motivation

Comparison between published search algorithms for NAS is therefore either very difficult

1)complex training protocol
2)different search spaces

contribution

Evaluation protocol.

1). Sample 8 architectures from the search space, uniformly at random, and use the method’s code to augment these architectures (same augment seed for all);
2). Use the method’s code to search for 8 architectures and augment them (different search seed, same augment seed)
3). Report mean and standard deviation of the top-1 test accuracy, obtained at the end of the augmentation, for both the randomly sampled and the searched architectures;

Relative Improvement(RI):
$RI = 100 \times (Accm - Accr)/Accr$

contribution

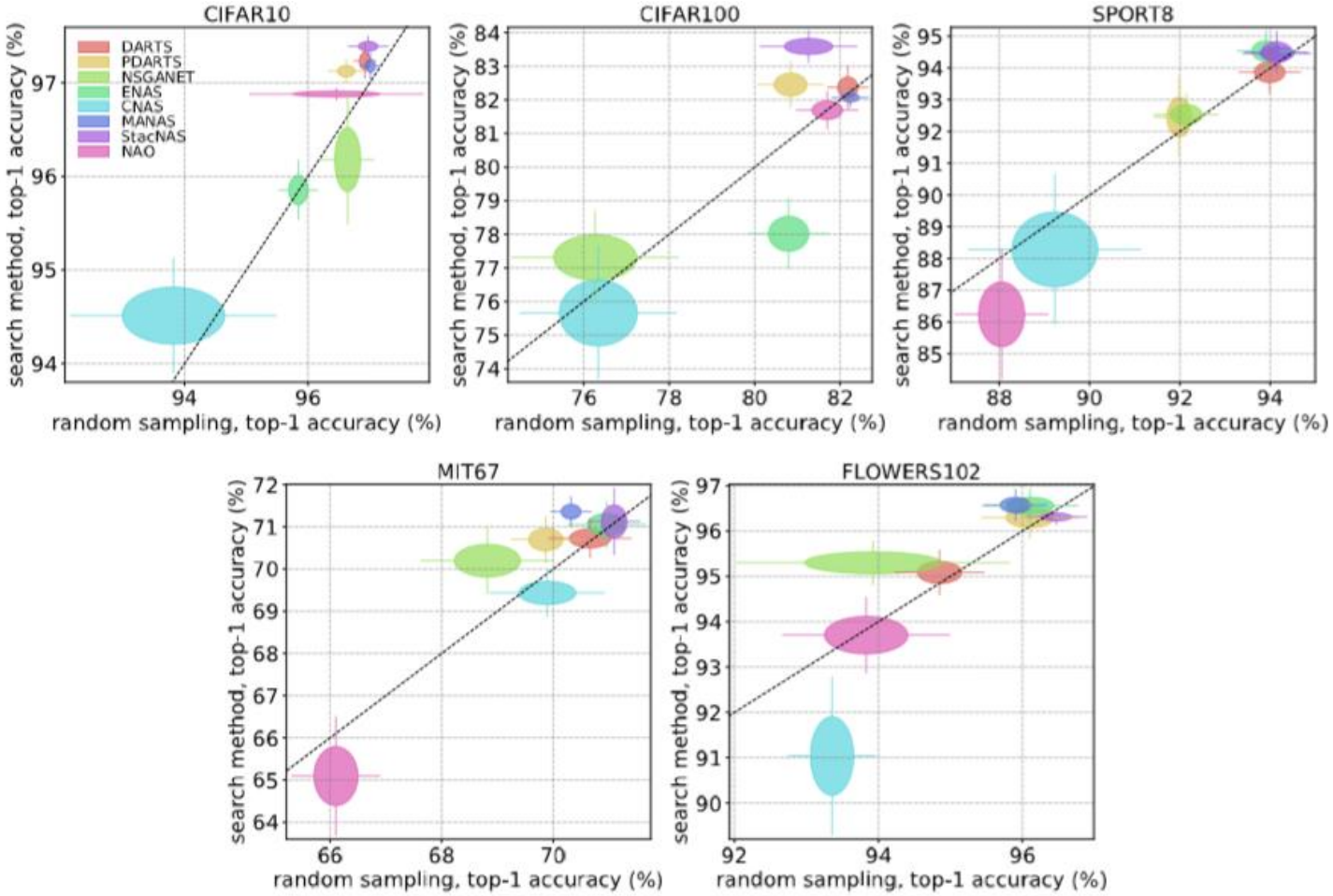


Figure 1: Comparison of search methods and random sampling from their respective search spaces. Methods lying in the diagonal perform the same as the average architecture, while methods above the diagonal outperform it. See also Table 1.

