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

Kakeru Mitsuno and Takio Kurita

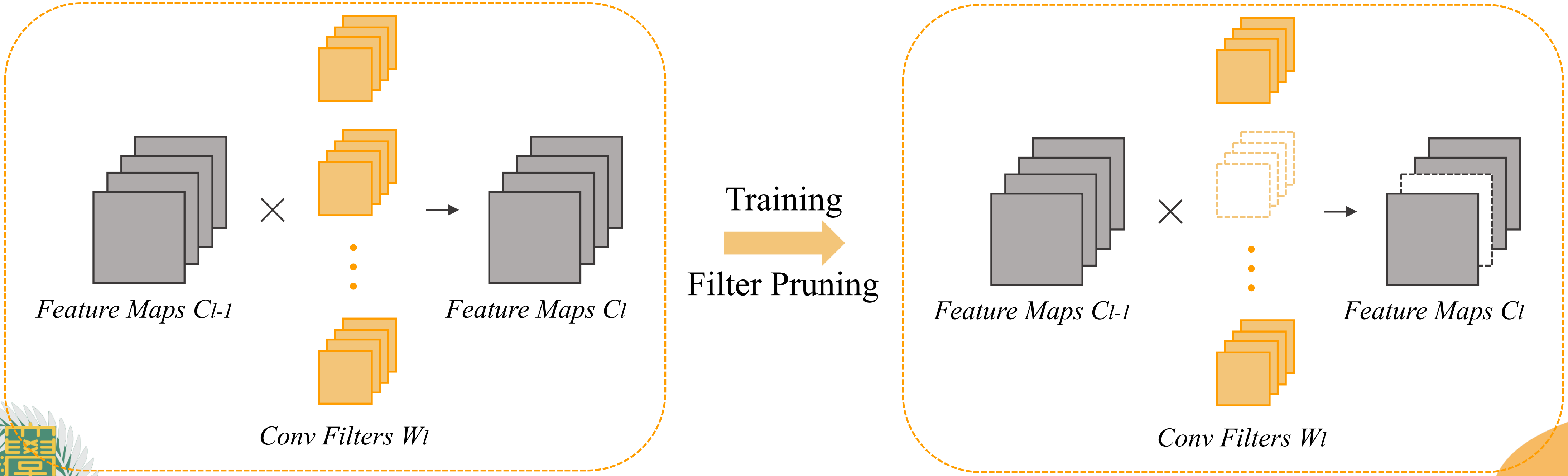
Hiroshima University, Japan

HIROSHIMA UNIVERSITY

Kurita Laboratory

Introduction

Filter pruning can reduce redundant parameters of large networks while preserving accuracy.

