BLOX

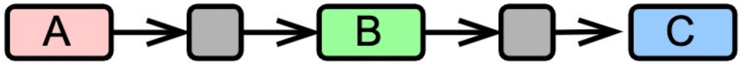
Macro Neural Architecture Search Benchmark and Algorithms

Thomas Chau[1*], Łukasz Dudziak[1*],
Hongkai Wen[1,3], Nicholas D. Lane[1,2], Mohamed S. Abdelfattah[4],

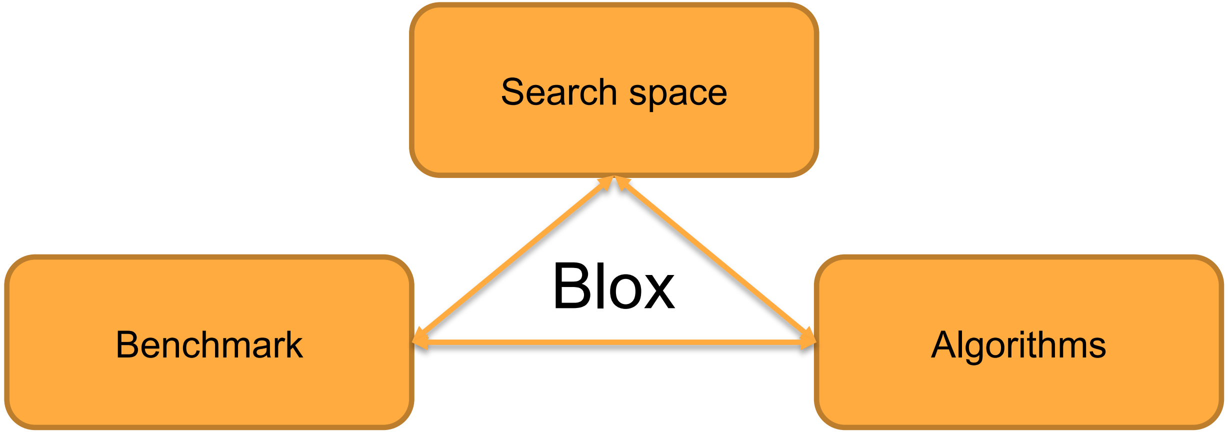
[1]*Samsung AI Center, Cambridge, UK*
[2]*University of Cambridge, UK*
[3]*University of Warwick, UK*
[4]*Cornell University, US*

