Reducing Neural Architecture Search Spaces with Training-Free Statistics and Computational Graph Clustering

Thorir Mar Ingolfsson, Mark Vero, Xiaying Wang, Lorenzo Lamberti, Luca Benini, Matteo Spallanzani

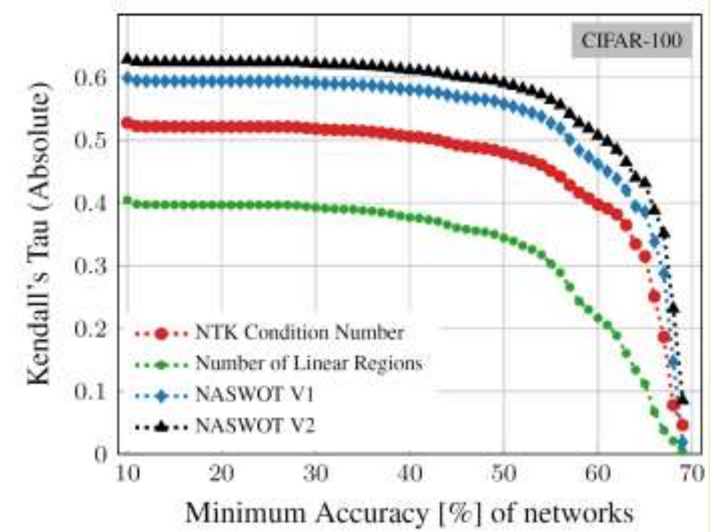
Neural Architecture Search (NAS) is the deep-learning-specific variant of model selection. The goal of NAS is **discovering those network topologies that have good task accuracy, i.e., that are most effective**.

NAS spaces are vast:
- many degrees of freedom (layers, connectivity, …);
- many options for each degree of freedom.

Several **stochastic and probabilistic algorithms** to explore NAS spaces have been proposed:
- Evolutionary Algorithms;
- Reinforcement Learning;
- Gradient-Based Learning;
- Bayesian Methods;
- Random Network Generation.

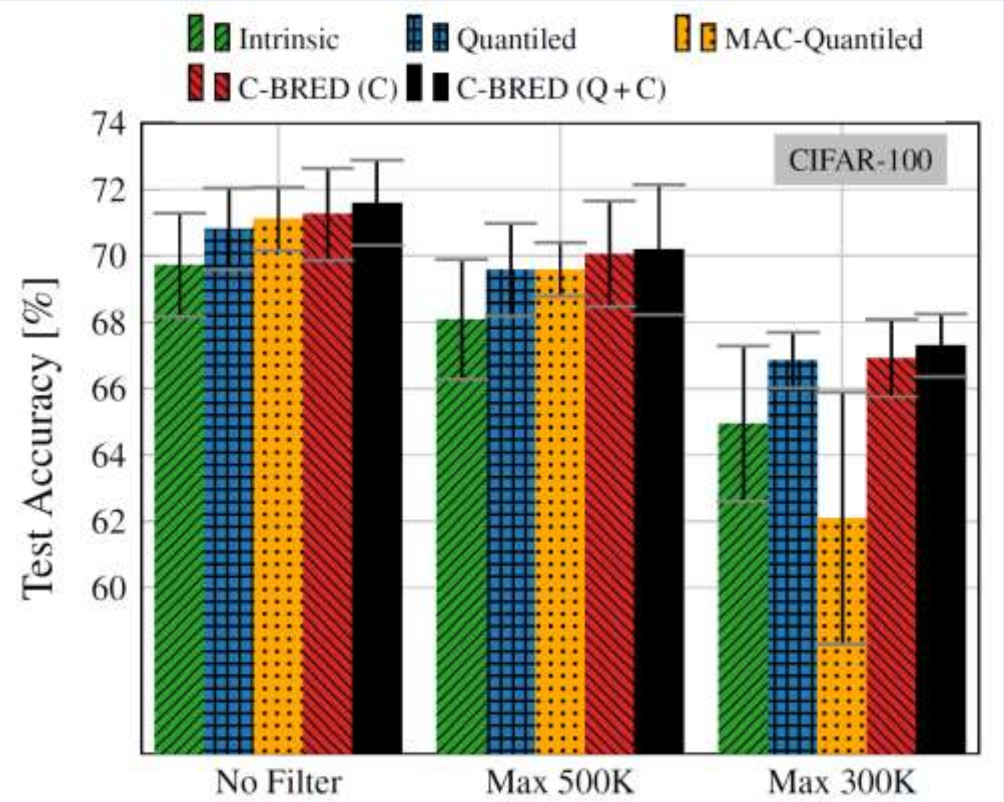
They **require training candidate networks** for several epochs: this is **time-consuming and computationally expensive**.

TF-Statistics give good **relational information** but not good pointwise information



We look at DNNs as **computational graphs** and cluster them to find **high performing** subspaces in an **unsupervised** way

The predictive power of TF statistics is enhanced by the cluster latent variable over different HW constraints.



Effective HW-constrained programs use non-destructive operators (e.g., no pooling).