One-Shot Neural Architecture Search via Self-Evaluated Template Network

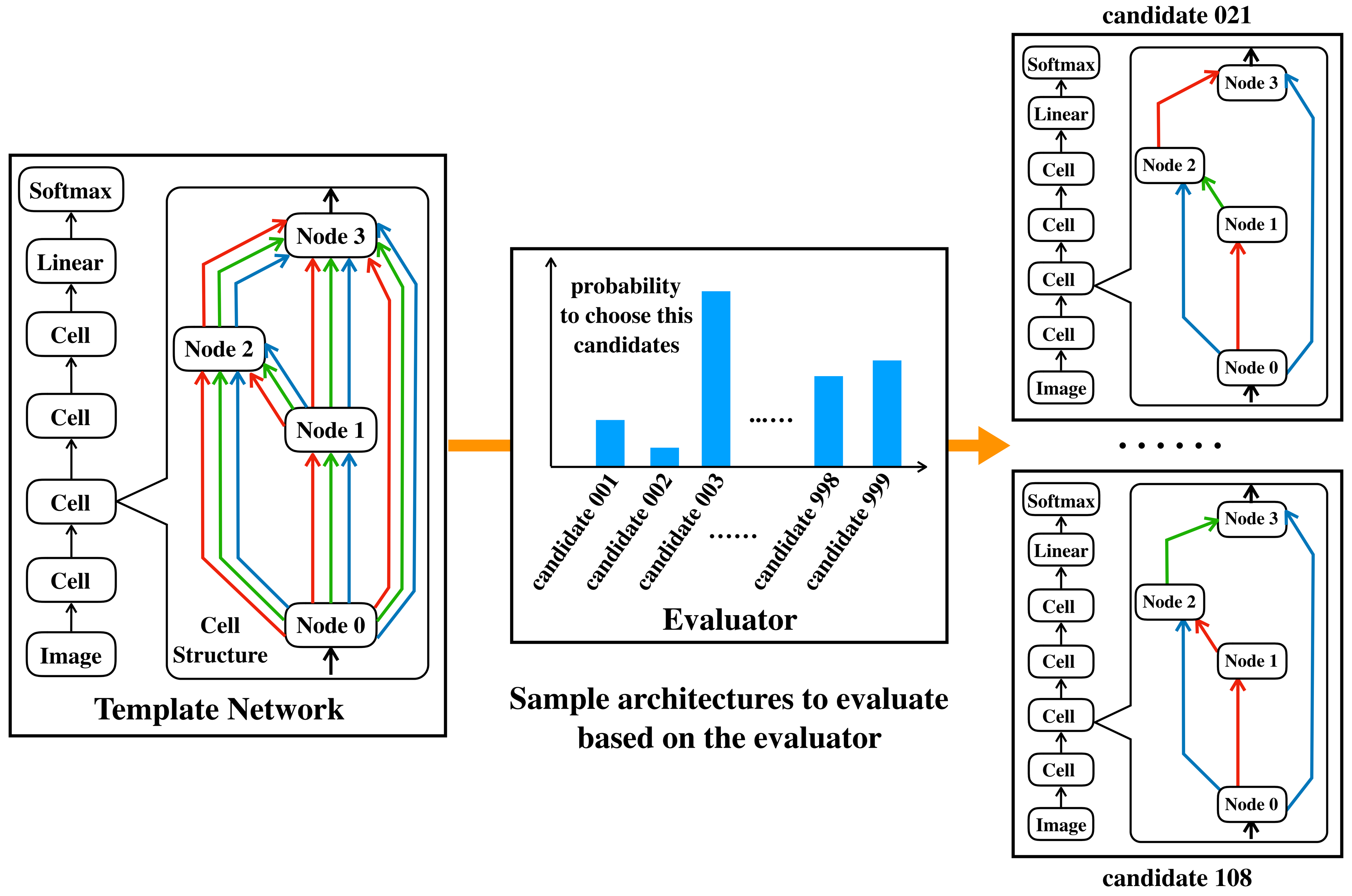
Xuanyi Dong[1,2], Yi Yang[2]

[1]Baidu Research

[2]ReLER, University of Technology Sydney

Overview

Neural Architecture Search (NAS) aims to automatically find an effective neural architecture from the search space.



Motivation: one-shot NAS methods have no sense of which candidate will perform better until evaluation.

