Differential Evolution for Neural Architecture Search

Noor Awad, Neeratyoy Mallik and Frank Hutter

Albert-Ludwigs-Universität Freiburg

Summary

Contributions

• Standardizing and benchmarking differential evolution (DE) as a *search strategy* for neural architecture search (NAS).

• We demonstrate that our DE yields state-of-the-art performance for NAS, comparing favorably to regularized evolution (RE) and Bayesian optimization.

Observations

• DE yields improved and more robust results for 13 tabular NAS benchmarks based on NAS-Bench-101, NAS-Bench1Shot1, NAS-Bench-201 and NAS-HPO bench.

• DE shows strong *final performance,* compared to RE, BOHB.

• DE appears to be robust to high-dimensional spaces and handle mixed-data types adeptly.

Canonical DE

DE is an evolutionary algorithm that is based on four steps:

- **Initialization:** Initialize a population space of *NP* individuals
$$pop_g = (X^1_{i,g}, X^2_{i,g}, ..., X^D_{i,g}), i = 1, 2, ..., NP$$
- **Mutation:** A new child/offspring is produced
$$V_{i,g} = X_{r_1,g} + F \cdot (X_{r_2,g} - X_{r_3,g})$$
- **Crossover:** Combine target and mutant to generate a trial
$$u^j_{i,g} = \begin{cases} v^j_{i,g} & \text{if } (rand < Cr) \text{ or } (j = j_{rand}) \\ x^j_{i,g} & \text{otherwise} \end{cases}$$
- **Selection:** Evaluates trial and compares to keep or discard
$$X_{i,g} = \begin{cases} U_{i,g} & \text{if } (f(U_{i,g}) \leq f(X_{i,g})) \\ X_{i,g} & \text{otherwise} \end{cases}$$

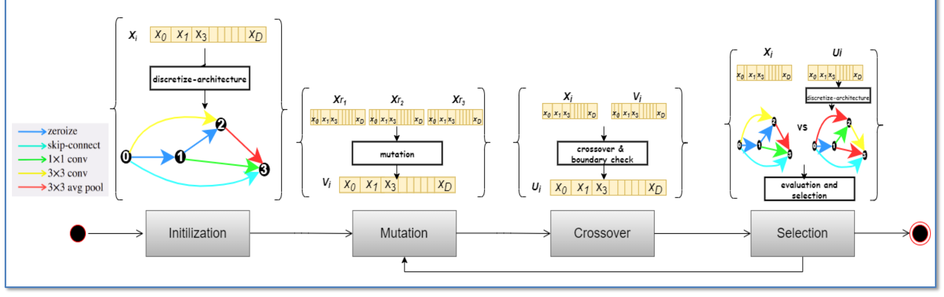
DE for NAS

• The figure below shows the general framework of our DE implementation for NAS.

• We scale all NAS parameters to [0, 1] to let DE work on individuals from a uniform, continuous space.

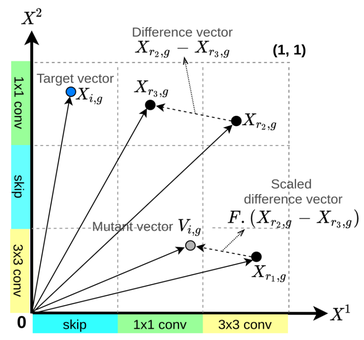
• We found the best way of applying DE when parameters are discrete or categorical is to keep the population in a continuous space, perform canonical DE, and only discretize copies of individuals to evaluate them.

• The mutation strategy selected was *rand1* and *binomial* crossover were selected as the DE strategies for this work.

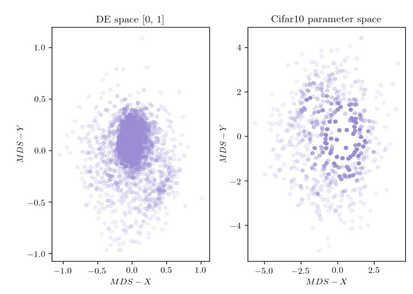


Parameter space mapping

- **Integer and float parameters**: $X_i \in [a_i, b_i]$ are retrieved as: $a_i + (b_i - a_i) \cdot U_{i,g}$, where the integer parameters are additionally rounded.
- **Ordinal and categorical parameters** $X_i \in \{x_1, ..., x_n\}$: the range [0, 1] is divided uniformly into n bins.