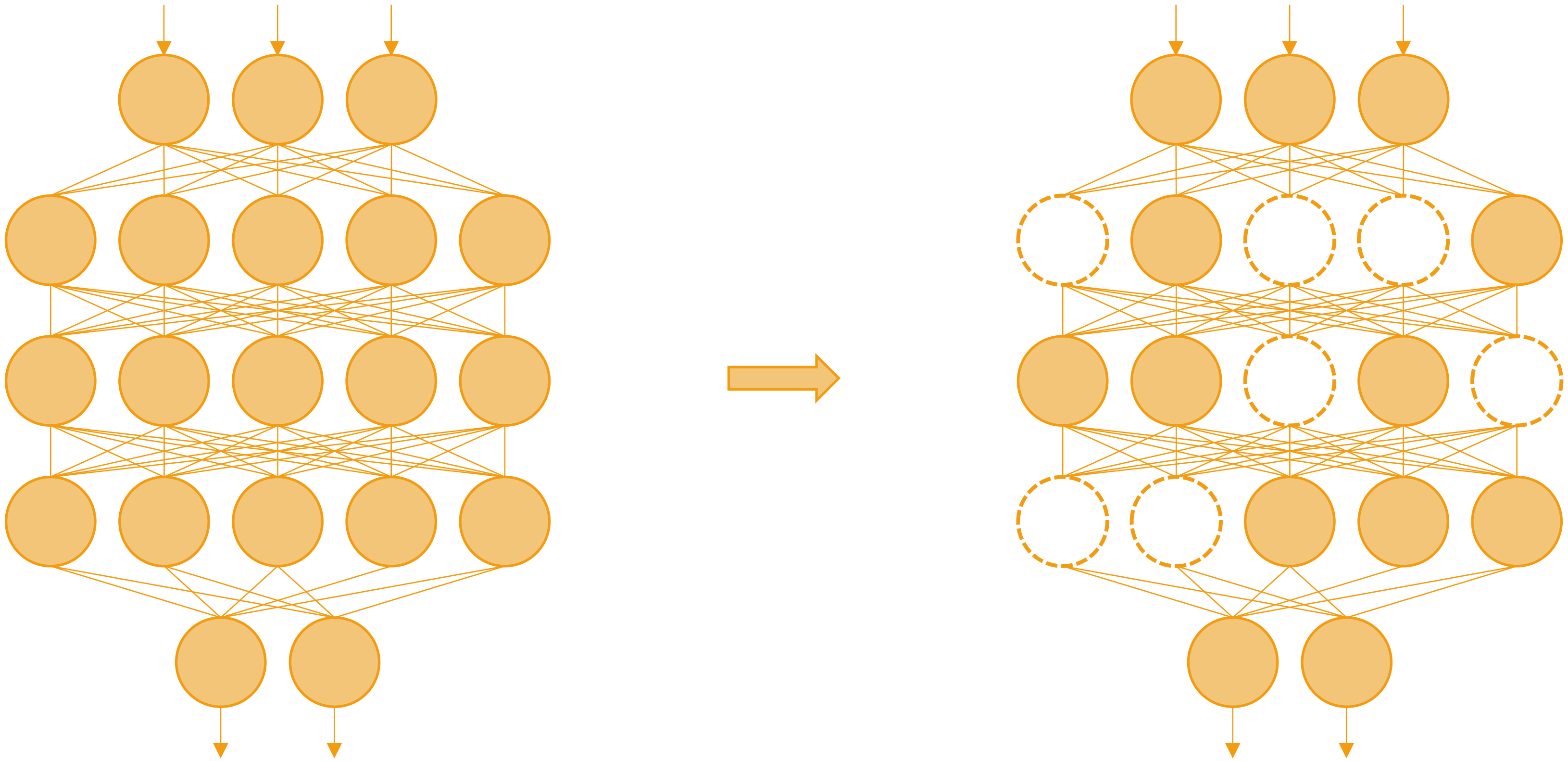


HIROSHIMA UNIVERSITY

Kurita Laboratory

Sparse regularization

Sparse regularization criteria can remove unnecessary parameters



HIROSHIMA UNIVERSITY

Kurita Laboratory

The structured sparse regularizations

The structured sparse regularization for structured pruning is defined as follows:

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

Typical structured sparse regularization for structured pruning

- Group lasso regularization[1][2]

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

- Exclusive sparse regularization[3][4]

$$r_{ES}(W) = \|W\|_1^2 = \left(\sum_i |W_i|\right)^2$$

- Group L1/2 regularization[5][6]

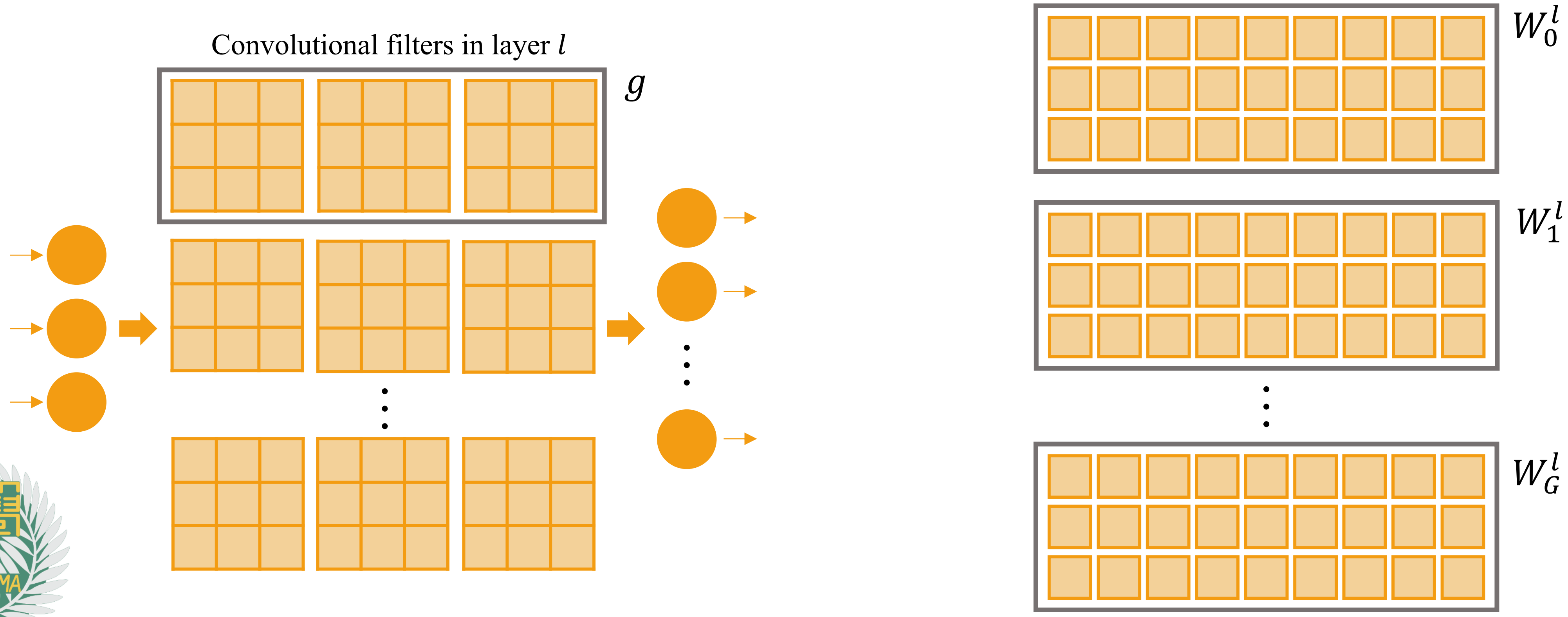
$$r_{GL_{1/2}}(W) = |W|_1^{1/2} = \sqrt{\sum_i |W_i|}$$

