Can we learn **compact** Neural Network Architectures, in a way that is **general purpose**?

Learning Compact, General Purpose Neural Network Architectures

Kale-ab Tessera, Benjamin Rosman

Introduction

In this work, we propose a unique approach to learning neural network architectures, by attempting to learn **compact, problem agnostic architectures.**

We hypothesize that the performance of sparse neural networks are not dependant on specific connections, but rather on the **number of connections** (weights), which define their capacity.

Motivation

- Current Methods → Primitive, limited by our biases.
- Wrong Focus → Optimizing weights θ, correct architecture f?
- Architectures → Extremely complicated.

Problems - Current NAS Methods

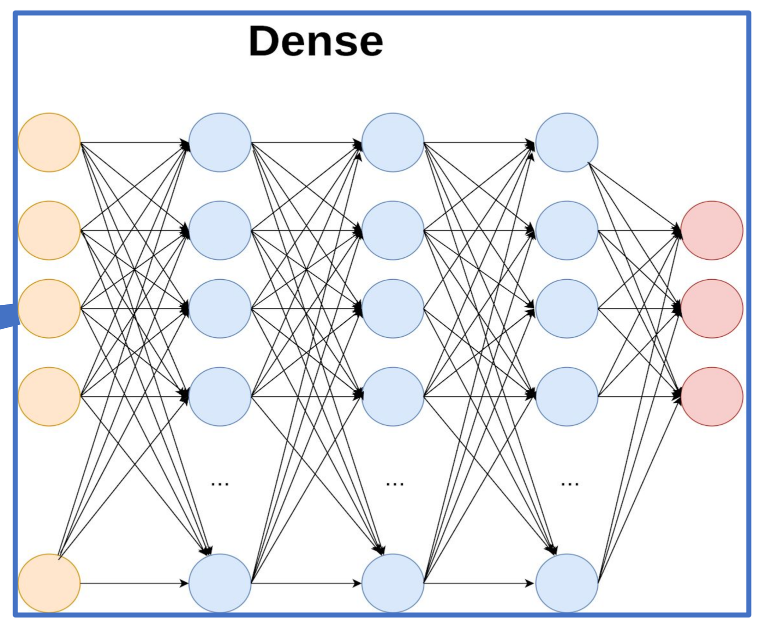
- **Computationally intensive** & **millions of parameters** [2,4,5].
- Still require **domain engineering** [4,5].
- **Restrictive search space** - only convolutional layers [1,2,3,5].

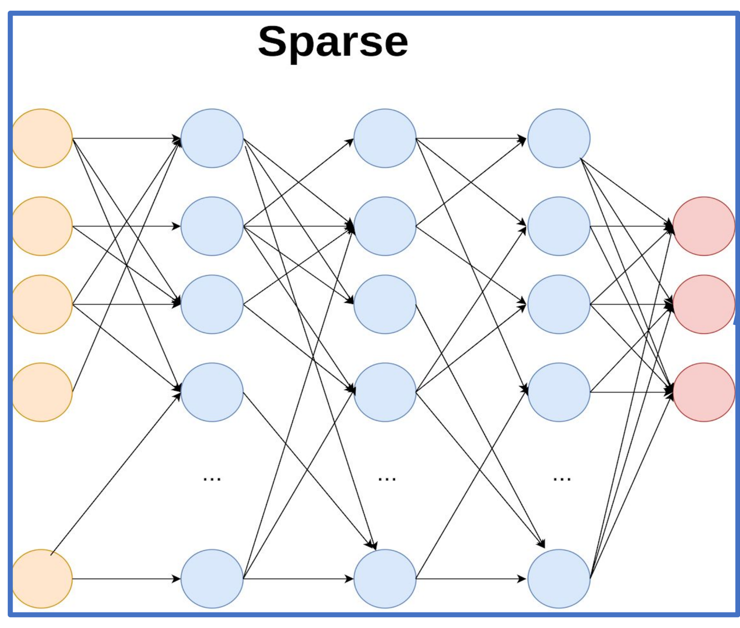
Methodology

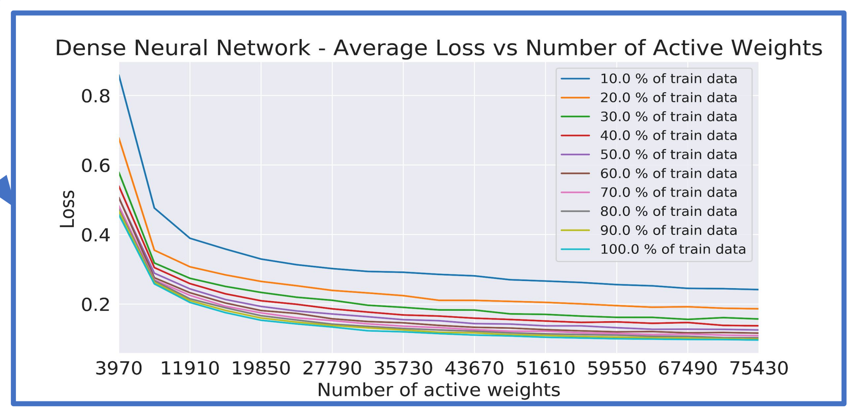
1. Define a concept called **Density** - percentage of active weights in a layer.
2. Allow specification of a **maximum depth** and **width** of a neural network and learn an efficient architecture within those bounds.
3. Use **Random Search** and **Bayesian Hyperparameter** optimisation to find the correct architectures using the above mentioned search space.
4. Use performance estimation techniques to **efficiently estimate** the performance of similar architectures.

