villar, V. Maven: a multimodal foundation model for supernova science. *Machine Learning: Science and Technology*, 5(4):045069, dec 2024. doi: 10.1088/2632-2153/ad990d. URL https://dx.doi.org/10.1088/2632-2153/ad990d.

Zhang, H., Wu, Q., Yan, J., Wipf, D., and Yu, P. S. From Canonical Correlation Analysis to Self-supervised Graph Neural Networks. *arXiv e-prints*, art. arXiv:2106.12484, June 2021. doi: 10.48550/arXiv.2106.12484.

Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., and Deng, L.-C. LAMOST spectral survey — An overview. *Research in Astronomy and Astrophysics*, 12(7):723–734, July 2012. doi: 10.1088/1674-4527/12/7/002.

A. Implementation Details

A.1. hyper-parameters

In big models, like DESA, hyperparameter tuning can be a very challenging task. This becomes even harder when the model consists of two steps - pre-training and alignment. We therefore chose to use a simple heuristic when defining the hyperparameters of our model. As the number of spectra samples is much larger than the number of light curve samples (6.5M vs 200K), we designed the individual encoders such that the spectra encoder has more parameters than the light curve encoder. This is motivated by neural scaling laws, a phenomenological relationship between dataset size, model parameters, and performance that was originally found in the vision and language domains (Kaplan et al., 2020), but recent works showed it also applies to astronomical data (Walmsley et al., 2024; Pan et al., 2024). Therefore, the dimension of the final spectra features was chosen to be 2048, and that of the light curve was chosen to be 256. The number of parameters in spectra and light curve encoders is about 500M and about 11M, respectively. During hybrid training, λ was chosen arbitrarily to be 0.5. The embedding dimension in dualformer was also chose to be 256, and the number of parameters in this module is also $\sim 11M$. The light curve encoder was trained using a learning rate decay scheduler with the cosine annealing method. The initial and final learning rate is $2 \cdot 10^{-5}$ decreasing to $2 \cdot 10^{-6}$. All other modules was trained with a constant learning rate of $2 \cdot 10^{-5}$. We trained all modules with AdamW optimizer (Loshchilov & Hutter, 2017). Lastly, we estimated the energy used to train the entire model using CodeCarbon[1] package. It is estimated to be 334 kWh for the entire model, out of which 204 kWh are for the pretraining stage.
All the code used for training and experiments is publicly available on https://github.com/IlayMalinyak/DESA.

A.2. Baseline Models

We compare our model with contrastive and regularized self-supervised methods that achieved state-of-the-art results in various tasks: VicReg, SimSiam, and MoCo. Each of the methods represents a different methodology - VicReg is a regularized method, SimSiam is a 'positive-only' contrastive method, and MoCo is a 'positive and negative' contrastive method. The use of positive and negative pairs in our scenario might be challenging because there are many samples with multiple spectra. This means that in a batch of samples, we might have off-diagonal positive pairs, which means that they would count as negative pairs. To overcome this, we created a version of MoCo with a special sampler that ensures the uniqueness of stars in each batch. We call this variant Moco-clean. To make sure that all models get the same information, we used the same pre-trained encoders (as specified in section 3.1) in all baselines. We also added the same conditional labels to all models and designed them to have at least the number of trainable parameters as in DESA, per task.

