Neural Architecture Search: Mapping the field via network text analysis and my journey through the field

Petra Vidnerová

Institute of Computer Science, The Czech Academy of Sciences

May 26, 2025

Outline

Introduction

- Neural Architecture Search (NAS)

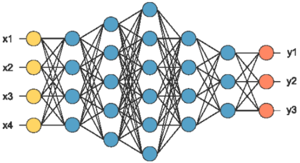
Overview of the field

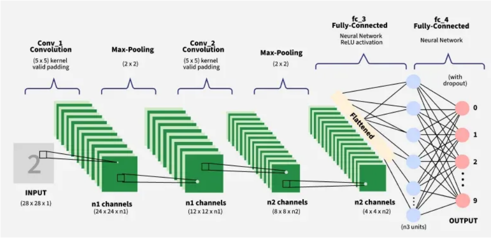
- Network text analysis applied to NAS papers

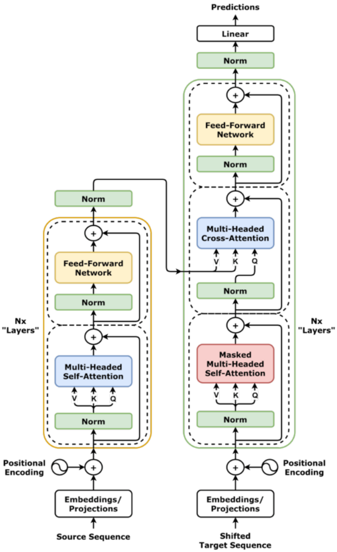
My results in the NAS field

- Evolutionary NAS approaches and Bayesian optimisation
- Performance Prediction

(Deep) Neural Networks





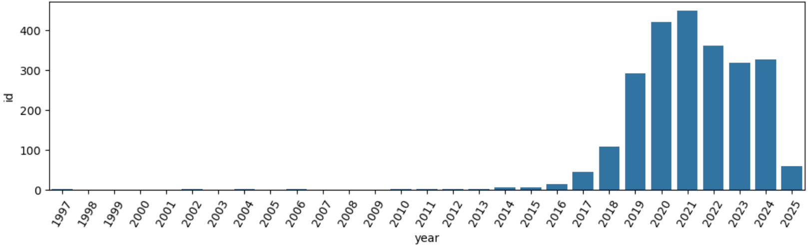


Neural Architecture Search

Neural Architecture Search (NAS)

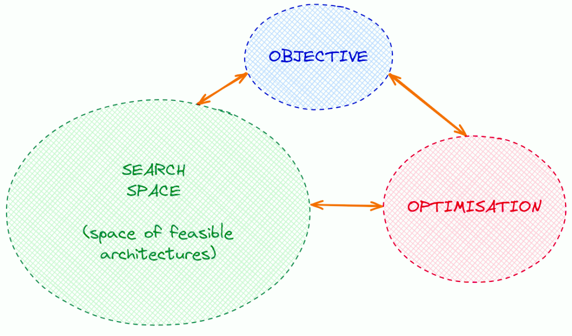
Automating the design of neural network architecture. Given a problem, NAS looks for an optimal architecture.

Research in NAS



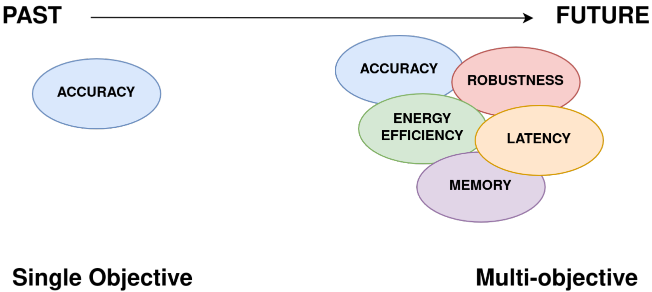
NAS – Optimisation Problem

- Minimize given objectives over the given search space



Objectives

- Measure the quality of a solution



Text Analysis

Dataset

- 10000 papers downloaded with the search query `Neural Architecture Search`, filtered using LLM
- 2 423 papers on NAS from ArXiV