Method: we propose self-evaluated template network (SETN). It consists of two components (1) *an evaluator*, learns to indicate the probability of each individual architecture being likely to have a lower validation loss. (2) *a template network*, which shares parameters among all candidates to amortize the training cost of generated candidates.

Algorithm

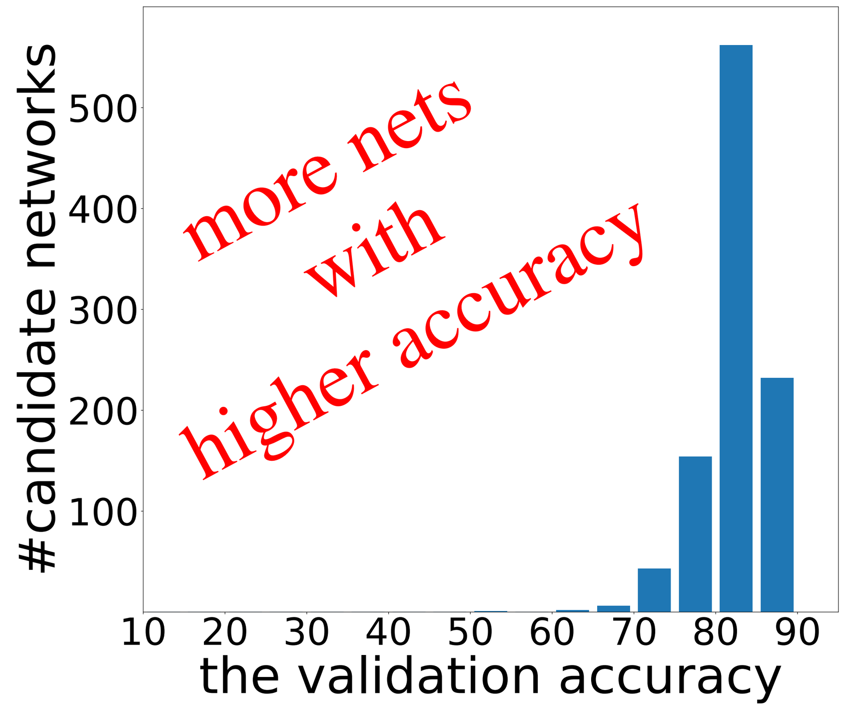
Inputs: a template network with shared parameters ω
an evaluator with α that indicates the *prob* of net
the training data

- Split the training data into two groups $\mathcal{D}_{tr}$ and $\mathcal{D}_{val}$
- **While** not converge **do**:
- optimize ω on $\mathcal{D}_{tr}$: *uniformly sampling* a candidate architecture to forward
- optimize α on $\mathcal{D}_{val}$: *aggregating* all candidate architectures via α to forward
- Using the learned evaluator to sample T "good" architectures to form $\mathcal{A}$
- Evaluating all candidates in $\mathcal{A}$ with the learned ω
- Selecting the candidate with the lowest loss on $\mathcal{D}_{val}$

SETN aims to solve the following three problems in NAS:

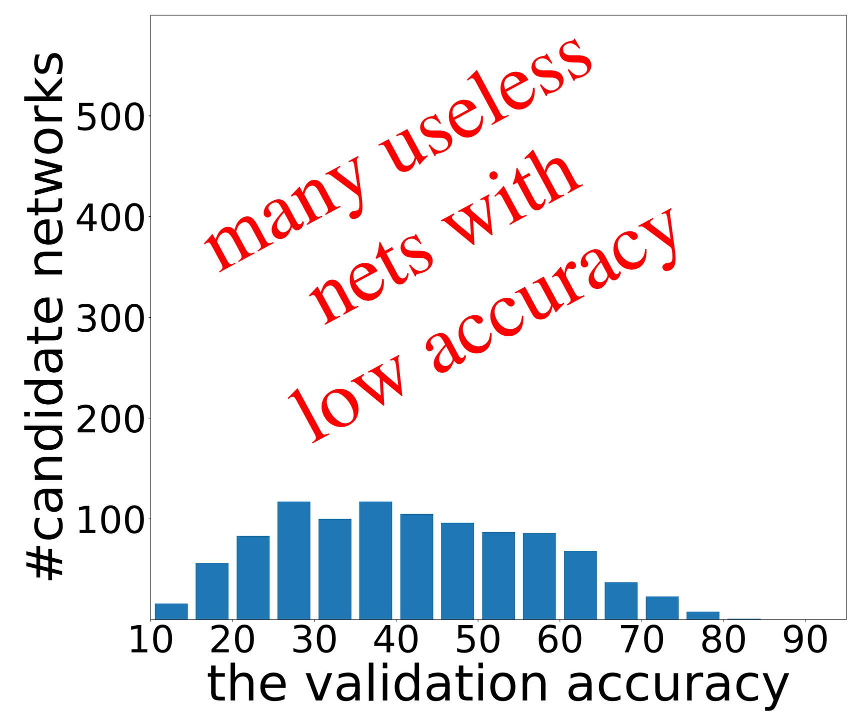
1. **Bias**. Previous methods could bias to some candidate architectures being "shallow" or having "less parameters".
2. **Unstable**. Previous methods use argmax(α) so that searched architectures in different runs may result in very diverse performance.
3. **Redundancy**. Two different candidates in the search space might be essentially the same architecture.

