Learning Compact, General Purpose Neural Network Architectures

Kale-ab Tessera, Dr. Benjamin Rosman
University of Witwatersrand

#DLIndaba2019 #SautiYetu

Neural Architecture Search - Why?

Deep Learning:

Data + **Architectures** + Non-architecture hyperparameters → Weights.

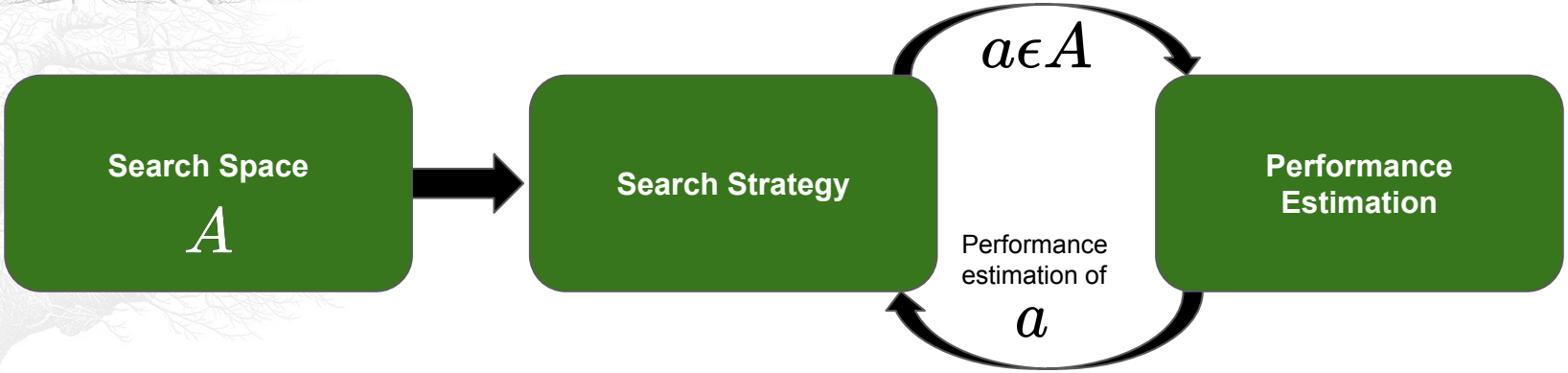
NAS:

Data + Non-architecture hyperparameters → **Architectures** + Weights.

Why?

- Current Methods - **Primitive, limited by our biases**.
- Focusing - Optimizing weights θ, correct function form f ?
- Architectures - Extremely **complicated**.

Neural Architecture Search - Current Methods?



Issues:

1. **Computationally intensive & millions of parameters** (Zoph and 6 [2016], Zoph et al. [2018], Real et al. [2017, 2019]).
2. Still require **domain engineering** (Zoph and Le [2016], Zoph et al.[2018]).
3. **Restrictive search space** - convolutional layers (Zoph et al. [2018], Real et al. [2019],Baker et al. [2016], Suganuma et al. [2017]).

Image: Adapted from Elsken et al.Neural architecture search: A survey.

