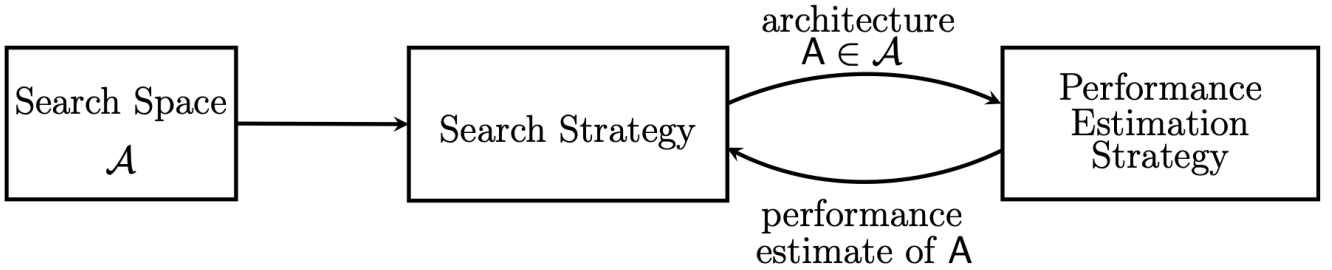
NLP Reading Group

Neural Architecture Search: A Survey

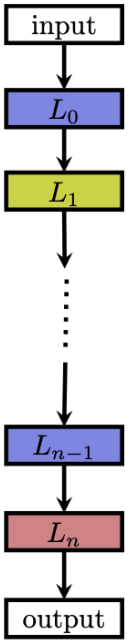
by Thomas Elsken, Jan Hendrik Metzen, Frank Hutter

Introduction



Search Space

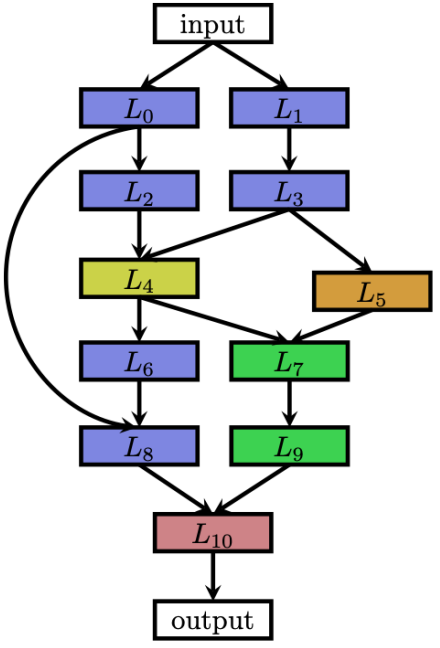
1. Chain-Structured Neural Networks



$$A = L_n \circ \dots L_1 \circ L_0$$

Search Space

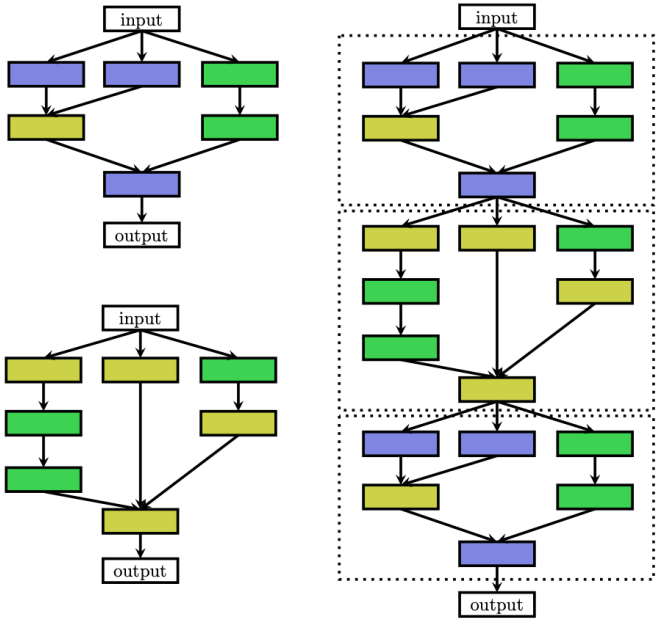
2. Multi-Branch Networks



Input of layer i : $g_i(L_{i-1}^{out}, \ldots, L_0^{out})$

Search Space

3. Cell Search Space : Presentation



Search Space

3.bis Cell Search Space : Pros and Cons

- ***Three Major Advantages :***

Size of the search space.

Easily transferable to other data sets.

