FreeREA: Training-Free Evolution-based Architecture Search Supplementary

Niccolò Cavagnero, Luca Robbiano, Barbara Caputo, Giuseppe Averta
Politecnico di Torino, Italy
{niccolo.cavagnero, luca.robbiano, barbara.caputo, giuseppe.averta}@polito.it

1. Fitting Function Ablation

With the aim of assessing the contribution of different terms composing our fitting function, we performed an ablation study by removing each one of the three components (leave one out). The different configurations are tested on all the datasets from the two benchmarks taken into account, NATS-Bench [1] and NAS-Bench-101 [3]. We list in Table 1 the results achieved by our method with a subset of metrics.

Interestingly, only the combination of all the three metrics demonstrated to achieve the highest test accuracy with the lowest variance on all the considered benchmarks. Remarkably the metrics we proposed, Skipped Layers and LogSynflow, seem to be the most important among the three, while the contribution of Linear Regions is limited and appreciable on NAS-Bench-101 only, probably because its discrimination is more effective on larger search spaces.

2. Search Algorithm Hyper-parameters Ablation

In addition, we also investigated different choices for N and n, population size and tournament size, following the couples explored in the original REA [2]. Table 2 reports on the experimental results. Numerical values suggest that almost all the configurations yield similar performances, vouching for an high robustness of the proposed method over the choices of the hyper-parameters. Interestingly, only the configuration (100, 2) has achieved a significantly lower test accuracy, likely caused by the large difference between N and n. It is also worth reporting that, since we allow the generation of offspring from the top two candidates in the sample, with $n = 2$ Tournament Selection is not happening, while the algorithm is simply evolving random networks in the population. These two factors, while improving the exploration of the search space, severely hinder the exploitation capability of the algorithm leading to poor results.

As a rule of thumb, then, it is reasonable to set a lower bound for $\frac{n}{N}$ such that $\frac{n}{N} \geq 0.20$. It is worth remarking that higher value of n and N may result in higher generalisation and exploration capabilities of the algorithm, but may slow down the search. However, in our experiments all the configurations are run for 45 seconds and proved to converge to the same optimal model, thus suggesting that a longer convergence time is not affecting the performance of the method for the NATS-Bench search space.

References

[1] Xuanyi Dong, Lu Liu, Katarzyna Musial, and Bogdan Gabrys. Nats-bench: Benchmarking nas algorithms for architecture topology and size. *IEEE transactions on pattern analysis and machine intelligence*, 2021.

[2] Esteban Real, Alok Aggarwal, Yanping Huang, and Quoc V Le. Regularized evolution for image classifier architecture search. In *Proceedings of the aaai conference on artificial intelligence*, volume 33, pages 4780–4789, 2019.

[3] Chris Ying, Aaron Klein, Eric Christiansen, Esteban Real, Kevin Murphy, and Frank Hutter. Nas-bench-101: Towards reproducible neural architecture search. In *International Conference on Machine Learning*, pages 7105–7114. PMLR, 2019.

Table 1: Ablation Study for different metrics composing the fitting function. Test Accuracy for the final model is reported. Each configuration has been run 30 times with a time limit of 45 seconds on NATS-Bench and 12 minutes on NAS-Bench-101.

	NATS-Bench			NAS-Bench-101
	CIFAR10	CIFAR100	ImageNet16-120	CIFAR10
Baseline	**94.36 ± 0.00**	**73.51 ± 0.00**	**46.34 ± 0.00**	**93.80 ± 0.02**
w/o Linear Regions	**94.36 ± 0.00**	**73.51 ± 0.00**	**46.34 ± 0.00**	93.55 ± 0.49
w/o Skipped Layers	93.76 ± 0.00	71.11 ± 0.00	41.44 ± 0.00	91.90 ± 1.69
w/o LogSynflow	94.30 ± 0.00	71.13 ± 0.00	44.48 ± 0.00	92.62 ± 0.08

Table 2: Ablation Study for different choices of N and n on NATS-Bench [1]. Test Accuracy for the final model is reported. Each configuration has been run 30 times with a time limit of 45 seconds.

N, n	CIFAR10	CIFAR100	ImageNet16-120
25, 5	**94.36 ± 0.00**	**73.51 ± 0.00**	**46.34 ± 0.00**
100, 2	93.95 ± 0.14	71.8 ± 1.85	45.65 ± 2.57
100, 50	**94.36 ± 0.00**	**73.51 ± 0.00**	**46.34 ± 0.00**
20, 20	**94.36 ± 0.00**	**73.51 ± 0.00**	**46.34 ± 0.00**
100, 25	94.34 ± 0.01	73.50 ± 0.00	**46.34 ± 0.00**
64, 16	**94.36 ± 0.00**	**73.51 ± 0.00**	**46.34 ± 0.00**