Large Neural Networks at a Fraction

Aritra Mukhopadhyay, Adhilsha Ansad, Subhankar Mishra
NISER Bhubaneswar and HBNI Mumbai, India

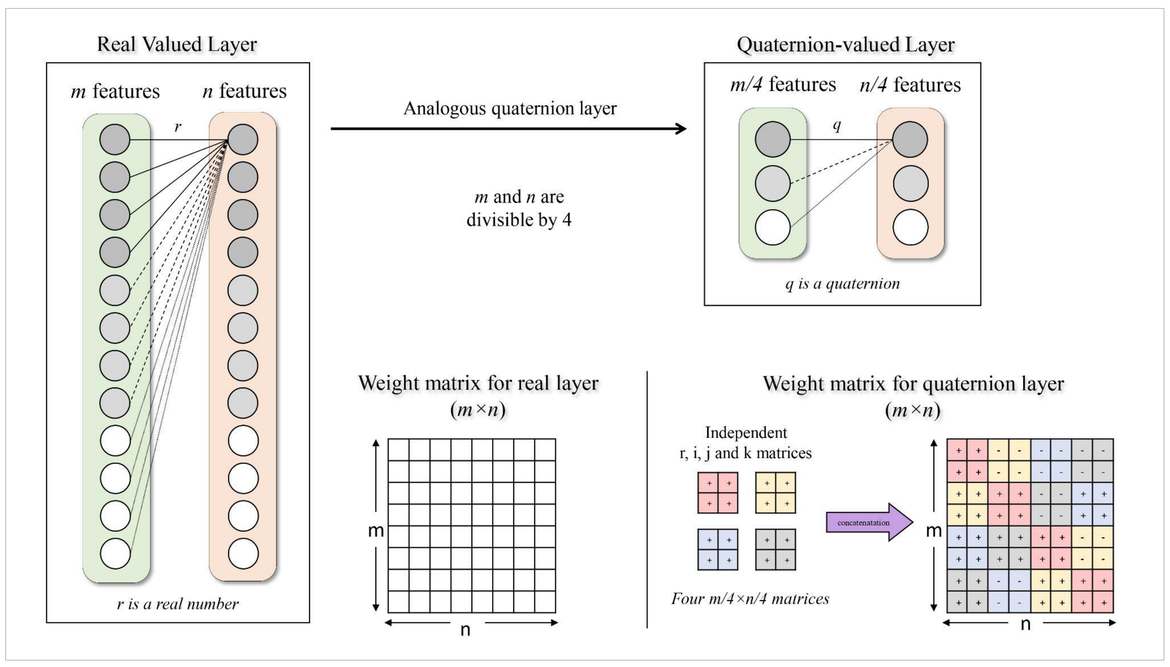
Overview

TL;DR: Experiments on the performance of quaternion models in extreme pruning conditions with larger models and larger datasets.

"The Lottery Ticket Hypothesis: A randomly-initialized, dense neural network contains a subnetwork that is initialized 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." [1]

A previous paper [2] explored the LTH performance of quaternion networks on small models (lenet-300-100, conv2, conv4 and conv6) and datasets (mnist, cifar10, cifar100). We are extending that work to see if the same observations hold even for larger models and larger datasets.

Quaternion Neural Network



Observation: Quaternion versions of real models have one-fourth the number of parameters, but perform almost as good. Moreover, upon pruning, performance of quaternion models remain intact till higher sparcities than real models.

Advantages: Less parameters ⇒ Less calculation ⇒ Faster Prediction speed ⇒ **Effective Deployment in low end devices**

Hyperparameters

Datasets Models	**CIFAR100** All ResNet models	**ImageNet64** ResNet18, 34 and 50	**ImageNet64** ResNet101 and 152
Training Epochs	100	15	15
Batch size	100	256	256
Optimizer	SGD	SGD	SGD
Learning Rate Scheduler (epochs)	0.1 (1-100)	0.1 (1-10), 0.01 (11- 12), 0.001 (13-15)	0.1 (1-10), 0.01 (11- 12), 0.001 (13-15)
Momentum	0.9	0.9	0.9
Weight Decay	1e-4	1e-4	1e-4
% Pruned per pruning iteration	70%	70%	50%