algorithm

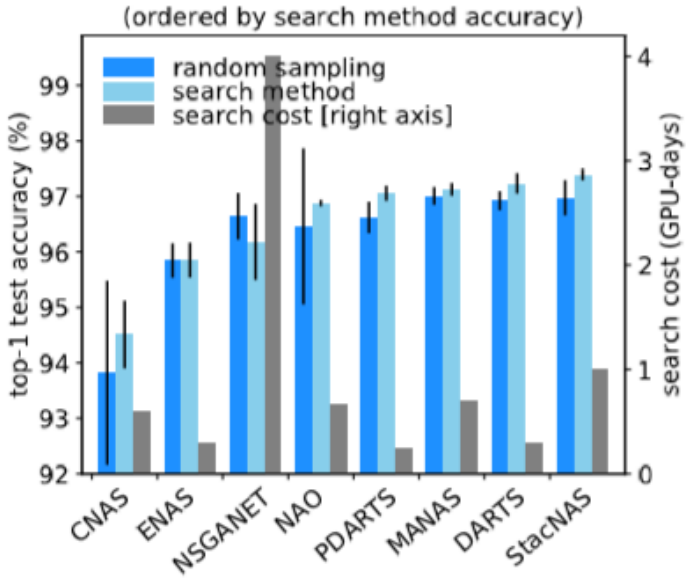


Figure 2: Performance and computational cost of the search phase on CIFAR10.

Table 1: Relative improvement metric, $RI = 100 \times (Acc_m - Acc_r)/Acc_r$ (in %), where Acc_m and Acc_r are the accuracies of the search method and random sampling baseline, respectively.

	C10	C100	S8	M67	F102
DARTS	0.32	0.23	-0.13	0.10	0.25
PDARTS	0.52	1.20	0.51	1.19	0.20
NSGANET	-0.48	1.37	0.43	2.00	1.47
ENAS	0.01	-3.44	0.67	0.13	0.47
CNAS	0.74	-0.89	-1.06	-0.66	-2.48
MANAS	0.18	-0.20	0.33	1.48	0.70
StacNAS	0.43	2.87	0.38	0.05	-0.16
NAO	0.44	-0.01	-2.05	-1.53	-0.13

