

Benchmark of neural network models with reduced number of parameters for image classification

Rares Folea

Faculty of Automatic Control and Computers
University Politehnica of Bucharest

The research objective:

Can we use traditional **deep neural architectures** and achieve good performance with a reduced number of trainable parameters?

What we mean by reduced number of parameters?

We want to investigate neural networks with at most 1M parameters

What methods are used for reducing the number of parameters?

Architectural changes
Pruning techniques

What datasets are being used?

CIFAR-10
CIFAR-100

Adjacent objective:

Validate or invalidate **The Lottery Ticket Hypothesis**

“A randomly-initialised, dense neural network **contains a subnetwork** that is initialised such that — when trained in isolation — it *can match the test accuracy* of the original network after training for at most the same number of iterations.”

(winning tickets: less than 10-20% of the initial size)

Our proposal:

Adapt SotA deep neural architectures (ResNet, VGG, ViT) to perform the task of image recognition with less than 1M parameters

Use **pruning** techniques on the SotA deep neural architectures to perform the task of image recognition

Constraints:

- Models with: **< 1M** parameters, compared with the full-size architecture (tens or hundreds of millions).
 - SotA: ViT-H/14, 632M parameters, 99.5% accuracy.
- Evaluate on **CIFAR-10** and **CIFAR-100**, same training/testing split for all the models.
- Train against various optimizers (SGDM, ADAM, AdaBound).

Architectural changes

ResNet18_S

VGG16_S

ViT(H/4)_S

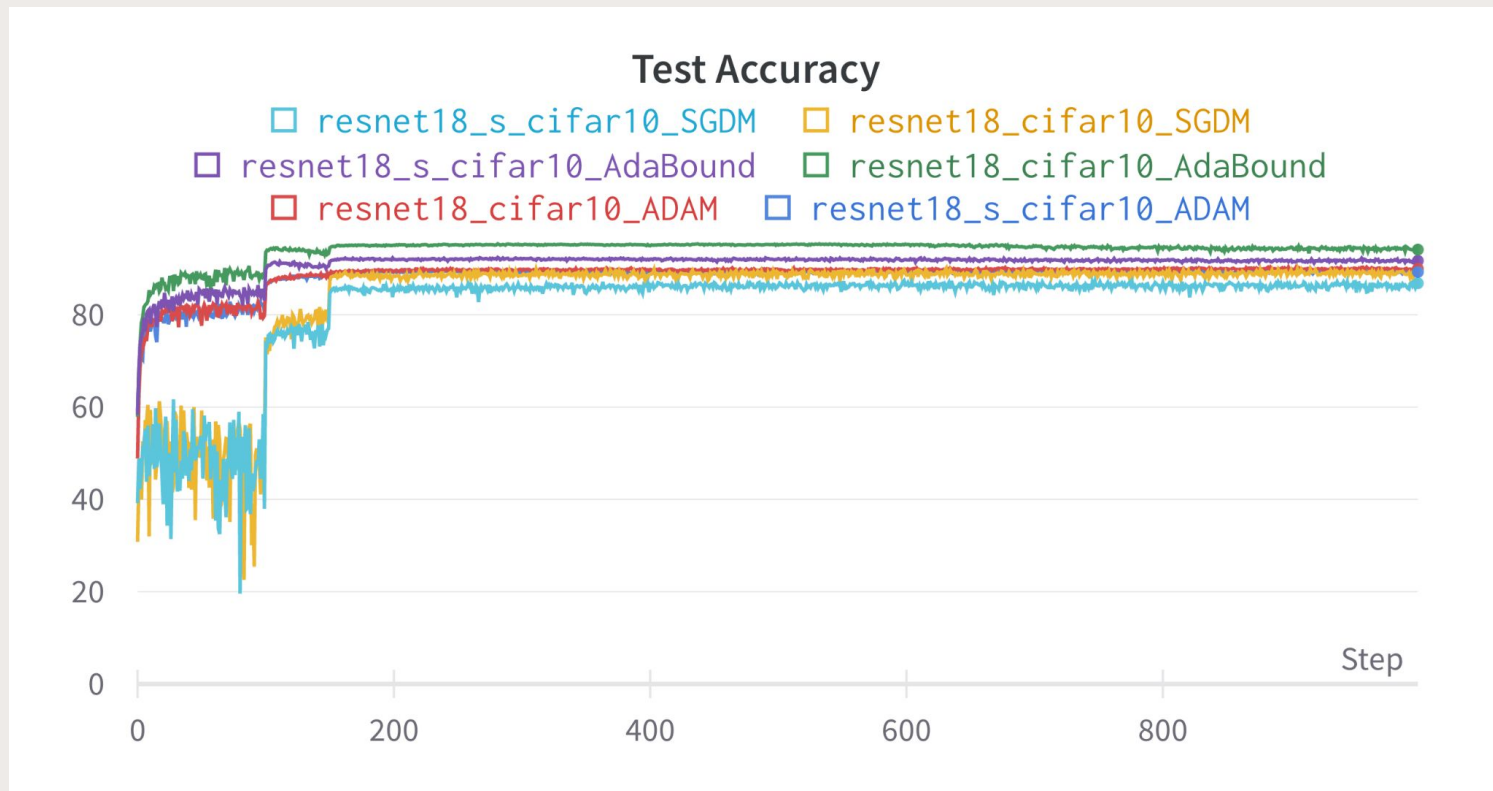
(and optimizer influence)