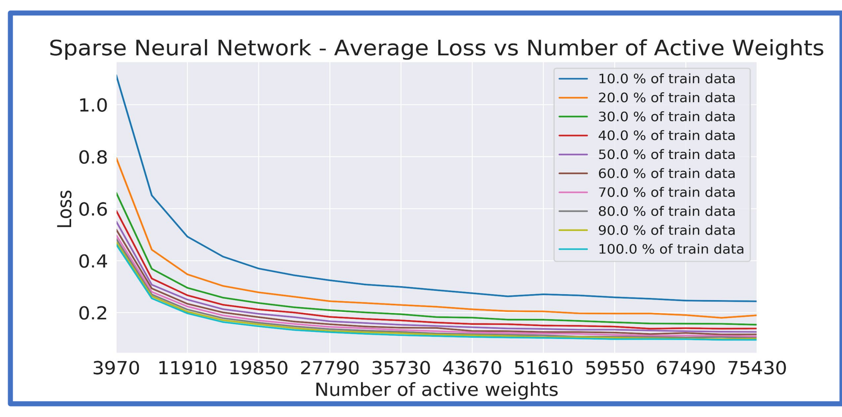
References

1. Baker, B., Gupta, O., Naik, N., and Raskar, R. (2016). Designing neural network architectures using reinforcement learning. CoRR, abs/1611.02167.
2. Real, E., Aggarwal, A., Huang, Y., and Le, Q. V. (2018). Regularized evolution for image classifier architecture search. arXiv preprint arXiv:1802.01548.
3. Suganuma, M., Shirakawa, S., and Nagao, T. (2017). A genetic programming approach to designing convolutional neural network architectures. In Proceedings of the Genetic and Evolutionary Computation Conference, pages 497–504. ACM.
4. Zoph, B. and Le, Q. V. (2016). Neural architecture search with reinforcement learning. arXiv preprint arXiv:1611.01578.
5. Zoph, B., Vasudevan, V., Shlens, J., and Le, Q. V. (2018). Learning transferable architectures for scalable image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 8697–8710.

