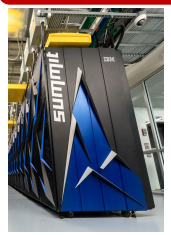
Neural Architecture Search using Deep Neural Networks and Monte Carlo Tree Search

Linnan Wang, Yiyang Zhao, Yuu Jinnai, Yuandong Tian, Rodrigo Fonseca

BROWN

facebook research

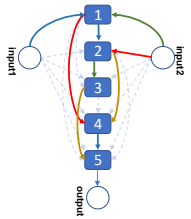
Motivation



Computing Cluster

Search Methods

NAS Agent



Target Architecture

Problems:

- Low efficiency
- Expensive training

Our solutions

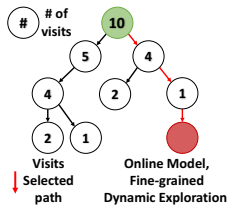
- Monte carol tree search
- Transfer learning

Why MCTS?

Methods	SMBO	sampling mechanism	scalable to $\|\Omega\| \sim 10^{20}$	global search
HyperBand (Li et al., 2016)	X	successive halving	✓	X
BOHB (Falkner et al., 2018)	✓	non-convex optimization	X	X
SMAC (Hutter et al., 2013)	✓	non-convex optimization	X	✓
TPE (Bergstra et al., 2011)	✓	non-convex optimization	X	✓
RE (Real et al., 2019)	X	top-k random	✓	X
Random (Li & Talwalkar, 2019)	X	random	✓	✓
MCTS (Wang et al., 2018)	✓	UCB and search tree	✓	✓

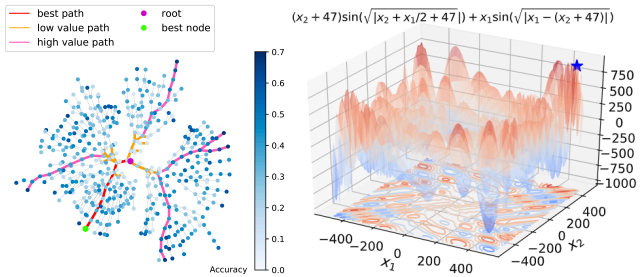
SMBO: Sequential Model Based Optimizations
$|\Omega|$ is the size of search space

1. Fine-grained adaptive exploration



2. Sample proposal:

MCTS traversal down the tree **V.S.** **Bayesian** Max $\phi(x), \forall x \in \Omega$, ϕ is acquisition, e.g. EI



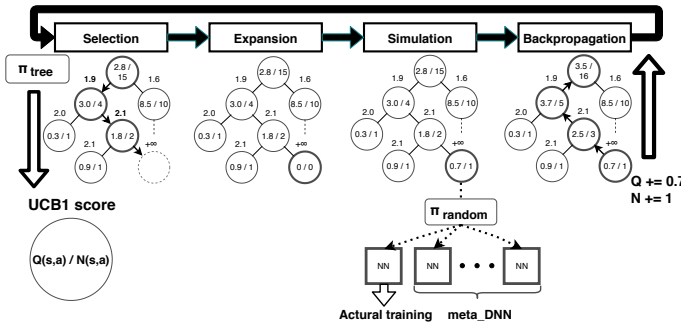
Easy reach of good region, regardless of dimensions

