Convolutional Neural Networks with Skeleton Kernel

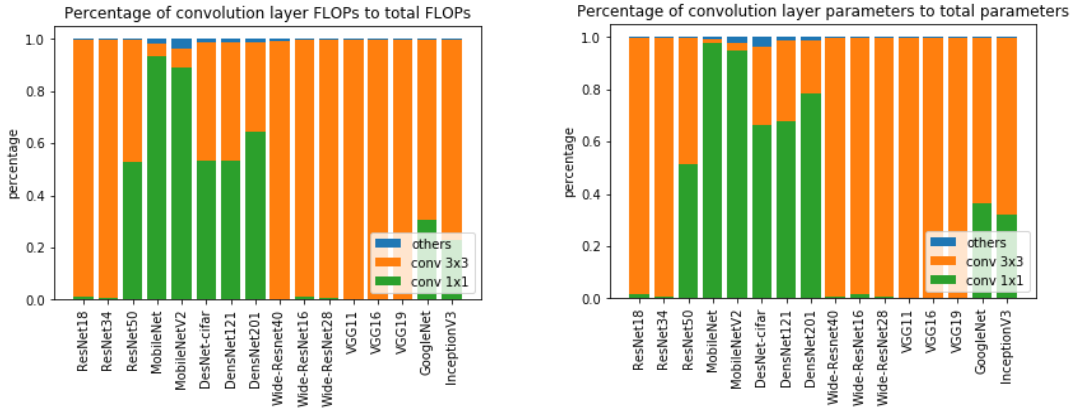
Weihao Zhuang, Tristan Hascoet, Ryoichi Takashima, Tetsuya Takiguchi, Yasuo Ariki

Contributions

- Reducing FLOPs and parameters of 3x3 convolution by 44%.
- Less than 1% dropping of accuracy training CIFAR-10/100 and ImageNet dataset of most models.

Introduction

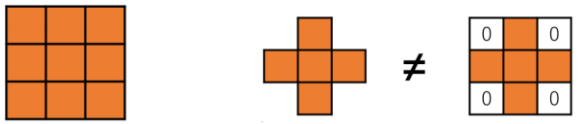
- Convolutional Neural Networks usually relay on 3x3 or larger kernel size to aggregate information.
- Parameters and FLOPs are growing quadratically concerning kernel size and take up most of the computation resources during the training step.
- Figure 2 indicates that the 3x3 convolution layer accounts for most parameters and FLOPs in many models.



(a) FLOPs (b) Parameters

Fig. 2 The percentage of total parameters and FLOPs occupied by convolution layer in Convolutional Neural Networks models.

Method



(a) Normal 3x3 kernel. (b) Proposed skeleton kernel.

Fig 1. Kernel shape between commonly used 3x3 kernel (a) and proposed skeleton kernel (b). Skeleton kernel takes up only five elements in memory, instead setting four elements to 0 on 3x3 kernel.

- In [1], the author found that the skeleton part of 3x3 kernel made the greatest contribution to the model after training CIFAR-10 datasets with ResNet50. Following their work, we asked the following question: Can we directly adopt skeleton part of 3x3 kernel to train our model?
- Replacing 3x3 convolution kernel to our skeleton kernel, which only using 5 element in computing devices.

[1] Ding, X., Guo, Y., Ding, G., & Han, J. (2019). Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1911-1920).

Experiments

Table 1 Results of CIFAR-10/100 training.

Models	C10 Acc	C10 Acc (ours)	C100 Acc	C100 Acc (ours)	Para (M)	Para(M) (ours)	MFLOPs	MFLOPs (ours)
ResNet50	96.62	94.85	77.66	76.14	23.51	18.49	1304.73	1035.51
MobileNetV2	93.69	93.08	62.56	61.64	2.30	2.27	94.60	91.60
DenseNet201	94.89	94.76	77.28	77.04	18.10	16.50	1379.50	1169.0
Wide-ResNet40	95.13	95.07	74.79	74.69	2.25	1.26	358.30	200.47
VGG16	94.27	94.15	72.68	72.60	14.73	8.19	314.03	174.71
GooGleNet	95.35	95.08	79.24	78.50	6.17	4.44	1529.42	1059.31
InceptionV3	95.55	95.50	79.29	79.16	22.13	18.65	3399.40	2838.04

Table 2 Results of ImageNet training.

Models	ImageNet Acc	ImageNet Acc (ours)	Para (M)	Para (M) (ours)	GFLOPs	GFLOPs (ours)
ResNet18	70.35	69.15	11.69	6.81	1.82	1.07
ResNet34	73.30	72.71	21.80	12.43	3.67	2.10
ResNet50	75.83	74.39	25.56	20.53	4.11	3.29

- Using skeleton kernel to replace 3x3 convolutional layer, the accuracy of most models has dropped to less than 1%, but the parameters and FLOPs are significantly reduced when training CIFAR-10/100 and ImageNet dataset.