Motivation

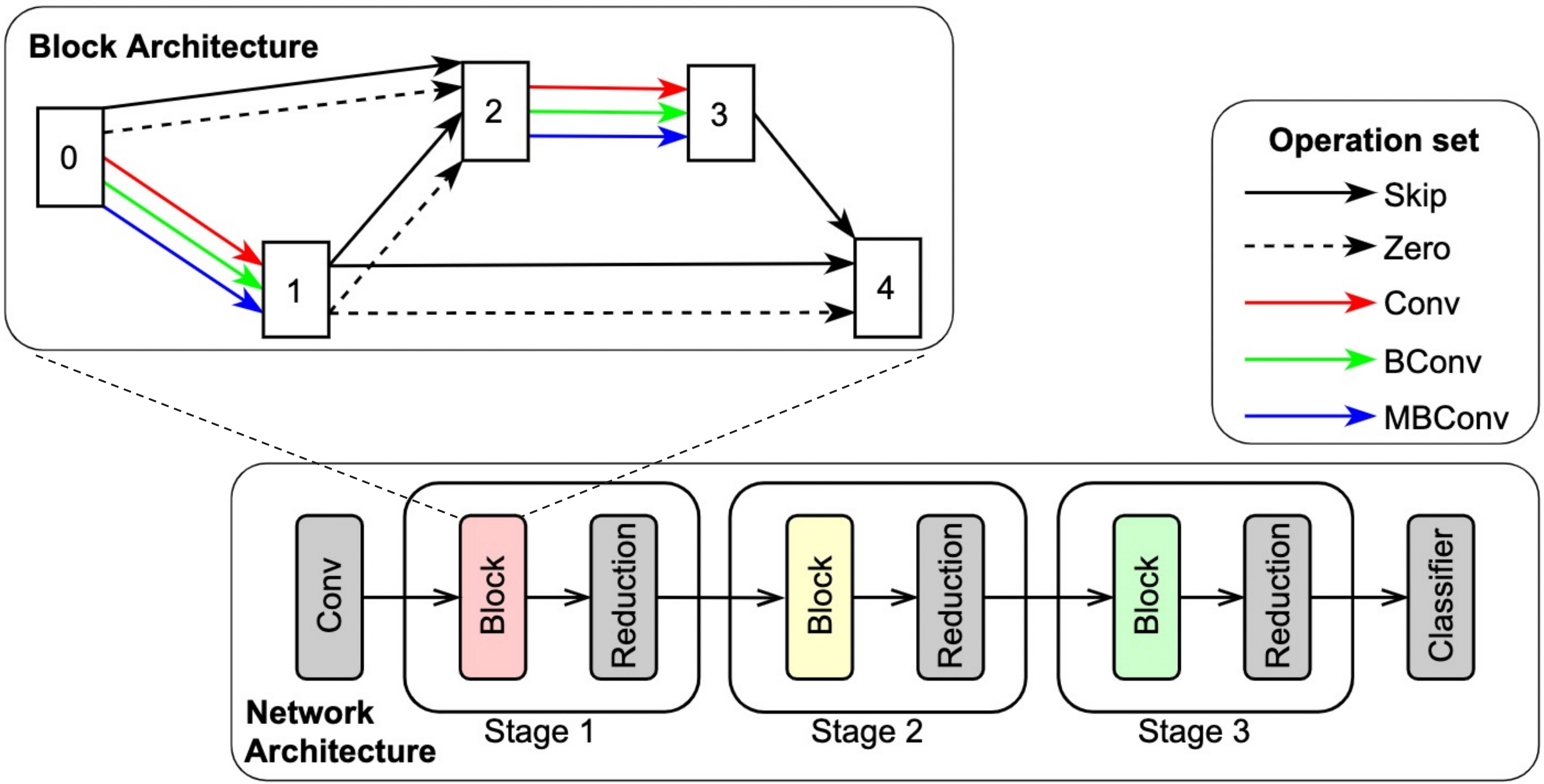
Cell-based Search Space



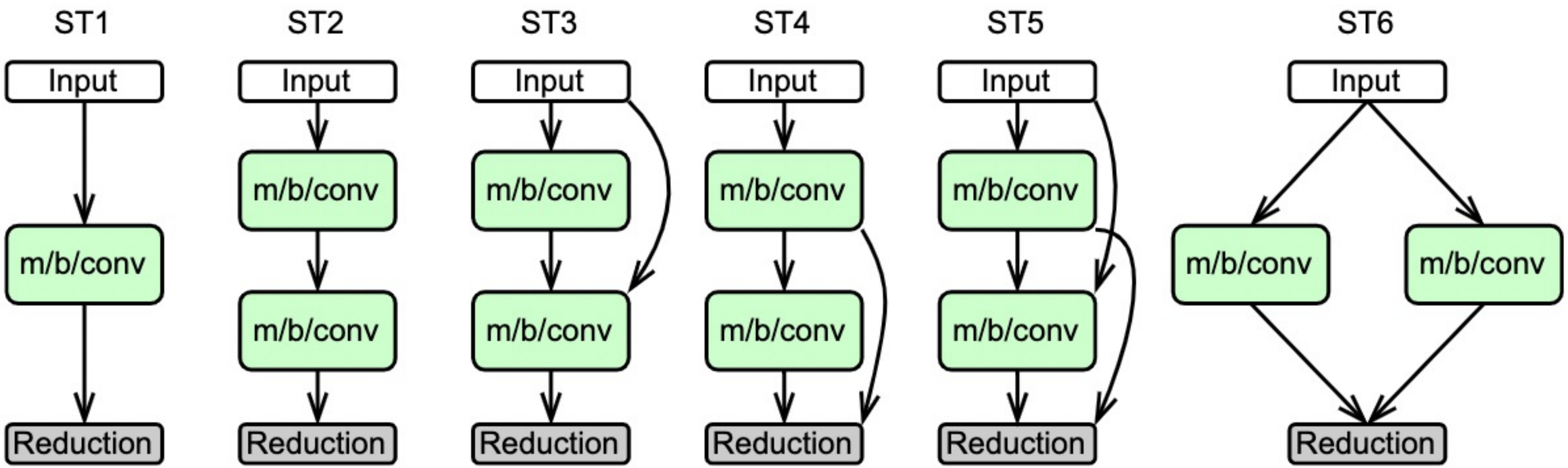
Macro Search Space



Search space