Experiments

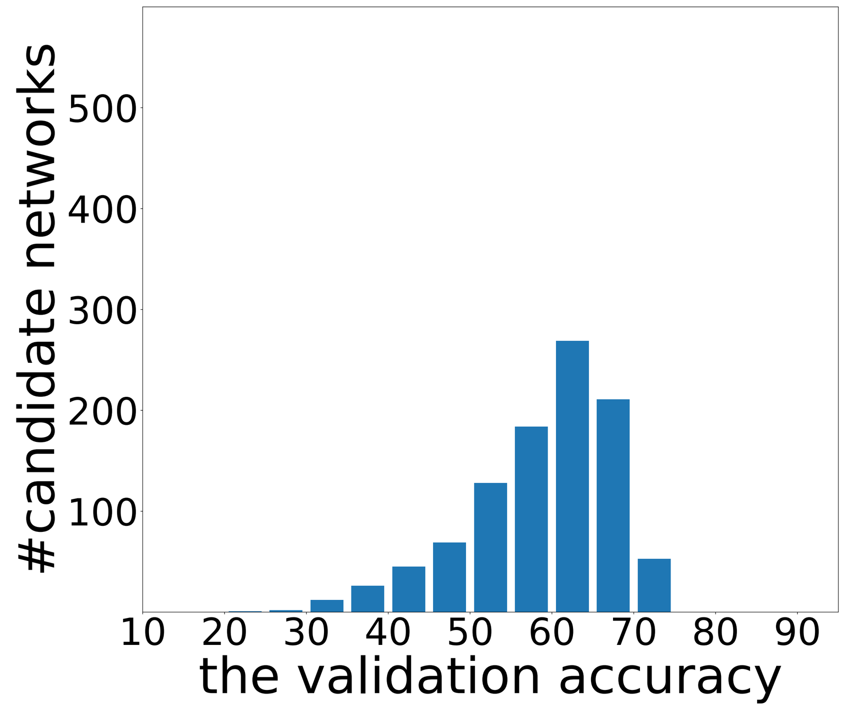


SETN

vs.



SETN without evaluator



SETN without uniformly sampling

CODES

x-axis: the accuracy on the validation set; y-axis: numbers of architectures

Method	GPU Days	M	C	Parameters	Error on CIFAR-10	Error on CIFAR-100
PyramidNet	–	–	–	26.0 MB	3.31	16.35
NASNet-A + CutOut	2000	20	32	3.3 MB	**2.65**	17.81†
PNAS + CutOut	150	11	48	3.2 MB	–	17.63
DARTS + CutOut	4	20	36	3.4 MB	2.83	–
GHN + CutOut	0.84	18	32	5.7 MB	2.84	–
ENAS + CutOut	0.45	20	36	4.6 MB	2.89	18.91†
GDAS + CutOut	0.84	20	36	3.4 MB	2.93	18.38
SETN-LR (T=1K) + CutOut	1.8	20	36	5.5 MB	2.81	17.88
SETN-NON (T=1K) + CutOut	1.8	20	36	3.7 MB	3.12	18.27
SETN (T=1) + CutOut	1.7	20	36	4.5 MB	3.41	18.12
SETN (T=1K) + CutOut	1.8	20	36	4.6 MB	2.69	**17.25**