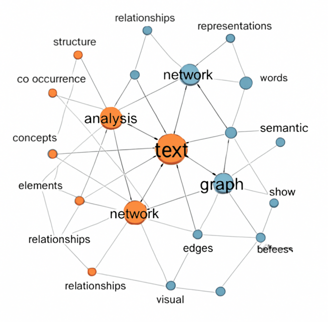
Keywords

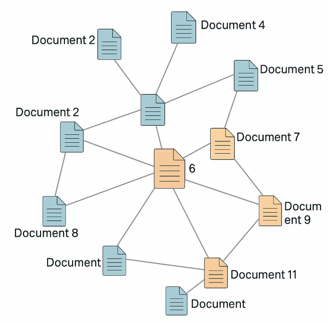
- Fixed set of keywords - documents classified based on abstract using LLM: `reinforcement learning, evolutionary algorithms, bayesian optimisation, multi-objective optimisation, supernet, weight sharing, differentiable optimisation, zero cost proxies or training free,hardware aware search, surrogate models`
- LLM asked to generate 5 keywords given an paper abstract

Network Text Analysis

Networks

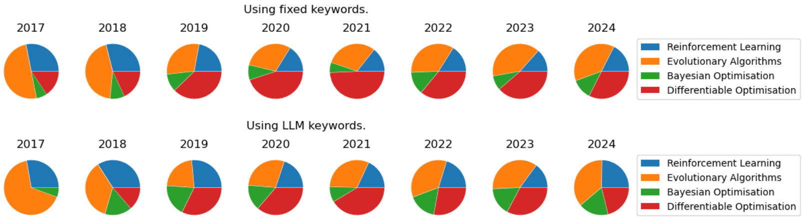
- Graphs (nodes, edges)
- Nodes documents, edges relationships
  - Citation networks, networks based on shared concepts, keywords
- Nodes keywords, edges shared documents



