Towards a Greener AI: Structured Pruning of Convolutional Neural Networks at Initialization

Faux, Sydney (School: Pace Academy)

Training deep convolutional neural networks (CNNs) requires immense time and energy, which often has a significant carbon footprint. Pruning CNNs increases their efficiency while maintaining their accuracy. However, pruning is often applied after networks are already trained. Therefore, this paper presents a novel algorithm, FlowPrune, that can prune unimportant filters in CNNs at their initialization. Treating CNNs like flow networks, the Boykov-Kolmogorov maximum flow algorithm is used to find the maximum flow of the network. The filters of the CNN that have the least impact on the "flow" of the model are then pruned to reveal a subnetwork that can achieve similar accuracy as the original network. Experiments on the LeNet-5 and AlexNet CNNs trained on the MNIST and CIFAR-10 datasets demonstrate that this method works and prunes over 60% of the computations necessary to run the CNNs while affecting their accuracy by less than 2% in both cases.

Awards Won:

Fourth Award of $500

Association for Computing Machinery: Fourth Award of $500

NC State College of Engineering: Award to attend NC State Engineering Summer Camp