Useful design principle in general.

- ***How to choose the Macro-Architecture?***

Optimizing macro-architectures : hierarchical search space introduced by Liu et al. (2018)

1. set of primitive operations,
2. different motifs that connect primitive operations via a directed acyclic graph,
3. Motifs that encode how to connect second-level motifs, and so on.

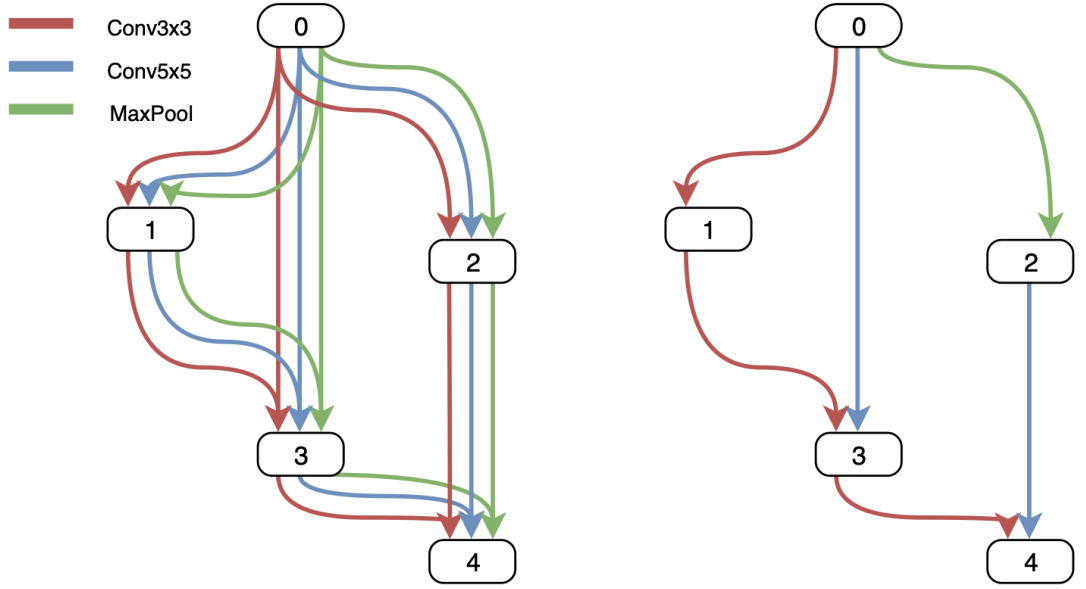
Search Strategy

- Reinforcement Learning : Baker et al. (2017), Zoph and Le (2017), Zhong et al. (2018), Zoph et al. (2018)
- Neuro-evolutionary Approaches : Real et al. (2017), Real et al. (2019), Liu et al. (2018), Elsken et al. (2019)...
- Comparison with Random Search : Real et al. (2019), Liu et al. (2018)
- Tree-based Models : Bergstra et al. (2013), Domhan et al. (2015), Mendoza et al. (2016), Zela et al. (2018)
- Other Discrete Search Methods : Negrinho and Gordon (2017), Wistuba (2017) and Elsken et al. (2017)
- Gradient-based optimization Liu et al. (2019)

Performance Estimation Strategy

Speed-up method	How are speed-ups achieved?	References
Lower fidelity estimates	Training time reduced by training for fewer epochs, on subset of data, downscaled models, downscaled data, ...	Li et al. (2017), Zoph et al. (2018), Zela et al. (2018), Falkner et al. (2018), Real et al. (2019), Runge et al. (2019)
Learning Curve Extrapolation	Training time reduced as performance can be extrapolated after just a few epochs of training.	Swersky et al. (2014), Domhan et al. (2015), Klein et al. (2017a), Baker et al. (2017b)
Weight Inheritance/ Network Morphisms	Instead of training models from scratch, they are warm-started by inheriting weights of, e.g., a parent model.	Real et al. (2017), Elsken et al. (2017), Elsken et al. (2019), Cai et al. (2018a,b)
One-Shot Models/ Weight Sharing	Only the one-shot model needs to be trained; its weights are then shared across different architectures that are just subgraphs of the one-shot model.	Saxena and Verbeek (2016), Pham et al. (2018), Bender et al. (2018), Liu et al. (2019b), Cai et al. (2019), Xie et al. (2019)

Illustration of One Shot Architecture Search :



Future of NAS

- Go beyond Image Classification
- Multi-task and multi-objective problems
- Defining more general search spaces

Thank you for your attention

Any questions?