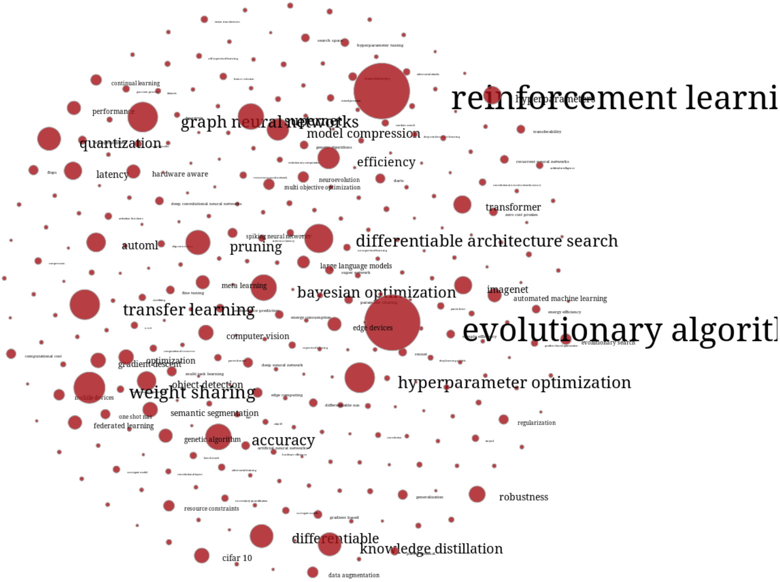


Optimisation

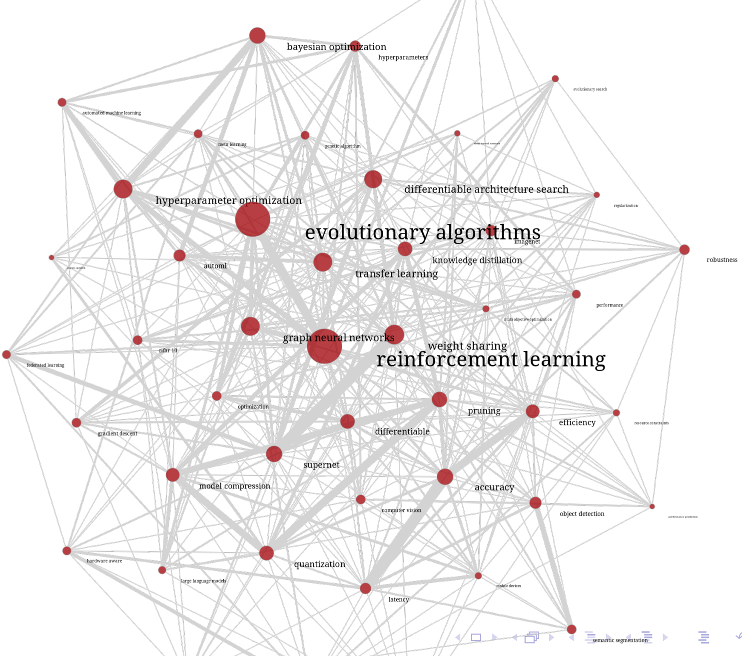
- Random search (baseline)
- Evolutionary and genetic algorithms
- Bayesian optimisation
- Reinforcement learning
- Differentiable techniques



