University of Utah

CS 6350 / MATH 7960 Deep Learning Course

Deep Learning Final Report

Ruyi Ma and Zejian Wu

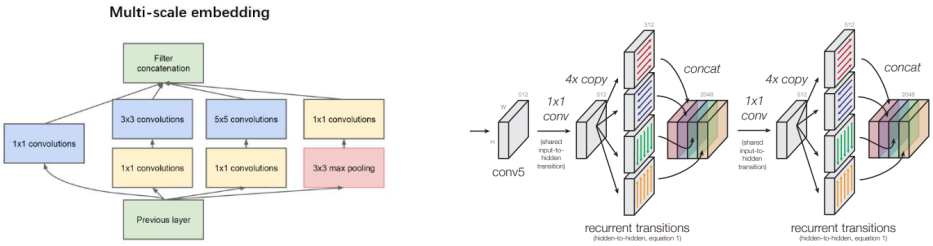
Instrutor: Prof. Berton Earnshaw

The Application of SKNet to Image Classification in Deep Learning

December 16, 2022

0.1 Introduction

In the neuroscience community, the receptive field size of the optic cortical neurons is regulated by the stimulus, i.e., the receptive field size should be different for different stimuli. Many previous works related to convolutional neural networks (CNNs) only optimize the model by improving the spatial structure of the network, such as the Inception module (Szegedy et al., 2015) by introducing different sizes of convolutional kernels to obtain information on different receptive fields. Or inside-outside networks (Bell et al., 2016) refer to spatial contextual information, etc. As shown in the following:



However, only one type of convolutional kernel is generally used in the same layer when building traditional CNNs, which means that the convolutional kernel size is determined for a specific model for a specific task, and the role of multiple convolutional kernels is rarely considered.

There are two common exciting questions:

1. Is it possible to incorporate feature dimension information into the network?

2. Is it possible to adjust the receptive field size based on multiple input information scales?

Based on the first idea, SENet (Squeeze-and-Excitation Networks) (Hu et al., 2018) was created and won the 2017 ImageNet classification competition. SKNet (Selective Kernel Networks) (Li et al., 2019) was born to solve such an exciting problem that we are focused on according to the second idea.

0.2 Related Work

Previous work on object recognition has been well-studied. Here are four typical categorizations which includes multi-branch convolutional networks, Grouped/depthwise/dilated convolutions, Attention mechanisms, and Dynamic convolutions.

Here is the summary of the related methods:

Related work	Typical Network
Multi-branch convolution networks	ResNet, shake-shake network, multi-residual network, FractalNets and Multilevel ResNets, InceptionNets
Grouped/depthwise/dilated convolutions	AlexNet, ResNeXts, IGCV1; IGCV2, IGCV3, Xception, MobileNetV1
Attention mechanisms	CNN, SENet, BAM, CBAM
Dynamic convolutions	Spatial transform networks, Dynamic filter, Active convolution, Deformable convolution networks

We are motivated by these concurrent methods to investigate deeper.

0.3 Dataset Overview

The CIFAR-10 (Alex, 2009) dataset has 60,000 color images, 32*32, and is divided into 10 classes with 6,000 images in each class. 50,000 of these are used for training, constituting 5 training batches of 10,000 images each; the other 10,000 are used for testing, constituting a separate batch:

- data – a 10000x3072 numpy array of uint8s. Each row of the array stores a 32x32 color image. The first 1024 entries contain the red channel values, the next 1024 the green, and the final 1024 the blue. The image is stored in row-major order so that the first 32 entries of the array are the red channel values of the first row of the image.
- labels – a list of 10000 numbers in the range 0-9. The number at index i indicates the label of the ith image in the array data.

0.4 Network Architectures

The main concern as mentioned previously, is that different sizes of receptive fields have different effects. Therefore, the goal is to make the network automatically exploit the information captured by receptive fields.

The authors of SKNet propose a dynamic selection mechanism for convolutional kernels in CNNs that allows each neuron to adaptively adjust the size of its receptive field (convolutional kernel) according to the multiscale of the input information. The inspiration is that the receptive field size of visual cortical neurons is adjusted according to the stimulus. Specifically, a building block called selective kernel unit (SK) is designed, in which multiple branches with different kernel sizes are fused using SoftMax guided by the information in these branches. A SKNet is formed from multiple SK units, and the neurons in the SKNet are able to capture target objects at different scales.