Results: CIFAR-10

ResNet_S has <10% of the total no. of params of ResNet18

| Architecture | Optimizer | Test accuracy |
|-------------------|-----------------|---------------|
| ResNet18 | <i>SGDM</i> | 89.4 |
| | <i>ADAM</i> | 89.97 |
| | <i>AdaBound</i> | 94.13 |
| ResNet18_S | <i>SGDM</i> | 86.85 |
| | <i>ADAM</i> | 89.27 |
| | <i>AdaBound</i> | 91.72 |

Results: Resnet18_S, CIFAR-10

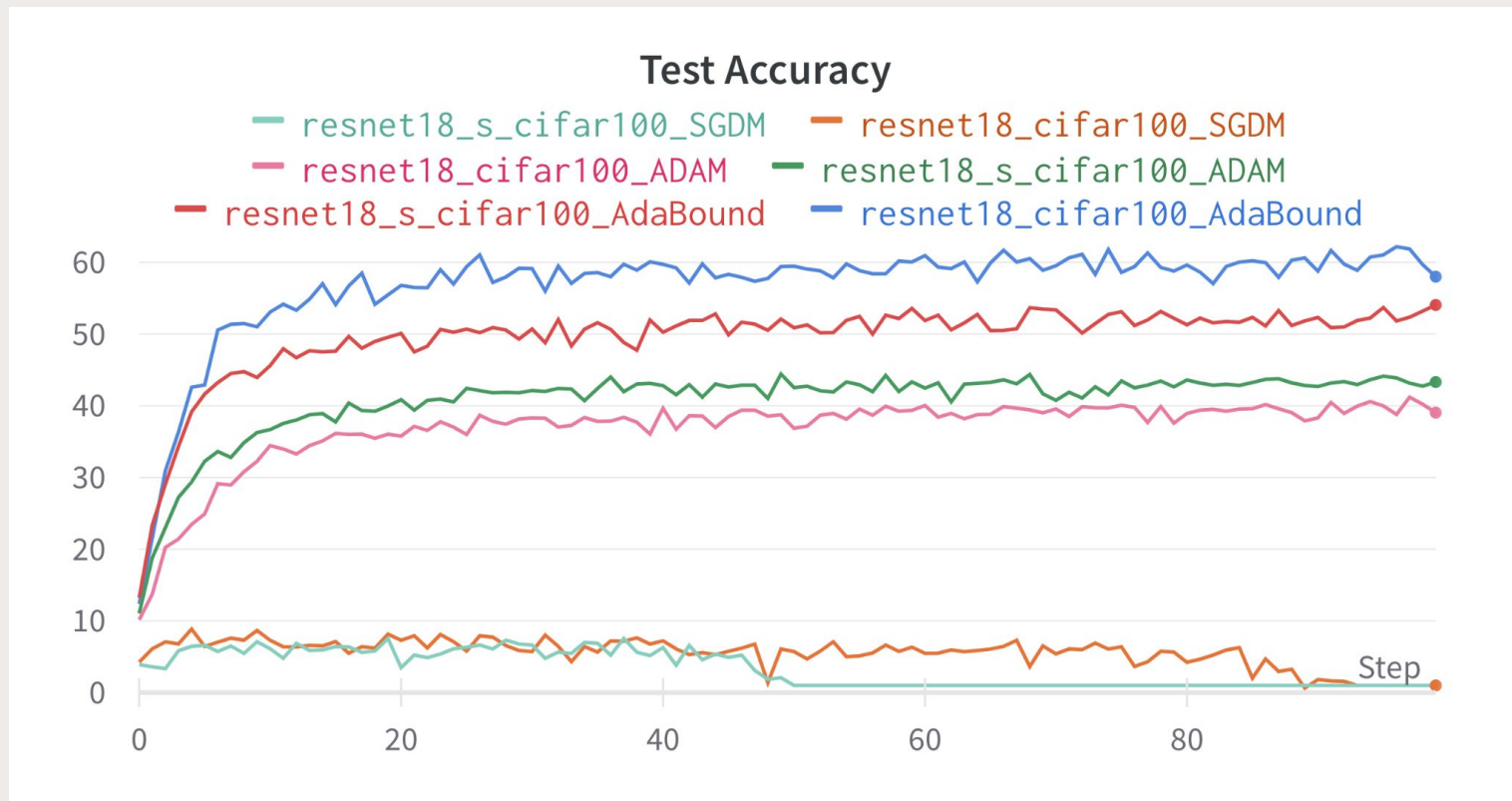


Results: CIFAR-100

ResNet_S has <10% of the total no. of params of ResNet18

| Architecture | Optimizer | Test accuracy |
|--------------|-----------------|---------------|
| ResNet18 | <i>SGDM</i> | 1 |
| | <i>ADAM</i> | 39.04 |
| | <i>AdaBound</i> | 58 |
| ResNet18_S | <i>SGDM</i> | 1 |
| | <i>ADAM</i> | 43.31 |
| | <i>AdaBound</i> | 54.05 |

