DEHB-WS: Joint Architecture and Hyperparameter Search with Weight Sharing

Zahra Padar[1], Sharat Patil[1], Sai Prasanna[1]
{padarz, patilsh, ramans}@cs.uni-freiburg.de
[1]University of Freiburg

1.Motivation

- Joint architecture (NAS) and hyperparameter (HPO) search has been shown to be essential for obtaining strong performance in deep neural networks.
- Weight sharing allows for faster search in NAS.
- **Idea**
  - Multi-fidelity optimization over joint architecture and hyperparameter space.
  - Using the Differential Evolution HyperBand algorithm.
  - Sampling architectural blocks from a shared supernet.

2.Approach

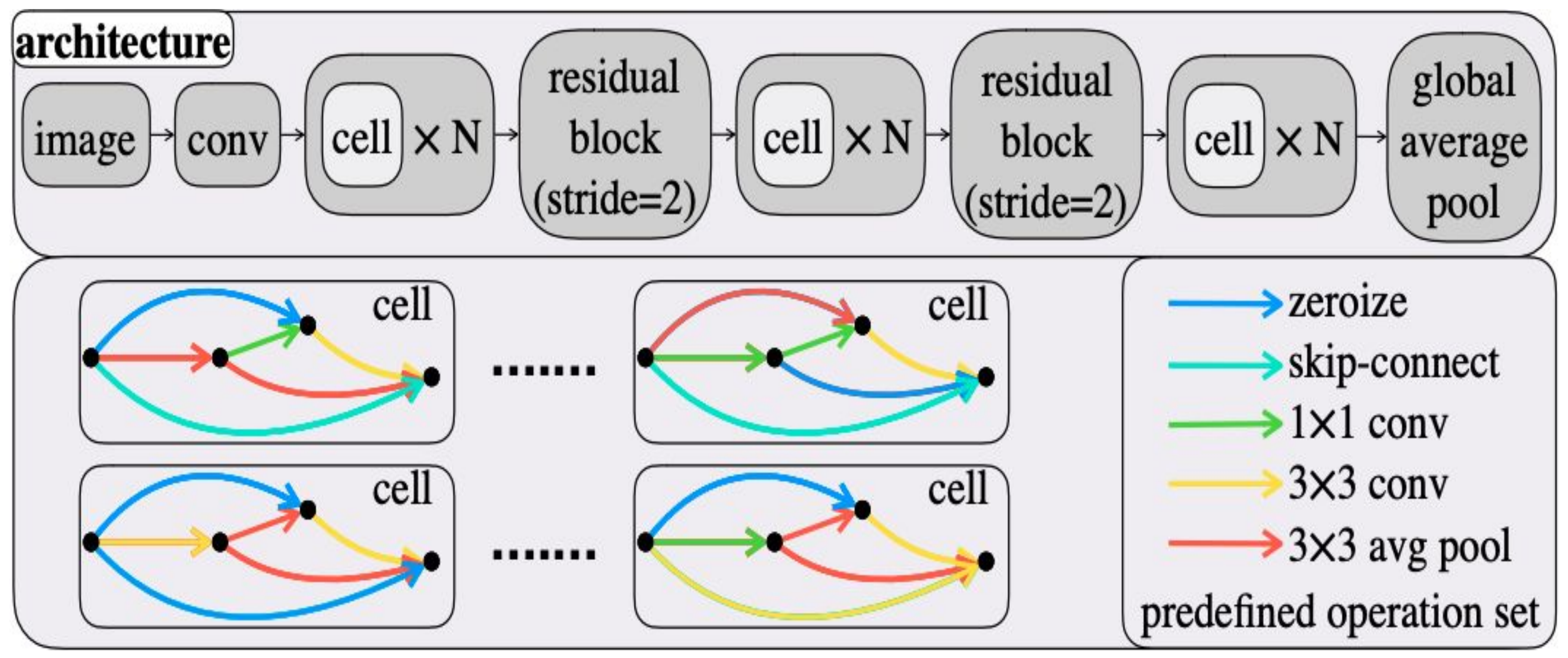
- **Supernet**
  - Once for All Supernet based CNN weight sharing.
  - Subnet architectures inherit weights and get trained independently.
  - Supernet is updated by averaging subnet weights.
- **DEHB - WS**
  - Differential Evolution (DE) component is not modified.
  - Cumulative budgets are used in the Successive Halving (SH) brackets.
  - Supernet update happens at the end of every SH rung using the DE selected configurations subnets.
- **Inherit and Finetune**
  - Incumbent architecture inherits the supernet weights.
  - The model is finetuned using the incumbent hyperparameters.

3.Search Space

JAHS-Bench-201 Search Space

Space	Property	Description
Architecture	Cell Search Space	NAS-Bench-201
Hyperparameter	Activation Learning Rate Weight Decay Trivial Augment	ReLU/Hardswish/Mish 10^{-3}-10^{0} 10^{-5}-10^{-2} On/Off
Fidelity	Epoch	0-200

- Fidelities for width and depth are set to maximum.
- NAS-Bench-201: 1x1 and 3x3 kernels do not share weights.



4.Experiments

- **Baselines**
  - Fixed CNN without Auto ML
  - Vanilla **DEHB**
  - **SMAC4MF**
- **JAHS-Bench-201** is queried for DEHB, SMAC4MF baselines and the final evaluations (Retrain 50)
- JAHS-Bench-201 training and evaluation pipeline are duplicated and used in DEHB-WS.
- **Dataset**: Upscaled 32x32 Fashion-MNIST used in JAHS-Bench-201.
- Function evaluations are used as the stopping criterion.

