ASTROFORMER: MORE DATA MIGHT NOT BE ALL YOU NEED, LEARNING TO PREDICT GALAXY MORPHOLOGIES WITH LIMITED DATA

Anonymous authors
Paper under double-blind review

ABSTRACT

Galaxy morphologies are crucial to understanding the processes by which galaxies form and evolve. Efficient methods to classify galaxy morphologies are required to extract physical information from modern-day astronomy surveys. Recent advancements in areas such as natural language processing and computer vision rely on intricate and massive models that have been trained using vast amounts of unlabelled or partly labeled data and training or deploying these state-of-the-art methods to resource constraint environments has been a challenge. In this paper, we introduce methods to learn from less amounts of data. We propose using a hybrid transformer-convolutional architecture drawing much inspiration from the success of CoAtNet and MaxViT. Concretely, we use the transformer-convolutional hybrid with a different stack design for the network and pair it with a careful selection of data augmentation and regularization techniques. Our approach sets a new state-of-the-art on predicting Galaxy Morphologies from images on the Galaxy10 DECals dataset which consists of 17736 labeled images achieving 94.86% top-1 accuracy, beating the current state-of-the-art for this task by 4.62%. Furthermore, we also find that models and training methods used for larger datasets would often not work very well in the low data regime. Our code and models will be released at a later date before the conference.

1 INTRODUCTION

It has been well shown that to fully comprehend the formation and evolution of galaxies, it is essential to accurately classify galaxy morphologies and we refer readers which have been explored more in Section A.2.

In practice, collecting high-quality labeled data or human annotators is very expensive. Synthetic data though shows significant promise, but is often a distorted version of the real data and any modeling or inference performed on synthetic data comes with additional risks. Models trained on synthetic data often need to be fine-tuned with real data before being deployed (Tremblay et al., 2018; Jordon et al., 2022). Furthermore, methods such as transfer learning will not fully solve the problem either due to bias in pre-training datasets that do not reflect environments. We also explored semi-supervised learning approaches for this task which showed great preliminary results and we believe is most certainly an interesting direction with more exploration on this front we could have potentially achieved competitive results. We however focused our efforts on building models when obtaining unlabeled data is costly as well Thus, in this paper, we focus on training in a supervised setting from scratch.

Our contributions can be summarized as: **(1)** We propose a hybrid transformer-convolutional architecture comparable to the approach employed by CoAtNet (Dai et al., 2021) with a different stack design and pair it with a careful selection of augmentation and regularization techniques which is able to learn and generalize well [1] in the low data regime. **(2)** With our approach we establish a new state-of-the-art for the task of identifying galaxy morphologies from images on the Galaxy10 DECals dataset, and to the best of our knowledge, we believe this is the first work to use a hybrid

[1] By generalization, we mean that we measured the gap between the training loss and the evaluation accuracy that is how well the model can generalize to unseen data.

transformer-convolutional model to improve models in this domain. **(3)** We explore approaches and techniques for designing non-transfer learned models for the low-data regime in general which can be applied to tasks other than the one we explore.

2 METHODOLOGY

We develop a variant of the CoAtNet (Dai et al., 2021) model using a different stack design and careful selection of augmentation and regularization techniques. The Transformer block makes use of the relative attention which efficiently combines depthwise convolutions (Sandler et al., 2018) and self-attention (Vaswani et al., 2017). A depthwise convolution uses a fixed kernel to extract features from a local region of the input data whereas self-attention allows the receptive field to be the global spatial space. Relative attention allows us to combine convolutions and self-attention.

$$y_i = \sum_{j \in \mathcal{G}} \frac{\exp\left(x_i^\top x_j + w_{i-j}\right)}{\sum_{k \in \mathcal{G}} \exp\left(x_i^\top x_k + w_{i-k}\right)} x_j \tag{1}$$

where $x_i, y_i \in \mathbb{R}^d$ are the input and output at position i, w_{i-j} represents the depthwise convolution kernel and $\mathcal{G}$ represents the global spatial space. Here, the attention weight $A_{i,j}$ is decided by both w_{i-j} and $x_i^\top x_j$. The update made to the attention weight is rather intuitive by simply summing a global static convolution kernel

$$\begin{aligned} A_{i,j} &= \sum_{k \in \mathcal{G}} \exp\left(x_i^\top x_k\right) \qquad \text{(standard self-attention)} \\ A_{i,j} &= \sum_{k \in \mathcal{G}} \exp\left(x_i^\top x_k + \boldsymbol{w_{i-k}}\right) \qquad \text{(relative self-attention)} \end{aligned} \tag{2}$$

To construct a network that uses a relative attention, we adopt an approach similar to CoAtNets by first down-sampling the feature map via a multi-stage network with gradual pooling to reduce the spatial size and then employ the global relative attention. To do so, CoAtNets propose using a network of 5 stages (`S0`, `S1`, `S2`, `S3`, `S4`) where `S0` is a simple 2-layer convolutional Stem and `S1` employs MBConv blocks with squeeze-excitation. In the work by Dai et al. (2021) they eliminate the possibility of using a C-C-C-T stack i.e. `S1`, `S2`, and `S3` employ MBConv blocks and `S4` employs a Transformer block, due to supposedly low model performance. However, in our experiments we find that a C-C-C-T design works much better than C-C-T-T which was adapted as the layout for CoAtNet. This is precisely due to **(1)** higher generalization capability of C-C-C-T and **(2)** highly unstable training of C-C-T-T and C-T-T-T architectures. These proposed architectural choices are summarized in Figure 1.

