香港中文大學
The Chinese University of Hong Kong

CENG5030

Part 2-6: Network Architecture Search

Bei Yu

(Latest update: April 9, 2019)

Spring 2019

These slides contain/adapt materials developed by

- Thomas Elsken, Jan Hendrik Metzen, and Frank Hutter (2018). “Neural architecture search: A survey”. In: *arXiv preprint arXiv:1808.05377*

Overview

Search Space Design

Blackbox Optimization

Beyond Blackbox Optimization

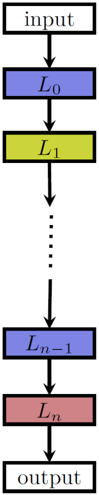
Overview

Search Space Design

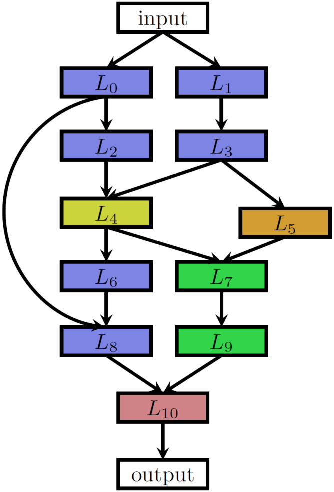
Blackbox Optimization

Beyond Blackbox Optimization

Basic Neural Architecture Search Spaces



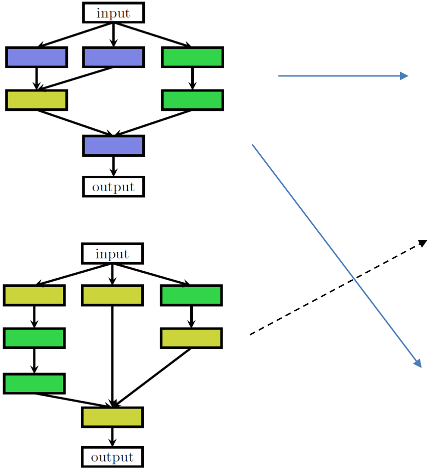
Chain-structured space
(different colours:
different layer types)



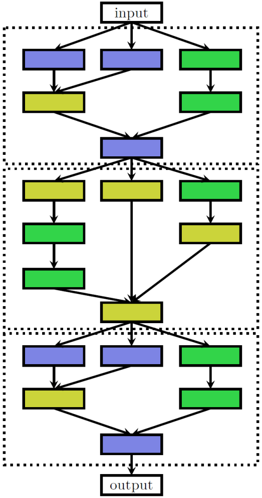
More complex space
with multiple branches
and skip connections

Cell Search Spaces

Introduced by Zoph et al [CVPR 2018]



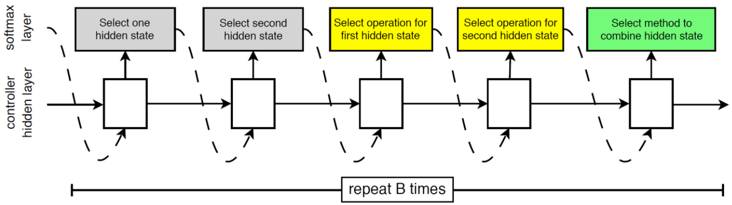
Two possible cells



Architecture composed of stacking together individual cells

NAS as Hyperparameter Optimization

- Cell search space by Zoph et al [CVPR 2018]



  – 5 categorical choices for Nth block:
    - 2 categorical choices of hidden states, each with domain {0, ..., N-1}
    - 2 categorical choices of operations
    - 1 categorical choice of combination method

    → Total number of hyperparameters for the cell: 5B (with B=5 by default)

- Unrestricted search space
  – Possible with conditional hyperparameters
    (but only up to a prespecified maximum number of layers)
  – Example: chain-structured search space
    - Top-level hyperparameter: number of layers L
    - Hyperparameters of layer k conditional on L >= k