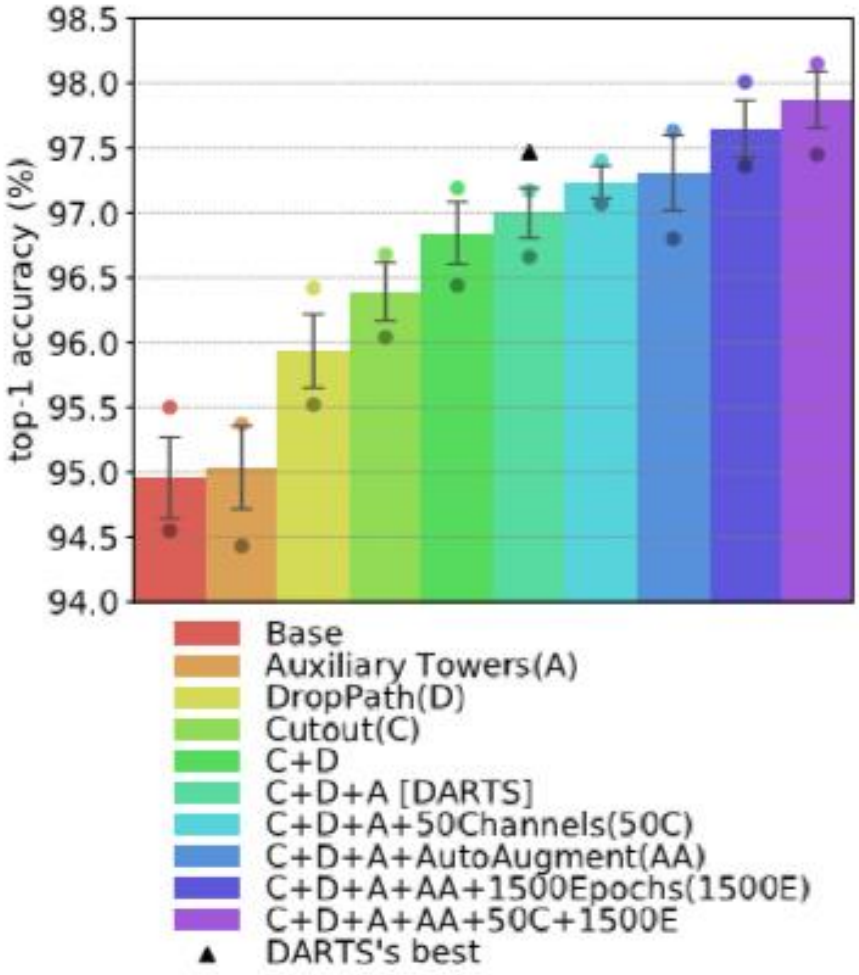
Training protocols

Method:

1) sample 8 random architectures on DARTS search space
2) Train them with different training protocols(details below)
3) Report mean, standard deviation and maximum of the top-1 test accuracy at the end of the training process.

training protocol:
Auxiliary Towers (A),
DropPath(D)
Cutout (C)
AutoAugment(AA)
extended training for 1500 epochs (1500E)
increased number of channels (50C).

Training protocols



Search space

Search space: DARTS
Number:200+
Trainning protocol:
Cutout+DropPath+Auxiliary Towers

Mean:97.03±0.2
Worts:96.18
Best:97.56

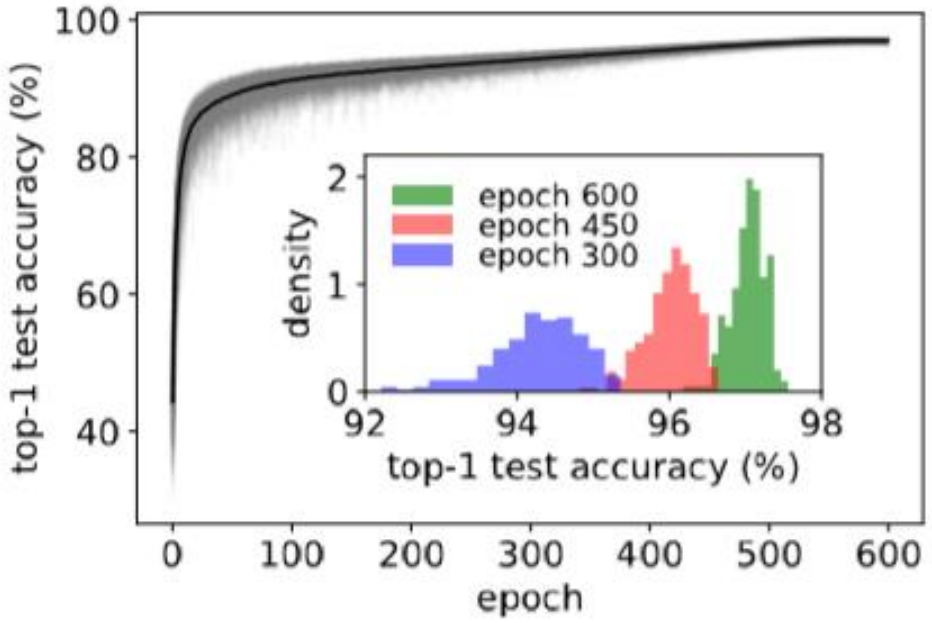


Figure 4: Training curves for the 214 randomly sampled architectures. Inset plot shows the histogram of accuracies at different epochs.