Results: Resnet18_S, CIFAR-100

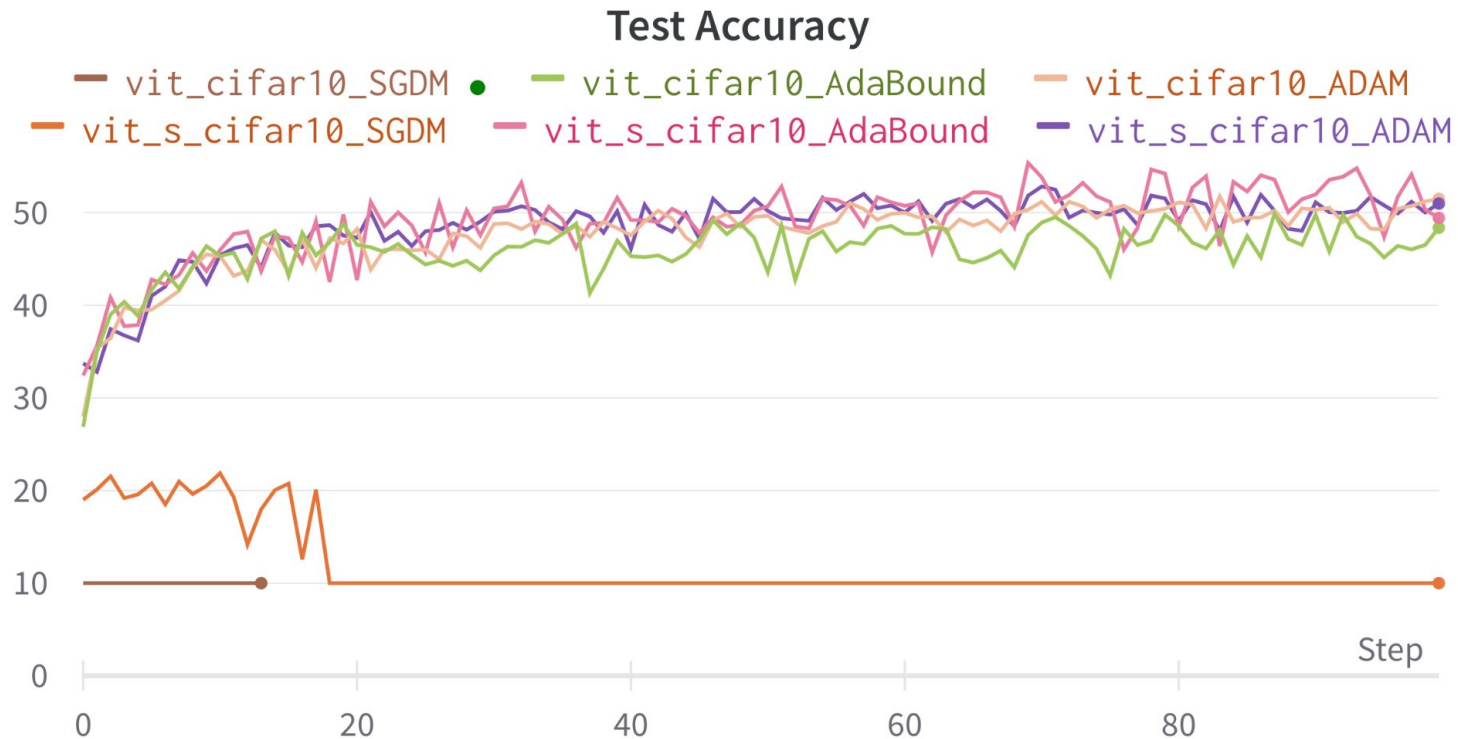


Results: CIFAR-10

ViT_S has ~5% of the total no. of params of ViT H/4
<0.1% of SotA ViT H/14

| Architecture | Optimizer | Test accuracy |
|--------------|-----------------|---------------|
| ViT | <i>SGDM</i> | 10 |
| | <i>ADAM</i> | 51.49 |
| | <i>AdaBound</i> | 48.38 |
| ViT_S | <i>SGDM</i> | 10 |
| | <i>ADAM</i> | 50.99 |
| | <i>AdaBound</i> | 49.44 |