Intractable for high dimensional tasks

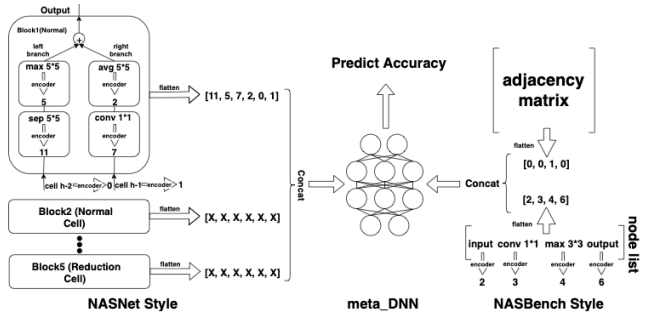
Methodology

1. MCTS agent

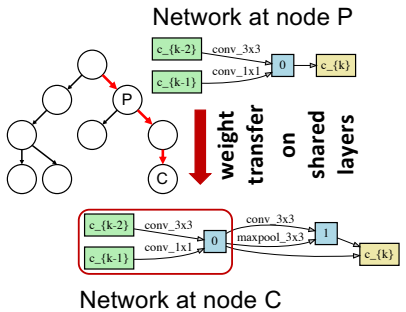
- Selection: traverse down with UCB.
- Expansion: add a new node into the tree.
- Evaluation: ***value function predictions*** and ***training***.
- Back-propagation: back-tracks to update #visits and value



2. Value function prediction:

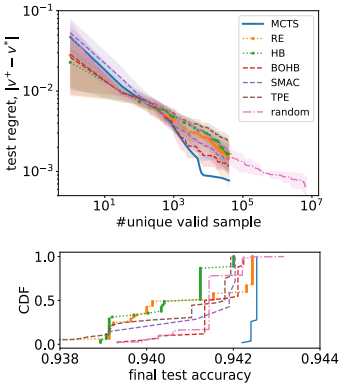


3. Transfer learning in training:



Experiment Results

1. NASBench evaluations



2. In practice

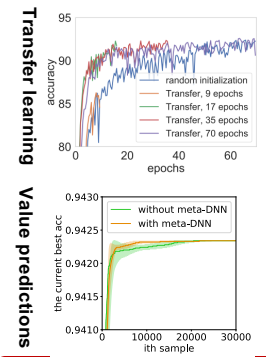
Cifar10

Model	Params	Err	GPU days	M
NASNet-A+cutout (Zoph et al. 2017)	3.3M	2.65	2000	20000
AmoebaNet-B+cutout (Real et al. 2018)	2.8M	$2.55_{\pm 0.05}$	3150	27000
DARTS+cutout (Liu et al. 2018)	3.3M	$2.76_{\pm 0.09}$	4	-
RENASNet+cutout (Chen et al. 2019)	3.5M	$2.88_{\pm 0.02}$	6	4500
AlphaX+cutout (32 filters)	2.83M	$2.54_{\pm 0.06}$	12	1000
PNAS (Liu et al. 2017a)	3.2M	$3.41_{\pm 0.09}$	225	1160
ENAS (Pham et al. 2018)	4.6M	3.54	0.45	-
NAONet (Luo et al. 2018)	10.6M	3.18	200	1000
AlphaX (32 filters)	2.83M	$3.04_{\pm 0.03}$	12	1000
NAS v3 (Zoph and Le 2016)	7.1M	4.47	22400	12800
Hier-EA (Liu et al. 2017c)	15.7M	$3.75_{\pm 0.12}$	300	7000
AlphaX+cutout (128 filters)	31.36M	$2.16_{\pm 0.04}$	12	1000

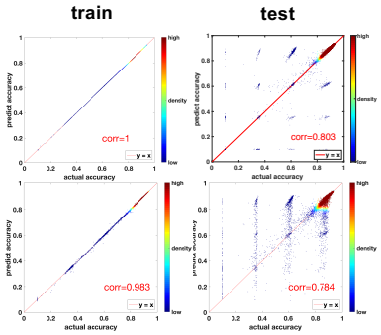
ImageNet

model	multi-adds	params	top1/top5 err
NASNet-A (Zoph et al. 2017)	564M	5.3M	26.0/8.4
AmoebaNet-B (Real et al. 2018)	555M	5.3M	26.0/8.5
DARTS (Liu et al. 2018)	574M	4.7M	26.7/8.7
RENASNet (Chen et al. 2019)	574M	4.7M	24.3/7.4
PNAS (Liu et al. 2017a)	588M	5.1M	25.8/8.1
AlphaX-1	579M	5.4M	24.5/7.8

3. Design validations



4. Ablations of predictors



Applications

1. Neural style transfer:

Content Style VGG AlphaX

2. Object detection:

MobileNet AlphaX MobileNet AlphaX