HIROSHIMA UNIVERSITY

Kurita Laboratory

The structured sparse regularizations

Grouping of the structured sparse regularization



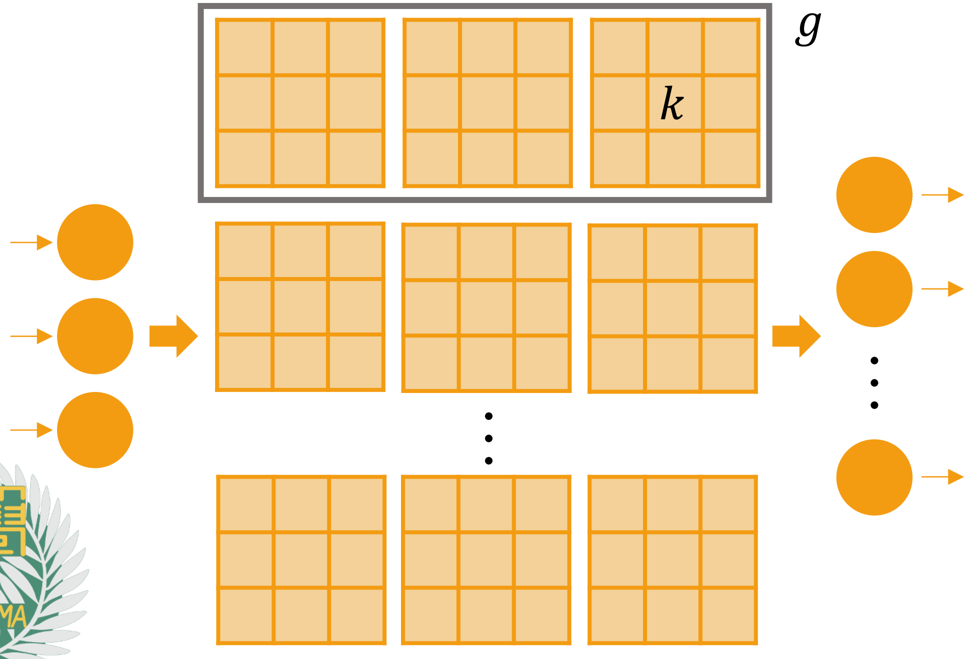
Continued…

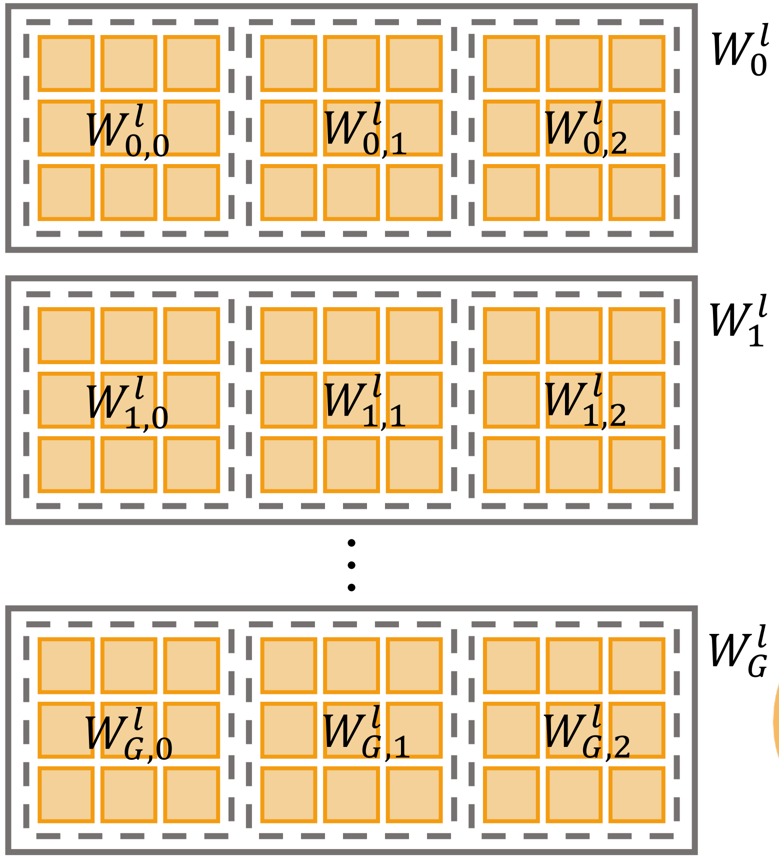
The hierarchical squared/square rooted group sparse regularization[7]