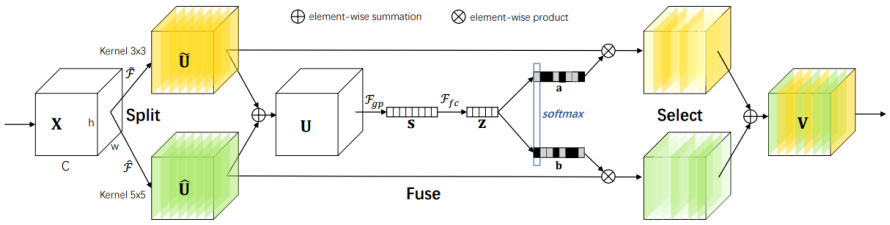


Figure 1. Selective Kernel Convolution.

This network is divided into three primary operations Split, Fuse, and Select.

- Split: generates multiple paths with different kernel sizes. The model in the figure above is designed with only two convolutional kernels of different sizes. It is possible to design multiple convolutional kernels with multiple branches.
- Fuse: combines and aggregates information from multiple paths to obtain a global and composite representation for the selection weights.
- Select: aggregates the feature maps of different size kernels based on the selection weights.

0.5 Evaluation and Results

We conducted an experiment on CIFAR-10 to evaluate the performance of SKNet. We compared the performances of SKNet and ResNeXt.

When training the model, the loss function is the MSE loss between generated images and the original images. We checked the accuracy rate when testing the model to evaluate our results.

We trained each model for 100 epochs. The results are shown in the following table.

Compared with ResNeXt, the loss of SKNet decreased from 74.05 to 56.00, and the accuracy rate of SKNet improved from 0.8657 to 0.8918 after going through 100 epochs, Therefore, SKNet outperforms ResNeXt in accuracy in our experiment.

We also compared the processing time of SKNet and ResNeXt. In our experiment, the processing time of SKNet was 3.4 times the processing time of ResNeXt.

Methods	Loss	Accuracy
ResNeXt	74.05	0.8657
SKNet	56.00	0.8918

0.6 Discussion and Future Improvement:

Discussion:

- SKNet enables the network to acquire information on different receptive fields, which may be a network structure with better generalization ability.
- The soft attention used in the Select of SKNet is similar to the weighting operation of the feature map in the Squeeze-and-Excitation block. The difference is that the Squeeze-and-Excitation block considers the weights between Channels. It may provide us with a new perspective to rethink the problem.
- Although the Inception network is well-structured and effective, it is a breakthrough that SKNet can adjust its structure to obtain information from different receptive fields.
- SKNet provides more accuracy than ResNeXt.
- Training SKNet costs more time, compared with ResNeXt.

Future Improvement:

- It took a long time to train SKNet. Although we used a server, the training speed is still slow. If we have more time, we can spend more time in testing different combinations of kernels.

0.7 References:

1. X. Li, W. Wang, X. Hu and J. Yang, Selective Kernel Networks, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019.
2. S. Xie, R. Girshick, P. Dollar, Z. Tu and K. He, Aggregated Residual Transformations for Deep Neural Networks, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
3. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, Going Deeper with Convolutions, 2015 IEEE Conference on Computer Vision and Pattern Recognition, 2015.
4. Bell S, Zitnick C L, Bala K, et al., Inside-Outside Net: Detecting Objects in Context with Skip Pooling and Recurrent Neural Networks, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.

5. J. Hu, L. Shen and G. Sun, Squeeze-and-Excitation Networks, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.

6. Alex Krizhevsky, Learning Multiple Layers of Features from Tiny Images, 2009.

7. K. He, X. Zhang, S. Ren and J. Sun, Delving Deep into Rectifiers: Surpassing Human-level Performance on ImageNet Classification, 2015 IEEE International Conference on Computer Vision, 2015.

8. K. He, X. Zhang, S. Ren and J. Sun, Deep Residual Learning for Image Recognition, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016

9. L. Itti and C. Koch, Computational Modelling of Visual Attention, Nature Reviews Neuroscience, 2001.

10. https://github.com/pppLang/SKNet, 2019.