In brief, the reasons this model performs so well on low-data-regime tasks, even when trained from scratch, and we believe it would work for other tasks in the low data regime are: **(1)** Careful selection of augmentation and regularization is very important especially in smaller datasets. **(2)** Our approach to train a hybrid transformer-convolutional model shows great generalizability and does not face the problem of highly unstable training and **(3)** The inherent translational equivalence helps make the training less prone to overfitting in the low data regime, more experiments on this front were done by Mohamed et al. (2020).

3 DISCUSSION

In this paper, we developed a supervised model to train from scratch on the Galaxy10 DECals dataset and explored other methods to train from scratch in the low data regime. We train a CoAtNet model and pair it with a careful selection of augmentation and regularization strategies as well as use a C-C-C-T design for the CoAtNet model which was earlier shown to have lower model performance than other stack designs. We apply this proposed model in the low-data regime and achieve state-of-the-art performance on the Galaxy10 DECals dataset. In the future, we hope to see more methods for training models with low data from scratch and we believe our model will enable more precise studies of galaxy morphology, and potentially a variety of other low-data-regime tasks.

REFERENCES

Kirsten N Blancato. *Decoding Starlight with Big Survey Data, Machine Learning, and Cosmological Simulations*. Columbia University, 2020.

Ronald J. Buta. Galaxy morphology, 2011. URL https://arxiv.org/abs/1102.0550.

Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey E Hinton. Big self-supervised models are strong semi-supervised learners. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 22243–22255. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/fcbc95ccdd551da181207c0c1400c655-Paper.pdf.

Yen Chen Chen. Classifying seyfert galaxies with deep learning. *The Astrophysical Journal Supplement Series*, 256(2):34, 2021.

MMClassification Contributors. Openmmlab's image classification toolbox and benchmark. https://github.com/open-mmlab/mmclassification, 2020.

Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 702–703, 2020.

Zihang Dai, Hanxiao Liu, Quoc V Le, and Mingxing Tan. Coatnet: Marrying convolution and attention for all data sizes. *Advances in Neural Information Processing Systems*, 34:3965–3977, 2021.

Ewa Deelman, James Blythe, Yolanda Gil, Carl Kesselman, Gaurang Mehta, Sonal Patil, Mei-Hui Su, Karan Vahi, and Miron Livny. Pegasus: Mapping scientific workflows onto the grid. In Marios D. Dikaiakos (ed.), *Grid Computing*, pp. 11–20, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg. ISBN 978-3-540-28642-4.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.

Terrance DeVries and Graham W. Taylor. Improved regularization of convolutional neural networks with cutout, 2017. URL https://arxiv.org/abs/1708.04552.

Arjun Dey, David J Schlegel, Dustin Lang, Robert Blum, Kaylan Burleigh, Xiaohui Fan, Joseph R Findlay, Doug Finkbeiner, David Herrera, Stéphanie Juneau, et al. Overview of the desi legacy imaging surveys. *The Astronomical Journal*, 157(5):168, 2019.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=YicbFdNTTy.

Stéphane d'Ascoli, Hugo Touvron, Matthew L Leavitt, Ari S Morcos, Giulio Biroli, and Levent Sagun. Convit: Improving vision transformers with soft convolutional inductive biases. In *International Conference on Machine Learning*, pp. 2286–2296. PMLR, 2021.

B. Flaugher, H. T. Diehl, K. Honscheid, T. M. C. Abbott, O. Alvarez, R. Angstadt, J. T. Annis, M. Antonik, O. Ballester, L. Beaufore, G. M. Bernstein, R. A. Bernstein, B. Bigelow, M. Bonati, D. Boprie, D. Brooks, E. J. Buckley-Geer, J. Campa, L. Cardiel-Sas, F. J. Castander, J. Castilla, H. Cease, J. M. Cela-Ruiz, S. Chappa, E. Chi, C. Cooper, L. N. da Costa, E. Dede, G. Derylo, D. L. DePoy, J. de Vicente, P. Doel, A. Drlica-Wagner, J. Eiting, A. E. Elliott, J. Emes, J. Estrada, A. Fausti Neto, D. A. Finley, R. Flores, J. Frieman, D. Gerdes, M. D. Gladders, B. Gregory, G. R. Gutierrez, J. Hao, S. E. Holland, S. Holm, D. Huffman, C. Jackson, D. J. James, M. Jonas, A. Karcher, I. Karliner, S. Kent, R. Kessler, M. Kozlovsky, R. G. Kron, D. Kubik, K. Kuehn, S. Kuhlmann, K. Kuk, O. Lahav, A. Lathrop, J. Lee, M. E. Levi, P. Lewis, T. S. Li, I. Mandrichenko, J. L. Marshall, G. Martinez, K. W. Merritt, R. Miquel, F. Muñ oz, E. H. Neilsen,