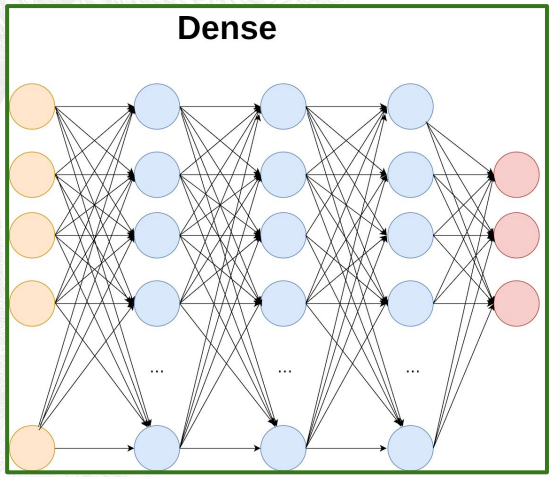
Our Work - Compact, General Purpose

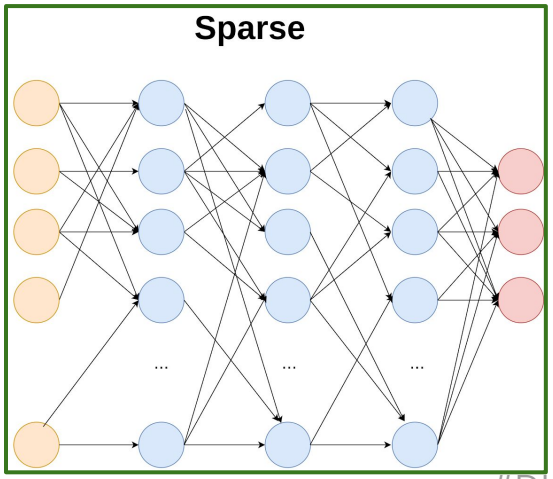
Goal:

Compact → Smaller Models, fewer weights.

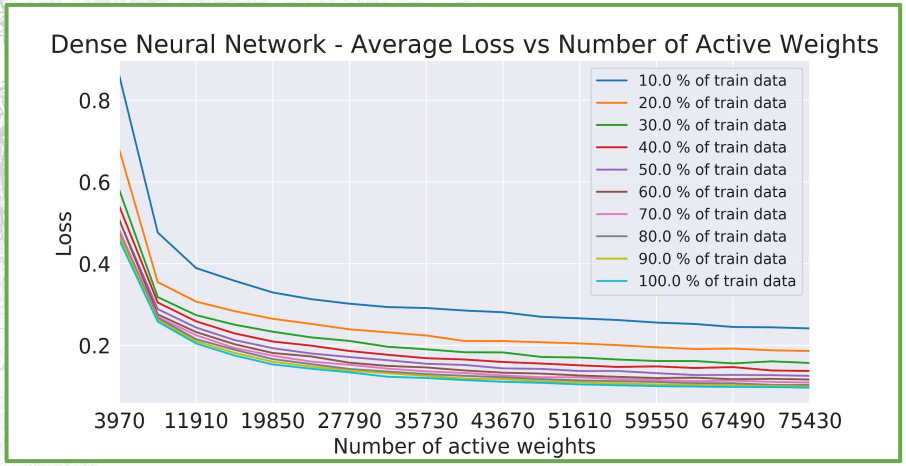
General Purpose → Only Neurons, Weights, Hidden layers.

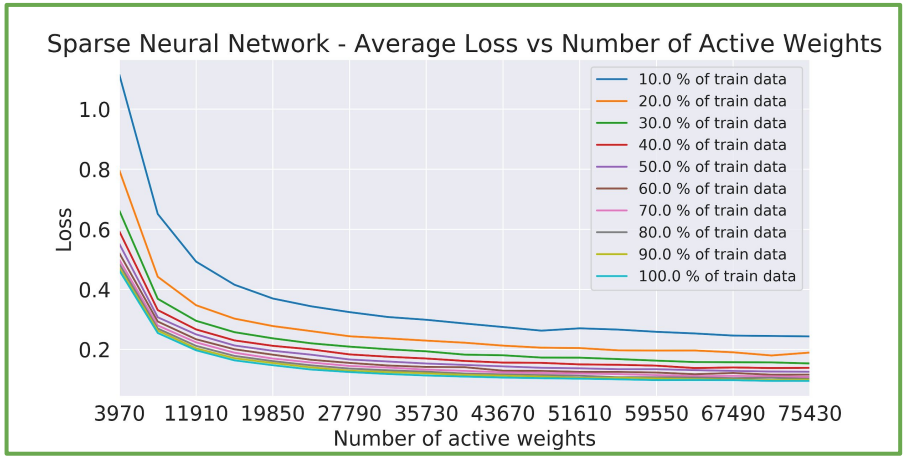
Dense vs Sparse Neural Networks:





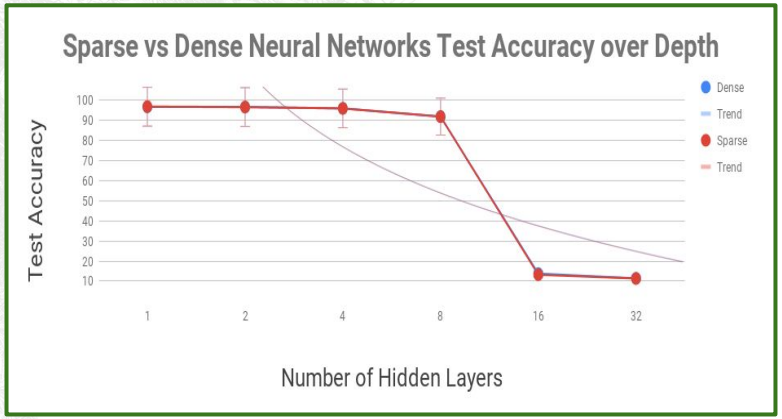
Our Work - Sparse vs Dense NN - 1 Hidden Layer

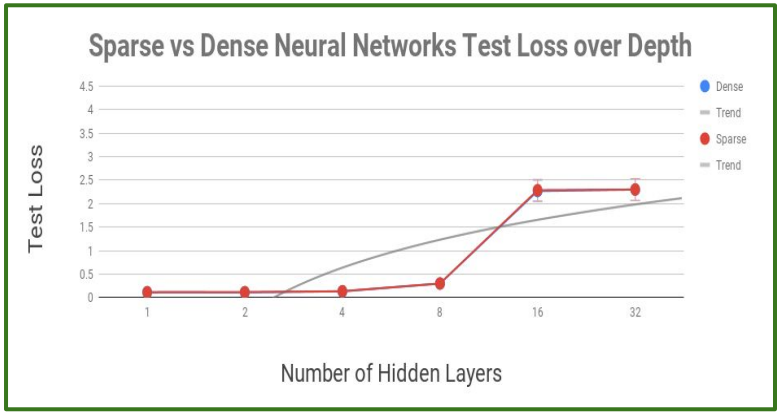




Our Work - Sparse vs Dense NN - Depth

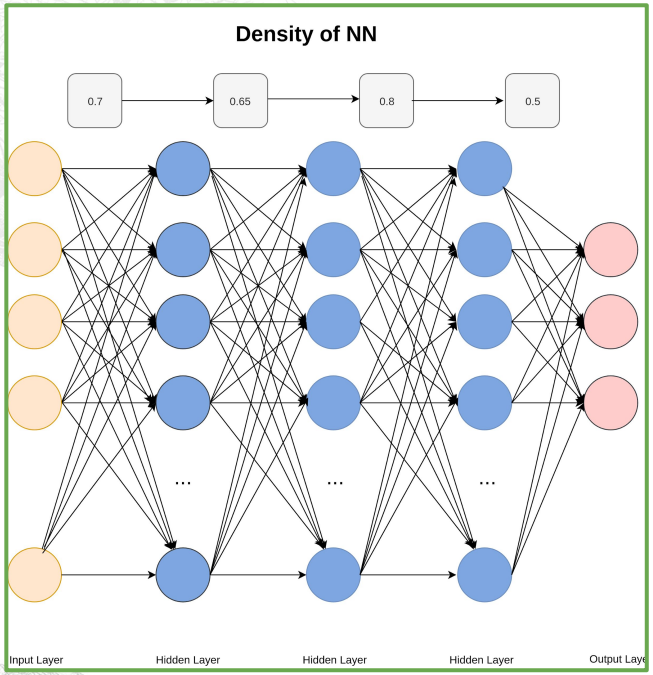
Deeper NN → Behaviour Holds ∴ **Number of weights = Capacity!**





Our Work - Conclusion

Density - % of active weights in a layer.



Preliminary Results - Mnist

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

Future

- Dynamic Neural Network Architectures.
- Performance Estimation.

References

- 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.
- Real, E., Aggarwal, A., Huang, Y., and Le, Q. V. (2018). Regularized evolution for image classifier architecture search. arXiv preprint arXiv:1802.01548.
- Baker, B., Gupta, O., Naik, N., and Raskar, R. (2016). Designing neural network architectures using reinforcement learning. CoRR, abs/1611.02167.
- Elsken, T., Metzen, J. H., and Hutter, F. (2018b). Neural architecture search: A survey. arXiv preprint arXiv:1808.05377