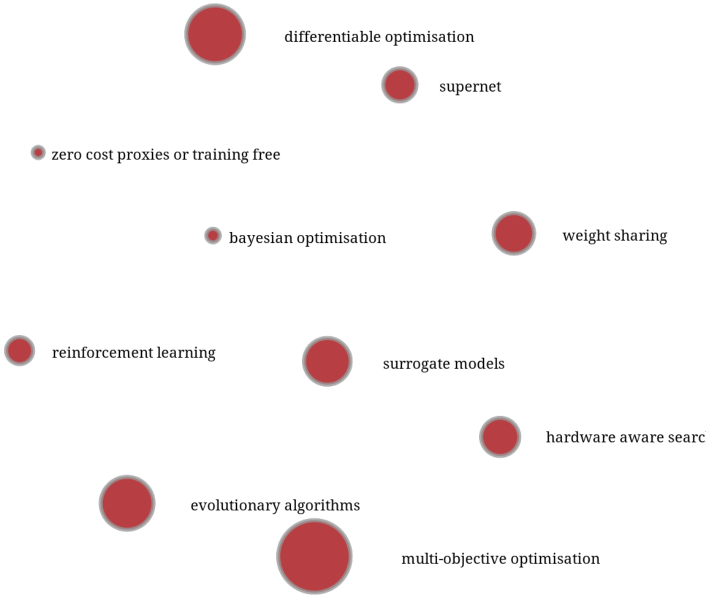
NAS Keywords by LLM



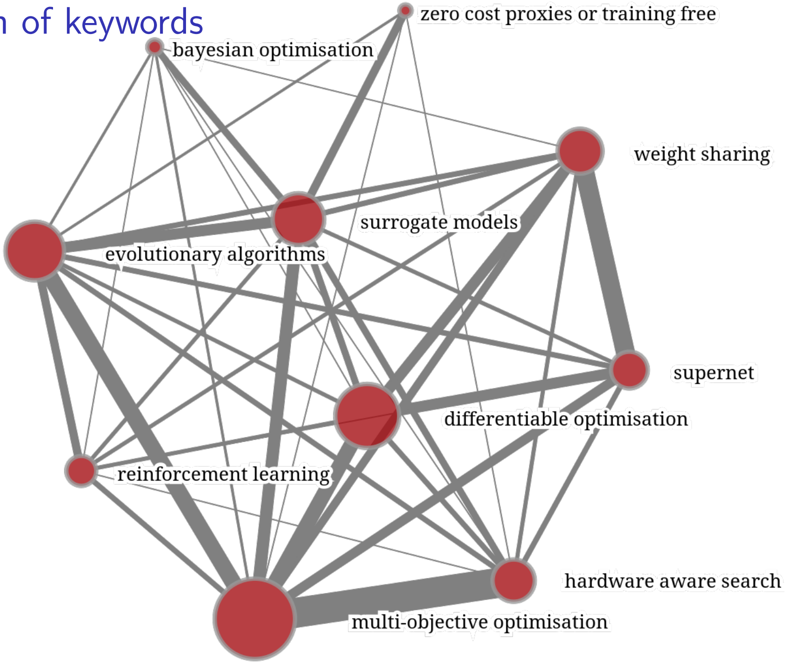
Graph of NAS Keywords by LLM



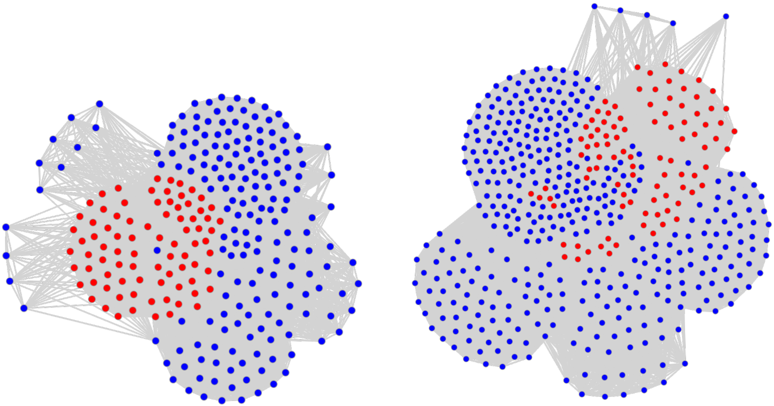
Fixed Selection of Keywords



Graph of keywords



Paper Graph

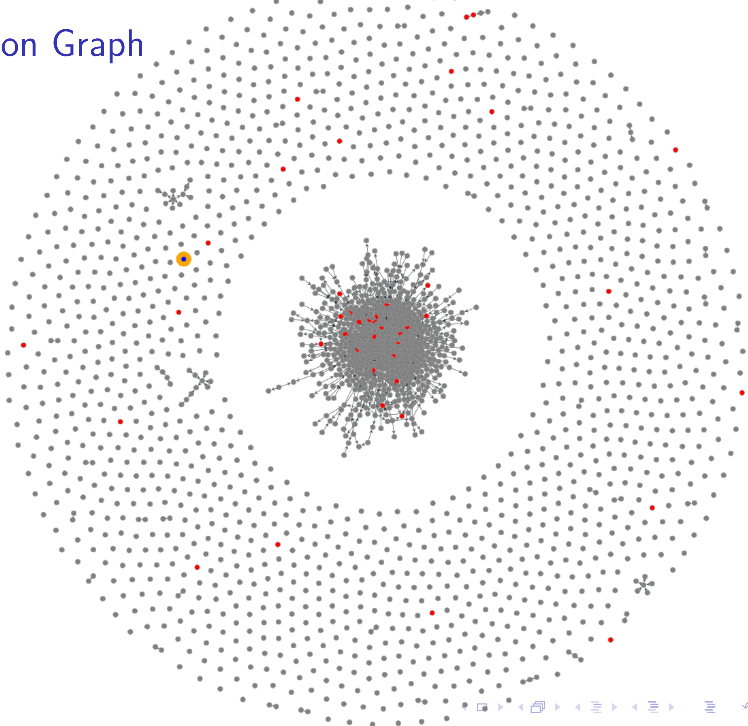


(left) Evolutionary papers

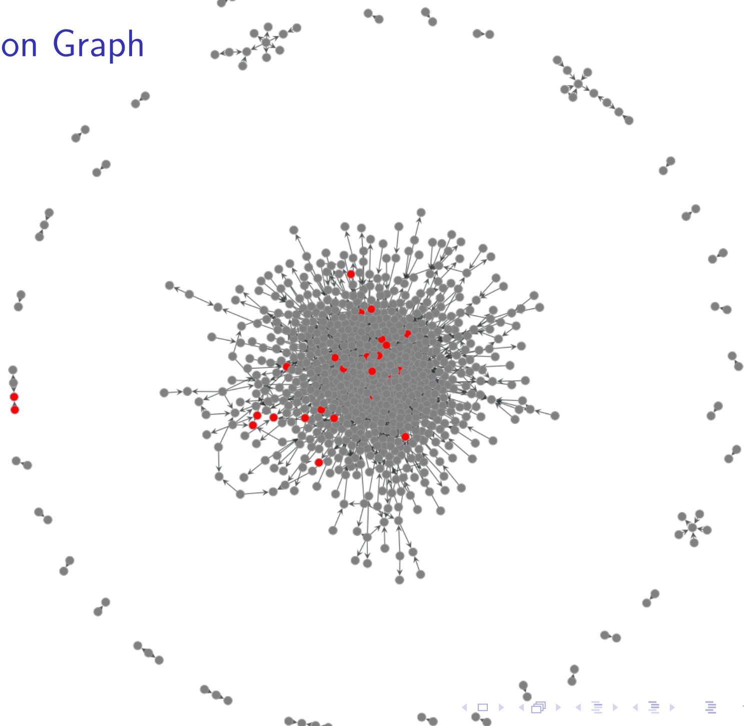
(right) Multi-objective optimisation papers