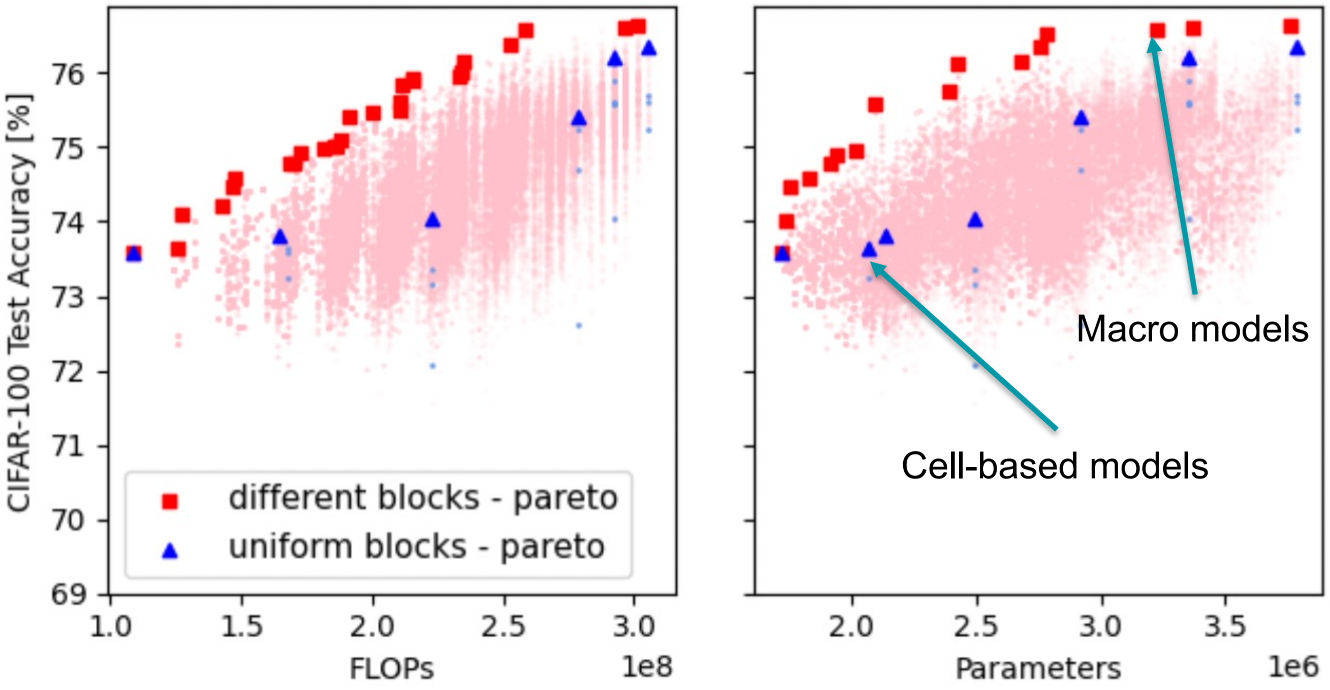
Block architectures



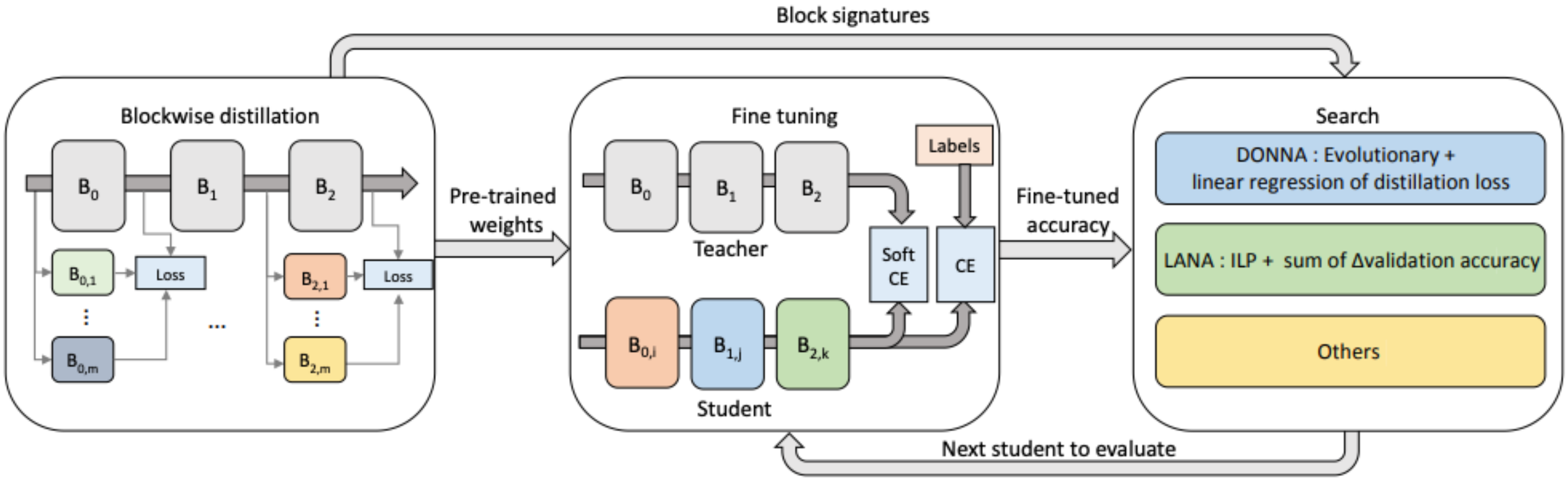
- Conv: VGG-style 3x3 convolutions.
- BConv: Resnet-style bottleneck with 5x5 depthwise separable convolutions.
- MBConv: EfficientnetV2 fused-inverted residual convolution including squeeze and excitation operation.

- 45 unique blocks → 91,125 models.