Model Statistics

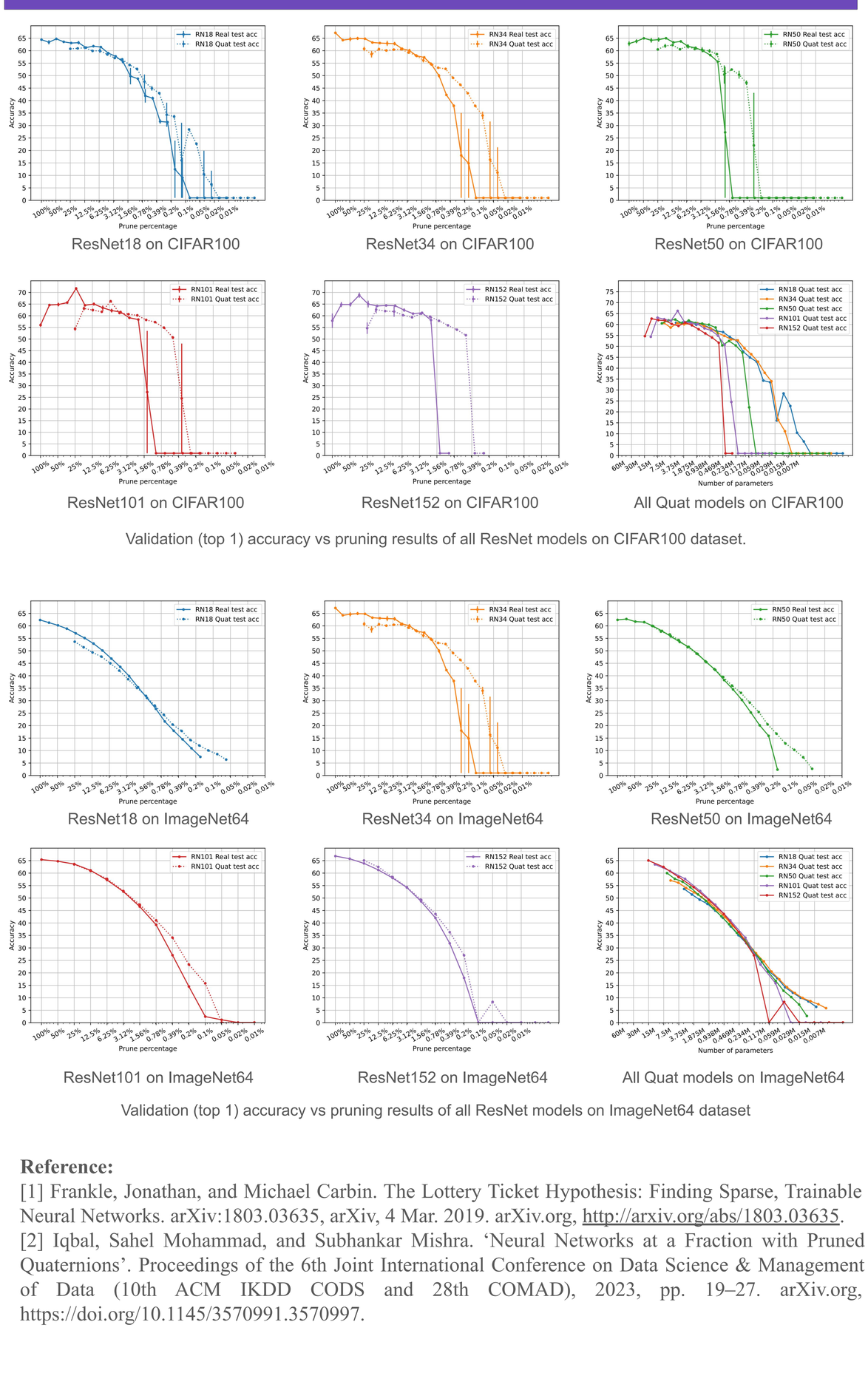
Model	**Number of Parameters**	
	Real Version	**Quaternion Version**
ResNet18	11.68M	2.93M
ResNet34	21.79M	5.46M
ResNet50	25.55M	6.43M
ResNet101	44.54M	11.22M
ResNet152	60.19M	15.16M

Dataset Statistics

Dataset	**Images**	**Image resolution**	**classes**
CIFAR100	60K	32x32 (coloured)	100
ImageNet64	1.28M	64x64 (coloured)	1000

Note: ImageNet64 dataset contains all the images in the original ImageNet (ILSVRC) dataset, just cropped down to 64x64 resolution.

Results



ResNet18 on CIFAR100 | ResNet34 on CIFAR100 | ResNet50 on CIFAR100

ResNet101 on CIFAR100 | ResNet152 on CIFAR100 | All Quat models on CIFAR100

Validation (top 1) accuracy vs pruning results of all ResNet models on CIFAR100 dataset.

ResNet18 on ImageNet64 | ResNet34 on ImageNet64 | ResNet50 on ImageNet64

ResNet101 on ImageNet64 | ResNet152 on ImageNet64 | All Quat models on ImageNet64

Validation (top 1) accuracy vs pruning results of all ResNet models on ImageNet64 dataset

Reference:
[1] Frankle, Jonathan, and Michael Carbin. The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks. arXiv:1803.03635, arXiv, 4 Mar. 2019. arXiv.org, http://arxiv.org/abs/1803.03635.
[2] Iqbal, Sahel Mohammad, and Subhankar Mishra. 'Neural Networks at a Fraction with Pruned Quaternions'. Proceedings of the 6th Joint International Conference on Data Science & Management of Data (10th ACM IKDD CODS and 28th COMAD), 2023, pp. 19–27. arXiv.org, https://doi.org/10.1145/3570991.3570997.