(red) Surrogate modelling papers (19% of EA, 15% Multi Obj.)

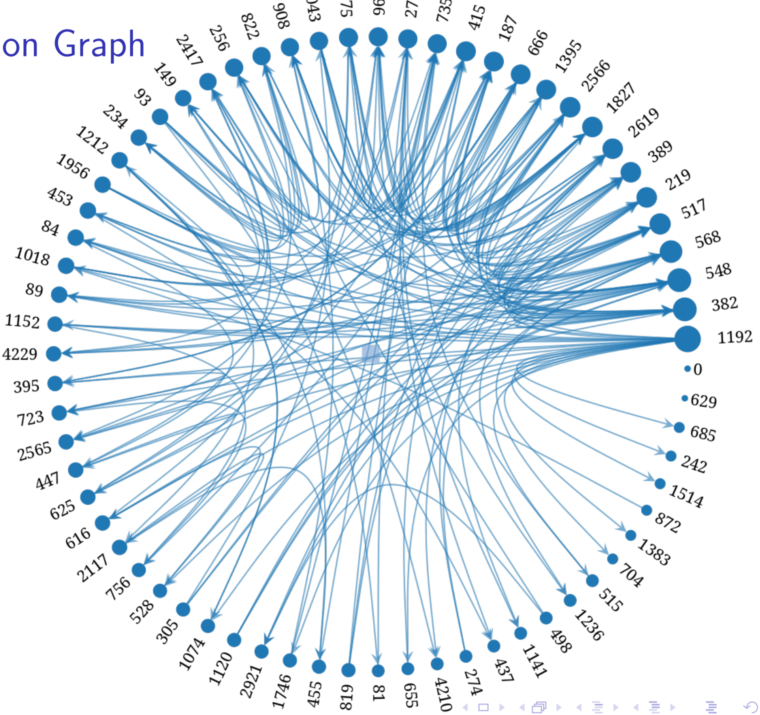
Citation Graph



Citation Graph



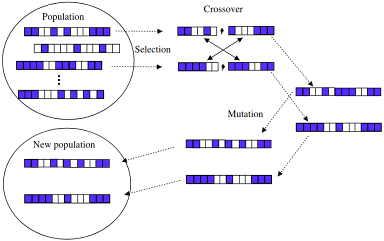
Citation Graph

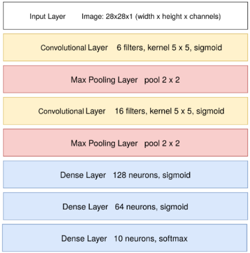


Our Work

Generated by AI.

Evolution of Deep Neural Networks





- Tested on tabular data (air pollution data set)
- Image datasets (MNIST, fashion-MNIST)
- Outperformed fixed networks and SVR

📖 P. Vidnerová, R. Neruda, *Evolving Keras Architectures for Sensor Data Analysis*, FedCSIS 2017

📖 P. Vidnerová, R. Neruda, *Evolution Strategies for Deep Neural Network Models Design*, ITAT 2017 🏆

Asynchronous Evolution

Parallelisation

- Parallel computation
- Individuals evaluated one by one
- No notion of generation
- As soon as there is an idle processor, new individual is created
- An arbitrary number of processors
- Slightly prefers smaller networks

📖 P. Vidnerová, R. Neruda, *Asynchronous Evolution of Convolutional Networks*, ITAT 2018

Challenges

- We still need penalise large networks → multi-objective design

Multi-objective approaches

Towards small networks

- Multi-objective evolution, NSGAII
- Found more compact networks with comparable accuracy