Macro vs cell-based models

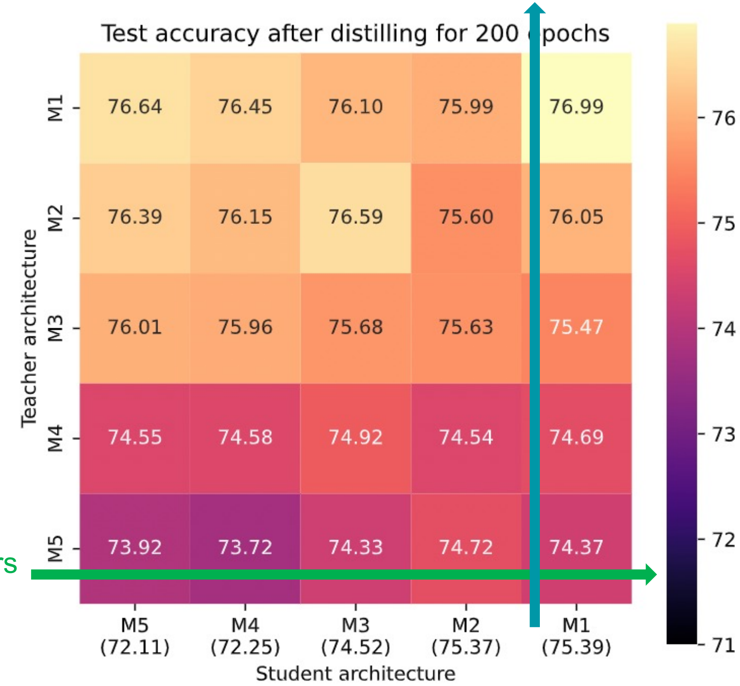


Blockwise NAS



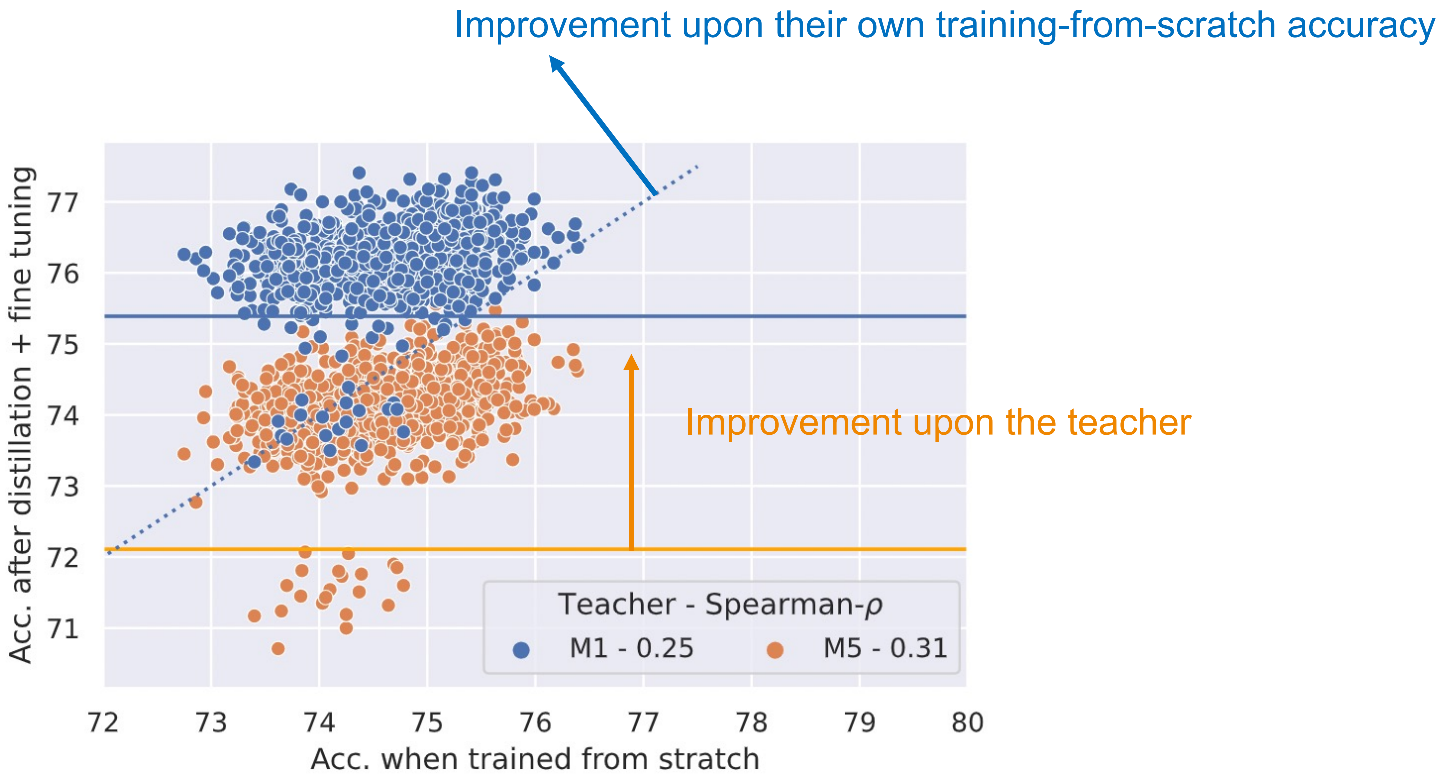
Does distillation give better performance vs normal training?