Search space

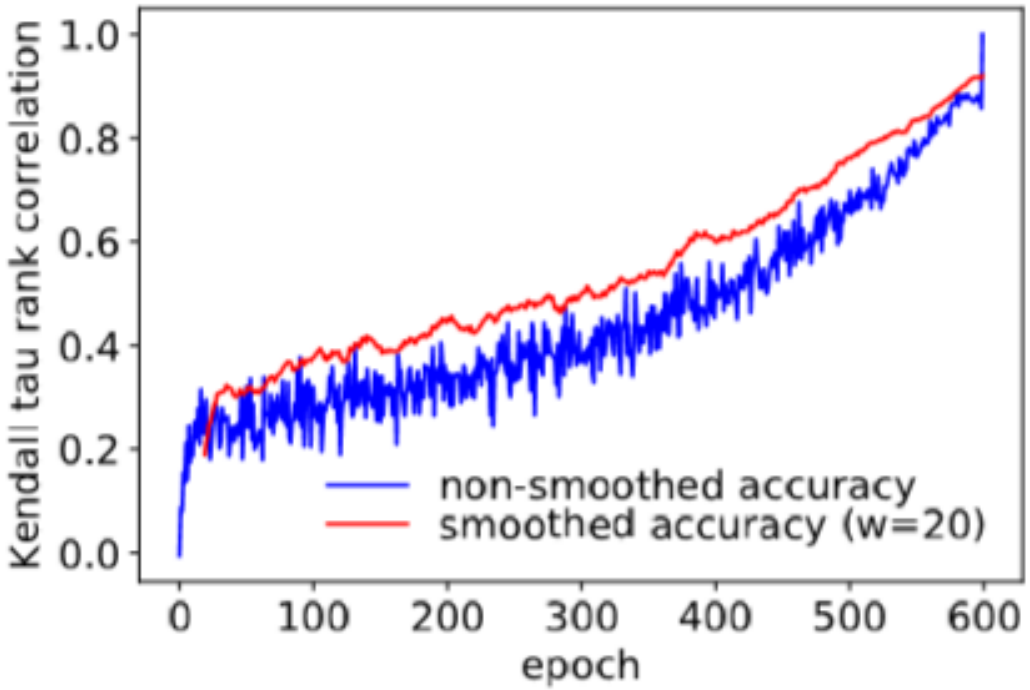


Figure 5: Correlation between accuracies at different epochs and final accuracy, using raw and smoothed accuracies over a window w.

Search space

operation

4Convolutions(1×1,3×3,7×7,
11×11)
2max pooling
operators (3×3, 5×5)
None
skip connect
operations

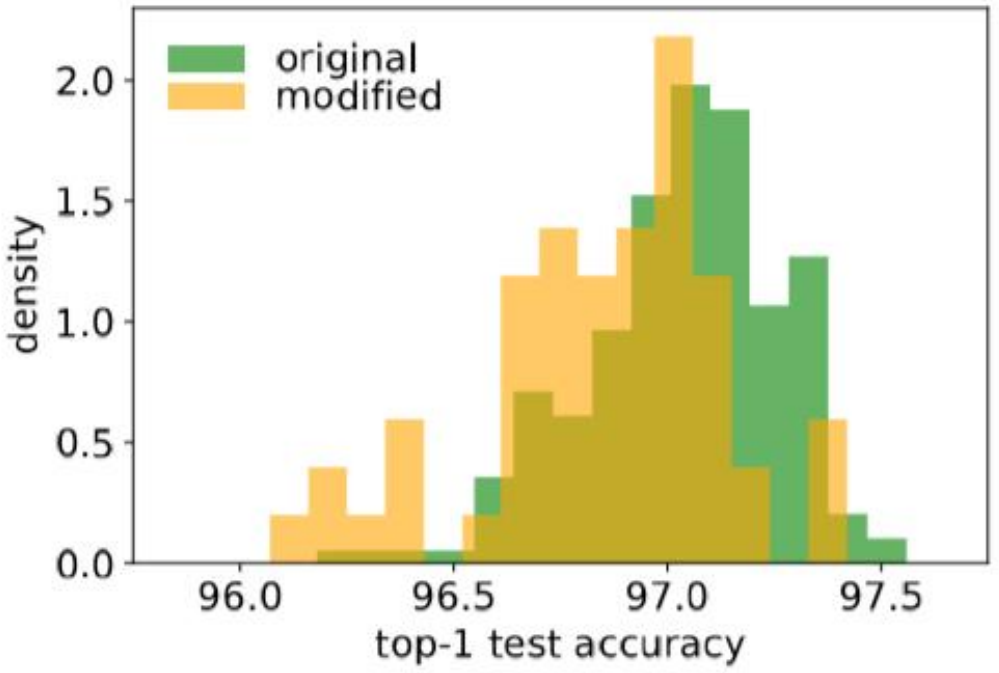


Figure 6: Histograms of the final accuracies (600 epochs) for architectures sampled from the DARTS search space (214 models) and our modified version (56 models).

