A Contrastive Approach to Weight Space Learning

by

D.J. Swanevelder

MSc Machine Learning/ Artificial Intelligence

Department of Applied Mathematics

Faculty of Science

Stellenbosch University

Supervisor: Ruan van der Merwe

November 2025

Contents

Chapter 1

Introduction

The prevalence of neural networks (NNs) has positioned them as a foundational technology in modern artificial intelligence. However, as their use has grown, so too has the focus on their inherent mechanistic limitations. Two of the most significant drawbacks of NNs are their lack of explainability—the ”black-box” effect—and the immense computational cost required for training. A novel field, weight space learning, emerges as an exploration of these very challenges.

Weight space learning primarily focuses on developing methods to represent the numerical weights of NN models in a latent space, which can then be used for various downstream tasks. The field mainly considers two types of tasks: generative and discriminative.

In the discriminative application, models use the weights of already-trained models as input to accurately predict meta-information about the original model. For example, previous work has used [x method] to predict [y].

In the generataive application, x method was used by y, z did this etc. etc.

The focus of this report is to extend the idea of laten weight representations, to a shared represeentation space. Where a model’s weights, results and the dataset it was trained on are all represented in a shared space. Figure 1.1 outlines the

What the planned design is, and how it contriubutes to the problem outlines in par 1

In Section 1 we will discuss x, next y, concluded by z in attempt to showcase that this report is the best

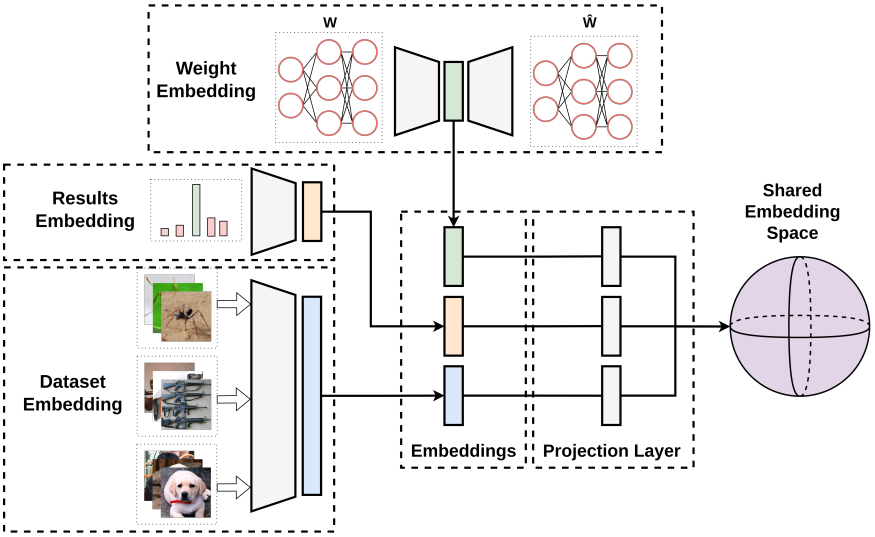

Figure 1.1: A representation of the process of embedding a dataset, model weights and results in a shared embedding space. A dataset embedded using a pretained CLIP encoder, averaging over all images. Model weights embedded using a autoregressive encoder. Model results embeddings and the shared embedding space created through contrastive learning [1]

Chapter 2

Summary and Conclusion

Bibliography

[1] G. E. Dahl, D. Yu, L. Deng, and A. Acero, “Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition,” *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 1, pp. 30–42, 2012.

Appendix A

Project Planning Schedule

This is an appendix.

Appendix B

Outcomes Compliance

This is another appendix.