# Filter Pruning using Hierarchical Group Sparse Regularization for Deep Convolutional Neural Networks

Kakeru Mitsuno and Takio Kurita

Hiroshima University, Japan

HIROSHIMA UNIVERSITY

Kurita Laboratory

## Abstract

Since the convolutional neural networks are often trained with redundant parameters, it is possible to reduce redundant kernels or filters to obtain a compact network without dropping the classification accuracy. Such methods are called structure pruning. In this paper, we propose a filter pruning method using the hierarchical group sparse regularization. It is shown in our previous work that the hierarchical group sparse regularization is effective in obtaining sparse networks in which filters connected to unnecessary channels are automatically close to zero. After training the convolutional neural network with the hierarchical group sparse regularization, the unnecessary filters are selected based on the increase of the classification loss of the randomly selected training samples to obtain a compact network. It is shown that the proposed method can reduce more than 50% parameters of ResNet for CIFAR-10 with only 0.3% decrease in the accuracy of test samples. Also, 34% parameters of ResNet are reduced for TinyImageNet-200 with higher accuracy than the baseline network.

## Related Works

- Structured Sparse Regularization

$$R(W^l) = \sum_{g \in G} r(W_g^l)$$

Ex. group lasso regularization

$$r_{GL}(W) = |W|_2 = \sqrt{\sum_i W_i^2}$$

- Hierarchical Squared Group Sparse Regularization

$$R_{SQ}(W^l) = \sum_{g \in G} \left( \sum_{k \in K} r(W_{g,k}^l) \right)^2$$

- Hierarchical Square rooted Group Sparse Regularization

$$R_{SQRT}(W^l) = \sum_{g \in G} \sqrt{\sum_{k \in K} r(W_{g,k}^l)}$$

## Proposed Method

The feature-wise filter pruning algorithm for deep convolutional neural networks

1. Train a large network as the initial network.
2. Train the network with the hierarchical group sparse regularization based on the feature-wise grouping to find unnecessary filters connected to input channels by enforcing their weights to be zero. (Fig. 1)
3. Prune the filters with smaller influence on the classification loss with the random sampled validation data
4. Train the obtained compact network from scratch.

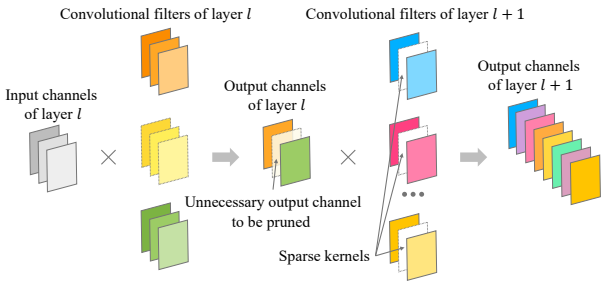
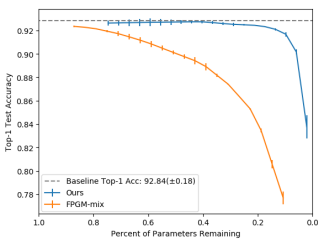
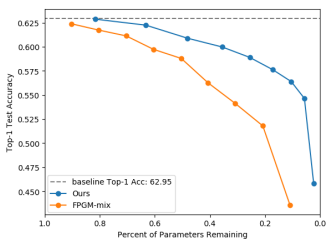



Fig. 1 An illustration of filter pruning

## Experiments

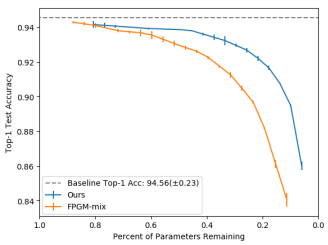



On CIFAR-10

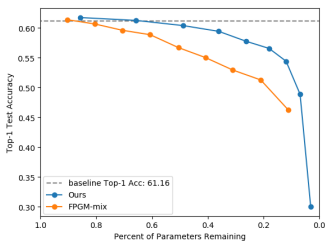



On TinyImageNet-200

Fig. 2 Results for VGG14

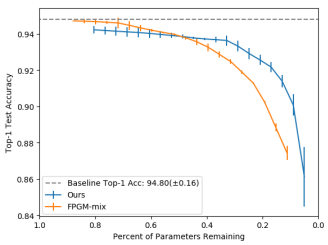



On CIFAR-10

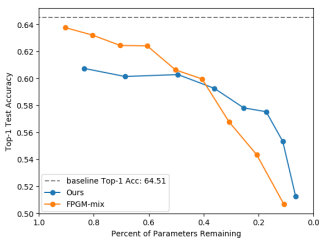



On TinyImageNet-200

Fig. 2 Results for ResNet20/18




On CIFAR-10




On TinyImageNet-200

Fig. 2 Results for ResNet32/34

## Conclusions

We propose a new filter pruning method with the hierarchical group sparse regularization based on the feature- wise grouping. The strategy of our pruning method is the step-wise pruning of the filters by searching the filter with the minimum loss increase. The performance of the pruned network is better than the state-of-the-art pruning method when more than 50% of the parameters are pruned.