📖 P. Vidnerová, Š. Procházka, R. Neruda, *Multiobjective Evolution for Convolutional Neural Network Architecture Search*, ICAISC 2020

📖 P. Vidnerová, R. Neruda, *Multi-objective Evolution for Deep Neural Network Architecture Search*, ICONIP 2020

Robustness

- Robustness against outliers, noise, adversarial examples

📖 P. Vidnerová, R. Neruda, *Vulnerability of classifiers to evolutionary generated adversarial examples*, Neural Networks 2020

📖 J. Kalina, A. Neoral, P. Vidnerová, *Effective Automatic Method Selection for Nonlinear Regression Modeling*, IJNS 2021

Towards Performance Prediction

NAS Computational Cost

- Evaluation of most objectives requires network training
- Each candidate network needs to be evaluated

Solutions

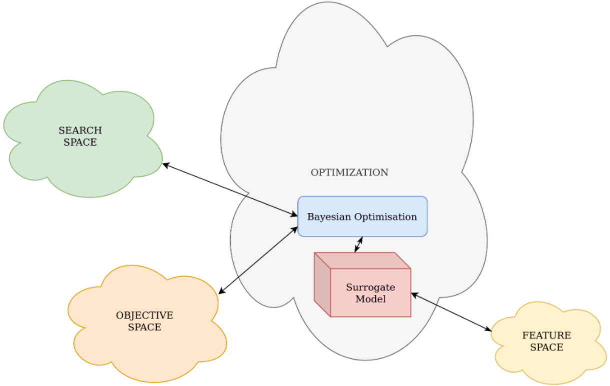
- Parallelisation - reduces time, but not cost
- Surrogate models, performance prediction

Performance prediction

- Imprecise prediction is enough (coarse to grain)
- Ranking is enough (who is the best)

Bayesian Optimisation

📖 P. Vidnerová, J. Kalina, *Multi-objective Bayesian Optimization for Neural Architecture Search*, ICAISC 2022



Performance Prediction

Zero-cost proxies

- Fast to compute metrics that correlate with accuracy
- Zero-cost ... because we don't train the network at all!

Our work

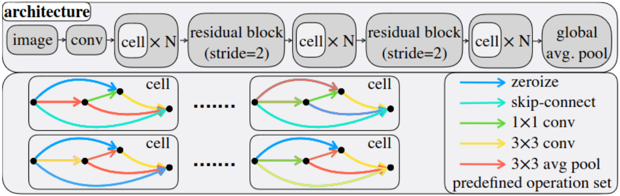
- Analyze zero-cost proxies as predictors
- Properties of the neural graph as a novel predictor
- Interpretability analysis of predictions
- Compare with predictors from related work

📖 G. Kadlecová, J. Lukasik, M. Pilát, P. Vidnerová, Petra, M. Safari, R. Neruda, F. Hutter, *Surprisingly Strong Performance Prediction with Neural Graph Features*, ICML 2024

Benchmarks and datasets

NAS Benchmarks

- Datasets of precomputed objectives on selected tasks
- NAS-Bench-101/201/301, NAS-Bench-201, NAS-Bench-301
- HW-NAS-Bench, TransNAS-Bench-101, robustness NB201



Source: NAS-Bench-201: Extending the scope of reproducible NAS. ICLR 2020.

Datasets

- CIFAR 10, 100, 10/100 classes, 60k images, 32x32 pixels
- ImageNet-16-120 (a downsampled variant, 1000 classes)

Methodology

Regression

- Random forest regressor
- Predict accuracy or other metrics

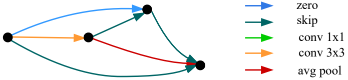
Input data – network encodings

- Zero cost proxies
- One hot encoding (of chosen operations)
- Graph properties

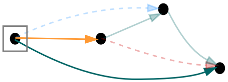
Experiments

- Analyze predictions
- Compare different network encodings, predictors

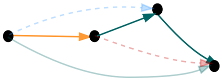
Properties of the neural graph



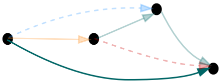
- We look at properties of the network graph – paths, counts, ...
- Node degree (c3x3, skip) means input degree counting only conv3x3 and skip
- Similarly, max path computes the maximum path over allowed operations