Fine-tuning does not always results in improvement of student models
e.g., M1 student drop from 75.39 to 74.37 when using M5 as teacher.

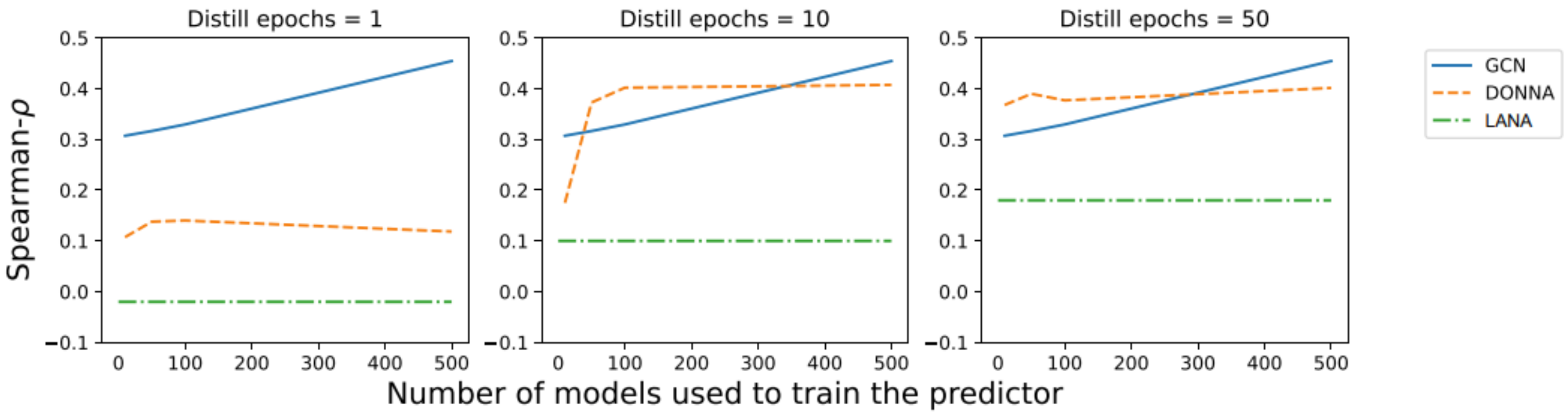


Student models improve upon their teachers
e.g., M5 teacher 72.11, students > 73.92

Does fine-tuning accuracy correlate to training-from-scratch accuracy?



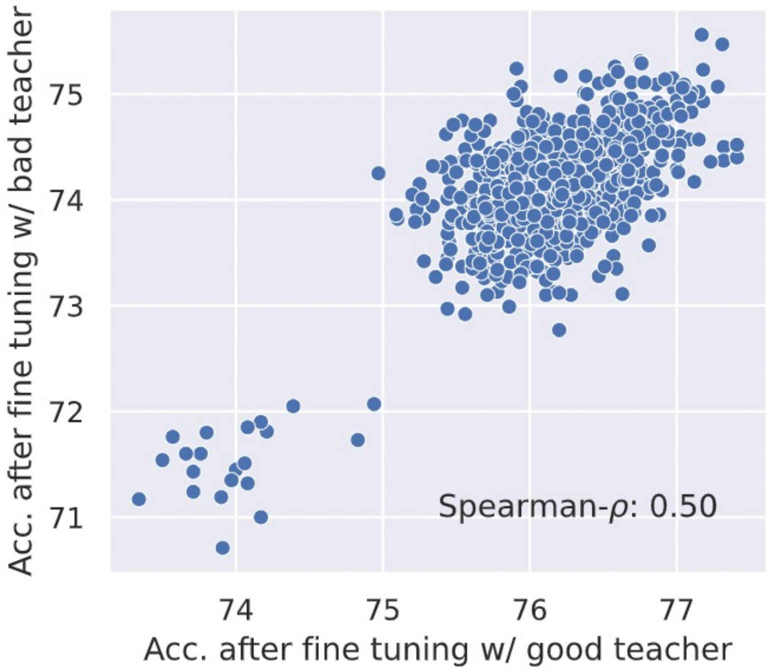
Can we use signatures to predict end-to-end performance?