R. C. Nichol, B. Nord, R. Ogando, J. Olsen, N. Palaio, K. Patton, J. Peoples, A. A. Plazas, J. Rauch, K. Reil, J.-P. Rheault, N. A. Roe, H. Rogers, A. Roodman, E. Sanchez, V. Scarpine, R. H. Schindler, R. Schmidt, R. Schmitt, M. Schubnell, K. Schultz, P. Schurter, L. Scott, S. Serrano, T. M. Shaw, R. C. Smith, M. Soares-Santos, A. Stefanik, W. Stuermer, E. Suchyta, A. Sypniewski, G. Tarle, J. Thaler, R. Tighe, C. Tran, D. Tucker, A. R. Walker, G. Wang, M. Watson, C. Weaverdyck, W. Wester, R. Woods, and B. Yanny and. THE DARK ENERGY CAMERA. *The Astronomical Journal*, 150(5):150, oct 2015. doi: 10.1088/0004-6256/150/5/150. URL https://doi.org/10.1088%2F0004-6256%2F150%2F5%2F150.

Hanan Gani, Muzammal Naseer, and Mohammad Yaqub. How to train vision transformer on small-scale datasets?, 2022. URL https://arxiv.org/abs/2210.07240.

Premanand Ghadekar, Kunal Chanda, Sakshi Manmode, Sanika Rawate, Shivam Chaudhary, and Resham Suryawanshi. Galaxy classification using deep learning. In Sridaran Rajagopal, Parvez Faruki, and Kalpesh Popat (eds.), *Advancements in Smart Computing and Information Security*, pp. 3–13, Cham, 2022. Springer Nature Switzerland. ISBN 978-3-031-23092-9.

Golnaz Ghiasi, Tsung-Yi Lin, and Quoc V Le. Dropblock: A regularization method for convolutional networks. *Advances in neural information processing systems*, 31, 2018.

Yasuhiro Hashimoto, Augustus Oemler Jr, Huan Lin, and Douglas L Tucker. The influence of environment on the star formation rates of galaxies. *The Astrophysical Journal*, 499(2):589, 1998.

Ali Hassani, Steven Walton, Nikhil Shah, Abulikemu Abuduweili, Jiachen Li, and Humphrey Shi. Escaping the big data paradigm with compact transformers, 2021. URL https://arxiv.org/abs/2104.05704.

JP HOLANDA and RLS SANTOS. The influence of greyscale in galaxies classification using convolutional neural networks. *Journal of Production and Automation*, 2022.

Wuyu Hui, Zheng Robert Jia, Hansheng Li, and Zijian Wang. Galaxy morphology classification with densenet. *Journal of Physics: Conference Series*, 2402(1):012009, dec 2022. doi: 10.1088/1742-6596/2402/1/012009. URL https://dx.doi.org/10.1088/1742-6596/2402/1/012009.

Forrest Iandola, Matt Moskewicz, Sergey Karayev, Ross Girshick, Trevor Darrell, and Kurt Keutzer. Densenet: Implementing efficient convnet descriptor pyramids. *arXiv preprint arXiv:1404.1869*, 2014.

James Jordon, Lukasz Szpruch, Florimond Houssiau, Mirko Bottarelli, Giovanni Cherubin, Carsten Maple, Samuel N. Cohen, and Adrian Weller. Synthetic data – what, why and how?, 2022. URL https://arxiv.org/abs/2205.03257.

Seung Hoon Lee, Seunghyun Lee, and Byung Cheol Song. Vision transformer for small-size datasets, 2021. URL https://arxiv.org/abs/2112.13492.

Henry W. Leung and Jo Bovy. Deep learning of multi-element abundances from high-resolution spectroscopic data. *MNRAS*, 483(3):3255–3277, March 2019. doi: 10.1093/mnras/sty3217.

Chris Lintott, Kevin Schawinski, Steven Bamford, Anå¾e Slosar, Kate Land, Daniel Thomas, Edd Edmondson, Karen Masters, Robert C. Nichol, M. Jordan Raddick, Alex Szalay, Dan Andreescu, Phil Murray, and Jan Vandenberg. Galaxy Zoo 1: data release of morphological classifications for nearly 900 000 galaxies. *MNRAS*, 410(1):166–178, January 2011. doi: 10.1111/j.1365-2966.2010.17432.x.

Chris J. Lintott, Kevin Schawinski, Anže Slosar, Kate Land, Steven Bamford, Daniel Thomas, M. Jordan Raddick, Robert C. Nichol, Alex Szalay, Dan Andreescu, Phil Murray, and Jan Vandenberg. Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *MNRAS*, 389(3):1179–1189, September 2008. doi: 10.1111/j.1365-2966.2008.13689.x.

Liyuan Liu, Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao, and Jiawei Han. On the variance of the adaptive learning rate and beyond. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=rkgz2aEKDr.

Yahui Liu, Enver Sangineto, Wei Bi, Nicu Sebe, Bruno Lepri, and Marco Nadai. Efficient training of visual transformers with small datasets. *Advances in Neural Information Processing Systems*, 34:23818–23830, 2021.

Jennifer M Lotz, Piero Madau, Mauro Giavalisco, Joel Primack, and Henry C Ferguson. The rest-frame far-ultraviolet morphologies of star-forming galaxies at z˜ 1.5 and 4. *The Astrophysical Journal*, 636(2):592, 2006.

Kaitlin Maile, Dennis G. Wilson, and Patrick Forré. Architectural optimization over subgroups for equivariant neural networks, 2022. URL https://arxiv.org/abs/2210.05484.

