HPOBench: A Collection of Reproducible Multi-Fidelity Benchmark Problems for HPO

Katharina Eggensperger

Philipp Müller

Neeratyoy Mallik

Matthias Feurer

René Sass

Aaron Klein

Noor Awad

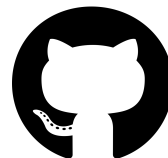Marius Lindauer

Frank Hutter

/automl/HPOBench

AutoML.org

UNI FREIBURG

Leibniz Universität Hannover

Available data is growing.

Applications of ML are growing.

Model size and complexity is growing.

We need efficient hyperparameter optimization methods!
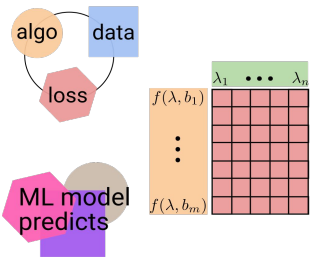
→ Multi-fidelity optimization

BUT

To **develop**, **improve**, **understand** and **compare** methods we need benchmark problems that are **realistic**, **efficient** and **available for a long time**.
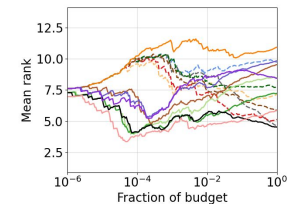
→ HPOBench

- The first collection of **containerized multi-fidelity** HPO benchmarks with **100+ benchmark problems**

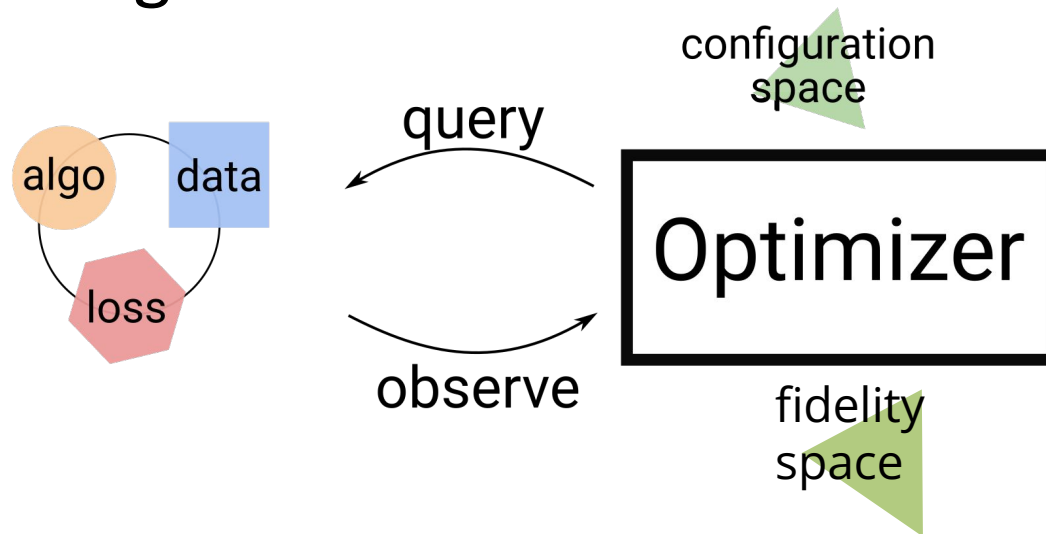Photo by frank mckenna on Unsplash

- The first set of HPO benchmarks that
  - are available as **raw and tabular** versions
  - which also support **multi-objective optimization** and **transfer-HPO across datasets**

- An exemplary **large-scale study** evaluating >10 optimization methods on all benchmarks

Benchmark ingredients:

configuration space

algo   data

loss

query

observe

Optimizer

fidelity space

Ideal features:

Efficiency

Reproducibility

Flexibility

raw
benchmark

tabular
benchmark

$f(\lambda, b_1)$
λ_1 $\bullet\bullet\bullet$ λ_n
$\vdots$
$f(\lambda, b_m)$

surrogate
benchmark

ML model
predicts

configuration
space

query

Optimizer

observe

fidelity
space

Unify interface

Collect
dependencies

Package
everything
in a container
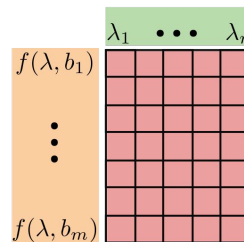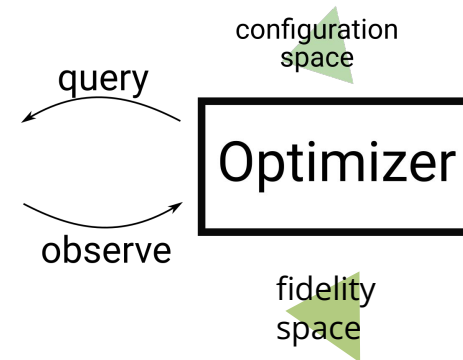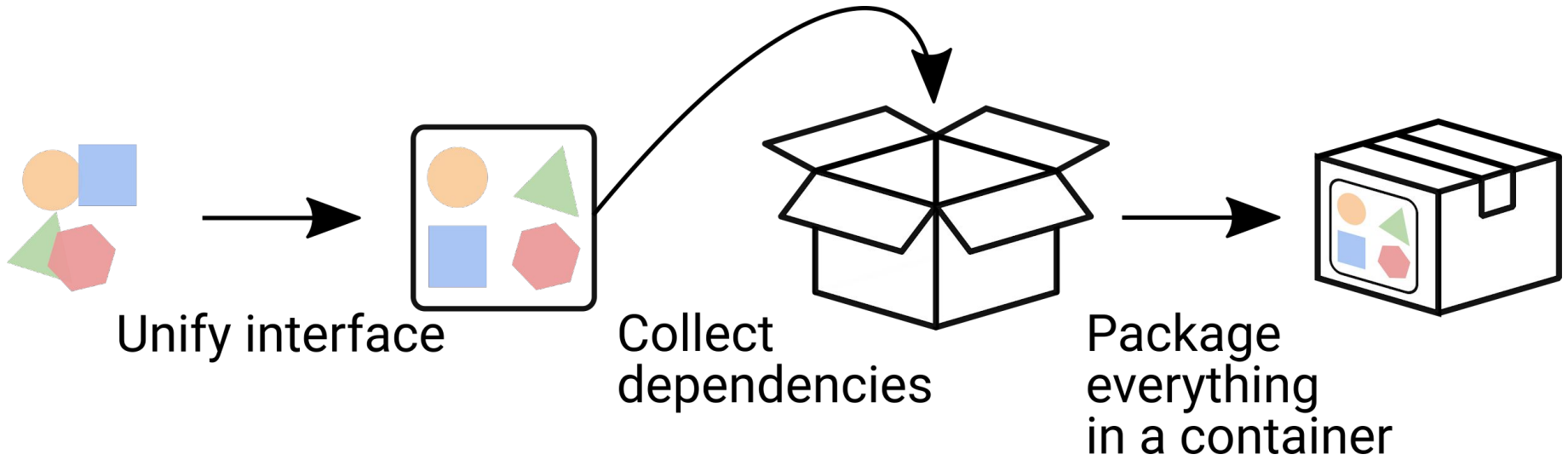
Family	#benchs	#cont(log)	#int(log)	#cat	#ord	fidelity	type
Cartpole	1	4(1)	3(3)	-	-	repetitions	✓
BNN	2	3(1)	2(2)	-	-	samples	✓
Net	6	5	1	-	-	time	✗
NBHPO	4	-	-	3	6	epochs	(✓)
NB101	3	-	-	26	-	epochs	(✓)
		-	-	14	-		
		21	1	5	-		
NB201	3	-	-	6	-	epochs	(✓)
NB1Shot1	3	-	-	9	-	epochs	(✓)
		-	-	9	-		
		-	-	11	-		
LogReg	20	2(2)	-	-	-	iter	✓, (✓)
SVM	20	2(2)	-	-	-	data	✓, (✓)
RandomForest	20	1	3(2)	-	-	#trees	✓, (✓)
XGBoost	20	3(2)	1(1)	-	-	#trees	✓, (✓)
MLP	8	2(2)	3(2)	-	-	epochs	✓, (✓)

✓	raw
(✓)	tabular
✗	surrogate

We ran >10 optimization methods on all benchmarks
and studied the following:

1. Do advanced methods improve
    over random baselines?

2. Do multi-fidelity methods
    improve over single-fidelity
    methods

→ Short answer: **Yes**
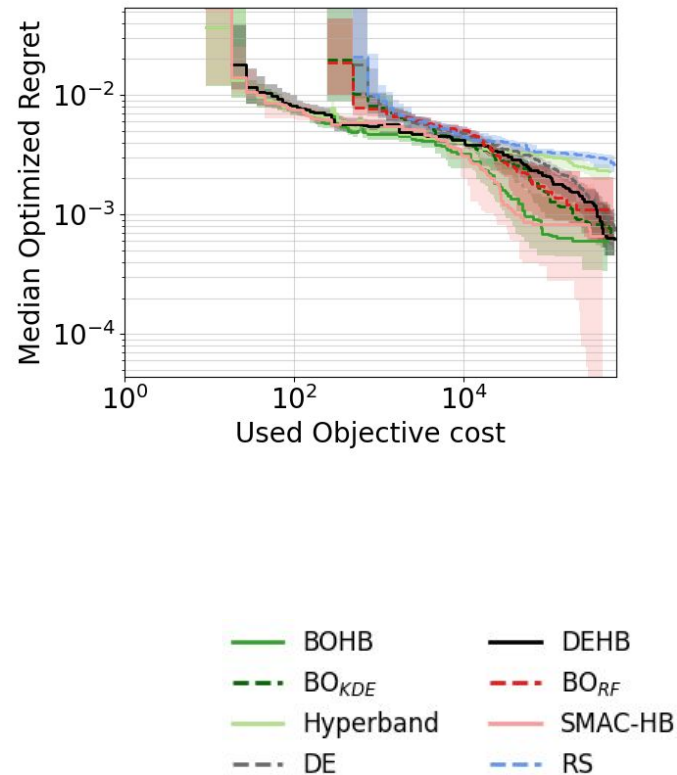
$\rightarrow$ **HPOBench** provides **>100 containerized benchmarks** for **multi-fidelity HPO**

What else you can do with HPOBench:

- multi-objective optimization and transfer-HPO across datasets
- compare raw, tabular and surrogate benchmarks
- ...

Thank you!

🐦 /AutoML_org
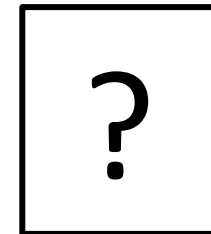
⬛ /automl/HPOBench

Katharina Eggensperger[1]

Philipp Müller[1]

Neeratyoy Mallik[1]

Matthias Feurer[1]

?

You

René Sass[2]

Aaron Klein[3]

Noor Awad[1]

Marius Lindauer[2]

Frank Hutter[1,4]

[1] Albert-Ludwigs-Universität Freiburg
[2] Leibniz Universität Hannover
[3] Amazon (work done prior to joining Amazon)
[4] Bosch Center for Artificial Intelligence