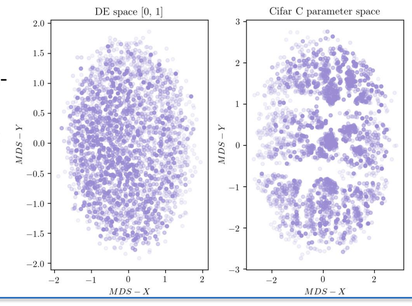


Search space visualization



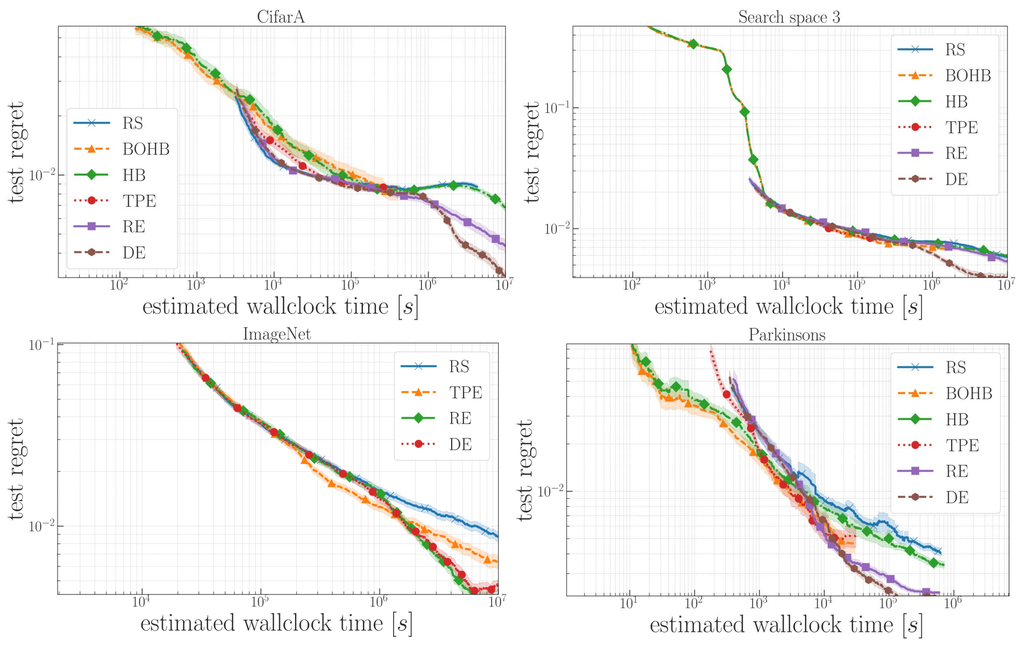
Multi-dimensional scaling (MDS) plots show the correspondence of the search trajectories between the DE space of [0, 1] and the original NAS parameter space.

Empty space artefacts may arise due to the non-linear mapping of MDS that tries to preserve the pairwise distances in the 2-dimensional projections for mixed parameter types.



Experiments

Benchmark Results



	NAS-Bench-101		
	CifarA	**CifarB**	**CifarC**
BOHB	0.0649 ± 0.00703	0.0648 ± 0.00203	0.065 ± 0.0023
RE	0.0612 ± 0.00342	0.0613 ± 0.00321	0.0637 ± 0.00378
DE	**0.0598 ± 0.00262**	**0.0611 ± 0.00225**	**0.0606 ± 0.00248**