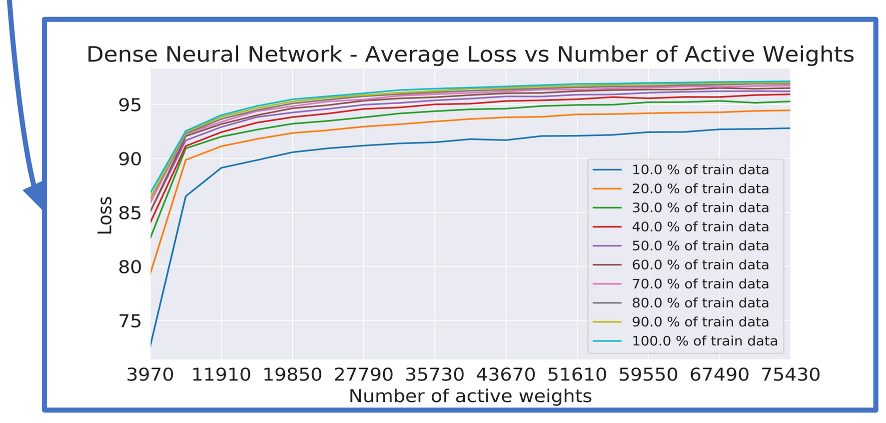


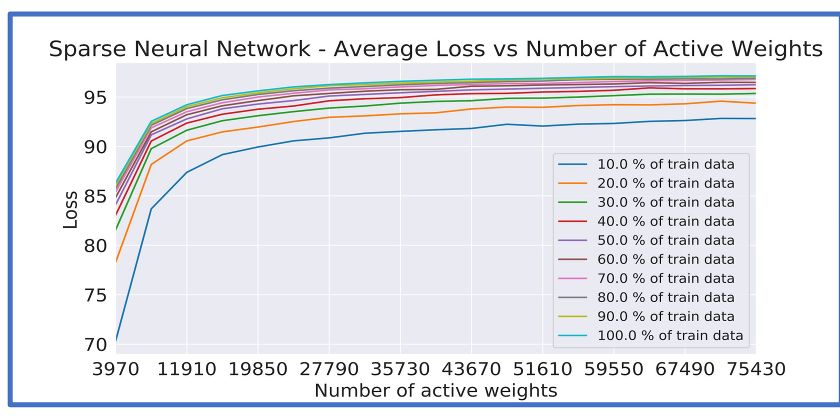




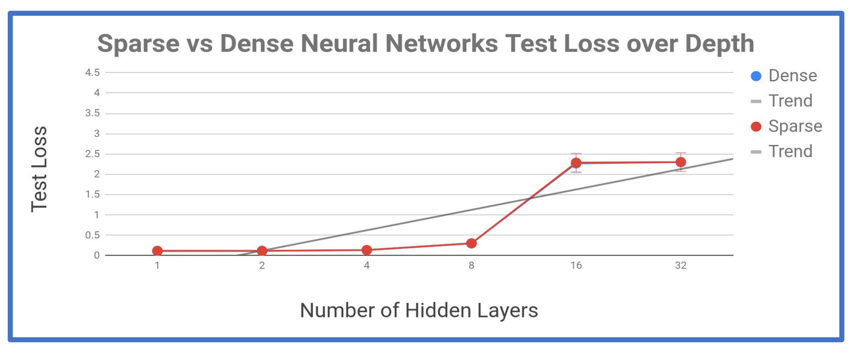


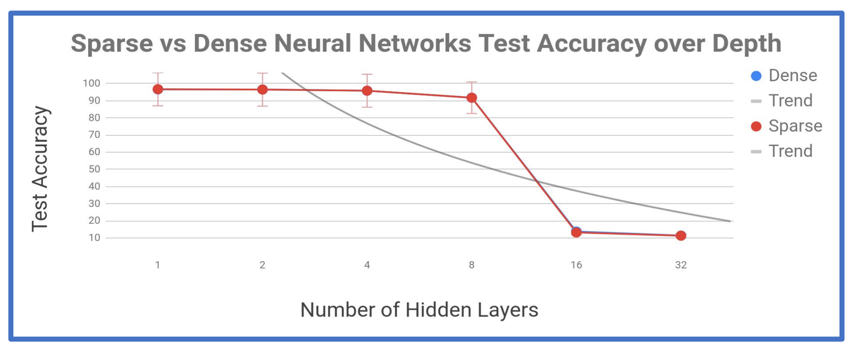
1. Similarity of Dense and Sparse Networks with 1 hidden layer.



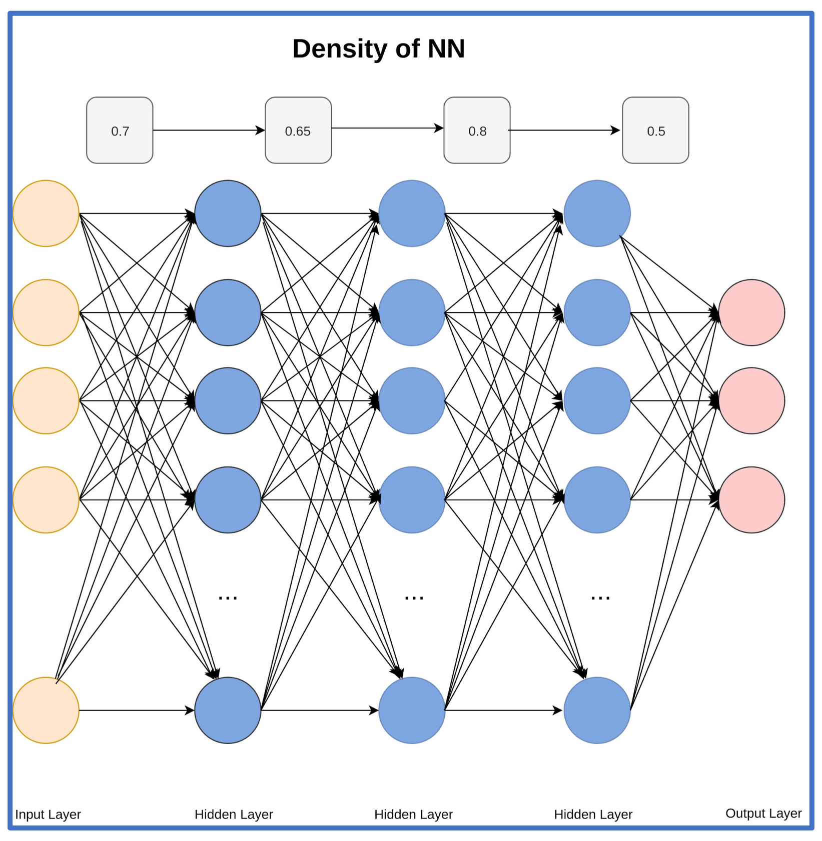


2. Similarity of Dense and Sparse Networks with 1 - 32 hidden layers.





3. Define Density.



4. Results achieved via Random Search.

Results - Mnist

- **97.86%** Accuracy.
- **0.072** Test Loss.
- Approx. **60 000 active weights**.
- **No tuning of non-architectural hyperparameters!**

RAIL LAB

WITS UNIVERSITY

MULTICHOICE ENRICHING LIVES

Scan me