[1] https://codecarbon.io/

B. supplementary graphs

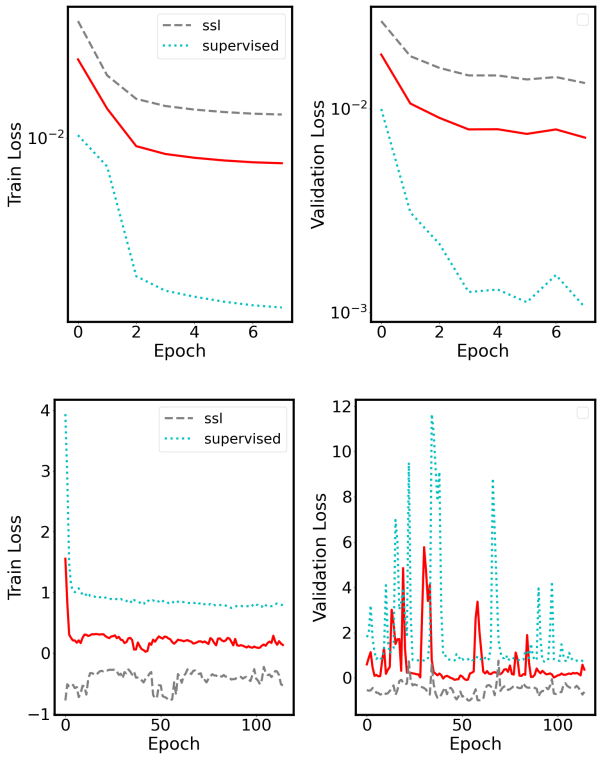

Figure 3. Training example of spectra encoder (upper panel) and light curve encoder (lower panel). The gray and cyan lines represents $\mathcal{L}_{ssl}$ and $\mathcal{L}_{supervised}$ respectively. The red line is the combined loss.

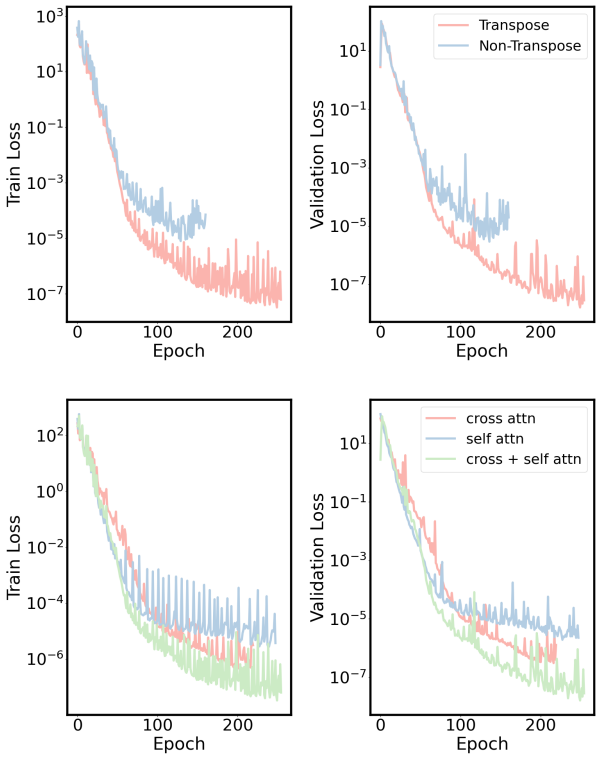

Figure 4. Results of ablation study of the dualFormer module. The upper panel shows a comparison between transposing and not transposing A. The lower panel compares different attention mechanisms.

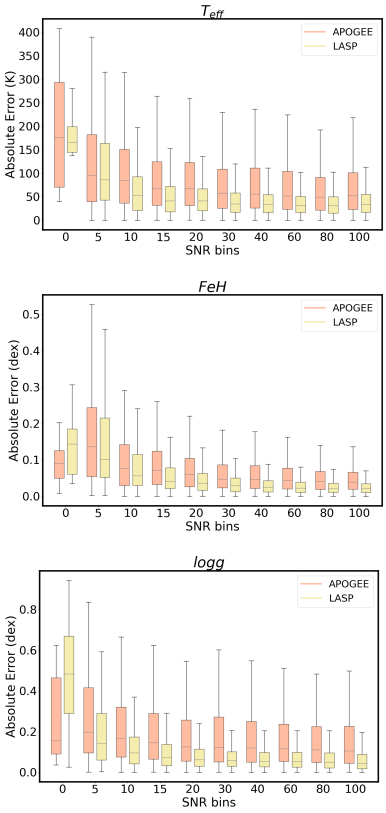

Figure 5. Box plots of MAE vs SNR for APOGEE test set and LAMOST test set.

