NAS-Bench-301 and the Case for Surrogate Benchmarks for Neural Architecture Search
Julien Siems*, Lucas Zimmer*, Arber Zela, Jovita Lukasik, Margret Keuper, Frank Hutter

**Contributions**
- First surrogate NAS benchmark:
  - Allows cheap benchmarking of NAS optimizers
  - Covers realistic space (10^16 architectures)
- We demonstrate that surrogate benchmarks can outperform tabular benchmarks
- Dataset of 60k architecture evaluations
- New insights into performance of Local Search on DARTS search space.

**Existing NAS-Benchmarks**
- **NAS-Bench-101** [Ying et al. 2019]
  - Exhaustively evaluated cell-search space on CIFAR-10 for 4 epoch budgets.
  - ~423k unique architectures
- **NAS-Bench-201** [Dong and Yi 2020]
  - Cell-search space with fixed cell connectivity evaluated on CIFAR-10/CIFAR-100/TinyImageNet.
  - ~6k unique architectures

  Benchmarks too small: Even Random Search competitive

**Tabular vs. Surrogate**
- Architecture evaluations are noisy
- Fit surrogate model on NB-101
- Compare to tabular benchmark with one evaluation w.r.t. MAE to remaining evaluations
  - Surrogate smoothes out noise
  - Surrogate yields strong predictive performance, even when trained on subset

**NAS-Bench-301 - Dataset**
- **Data Collection**
  - Objective: Cover space efficiently
- **Data Visualization**
  - Public Dataset: ~60k
  - Good coverage
  - Wide performance range

**NAS-Bench-301 - Surrogate Models**
- **Data Fit**
  - Model: Test
  - MAE
  - $\mu$-SVR: 0.709
  - MLP (Path cnc.): 0.704
  - RF: 0.679
  - c-SVR: 0.675
- **Noise Modelling**
  - Deep ensembles
  - Evaluate 500 architectures with 5 different seeds
  - Reproduce results from Tabular vs. Surrogate experiment

**NAS-Bench-301 - Benchmark**
- **Blackbox Optimizers**
- **One-Shot Optimizers**

**NAS-Bench-301 - Case Study**
- **Local Search result by White et al. 2020**
  - Not competitive on DART Search Sp.
  - Tested with small compute budget
  - Fast NB-301 analysis suggests it is competitive.
  - LS-GIN, LS-XGB
  - Verified by extensive empirical evaluation
  - LS-GT

  Demonstration of how NAS-Bench-301 can be used to cheaply check research hypothesis.

**References**

Benchmark Publicly Available: https://github.com/automl/nasbench301