Regression with Image and Ancillary Data

Ryan Samman
School of Computer Science
University of Waterloo
Waterloo, ON, N2L 3G1
rsamman@uwaterloo.ca
report due: August 12

Abstract

In this project, I explore regression prediciton using both images and ancilliary data. I experiment with different approaches, including CNN ensembles, and pre-trained models like DinoV2 with XGBoost. All code and data are available on GitHub for reproducibility (link).

1 Introduction

The task involves regression prediction with both image data and ancillary data. Our goal is to explore various methodologies, from baseline models to advanced techniques like pre-trained networks, to achieve the best performance.

2 Related Works

This work relies on the work originally created to train an ensemble model to predict plant traits. Furthermore, it draws from ResNet50, DinoV2, and XGBoost.

3 Main Results

We began with a baseline model using only ancillary data, followed by min-max scaling and outlier removal. Furthermore, we added simple Gaussian noise to regularize the data.

Submitted to CS480/680, 2024 Spring. Do not distribute.

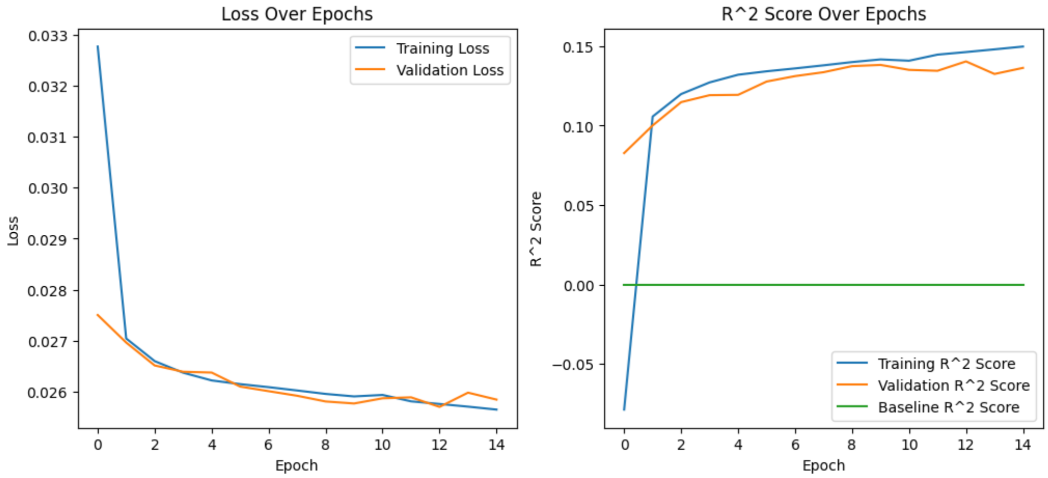

Figure 1: Auxilliary Data Only

Then, we had moved onto ResNet50, where I had attempted to add several layers ontop of the pre-trained model. However, the performance was significantly lacking. Again, I add gaussian noise, as well as perform several operations on the image like Affine/Flips, Color Jittering, and normalization to the mean.

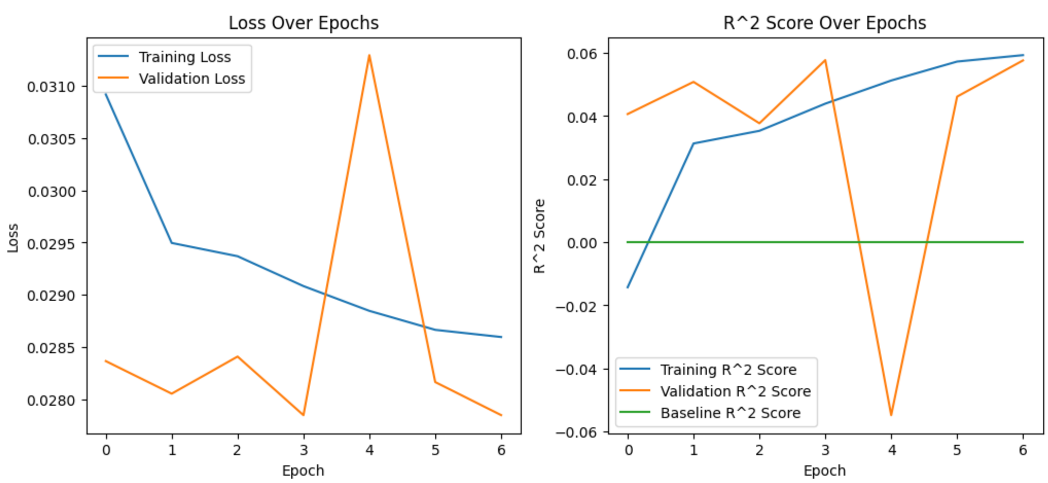

Figure 2: ResNet + Images Only

Afterwards, I had created an ensemble of the two, as was done in the paper predicting plant traits. However, the performance was still significantly lacking.