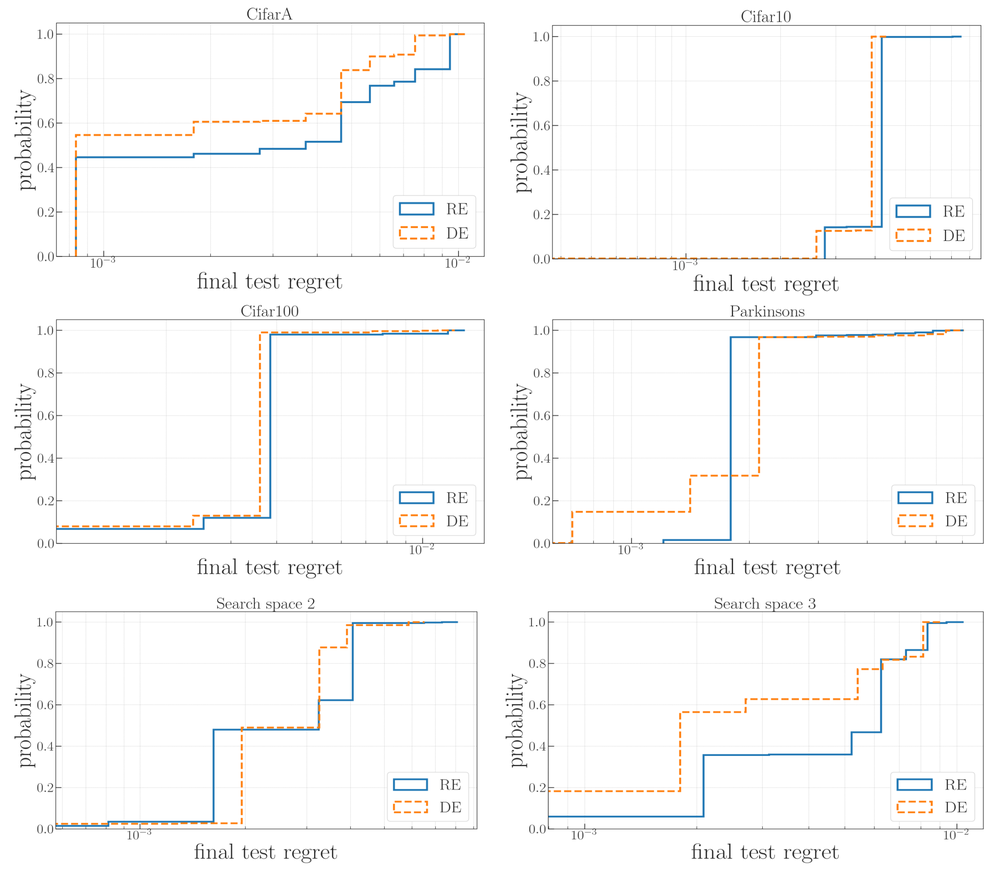
For NAS-101: DE is able to exploit high-dimensional spaces well and handle mixed-types better.

	NAS-Bench-1Shot1		
	Search space 1	**Search space 2**	**Search space 3**
BOHB	0.0599 ± 0.00271	0.0606 ± 0.00215	0.0602 ± 0.00213
RE	**0.0566 ± 0.00076**	0.0607 ± 0.00122	0.0588 ± 0.00261
DE	**0.0569 ± 0.00097**	0.0605 ± 0.00113	**0.0573 ± 0.00303**

For NAS-1shot1: For search space 3, the most complex, largest space, DE performs best and converges fastest.

	NAS-HPO			
	Protein	**Slice**	**Naval**	**Parkinsons**
BOHB	0.2208 ± 0.00446	0.00019 ± 6.82e-05	5.73e-05 ± 2.3e-04	0.0089 ± 0.00685
RE	**0.2155 ± 0.00028**	**0.00016 ± 2.06e-06**	3.59e-05 ± 5.62e-06	0.0065 ± 0.00056
DE	0.2156 ± 0.00048	**0.00016 ± 3.54e-06**	**3.58e-05 ± 3.81e-06**	**0.0064 ± 0.00078**

Robustness



• For NAS-Bench-101, DE is robust in solving CifarA and CifarC while RE is better in solving CifarB.

• For NAS-Bench-1Shot1, DE is more robust to solve the three search spaces while we can say that RE is competitive in search space 2.

• For NAS-201,RE is more robust than DE in ImageNet while DE is competitively robust to RE in Cifar10 and Cifar100.

• For NAS-HPO, DE shows more robust performance in Slice and Parkinsons datasets. For Protein and Naval datasets, DE is competitively robust to RE.

Implementation Publicly Available: https://github.com/automl/DE-NAS

AutoML.org

@automlfreiburg