Zero-shot AutoML with Pretrained Models

Ekrem Öztürk*[1], Fabio Ferreira*[1], Hadi S. Jomaa*[2], Lars-Schmidt Thieme[2], Josif Grabocka[1], Frank Hutter[1,3]
*equal contribution, [1]University of Freiburg, [2]University of Hildesheim, [3]Bosch Center for Artificial Intelligence

UNI FREIBURG

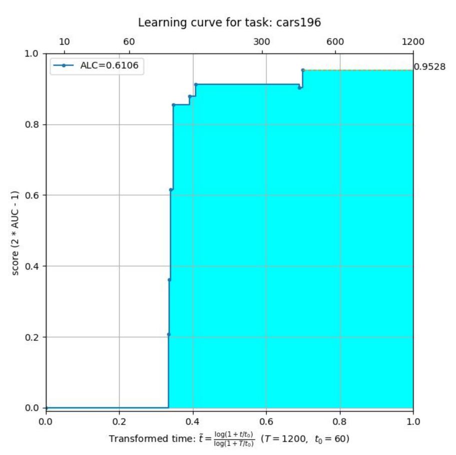
In a Nutshell

- Given a new dataset D, **how should we choose a pre-trained model** to fine-tune to D, **and set the fine-tuning hyperparameters?**
  - We **meta-learn a domain-independent zero-shot surrogate model** that solves this task given only **trivial meta-features.**
- We **release a large DL meta-dataset for image classification** that is over **1000 times larger** than previous meta-datasets.
- We evaluate our approach on the ChaLearn **AutoDL challenge benchmark**, clearly **outperforming all challenge contenders**.

Anytime Performance Setting

Model initialization: 20min budget, no dataset access
Training & prediction: 20min budget starting after initialization
Evaluation metric: area under the learning curve (ALC)

Evaluation starts at training time 0, i.e. **probing is very costly**. The model needs to start making predictions as early as possible without sacrificing performance to maximize the turquoise area.



Related Work

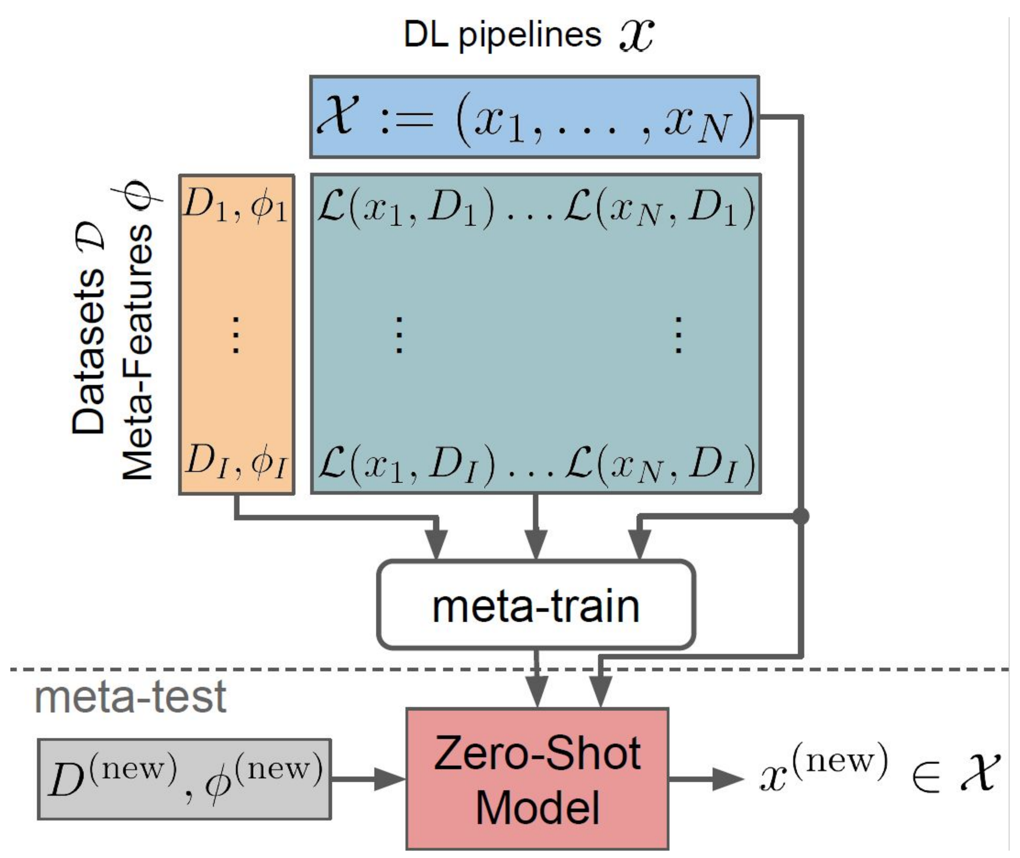
To solve the **algorithm selection** problem (Rice, 1976), we use **AutoFolio** (Lindauer et al., 2015) that chooses between regression, classification, etc. using algorithm configuration (Hutter et al., 2011). **Transfer HPO** (Wistuba & Grabocka, 2021; Jomaa et al., 2021a; Salinas et al., 2020) leverages prior knowledge with few observations on the target dataset, whereas **zero-shot HPO** (Wistuba et al., 2015; Winkelmolen et al., 2020) requires no observations on the target dataset. Related to our work, **XAS** (Tornede et al., 2020) describes datasets and pipelines as joint feature vectors, and **RankML** (Laadan et al.) uses dataset meta-features to predict the ranking of ML pipelines.