Search space

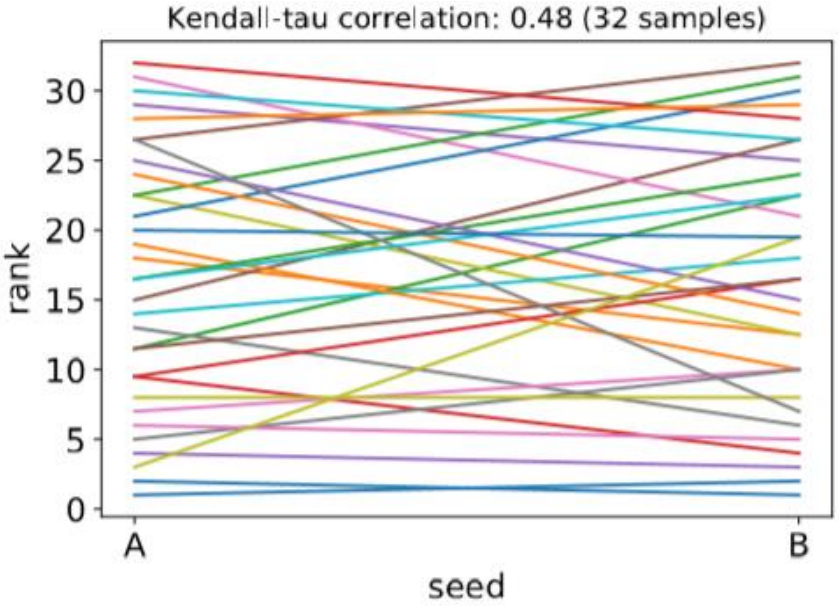


Figure 8: Changes in the ranking of different architecture when trained with two different seeds (A and B).

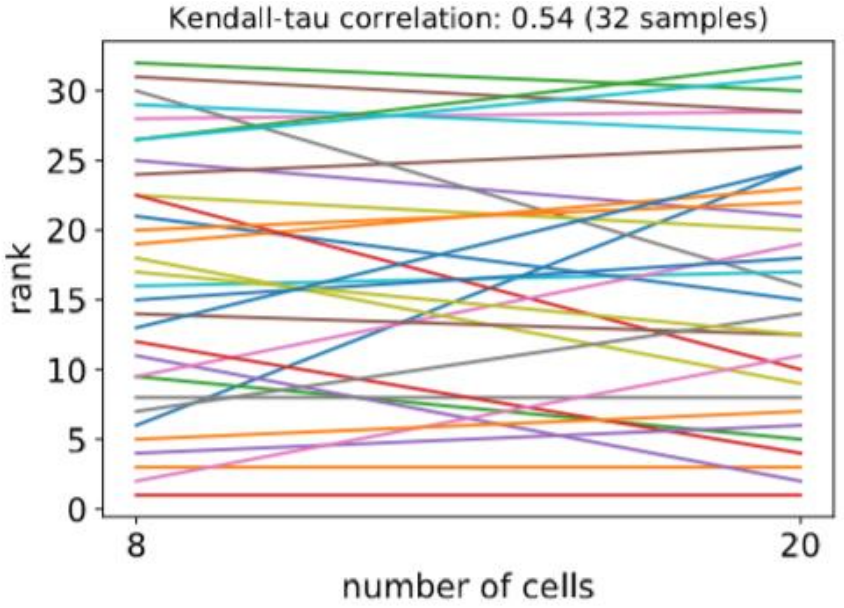


Figure 9: Changes in the ranking of different architecture when trained with different numbers of cells (same seed).

$0.13\% \pm 0.08$

Conlude

Augmention tricks:

We therefore suggest that both results, with and without training tricks, should be reported

Search Space:

We hope that future works will attempt to develop more expressive search spaces, capable of producing both good and bad network designs

future research could investigate the optimal wiring at a global level

Multiple datasets:

The best solution for this is likely to test NAS algorithms on a battery of datasets

Investigating hidden components:

We suggest that proper ablation studies can lead to better understanding of the contributions of each element of the pipeline

Conlude

The importance of reproducibility:

Release corresponding seed，training protocol

NAS-Bench-101 (Ying et al., 2019), a dataset mapping architectures to their accuracy, can be extremely useful, as it allows the quality of search strategies to be assessed in isolation from other NAS components

Hyperparameter tuning cost:

we argue that either (i) hyperparameters are general enough so that they do not require tuning for further tasks, or (2) the cost is included in the search budge