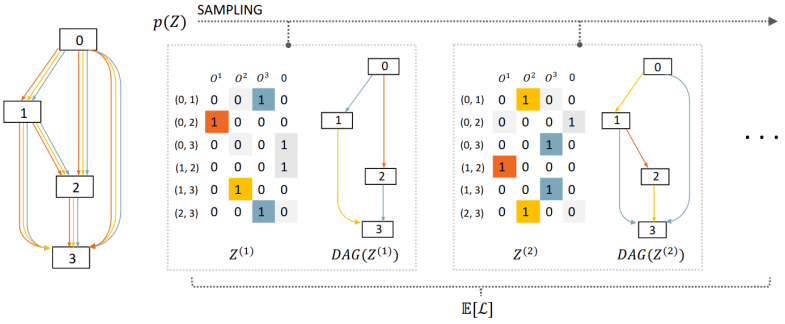
AutoML: Neural Architecture Search

Part 2: One-shot Neural Architecture Search

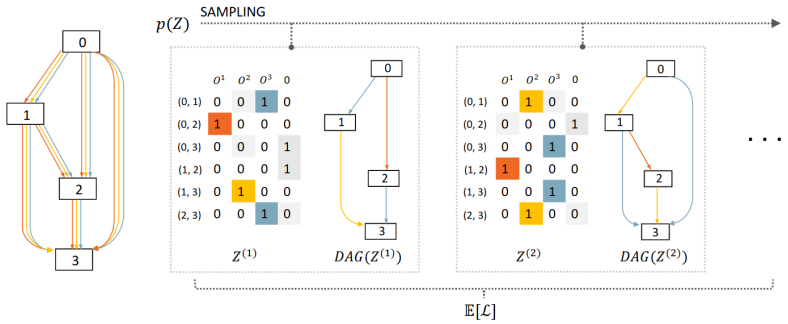
Bernd Bischl Frank Hutter Lars Kotthoff
Marius Lindauer Joaquin Vanschoren

Basic Principle



- The **one-shot model** is a multi-graph containing all possible DAGs
  - Every DAG represents a single architecture $Z^{(\cdot)}$ in the search space
  - Nodes represent aggregating operations (e.g. summation, concatenation) for incoming tensors.
  - Edges represent operations O^i (in the figure: one color per operation)
- The row labels in the matrix above represent a pair of nodes (j, k) in the graph and the column labels the operations O^i. A value of 1 means that that operaration is active in the edge connecting node j to k.

Basic Principle



- The most important principle in one-shot models is **weight-sharing** between graphs.
  - The one-shot model is trained as a normal neural network, i.e. with mini-batch training. The question is how to distinguish single architectures in the one-shot model during this training?
  - One way is that for each sampled mini-batch also sample stochastically an architecture (DAG) and update only the parameters of that architecture.
  - For all subsequent iterations in case a new sampled architecture has common edges (i.e. some entries in the matrices are the same) in the DAG, the weights are shared.

Convolutional Neural Fabrics [Saxena and Verbeek, 2017]

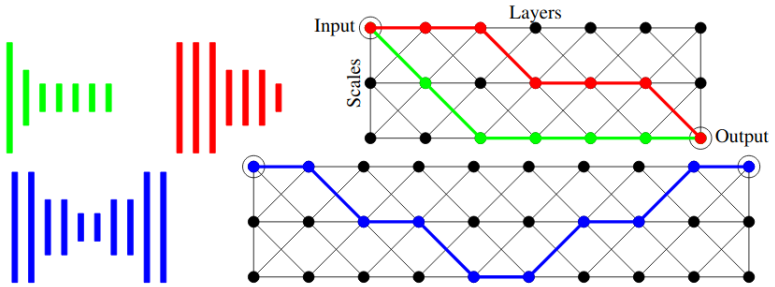


Figure 1: Fabrics embedding two seven-layer CNNs (red, green) and a ten-layer deconvolutional network (blue). Feature map size of the CNN layers are given by height. Fabric nodes receiving input and producing output are encircled. All edges are oriented to the right, down in the first layer, and towards the output in the last layer. The channel dimension of the 3D fabric is omitted for clarity.

- One path from the input to the output of the lattice determines the network structure.
- Paths (architectures) that overlap also share the parameters.

Impact of DropPath [Bender et al., 2018]

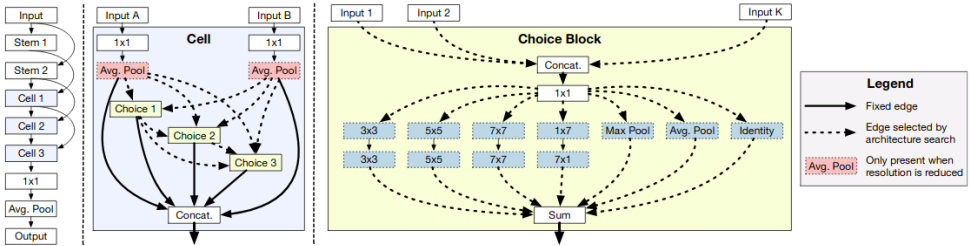


Figure 3. Diagram of the one-shot architecture used in our experiments. Solid lines indicate components that are present in every architecture, while dashed lines indicate optional components that are part of the search space.

- One other way to distinguish the single architectures in the one-shot model is as follows:
  1. Train the one-shot model as a normal network without sampling any individual path (a matrix with only ones "Basic Principle" slide).
  2. Sample K individual architectures after training the one-shot model and evaluate those on the validation set with the one-shot model weight.
  3. Choose the best on validation and re-train that from scratch and return the test error.

Questions to Answer for Yourself / Discuss with Friends

- Repetition:
  What are some pros and cons of the cell search space compared to the basic one?
- Repetition:
  Explain the way in which level-3 motivs in the hierarchical search space use level-2 motivs.
- Repetition:
  What are some pros and cons of the hierarchical search space compared to the other ones?