Problem Definition and Methodology

Given a set of N DL pipelines $\mathcal{X} := \{x_n\}_{n=1}^{N}$ and a collection of I datasets $\mathcal{D} = \{D_i\}_{i=1}^{I}$ with meta-features ϕ_i for dataset $D_i \in \mathcal{D}$, and a $N \times I$ matrix of costs $C(x_n, D_i)$ representing the cost of pipeline x_n on dataset D_i, the problem of zero-shot AutoML with pretrained models (ZAP) is to find a mapping $f : \Phi \rightarrow \mathcal{X}$ that yields minimal expected cost over $\mathcal{D}$:

$$argmin_f \; \mathbb{E}_{i \sim \{1,\ldots,I\}} \left[C(f(\phi_i), D_i) \right]$$

- **ZAP-AS:** Formulate the problem as algorithm selection and employ *AutoFolio* as the zero-shot model to meta-learn over the the cost matrix.
- **ZAP-HPO**
  - Project the DL pipelines into the geometric space $\mathcal{X} \subseteq \mathcal{M} \times \Theta$
  - Employ a parametric model f_ψ
    $$f(\psi)_{i,j} := f(x_j, \phi_i; \psi) : \mathcal{M} \times \Theta \times \Phi \rightarrow \mathbb{R}_+$$
  - Meta-learn f_ψ over the cost matrix with the ranking loss objective below
    $$argmin_\psi \sum_{(i,j,k) \in \mathcal{E}} \log\left(\sigma\left(f(\psi)_{i,j} - f(\psi)_{i,k}\right)\right)$$

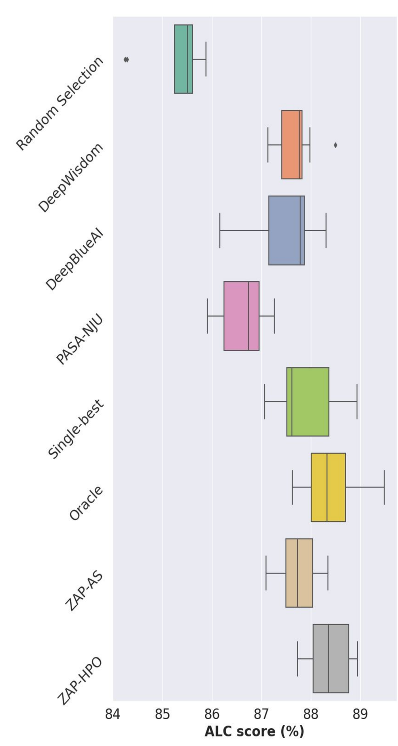


Evaluation Protocol & Experiments

- Baselines: **Random Selection** among the pipelines, **Single-best** pipeline on average across the datasets, **Oracle** pipelines which are the best ones per dataset, and the ChaLearn AutoDL challenge top-3 solutions **DeepWisdom**, **DeepBlueAI**, **PASA-NJU**.
- ZAP benchmark evaluation: 35-fold cross-validation over 525 datasets (meta-train on 510, test on the remaining 15).
- AutoDL benchmark evaluation: Meta-train on the whole meta-dataset and evaluate on the official platform.
- All the results are averaged over 10 repetitions.

Solution	Augmentation	ML technique
DeepWisdom	FAA	ResNet18/9 Meta-trained solution agents
DeepBlueAI	FAA	ResNet18 Adaptive ensemble learning
PASA NJU	Simple	ResNet18/SeResnext50 Data adaptive preprocessing

FAA: Fast AutoAugment (Lim et al., 2019)



ZAP benchmark
525 datasets

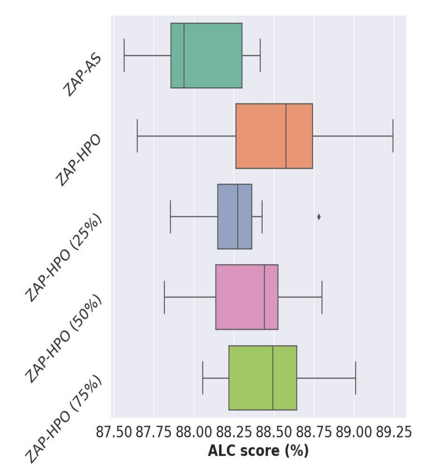
Solution	Rank (ALC)	Rank(Normalized AUC)
DeepWisdom	2.53 ± 0.06	2.63 ± 0.03
DeepBlueAI	2.64 ± 0.04	2.73 ± 0.03
PASA NJU	2.76 ± 0.03	2.56 ± 0.03
ZAP-HPO	$\mathbf{2.07 \pm 0.05}$	$\mathbf{2.08 \pm 0.03}$

Solution	Rank (ALC)
DeepWisdom	2.46 ± 0.13
DeepBlueAI	2.76 ± 0.08
PASA NJU	2.62 ± 0.11
ZAP-HPO	$\mathbf{2.16 \pm 0.15}$

AutoDL benchmark
5 datasets

Ablation study: Meta-learning from sparsely filled cost-matrix also outperforms the competition winners.

Solution	75% filled	50% filled	25% filled
DeepWisdom	2.52 ± 0.06	2.52 ± 0.05	2.52 ± 0.05
DeepBlueAI	2.64 ± 0.05	2.64 ± 0.04	2.63 ± 0.04
PASA NJU	2.75 ± 0.03	2.75 ± 0.04	2.75 ± 0.04
ZAP-HPO (sparse)	$\mathbf{2.09 \pm 0.06}$	$\mathbf{2.09 \pm 0.05}$	$\mathbf{2.10 \pm 0.04}$



AutoML.org

@automlfreiburg

Implementation Publicly Available: `bit.ly/ZAPhpo`