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

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

Convolutional filters in layer l





HIROSHIMA UNIVERSITY

Kurita Laboratory

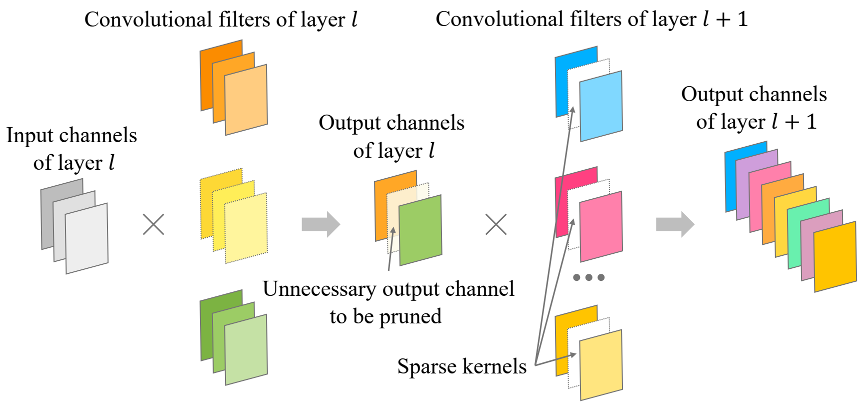
Proposed Method

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

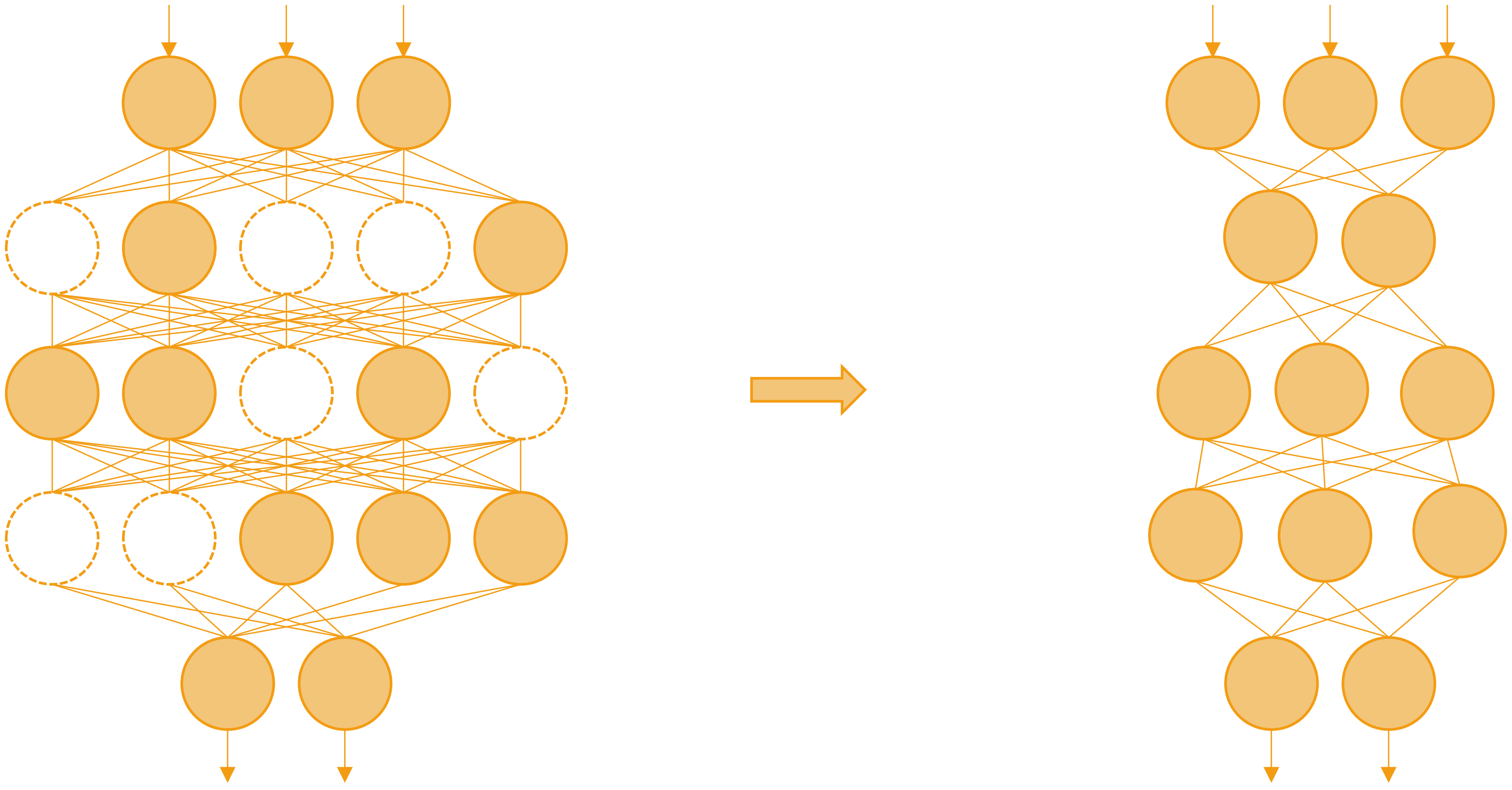
1. Train a large network as the initial network.
2. Train the network with the hierarchical group sparse regularization based on the feature-wise grouping to find unnecessary filters connected to input channels by enforcing their weights to be zero.

The feature-wise group sparse regularization is defined as $R(W^l) = \sum_{j=1}^{c_l-1} r(W^l_{,j,,})$

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.