Overview

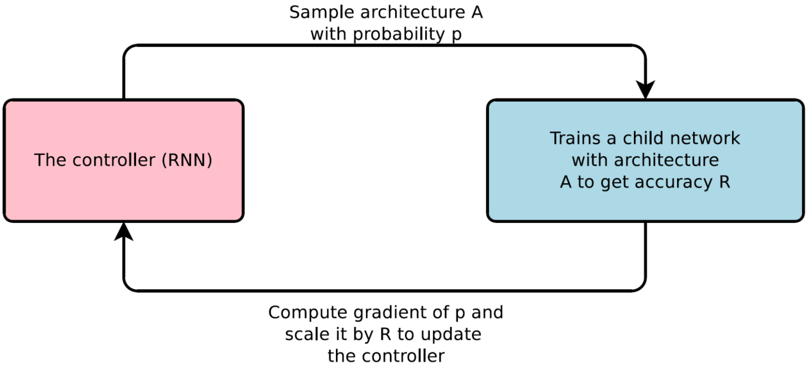
Search Space Design

Blackbox Optimization

Beyond Blackbox Optimization

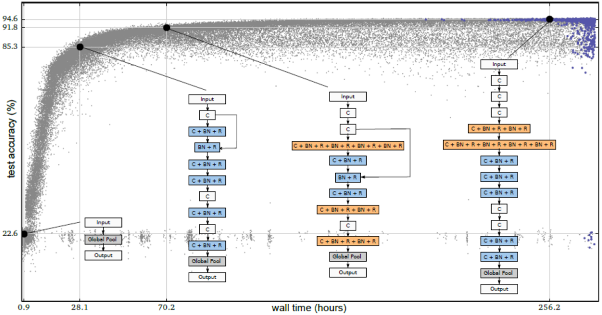
Reinforcement Learning

- NAS with Reinforcement Learning [Zoph & Le, ICLR 2017]
  - State-of-the-art results for CIFAR-10, Penn Treebank
  - Large computational demands
    - **800 GPUs for 3-4 weeks, 12.800 architectures evaluated**



Evolution

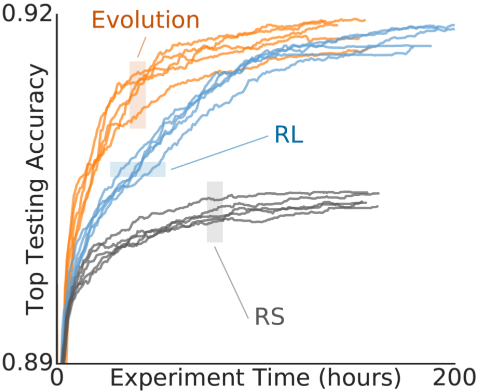
- Neuroevolution (already since the 1990s)
  - Typically optimized both architecture and weights with evolutionary methods
    [e.g., Angeline et al, 1994; Stanley and Miikkulainen, 2002]
  - Mutation steps, such as adding, changing or removing a layer
    [Real et al, ICML 2017; Miikkulainen et al, arXiv 2017]



Regularized / Aging Evolution

- Standard evolutionary algorithm [Real et al, AAAI 2019]
  - But oldest solutions are dropped from the population (even the best)
- State-of-the-art results (CIFAR-10, ImageNet)
  - Fixed-length cell search space

Comparison of
evolution,
RL and
random search



Bayesian Optimization

- Joint optimization of a vision architecture with 238 hyperparameters with TPE [Bergstra et al, ICML 2013]
- Auto-Net
  - Joint architecture and hyperparameter search with SMAC
  - First Auto-DL system to win a competition dataset against human experts [Mendoza et al, AutoML 2016]
- Kernels for GP-based NAS
  - Arc kernel [Swersky et al, BayesOpt 2013]
  - NASBOT [Kandasamy et al, NIPS 2018]
- Sequential model-based optimization
  - PNAS [Liu et al, ECCV 2018]

Overview

Search Space Design

Blackbox Optimization

Beyond Blackbox Optimization

Main approaches for making NAS efficient

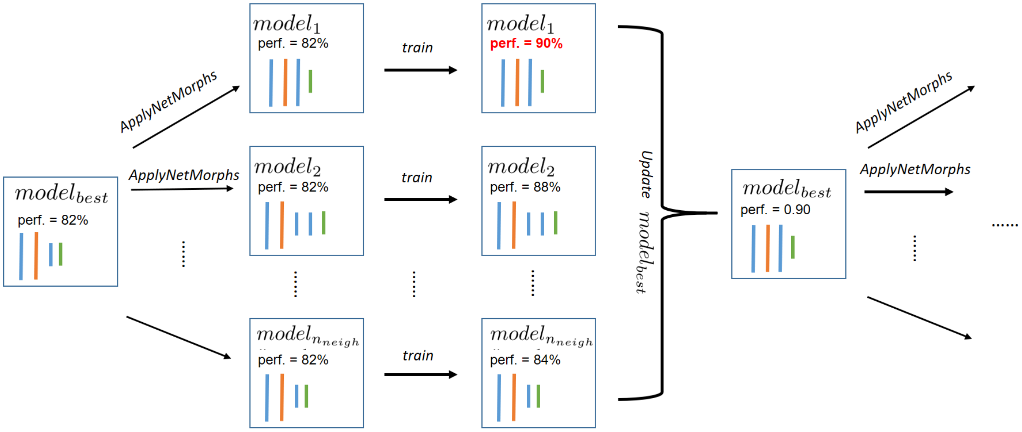
- Weight inheritance & network morphisms
- Weight sharing & one-shot models
- Multi-fidelity optimization
  [Zela et al, AutoML 2018, Runge et al, MetaLearn 2018]
- Meta-learning [Wong et al, NIPS 2018]