Sachin Mehta and Mohammad Rastegari. Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=vh-0sUt8HlG.

Mirgahney Mohamed, Gabriele Cesa, Taco S. Cohen, and Max Welling. A data and compute efficient design for limited-resources deep learning, 2020. URL https://arxiv.org/abs/2004.09691.

O. H. Parry, V. R. Eke, and C. S. Frenk. Galaxy morphology in the ΛCDM cosmology. *Monthly Notices of the Royal Astronomical Society*, 396(4):1972–1984, 07 2009. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2009.14921.x. URL https://doi.org/10.1111/j.1365-2966.2009.14921.x.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

Hieu Pham, Zihang Dai, Qizhe Xie, and Quoc V. Le. Meta pseudo labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11557–11568, June 2021.

Priyanka S. Radhamani, Mhd Saeed Sharif, and Wael Elmedany. An effective galaxy classification using fractal analysis and neural network. In *2022 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, pp. 19–24, 2022. doi: 10.1109/3ICT56508.2022.9990776.

Vicente Rodriguez-Gomez, Laura V Sales, Shy Genel, Annalisa Pillepich, Jolanta Zjupa, Dylan Nelson, Brendan Griffen, Paul Torrey, Gregory F Snyder, Mark Vogelsberger, et al. The role of mergers and halo spin in shaping galaxy morphology. *Monthly Notices of the Royal Astronomical Society*, 467(3):3083–3098, 2017.

Allan Sandage. Star formation rates, galaxy morphology, and the hubble sequence. *Astronomy and Astrophysics*, 161:89–101, 1986.

Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520, 2018.

Chen Sun, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta. Revisiting unreasonable effectiveness of data in deep learning era. In *Proceedings of the IEEE international conference on computer vision*, pp. 843–852, 2017.

Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pp. 6105–6114. PMLR, 2019.

Mingxing Tan and Quoc Le. Efficientnetv2: Smaller models and faster training. In *International conference on machine learning*, pp. 10096–10106. PMLR, 2021.

Hugo Touvron, Matthieu Cord, Alaaeldin El-Nouby, Piotr Bojanowski, Armand Joulin, Gabriel Synnaeve, and Hervé Jégou. Augmenting convolutional networks with attention-based aggregation, 2021. URL https://arxiv.org/abs/2112.13692.

Jonathan Tremblay, Aayush Prakash, David Acuna, Mark Brophy, Varun Jampani, Cem Anil, Thang To, Eric Cameracci, Shaad Boochoon, and Stan Birchfield. Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

KA Venn, Sebastien Fabbro, Adrian Liu, Yashar Hezaveh, L Perreault-Levasseur, Gwendolyn Eadie, Sara Ellison, Joanna Woo, JJ Kavelaars, KM Yi, et al. Lrp2020: Machine learning advantages in canadian astrophysics. *arXiv preprint arXiv:1910.00774*, 2019.

Devesh Walawalkar, Zhiqiang Shen, Zechun Liu, and Marios Savvides. Attentive cutmix: An enhanced data augmentation approach for deep learning based image classification, 2020. URL https://arxiv.org/abs/2003.13048.

Mike Walmsley, Chris Lintott, Tobias Géron, Sandor Kruk, Coleman Krawczyk, Kyle W. Willett, Steven Bamford, Lee S. Kelvin, Lucy Fortson, Yarin Gal, William Keel, Karen L. Masters, Vihang Mehta, Brooke D. Simmons, Rebecca Smethurst, Lewis Smith, Elisabeth M. Baeten, and Christine Macmillan. Galaxy Zoo DECaLS: Detailed visual morphology measurements from volunteers and deep learning for 314 000 galaxies. *MNRAS*, 509(3):3966–3988, January 2022. doi: 10.1093/mnras/stab2093.

Ross Wightman. Pytorch image models. https://github.com/rwightman/pytorch-image-models, 2019.

Qizhe Xie, Minh-Thang Luong, Eduard Hovy, and Quoc V Le. Self-training with noisy student improves imagenet classification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10687–10698, 2020.

Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6023–6032, 2019.

Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017.

Michael Zhang, James Lucas, Jimmy Ba, and Geoffrey E Hinton. Lookahead optimizer: k steps forward, 1 step back. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/file/90fd4f88f588ae64038134f1eeaa023f-Paper.pdf.

Hu Zou, Xu Zhou, Xiaohui Fan, Tianmeng Zhang, Zhimin Zhou, Xiyan Peng, Jundan Nie, Linhua Jiang, Ian McGreer, Zheng Cai, Guangwen Chen, Xinkai Chen, Arjun Dey, Dongwei Fan, Joseph R. Findlay, Jinghua Gao, Yizhou Gu, Yucheng Guo, Boliang He, Zhaoji Jiang, Junjie Jin, Xu Kong, Dustin Lang, Fengjie Lei, Michael Lesser, Feng Li, Zefeng Li, Zesen Lin, Jun Ma, Moe Maxwell, Xiaolei Meng, Adam D. Myers, Yuanhang Ning, David Schlegel, Yali Shao, Dongdong Shi, Fengwu Sun, Jiali Wang, Shu Wang, Yonghao Wang, Peng Wei, Hong Wu, Jin Wu, Xiaohan Wu, Jinyi Yang, Qian Yang, Qirong Yuan, and Minghao Yue. The third data release of the beijing–arizona sky survey. *The Astrophysical Journal Supplement Series*, 245(1):4, oct 2019. doi: 10.3847/1538-4365/ab48e8. URL https://doi.org/10.3847%2F1538-4365%2Fab48e8.