Continued…



Train from scratch

HIROSHIMA UNIVERSITY

Kurita Laboratory

Experimental setting

Networks and datasets

- VGG14 and ResNet20/32 on CIFAR-10
- VGG14 and ResNet20/32 on CIFAR-100
- VGG14 and ResNet18/34 on TinyImageNet-200

Sparse Regularization

- Hierarchical squared group L1/2 regularization

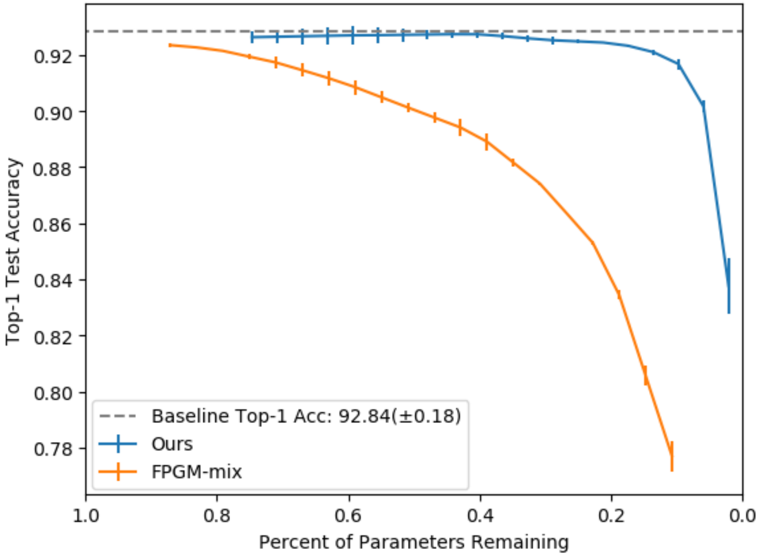
Comparison Method

- FPGM-mix[8] (one of the state-of-the-art method)