Network morphisms

- Network morphisms [Chen et al, 2016; Wei et al, 2016; Cai et al, 2017]
  - Change the network structure, but not the modelled function
    - I.e., for every input the network yields the same output as before applying the network morphism
  - Allow efficient moves in architecture space

Weight inheritance & network morphisms

[Cai et al, AAAI 2018; Elsken et al, MetaLearn 2017; Cortes et al, ICML 2017; Cai et al, ICML 2018]



→ enables efficient architecture search

Weight Sharing & One-shot Models

- Convolutional Neural Fabrics [Saxena & Verbeek, NIPS 2016]
  - Embed an exponentially large number of architectures
  - Each path through the fabric is an architecture

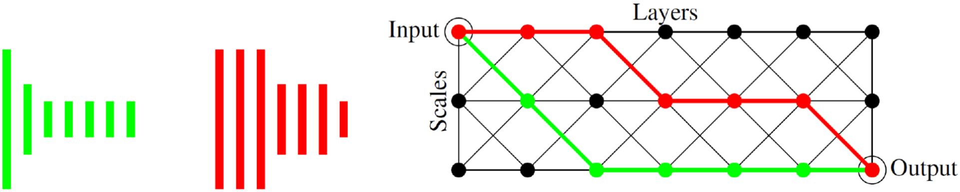
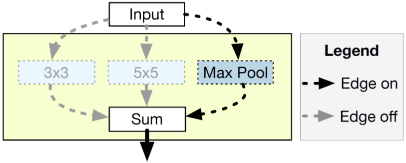


Figure: Fabrics embedding two 7-layer CNNs (red, green).
Feature map sizes of the CNN layers are given by height.

Weight Sharing & One-shot Models

- Simplifying One-Shot Architecture Search [Bender et al, ICML 2018]
  - Use path dropout to make sure the individual models perform well by themselves



- ENAS [Pham et al, ICML 2018]
  - Use RL to sample paths (=architectures) from one-shot model

- SMASH [Brock et al, MetaLearn 2017]
  - Train hypernetwork that generates weights of models

DARTS: Differentiable Neural Architecture Search

[Liu et al, Simonyan, Yang, arXiv 2018]

- Relax the discrete NAS problem
  - One-shot model with continuous architecture weight α for each operator
  - Use a similar approach as Luketina et al [ICML'16] to interleave optimization steps of α (using validation error) and network weights

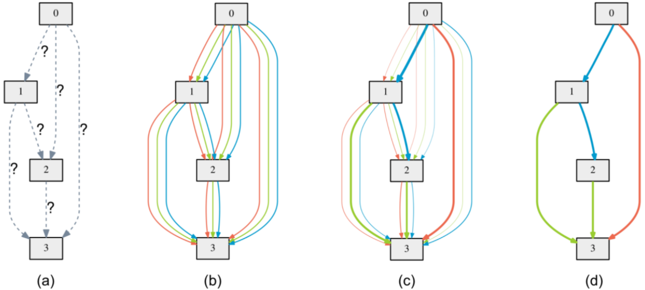


Figure 1: An overview of DARTS: (a) Operations on the edges are initially unknown. (b) Continuous relaxation of the search space by placing a mixture of candidate operations on each edge. (c) Joint optimization of the mixing probabilities and the network weights by solving a bilevel optimization problem. (d) Inducing the final architecture from the learned mixing probabilities.

Further Reading List

- Tianqi Chen, Ian Goodfellow, and Jonathon Shlens (2016). “Net2Net: Accelerating Learning via Knowledge Transfer”. In: *Proc. ICLR*
- Shreyas Saxena and Jakob Verbeek (2016). “Convolutional neural fabrics”. In: *Proc. NIPS*, pp. 4053–4061
- Andrew Brock et al. (2018). “SMASH: one-shot model architecture search through hypernetworks”. In: *Proc. ICLR*
- Hanxiao Liu, Karen Simonyan, and Yiming Yang (2019). “DARTS: Differentiable architecture search”. In: *Proc. ICLR*