Aleksandra Ćiprijanović, Ashia Lewis, Kevin Pedro, Sandeep Madireddy, Brian Nord, Gabriel N. Perdue, and Stefan M. Wild. Semi-supervised domain adaptation for cross-survey galaxy morphology classification and anomaly detection, 2022. URL https://arxiv.org/abs/2211.00677.

A APPENDIX

A.1 URM STATEMENT

The authors acknowledge that at least one key author of this work meets the URM criteria of ICLR 2023 Tiny Papers Track.

A.2 GALAXY MORPHOLOGIES

A galaxy's morphology describes its physical appearance, and encodes important information about the physical processes that have shaped its growth (Parry et al., 2009). Galaxy morphologies are influenced heavily by the physical properties of galaxies, like star formation (Hashimoto et al., 1998; Sandage, 1986; Lotz et al., 2006) and galaxy mergers (Rodriguez-Gomez et al., 2017) among others (Buta, 2011). Thus understanding galaxy morphologies leads to a better understanding on these fronts as well. One such instance is where scientists have used the disc and spheroid structures that make up a galaxy to figure out crucial components of its formation (Deelman et al., 2004). As such, a galaxy's morphology is a culmination of internal physical processes (e.g., star formation, stellar dynamics such as active galactic nuclei and dark matter, feedback), as well as its environment and interactions with other galaxies.

A.3 HYBRID MODELS

Recently, many hybrid transformer-convolutional models have gained a lot of popularity in vision tasks especially with the success of models like MobileViT (Mehta & Rastegari, 2022), ResNet-ViT (Dosovitskiy et al., 2021), ConViT (d'Ascoli et al., 2021), PatchConvNet (Touvron et al., 2021), and CoAtNet (Dai et al., 2021) have shown. On one hand the spatial inductive biases of convolutional neural networks allow them to learn representations with fewer parameters across different vision tasks as well as enable sample-efficient learning however they often have a potentially lower performance. On the other hand, transformers learn global representations and have shown a lot of success in vision tasks, and have outperformed convolutional neural nets for image classification. Though transformers tend to have larger model capacities, their generalization can be worse than convolutional neural networks. However, transformers are very data hungry (Hassani et al., 2021) and often require pre-training on large datasets. Typically Vision Transformers (ViTs) show great performance when trained on ImageNet-1k/22k (Deng et al., 2009) or JFT-300M (Sun et al., 2017) datasets; however, the absence of such large-scale pre-training is very detrimental to its performance.

A.4 DATA

We use the Galaxy10 DECals dataset introduced by Leung & Bovy (2019) which contains $\sim$ 17.7k labeled images (see raw images in Appendix A.8, Figure 4). These images come from the DESI Legacy Imaging Surveys which include: two DECals campaign (Dey et al., 2019; Walmsley et al., 2022), the Beijing-Arizona Sky Survey (Zou et al., 2019) as well as the Mayall z-band Legacy Survey (Dey et al., 2019). The labels for the galaxy morphologies come from the GalaxyZoo Data Release 2 (Lintott et al., 2008; 2011). This dataset includes 10 strictly exclusive morphology classes: disturbed galaxies, merging galaxies, round smooth galaxies, in-between round smooth galaxies, cigar shaped smooth galaxies, barred spiral galaxies, unbarred tight spiral galaxies, unbarred loose spiral galaxies, edge-on galaxies without bulge, edge-on galaxies with bulge. Each of these classes is strictly exclusive from the other. The dataset is imbalanced, in that not all classes have similar number of images. In particular, there are only 334 labeled images present for "cigar shaped smooth galaxies", so during training, we use stratified sampling. The size of this dataset also allows us to explore training techniques and develop approaches that work well in the low data regime and generalize well.

The data collected for the DECals campaign uses Dark Energy Camera Flaugher et al. (2015) on the Blanco 4m telescope and covers both the North Galactic Cap region at Dec $\leq 32^\circ$ and the South Galactic Cap region at Dec $\leq 34^\circ$. The DECals survey utilizes a method of tiling the sky that involves three separate passes. These passes are slightly shifted in relation to one another, with an approximate offset range of 0.1° - 0.6°. The specific pass and duration of exposure for each observation are determined in real-time based on multiple factors allowing for a nearly uniform

level of depth across the survey. The Beijing-Arizona Sky Survey covers 5000 square degrees of the Northern Galactic Cap, using Steward Observatory's 2.3m Bok Telescope. The Mayall z-band Legacy Survey imaged the Dec $\geq 32^{\circ}$ region using the 4-m Mayall telescope (Dey et al., 2019).

A.5 Related Work

Classifying galaxy morphologies on the Galxy10 DECals is a rather well-established task and there have been multiple works in the past by Venn et al. (2019); Blancato (2020); Chen (2021); Ghadekar et al. (2022); Hui et al. (2022); Radhamani et al. (2022); HOLANDA & SANTOS (2022); Ćiprijanović et al. (2022); Maile et al. (2022) have employed multiple techniques for this task. However, no work has explored using hybrid models for this task which allowed us to set a new state-of-the-art for Galaxy10 DECals.