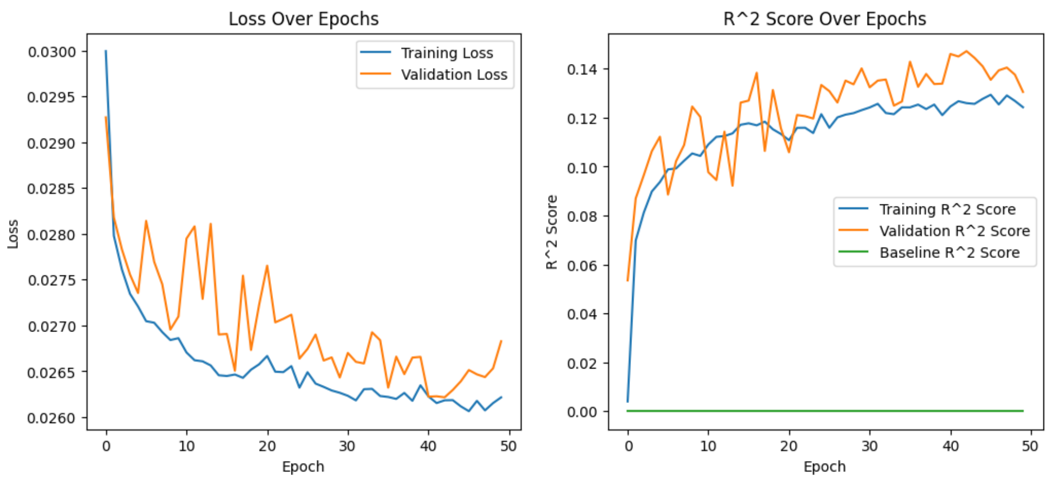

Figure 3: Ensemble of Resnet and Auxillary Data

After this attempt, I decided to experiment with XGBoost. To my surprise, I was able to reach an R2 0.1995 with the auxilliary data alone.

Afterwards, I had decided to use a pretrained model such as dinov2. However, my approach which I had taken so far of training dinov2 had not worked.

Instead, understanding that Dinov2 was an unsupervised model used to create image embeddings, I had generated the embeddings of the data using several model sizes of Dinov2, and used XGBoost to regression fit several boosted weak classifiers to the plant trait predictions. Furthermore, I had performed hyperparameter tuning to optimize the results of XGBoost, by brute forcing several configurations until I had seen no further improvements.

Table 1: Model Comparison on Regression Task with Ancillary and Image Data.

Model	$R^2\uparrow$
Auxilliary Network	0.12
CNN Ensemble	0.13
DinoV2 (small)	0.30
DinoV2 (base)	0.38

As seen in Table 1, the XGBoosted DinoV2 model, especially the larger variants, outperformed the CNN ensemble considerably.

4 Conclusion

In conclusion, while DinoV2 and XGBoost showed potential for regression tasks, significant challenges remain, particularly in effectively harnessing large image encoders.

Future work could explore larger transformers of DinoV2, or using different Image Transformers. Furthermore, we could randomly perturb our images to generate more varied embeddings to fit to our XGBoost. We can experiment further with freezing of pre-trained layers for better performance.

Acknowledgement

I would like to thank the creators of Resnet and Pytorch, of which the first iterations of my project were based upon. Then, I would like to acknowledge the pretrained model DinoV2, and the XG-Boost library.

References

Chen, T. and C. Guestrin (Aug. 2016). "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. ACM.

He, K., X. Zhang, S. Ren, and J. Sun (2015). "Deep Residual Learning for Image Recognition". arXiv: `1512.03385 [cs.CV]`.

Oquab, M. et al. (2023). "DINOv2: Learning Robust Visual Features without Supervision". *arXiv preprint arXiv:2304.07193*.

Schiller, C., S. Schmidtlein, C. Boonman, Á. Moreno-Martínez, and T. Kattenborn (2021). "Deep learning and citizen science enable automated plant trait predictions from photographs". *Scientific Reports*, vol. 11, no. 1, p. 16395.