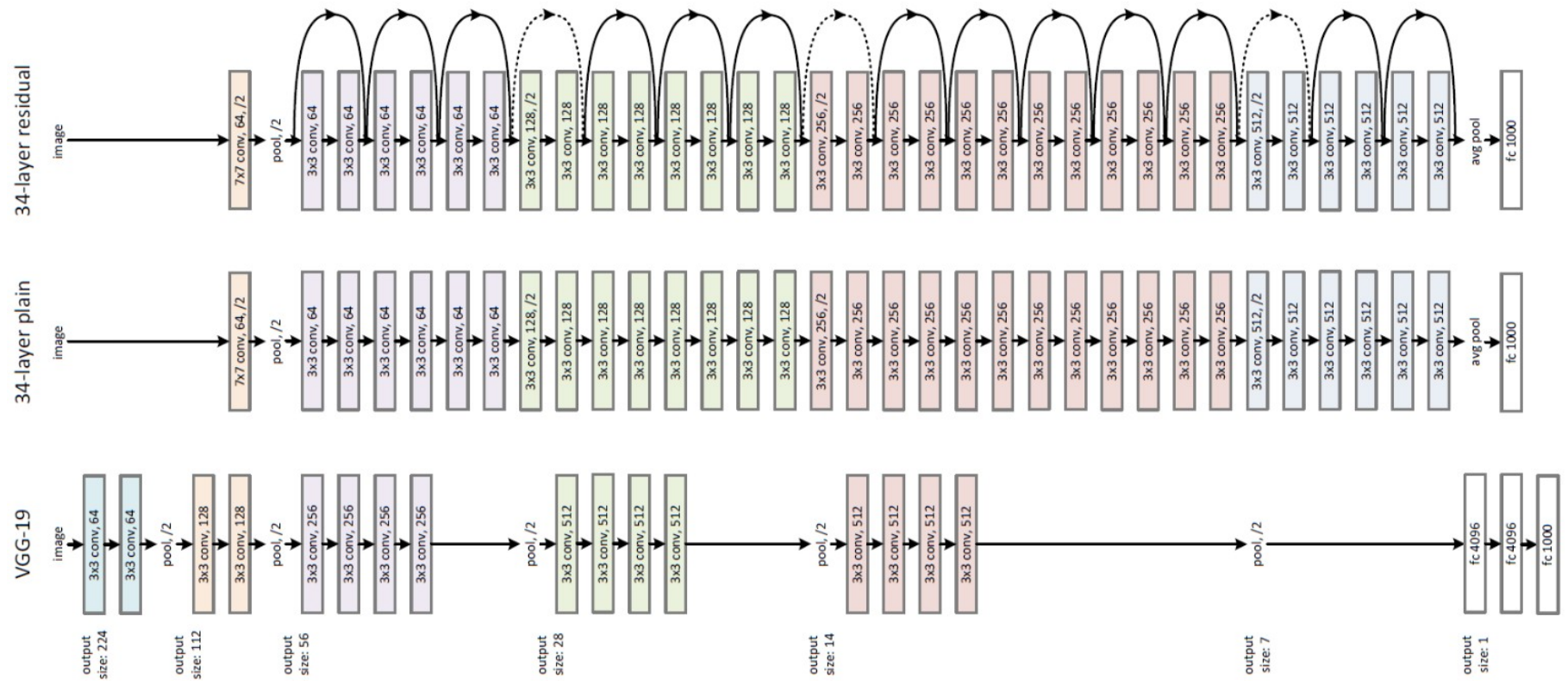
Neural Architecture Search : Introduction Presentation

Ankith Kumar (M.Tech, CSA)
Ashish Kumar Jayant (M.Tech Research, CSA)
Somyajit Guin (PhD, CSA)

ResNet (Microsoft): Winner of ImageNet

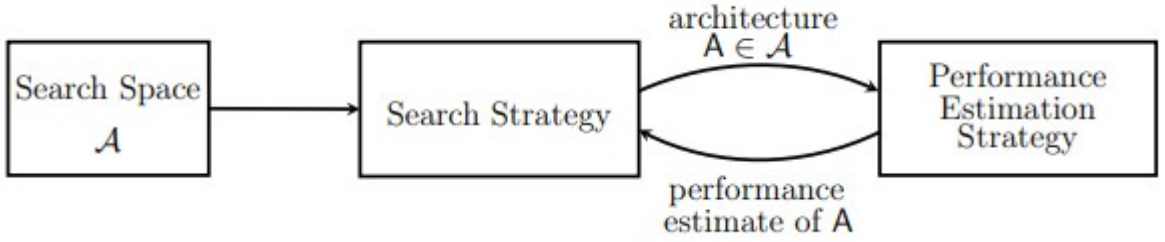


Design of Deep Neural Network Architectures

- State of the art deep neural network architectures are handcrafted by team of researchers and engineers.
- These architecture are generally enormous and highly complex.
- But they are dataset specific.
- How can someone come up with this for any dataset they have, provided we can trade some compute and time?
- Can we design a efficient search strategy to search through space of all neural network architectures?
- Can we design a generalised approach that works across datasets?

Neural Architecture Search

- Process of automating the architecture design part.
- Performance should be comparable to state of the art results.
- Search strategy should be efficient.
- Usability across datasets.