Results: ViT(H/4)_S CIFAR-10

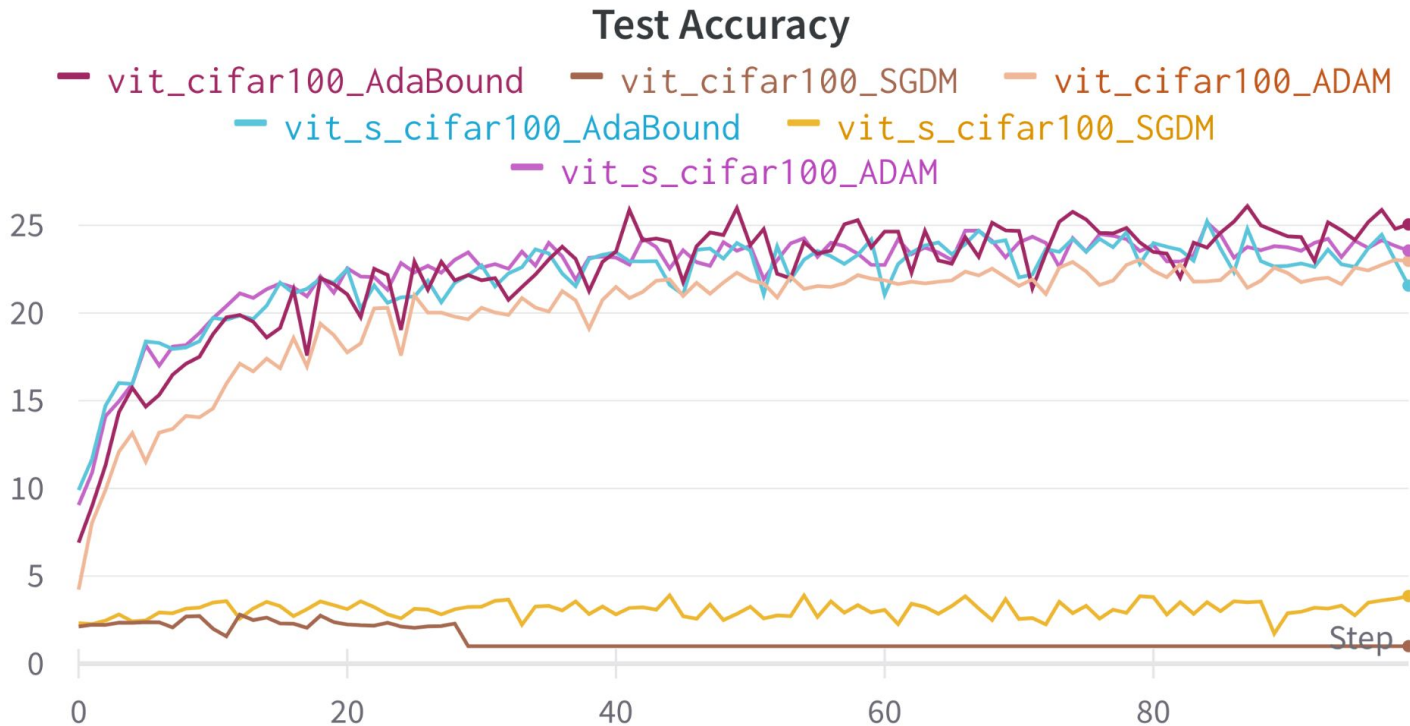


Results: CIFAR-100

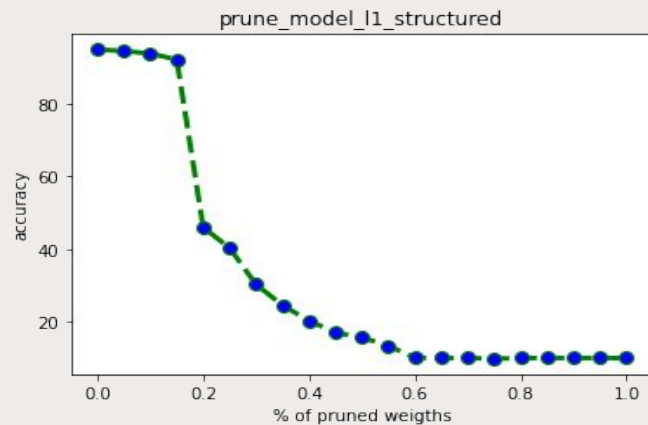
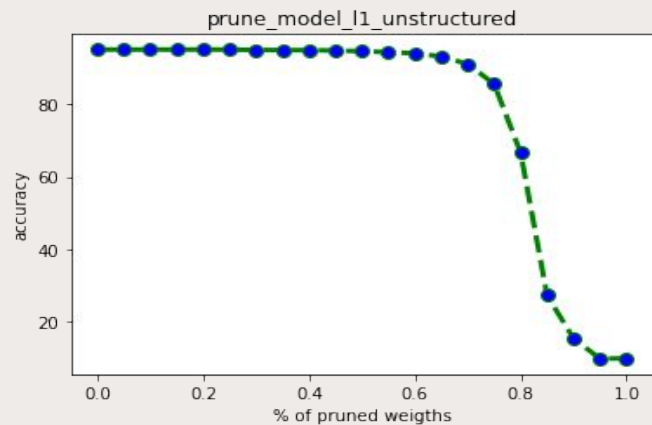
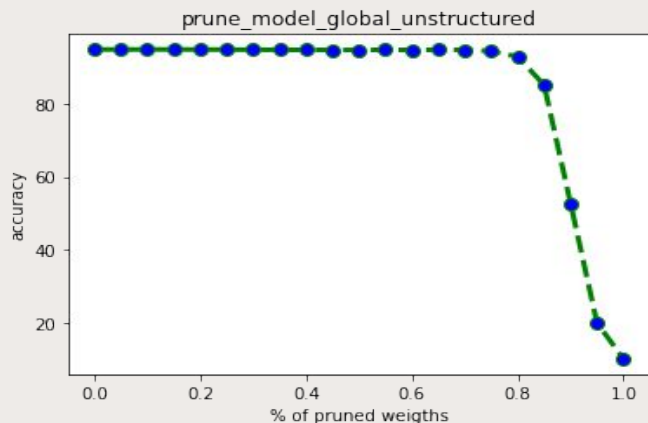
ViT_S has ~5% of the total no. of params of ViT H/4
<0.1% of SotA ViT H/14

| Architecture | Optimizer | Test accuracy |
|--------------|-----------------|---------------|
| ViT | <i>SGDM</i> | 1 |
| | <i>ADAM</i> | 22.97 |
| | <i>AdaBound</i> | 25.05 |
| ViT_S | <i>SGDM</i> | 3.86 |
| | <i>ADAM</i> | 23.56 |
| | <i>AdaBound</i> | 21.56 |

Results: ViT(H/4)_S CIFAR-100



ResNet18 spotlight: Pruning



Thank you!

Follow **the research** development:

<https://github.com/raresraf/benchmark-dnn-small/>

Follow **the experimental results**:

<https://wandb.ai/raresraf/dnn>

<https://wandb.ai/raresraf/dnn-cifar100>