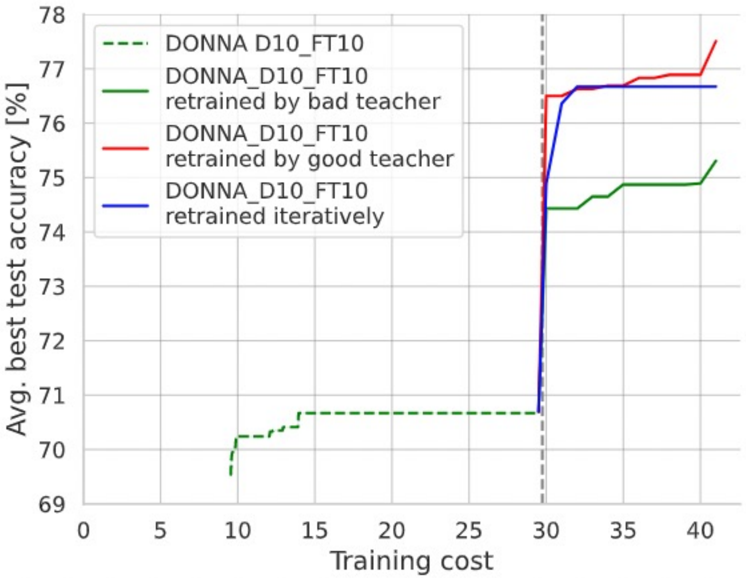


* DONNA: a linear regression model with second-order terms.
* HANT: a simple sum of signatures is used as a proxy.
* GCN: a graph convolutional network to capture graph topology and predict performance of a model.

Can we search for good models without prior knowledge of a good teacher?

The iterative approach has significantly improved the model accuracy without knowing a good teacher in advance.





Comparison of different NAS methods

Method	Acc. @ cost=40	Cost @ acc.=76.6
Conventional NAS		
Regularized Evolution	76.10	✗
BRP-NAS	76.40	400
DARTS-PT	74.52†	✗
Blockwise NAS *assuming good teacher (M1)*		
FT200	76.90	25
FT10	73.47	✗
FT10 + FT200	77.66	30
Blockwise NAS *assuming bad teacher (M5)*		
FT10	70.67	✗
FT10 + FT200	74.90	✗
FT10 + FT200 iter.	76.67	30
FT10 + FT200 M1	76.88	30

Standard blockwise with good teacher

Reduced fine-tuning followed by full fine-tuning with good teacher

Iterative fine-tuning approach

More about our work

- https://github.com/SamsungLabs/blox

BLOX: Macro Neural Architecture Search Benchmark and Algorithms

Thomas Chau[1*], Łukasz Dudziak[1*], Hongkai Wen[1,3]
Nicholas D. Lane[1,2], Mohamed S. Abdelfattah[4]

[1] Samsung AI Center, Cambridge, UK
[2] University of Cambridge, UK [3] University of Warwick, UK
[4] Cornell University, USA
[*] *Indicates equal contributions*

{thomas.chau, l.dudziak, hongkai.wen, nic.lane}@samsung.com
mohamed@cornell.edu