Transformer-convolutional hybrids are a recent innovation and in the past multiple models have proposed different approaches to constructing such hybrids namely: like MobileViT (Mehta & Rastegari, 2022), ResNet-ViT (Dosovitskiy et al., 2021), ConViT (d'Ascoli et al., 2021), PatchConvNet (Touvron et al., 2021), and CoAtNet (Dai et al., 2021). However they do not explore the performance or modification that could be made to these methods to perform well for lower amounts of data with training from scratch. The work by Gani et al. (2022) extensively explores training ViTs in the low data regime and have shown success on small datasets. Their work based on learning self-supervised inductive biases from small-scale datasets use these biases as a weight initialization scheme for fine-tuning. The work by Lee et al. (2021) explores modifying ViTs to learn locality inductive bias. In this paper, we explore training in low data regimes with hybrid models and have motivated the use of hybrid models in the low data regime, unlike these works.

A.6 Results

Table 1: Model Performance on Galaxy10 DECals dataset. `Galaxy10 only` denotes training on the Galaxy10 DECals dataset only; `Galaxy10 + Zoo` denotes the use of extra unlabeled data from GalaxyZoo, though the models that use extra unlabeled data are not the focus of this work, we present those results as well. The rows under the horizontal line in this table represents experiments performed in this work.

Method Description	**Galaxy10 DECals top-1 accuracy(↑)**	
	Galaxy10 only	Galaxy10 + Zoo
Random Baseline	14.77	-
Fractal Analysis (Radhamani et al., 2022)	73.45	-
Architectural Optimization Over Subgroups (Maile et al., 2022)	77.00	-
DeepAstroUDA (Ćiprijanović et al., 2022)	79.00	-
Deep Galaxies CNN (Ghadekar et al., 2022)	84.04	-
EfficientNet (Tan & Le, 2019), (presented by Hui et al. (2022))	86.00	-
Luma (HOLANDA & SANTOS, 2022)	86.20	-
DenseNet 121 (Iandola et al., 2014), (presented by Hui et al. (2022))	88.64	-
EfficientNetv2 (Tan & Le, 2021)	90.24	-
Noisy Student (Xie et al., 2020)	-	91.80
SimCLRv2 (Chen et al., 2020)	-	90.24
Standard CoAtNet-4 (Dai et al., 2021)	81.55	-
Ours (CoAtNet-4 C-C-C-T + AugReg)	**94.86**	-

We report all results and compare against previous models and a random baseline (equivalent to making a guess) in Table 1. We also provide a few random ground truth comparisons to our predictions in Figure 2. The performance of models is calculated using the metrics that are typical for an image classification problem, top-1 accuracy. Additionally, we find that using any of the other larger variants of CoAtNet is detrimental for this task and the model starts heavily overfitting even with

our design changes, augmentation, and regularization. We also explore semi-supervised learning for this task, our very preliminary experiments included applying Noisy Student (Xie et al., 2020), SimCLRv2 (Chen et al., 2020), and Meta Pseudo Labels (Pham et al., 2021) to the Galaxy10 DECals dataset as well as using massive amounts of unlabeled images from the GalaxyZoo Dataset though these are not the focus of this paper. These are also summarized in table 1.

A.7 Implementation Details

In this section, we explain the implementation details of the experiments and our proposed model.

Augmentation. To evaluate the benefits partially-occluded augmentation methods like CutMix (Yun et al., 2019), DropBlock (Ghiasi et al., 2018), and Cutout (DeVries & Taylor, 2017) hold for this task we run some experiments. We find that these regional dropout-based augmentation techniques have a strong detrimental effect on datasets such as the one we use [2] like the Galaxy10 DECals, this is mainly due to the nature of the task, one example we observe is that even some minor augmentations on an image with the class "edge-on galaxies with bulge" could cause the ground-truth (as identified by a human) of the augmented image to shift to the class "edge-on galaxies without bulge" [3] which is detrimental to the model performance. This example is visually in Appendix A.8, Figure 3.

Finally, for augmentation, we consider a combination of Mixup (Zhang et al., 2017) and RandAugment (Cubuk et al., 2020). As for regularization strategies, we make use of stochastic depth regularization, weight decay, and label smoothing. The hyperparameters values for these regularizations are listed in Appendix A.7. Surprisingly, we find that strong augmentations techniques give much higher performance gains than stronger regularization. Overall we believe, that judiciously choosing augmentation and regularization strategies is crucial to model performance in low data regimes and careful selection of the associated hyperparameters for augmentation and regularization is equally important.

Sampling. There is a class imbalance in the dataset, meaning that not all classes have a similar number of images. For this reason, we follow a stratified sampling strategy during data loading to ensure each batch contains $10 \pm 4\%$ instances of each label class.

Model backbone. Throughout this work we use CoAtNet as the model backbone due to their success in efficiently unifying depthwise convolutions and self-attention through relative attention and their approach of vertically stacking convolution layers and attention layers. Since the Galaxy10 DECals dataset does not contain a large amount of data, other transformer-based models did not show great results for this task whereas a hybrid model with regularization and augmentation techniques, was able to generalize well and achieved better results.