Node degree (c3x3, skip): 2



Max path (c3x3, skip): 3



Max path (skip): 1

Properties of the neural graph

- Number of operations
- Min path from input over operations O
- Max path from input to over operations O
- Out degree of the input node counting only operations O
- In degree of the output node counting only operations O
- Mean in/out degree of intermediate nodes counting only operations O

Advantages

- Simple, interpretable, fast to compute

Disadvantages

- Highly correlated, dependent on search space

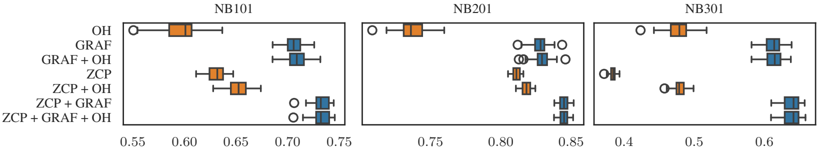
Accuracy prediction

Usage in performance prediction

- Gathering all properties, we use them as input data to a random forest regressor
- We compare with other network encodings – all ZCP, one-hot encoding (OH), and their combinations

Results

- Our results (GRAF) are better than ZCP and OH
- ZCP combined with network properties is the best

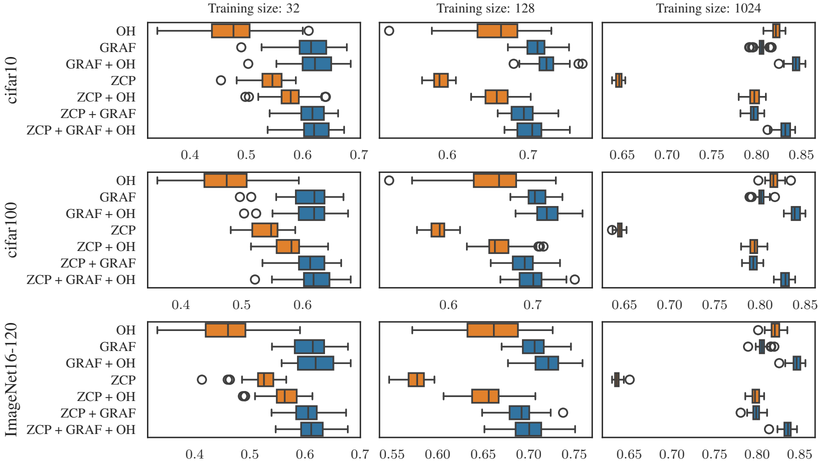


Interpretable prediction

- For network properties and ZCP, we compute Shapley coefficients (considers feature set importance)
- We look at the most important features
- Results – different features are important for diverse tasks
- nwot is important for CIFAR10, but not for autoencoder
- autoencoder needs skip connections – result from related work!

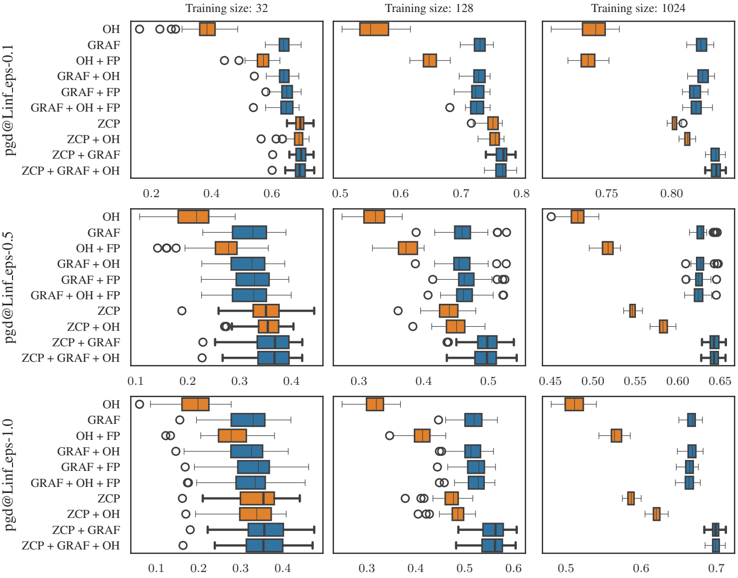
NB201 - `cifar-10`		TNB101-micro - `autoencoder`	
Feature name	Mean rank	Feature name	Mean rank
jacov	0.00	min path over skip	0.00
nwot	1.12	jacov	1.00
flops	3.62	fisher	2.00
synflow	4.08	min path over [skip,C3x3]	5.50
min path over [skip,C3x3,C1x1]	4.78	snip	5.58
params	5.04	min path over [skip,C1x1]	5.64
epe_nas	6.04	grad_norm	6.64
zen	6.36	zen	8.08
min path over [skip,C3x3]	11.08	grasp	9.34
min path over skip	11.88	l2_norm	9.74