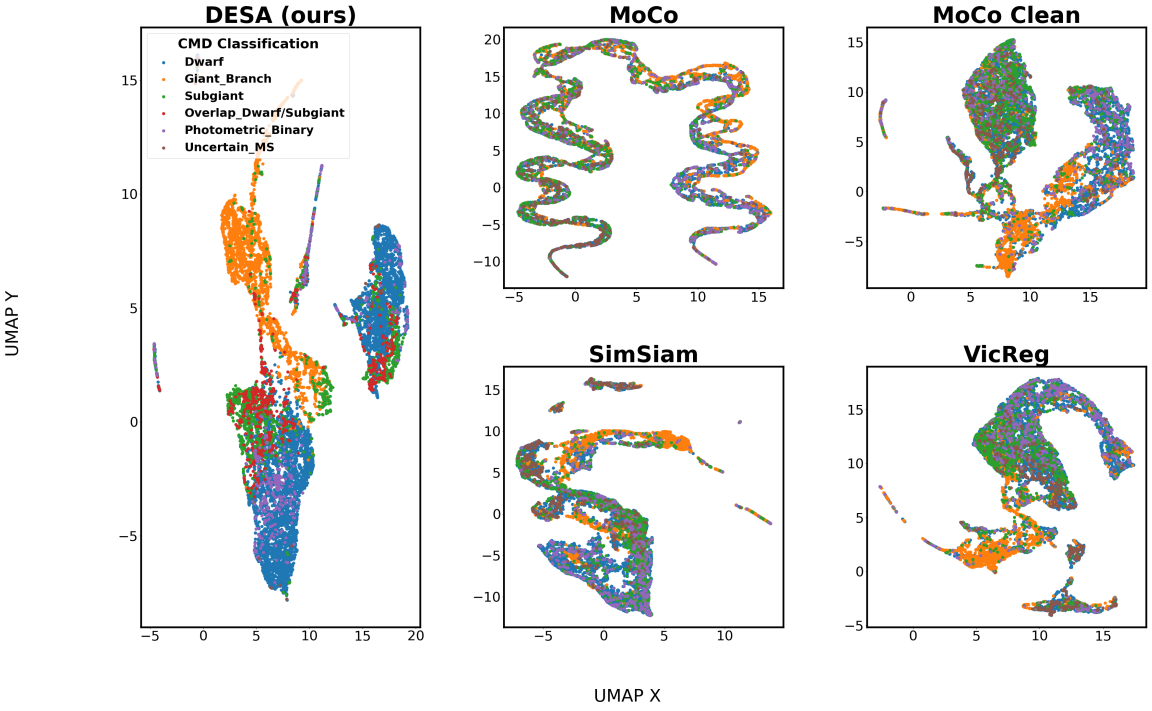

Figure 6. UMAP of the final features of DESA model and all baselines. Colors are Color Magnitude classes from (D. et al., 2025).

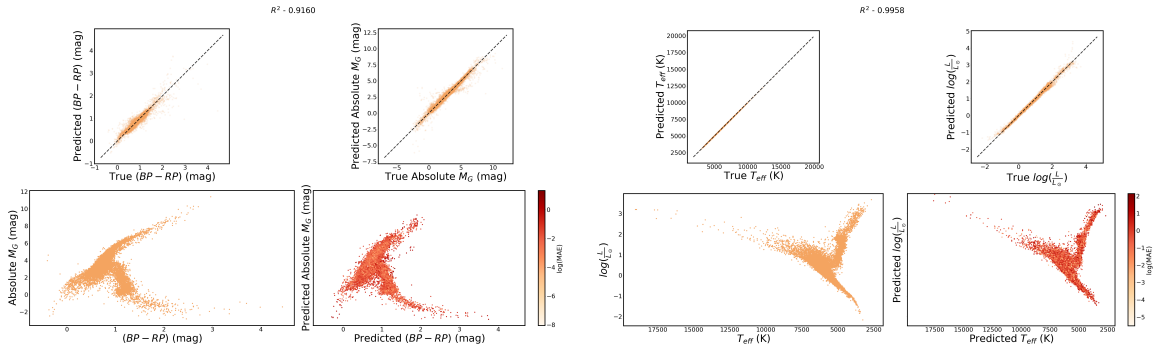

Figure 7. Few-shot results of DESA. Left - prediction of $BP-RP$ color and G band magnitude (upper panel) and a comparison between the true and predicted color magnitude diagram (lower panel). Right - the same for $T_e f f$ and $log(\frac{L}{L_\odot})$.