Baseline model. We established a naive baseline with random guesses. The baseline model we chose was training CoAtNet-4 without any design modifications. This gets to a top-1 accuracy of 81.55% on the Galaxy10 DECals dataset. To train this model, we employ standard augmentations (MixUp and CutMix) and train the network for 300 epochs using the default settings in timm with a batch size of 256.

Loss function. We adopt the standard cross entropy loss with smoothing. We also performed some preliminary experiments using Dense Relative Localization Loss (Liu et al., 2021) and we believe this might be a potentially promising direction as well.

Code. Our code is in PyTorch 1.10 (Paszke et al., 2019). We use a number of open-source packages to develop our training workflows. Most of our experiments and models were trained with timm (Wightman, 2019) and we also use mmclassify (Contributors, 2020) for some of the

[2] We also explore Attentive CutMix (Walawalkar et al., 2020) and though it identifies the most discriminative regions based on the intermediate attention maps from a feature extractor, we do not observe a significantly higher generalization capability.

[3] More information on how galaxy morphologies are manually classified can be found at `https://data.galaxyzoo.org/gz_trees/gz_trees.html`

experiments. Our hardware setup for the experiments included either four NVIDIA Tesla V100 GPUs or a TPUv3 cluster. We utilized mixed-precision training with PyTorch's native AMP (through `torch.cuda.amp`) for mixed-precision training and a distributed training setup (through `torch.distributed.launch`) which allowed us to obtain significant boosts in the overall model training time. We report the number of parameters and FLOPS of the final model in Table 3.

Hyperparameters. The choice of hyperparameters for training the CoAtNet-4 C-C-C-T + AugReg model is shown in Table 2. The rest of the hyperparameters were kept to their defaults as provided in timm. The hyperparameters related to Lookahead are used at their default values as suggested in the paper.(Wightman, 2019).

Table 2: Hyper-parameters used to train CoAtNet-4 C-C-C-T + AugReg

Hyper-parameter	**Values**
Stochastic depth rate	0.2
Center crop	False
Mixup Alpha	0.8
Train epochs	300
Label smoothing	0.1 (timm default)
Train batch size	256
Optimizer type	Lookahead (Zhang et al., 2019) + RAdam (Liu et al., 2020)
LR decay schedule	Cosine
Base learning rate	2e-5
Warmup learning rate	1e-5 (timm default)
Warmup	5 epochs
Weight decay rate	1e-2
Gradient clip	None
EMA decay rate	None
RandAugment layers	2

Table 3: Model parameters and FLOPS for the final model we propose.

Model	**Eval Size**	**# Params**	**# FLOPs**	**top-1 accuracy**
EfficientNetv2	384^2	117.24M	12.40B	90.24
Standard CoAtNet-4	224^2	274.06M	62.65B	81.55
CoAtNet-4 C-C-C-T + AugReg	224^2	271.54M	60.54B	94.86

A.8 Supplemental Figures

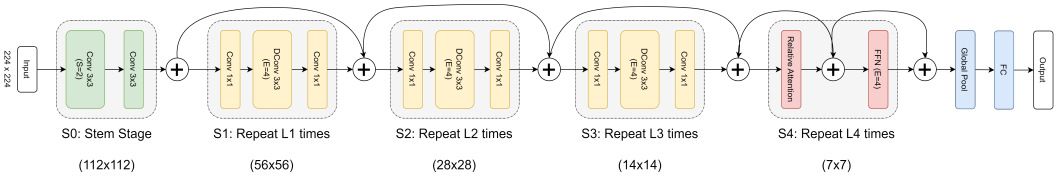

Figure 1: Overview of the proposed model, notice the C-C-C-T design we employ.

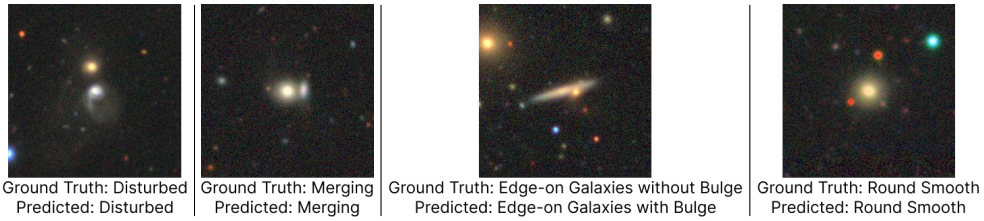

Figure 2: Random selection of ground truth comparison against our model predictions on the hold-out test set. We also include one example where the model predicted incorrectly.