HW metrics prediction



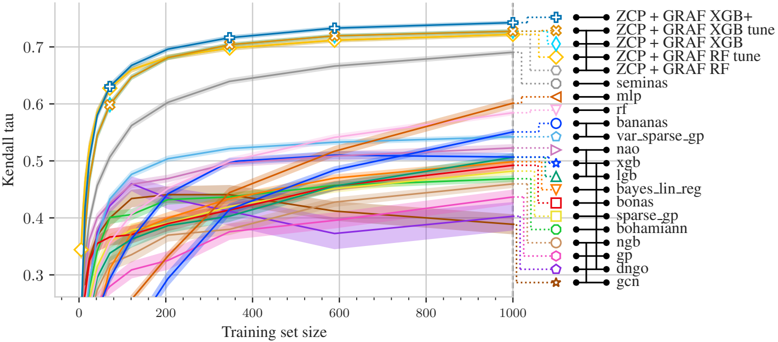
- HW metrics differ in difficulty of prediction
- edgegpu energy is one of difficult tasks

Robustness prediction



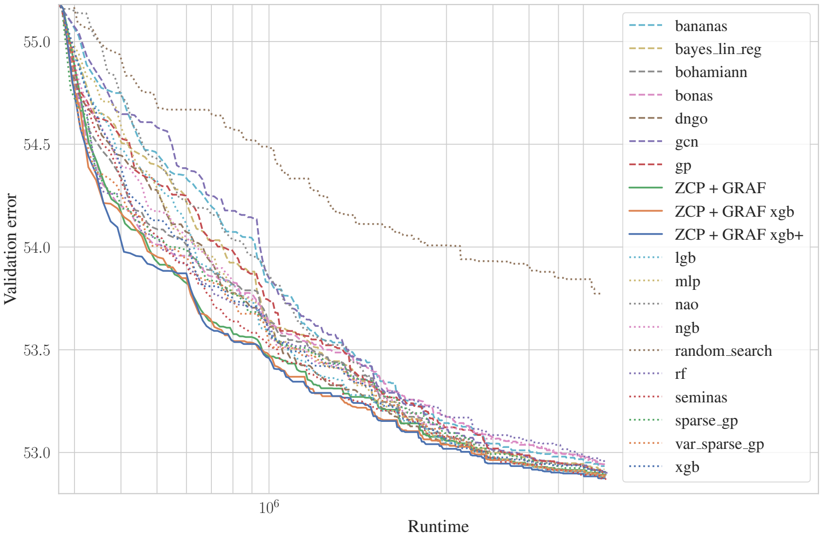
Comparison with existing predictors – NB101, CIFAR10

- Same experiment as in a predictor survey [1]
- Outperforms all available predictors
- Some predictors take much longer to train, e.g. graph neural networks!



[1] Colin White, Arber Zela, Binxin Ru, Yang Liu, & Frank Hutter (2021). How Powerful are Performance Predictors in Neural Architecture Search?. In Advances in Neural Information Processing Systems.

Usage in the NAS process – ImageNet16-120 search



Discussion and future work of GRAF

Pros

- Better and faster than most predictors
- Great interpretability
- Works across tasks and search spaces
- Baseline for complex predictors

Cons

- Properties need to be used with ZCP for best performance
- Some graph neural networks with ZCP can be better

Thank you! Questions?