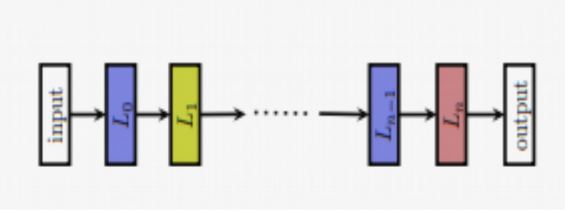


Search Spaces

- Set of architectures which our approach will search through.
- Prior knowledge of some components can reduce our search space, however it has some human bias.
- Common search spaces in recent work : Chain structured , Chain structures with skip connections etc. [1]

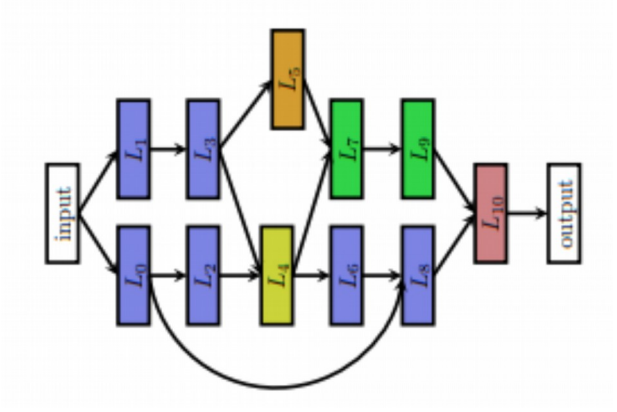
Search Spaces Approach 1 : Macro - Architectures [1]

- Chain-structured Neural Network
  - Sequence of n layers, where i-th layer receives input from i-1th layer
  - Parameters :
    - No of layers (n)
    - Type of operation every layer execute (e.g – pooling,convolution etc)
    - Hyperparameter associated with each operation (e.g – filter size,kernel size etc) *size of this depends on type of operations, hence this is a conditional space



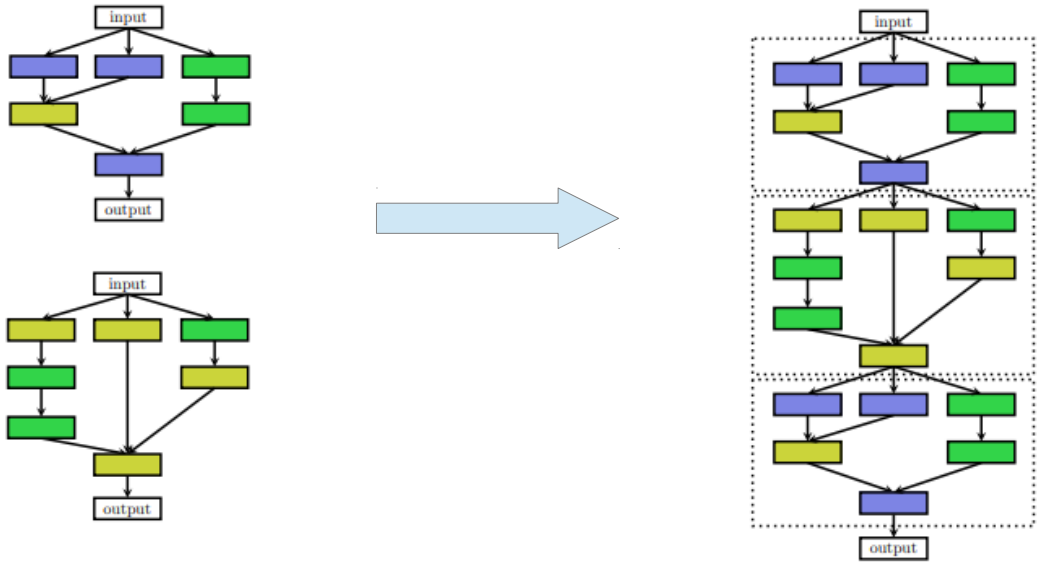
Search Spaces Approach 1 : contd

- Multi-branch networks (more general)
  - Input of ith layer can be written as function of previous layer inputs i.e, input(L_i) = f(L_{i-1}^{out},L_{i-2}^{out},.....L_0^{out})
  - Results in more degrees of freedom



Search Space Approach 2 : Micro-Architectures

- Search for cells rather than whole architectures.
- Final architectures is stacking these cells in predefined manner.



Advantages of Approach 2

- Size of search space is drastically reduced since cells usually consist of significantly less layers.
- Architectures built from cells can be easily transferred to other datasets by simply varying no of cells & parameters within cell.
- Zoph et al. (2018) transfer cells optimized on CIFAR-10 to ImageNet and achieved state of the art performance,

Approach 2 contd : Design choices

- How many no of cells? - Unbounded search space
- How they should be connected? - Arbitirary in principle
- This leads to the fact if we just optimize cells, we may have to do manual engineering of final architecture part.
- Idea – Jointly optimize macro and micro architectures.
  - Hierarchical search space : Liu et al. (2018b)

Search Strategy

- Random Search
- Bayesian Optimization
- Evolutionary Methods
- Reinforcement Learning
- Gradient- based methods

Performance Estimation Strategy

- Train-Validation Approach : Time consuming
- Low fidelity Appproximations approaches
  - Training on subset of data
  - Training on lower dimensional data (e.g low resolution images)
  - Less complicated models
- Learning curve extrapolation
- Weight Inheritance : warm start by inheriting weights from paraent models
- One Shot models : one – shot model trained and weights shared to its all subgraph models

References

1) Neural Architecture Search: A Survey

https://arxiv.org/pdf/1808.05377.pdf