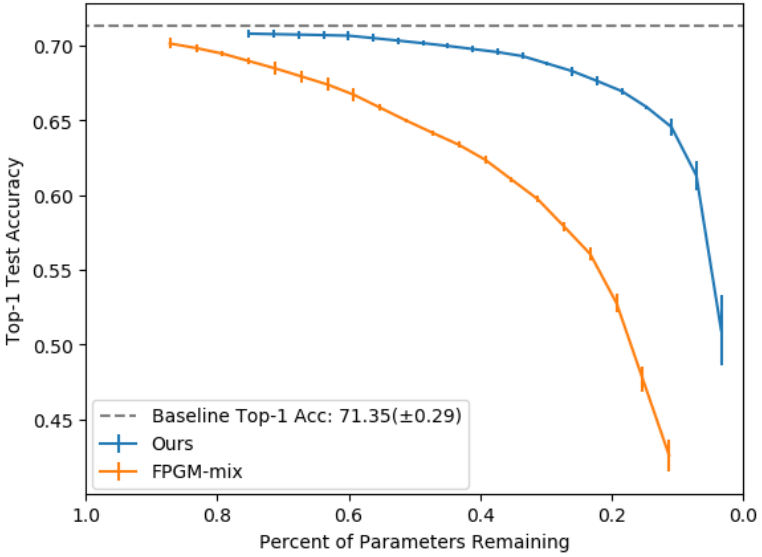
Results

VGG14

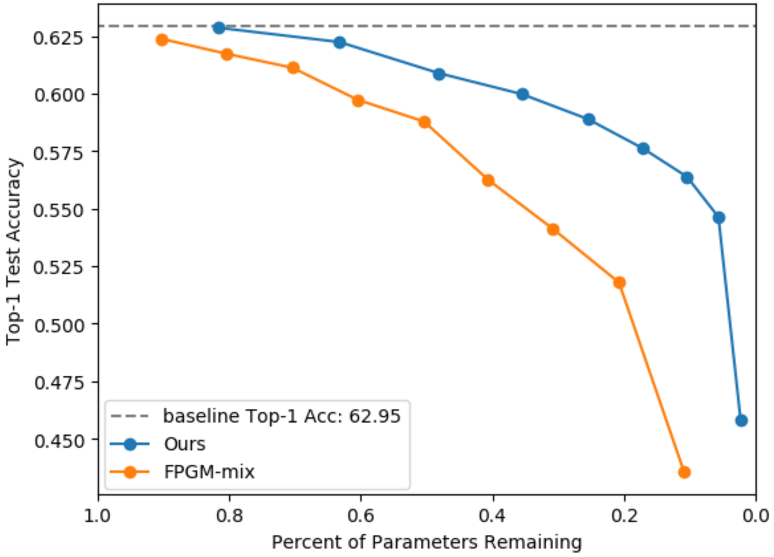
Ours
FPGA-mix



CIFAR-10



CIFAR-100



TinyImageNet-200

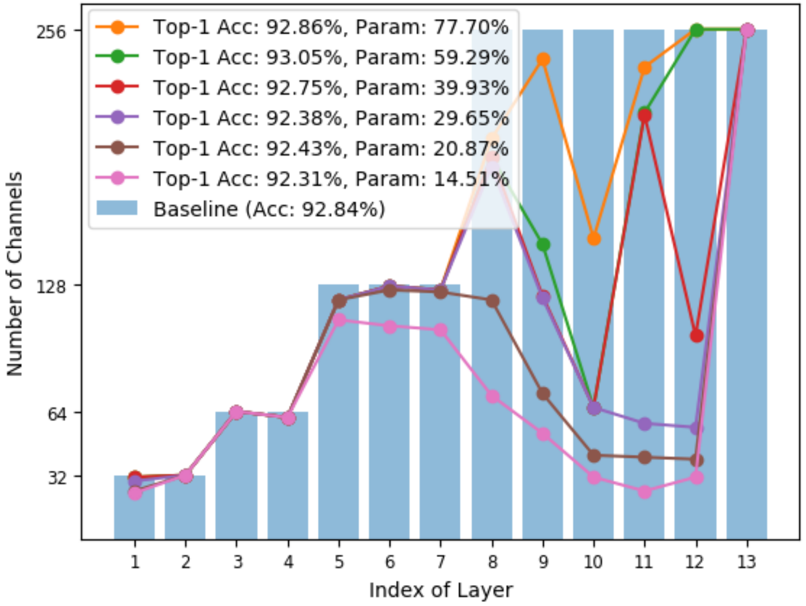
HIROSHIMA UNIVERSITY

Kurita Laboratory

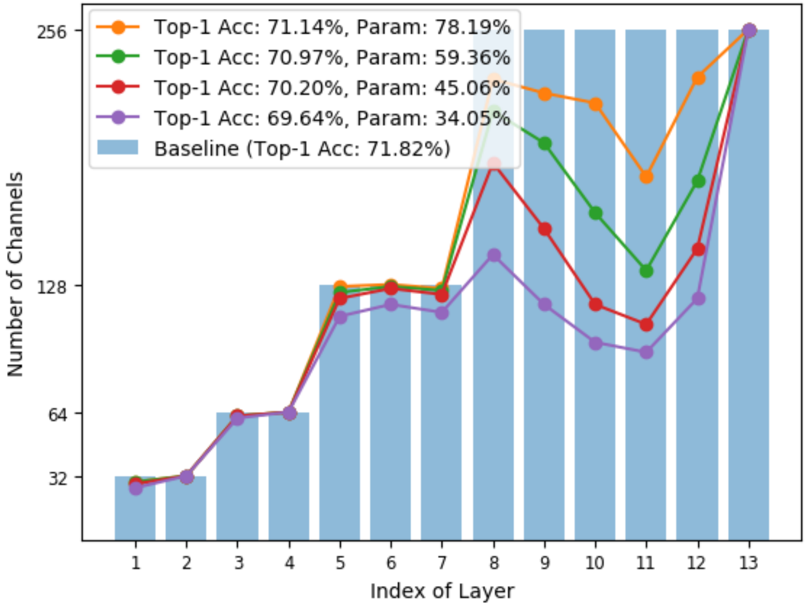
Results

VGG14

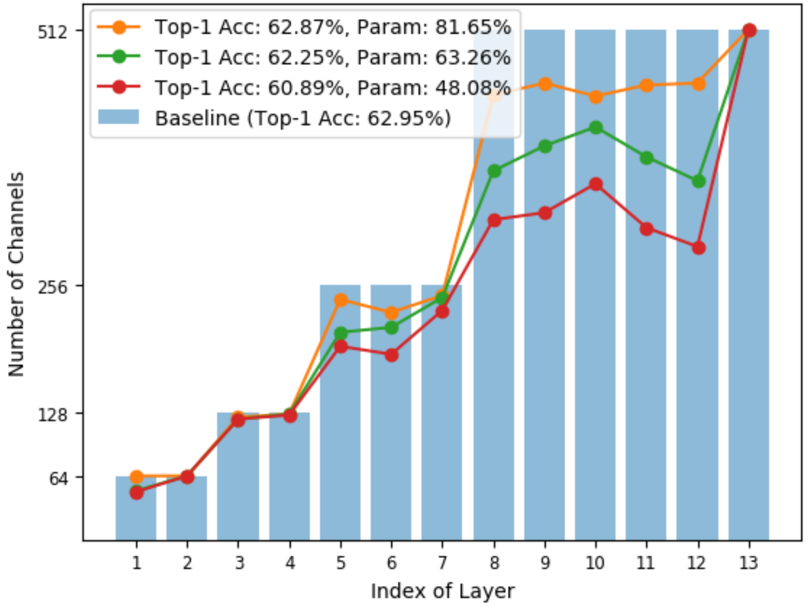
The numbers of the pruned channels in each layer