5.Results & Key Insights

Search performance:
- DEHB-WS has the lowest validation regret (score) during search.

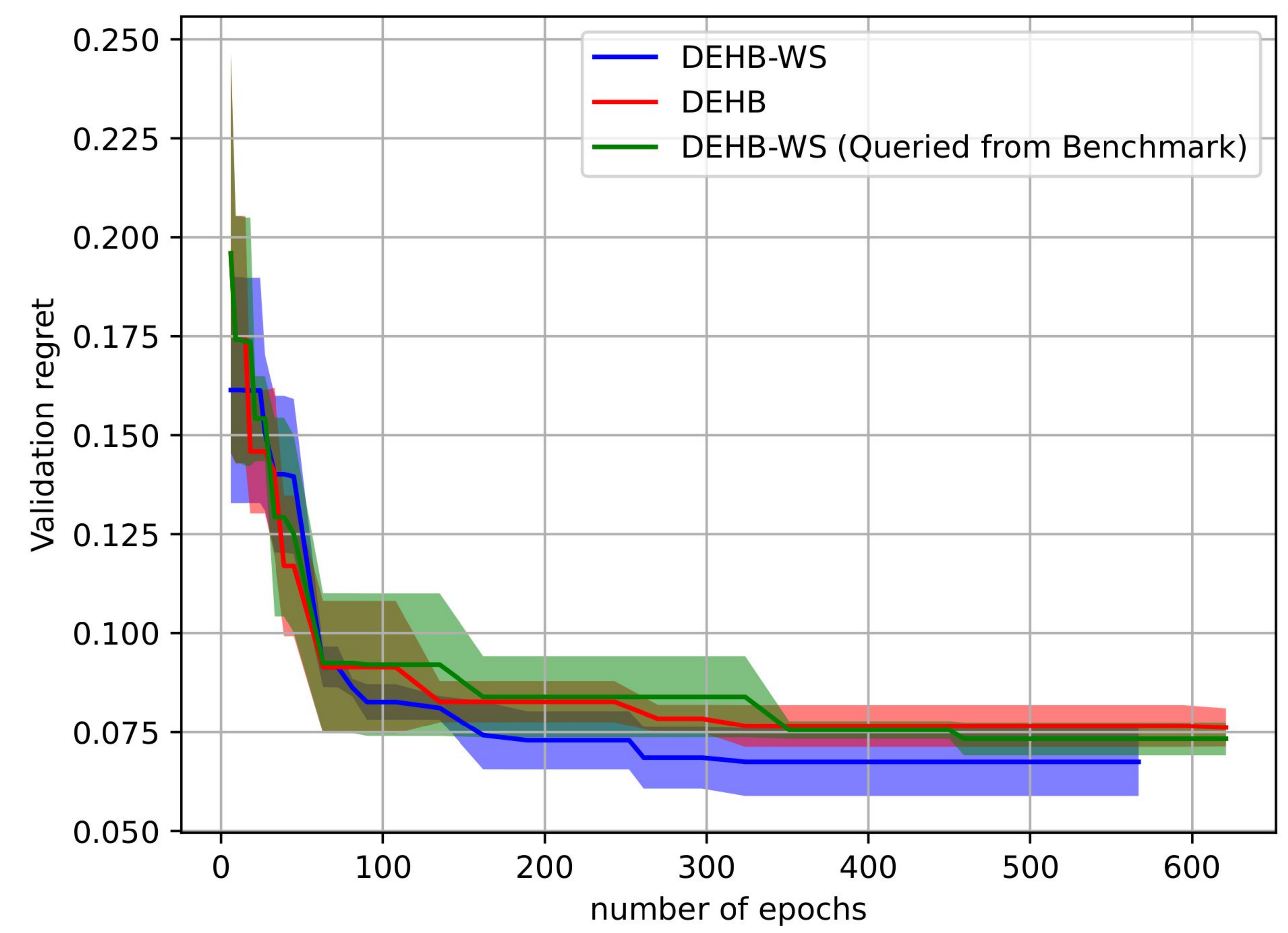
Effect of Inheriting weights + Finetuning:
- DEHB-WS incumbents achieve the best test accuracy, with inheritance + finetuning vs. trained from scratch.

Approach	Test Acc. (3 seeds)
Fixed CNN Baseline	90.45 ± 0.47
DEHB (Retrain 50)	92.62 ± 0.80
SMAC4MF (Retrain 50)	91.01 ± 1.46
DEHB-WS (Retrain 50)	92.11 ± 0.73
DEHB-WS (Inherit+Finetune 10)	92.43 ± 1.52
DEHB-WS (Inherit+Finetune 15)	**93.37 ± 1.16**

Approach	Val. Regret (3 seeds)
DEHB	0.0762±0.0048
SMAC4MF	0.0746±0.01
DEHB-WS	**0.0675±.0086**

Faster Search:
- DEHB-WS reaches lower validation regret in fewer epochs vs DEHB.

Better-performing Incumbents overall:
- DEHB-WS incumbents -with or without weight inheritance- reach lower validation regret vs. DEHB's.



Code: https://github.com/automl-classroom/final-project-autobots/tree/master/jahs_fashion