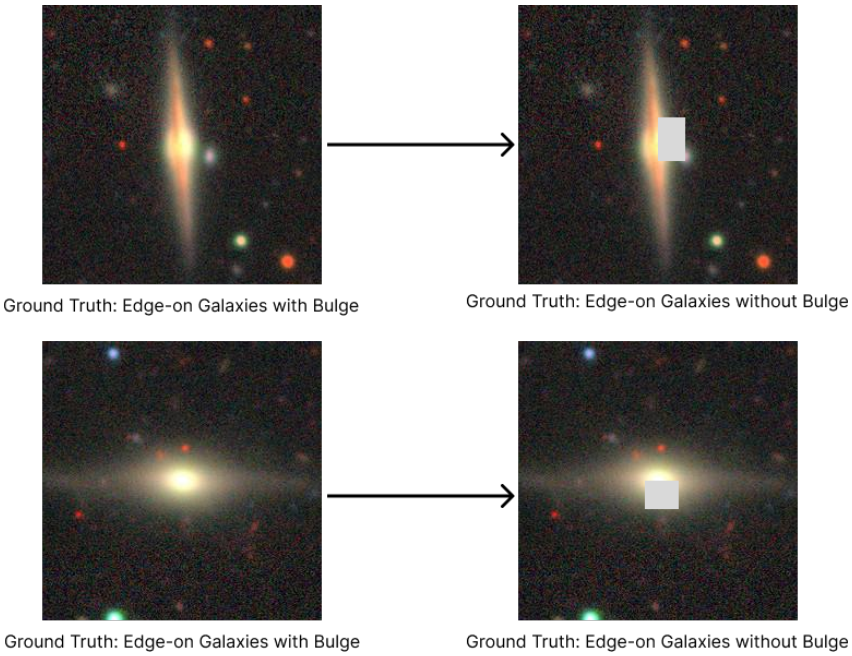

Figure 3: Random selection of augmented images after applying regional dropout-based augmentation techniques. We note the visible differences in the galaxy morphologies as well as observe visually why partial-occluded augmentation techniques do not work well for this dataset.

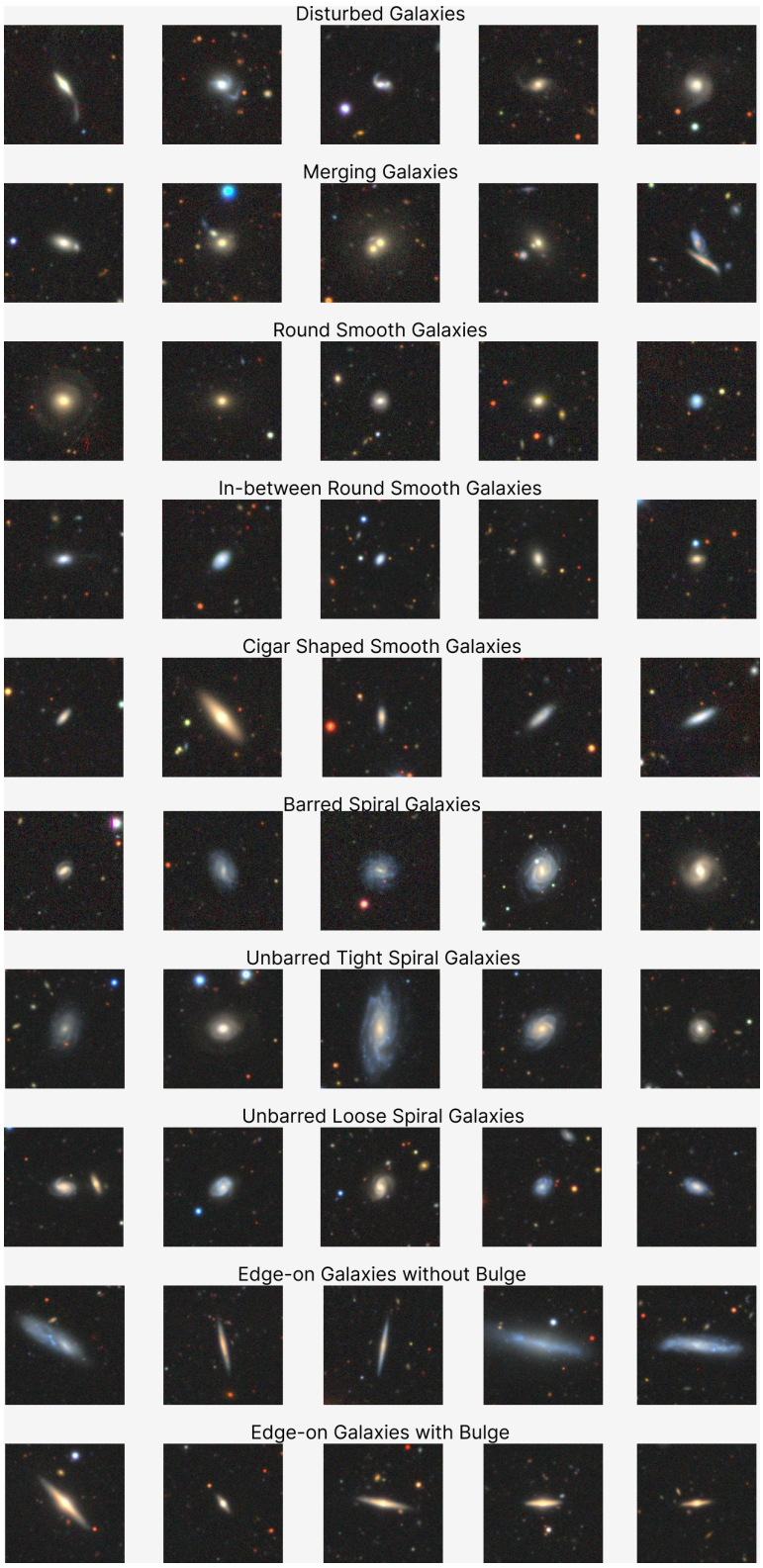

Figure 4: Raw data (galaxy images and labels) from the Galaxy10 DECals dataset. All images are a random sample from the training set.