CIFAR-10



CIFAR-100



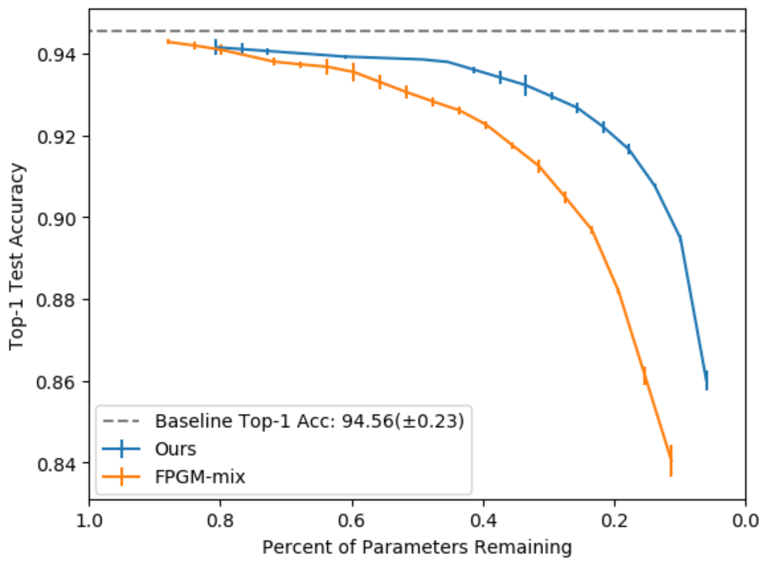
TinyImageNet-200

HIROSHIMA UNIVERSITY

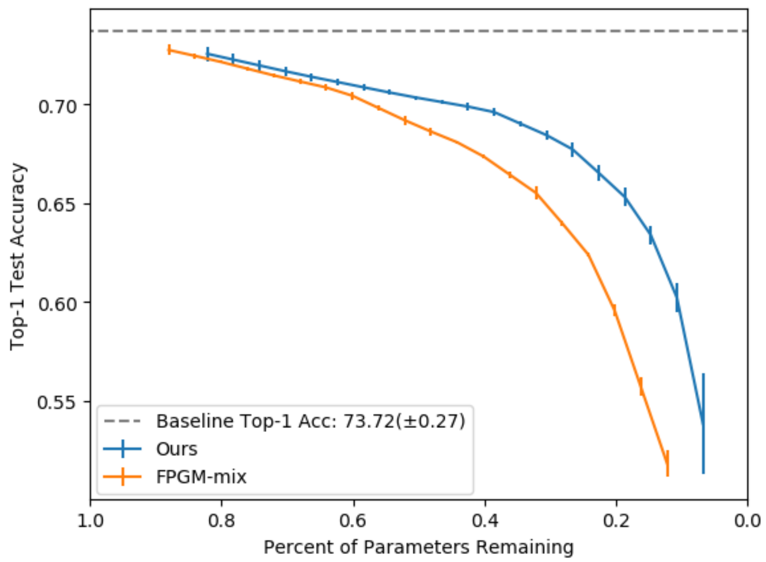
Kurita Laboratory

Results

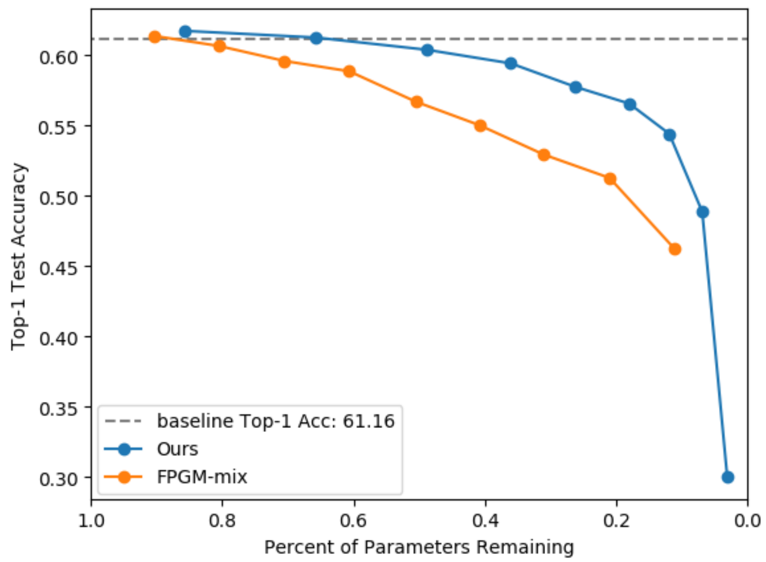
ResNet18/20



CIFAR-10



CIFAR-100



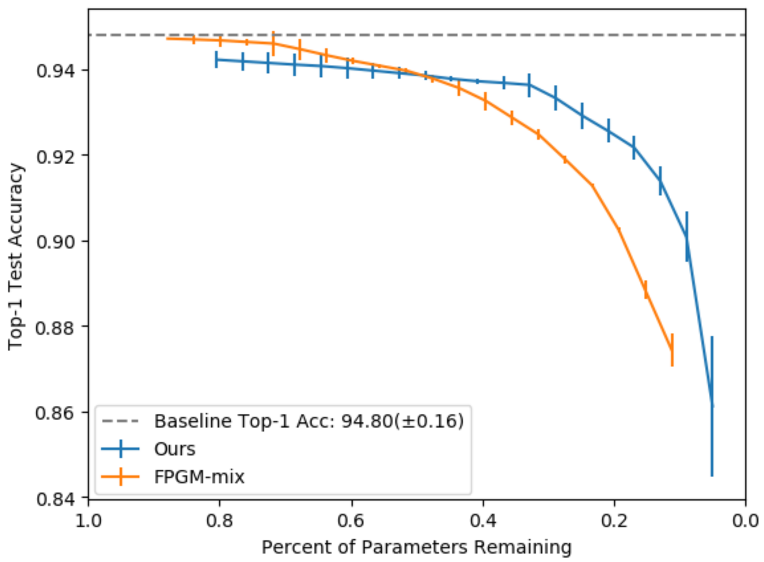
TinyImageNet-200

HIROSHIMA UNIVERSITY

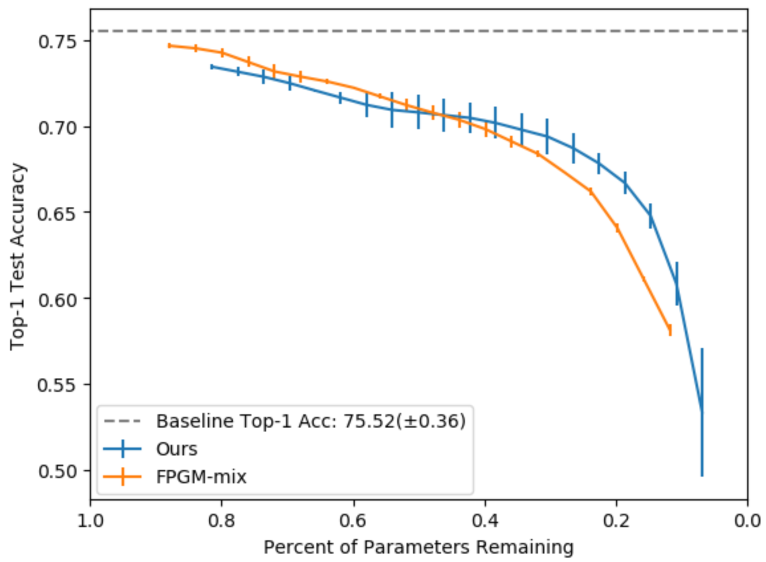
Kurita Laboratory

Results

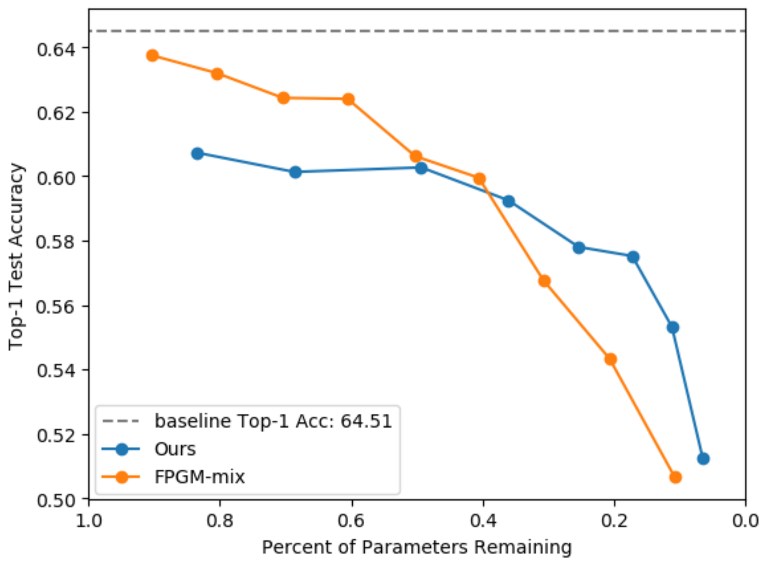
ResNet32/34



CIFAR-10



CIFAR-100



TinyImageNet-200

HIROSHIMA UNIVERSITY

Kurita Laboratory

Conclusions

- Proposed a new **filter pruning method** with the hierarchical group sparse regularization based on the **feature- wise grouping**.

- The strategy of 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, especially when **more than 50% of the parameters are pruned**.

References

[1] Yuan et al. Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68(1):49–67, 2006.

[2] Mark Schmidt. Graphical model structure learning with l_1-regulariza-tion.University of British Columbia, 2010.

[3] Zhou et al. Exclusive lasso for multi-task feature selection. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pages988–995, 2010.

[4] Kong et al. Exclusive feature learning on arbitrary structures via $l_{1,2}$-norm. In Advances in Neural Information Processing Systems, pages 1655–1663,2014.

[5] Li et al. Smooth group $l_{1/2}$ regularization for input layer of feedforward neural networks. Neurocomputing, 314:109–119, 2018

[6] Alemu et al. Group $l_{1/2}$ regularization for pruning hidden layer nodes of feedforward neural networks. IEEE Access, 7:9540–9557, 2019.

[7] Mitsuno et al. Hierarchical group sparse regularization for deep convolutional neural networks. In *Proceedings of the international joint conference on neural networks*, 2020

[8] Heet al. Filter pruning via geometric median for deep convolutional neural networks acceleration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4340–4349, 2019.

HIROSHIMA UNIVERSITY

Kurita Laboratory

Thank you

HIROSHIMA UNIVERSITY

Kurita Laboratory