Reading summary: ImageNet Classification with Deep Convolutional Neural Networks
T11 - PangCloud Members: Su Pu, Yicheng Wang, Tianyang Liu
Link:https://docs.google.com/document/d/1qwXwKH9BwFdXcISdsVlHXu3w2YdDGv-hx2G_rmdXA7Q/edit

Introduction (Tianyang Liu):

This paper is proposed in 2012 ImageNet, the CNN method in this paper achieved a wining top-5 rest error rate of 15.3%, and the second-best entry is 26.2%. The machine learning methods are currently very popular for object recognition. However, in order to learn about thousands of object form millions of images, we need a model with a large learning capacity, however, the ImageNet is so large that the immense complexity of the objects in this database cannot be easily processed. The convolution neural networks have many models which contains different depth and breadth, at the same time, they make the mostly correct assumptions about the nature of images, in a word, the CNN have fewer connections and they are more easily to be trained.

The paper tested their method on the largest convolutional neural on the ILSVRC-2012 neural network. In order to accelerate the training process and process more data at the same time, the paper implement their neural network on the GPU.

The architecture adopts many advanced technologies in recent machine learning technology, the Rectified Linear Units (ReLUs) activation method proposed by Nair and Hinton [1], the non-saturating nonlinearity structure of ReLUs are faster than the saturation nonlinearities structures of other structure.

The structure contains five convolution layers, each convolutional layers followed by an overlap poling layer, the convolutional layers extract the features of the images, the kernels of each convolution layer slide through the whole picture and calculate the convolution result of the kernel and the picture, the pooling layers extract the important convolution layer. After these process, the parameters go through a fully connected layer, then the paper calculate the loss and update the weights in the neuron network.

The paper also introduces ways to reduce overfitting, the first is data augmentation, the paper artificially enlarge the dataset using label-preserving transformations. The second method is dropout, the drop out means to set zero of the output of each layer with the probability of 0.5.

The result is very promising and achieves top-1 and top-5 test set error rate of 37.5% and 17.0%.

Three strong points of the paper (Su Pu):

1. First implemention of deep learning model with large database.

It is the first method which use the convolution network computation for image classification. The traditional like support vector mahine (SVM), k-nearest neighbors algorithm(KNN) and other method, they only contains one or two layers and the performance is not very good, with the support of advanced calculation technology, the deep convolution neuro network contains more features and gather a more accuracy result.

2. Rectified Linear Units (ReLUs)

The ReLUs which have the formula f(x) = max(0,x) has larger gradient descent than the tanh's and similar method which have the formula f(x) = tanh(x), therefore the ReLUs can accelerate the training process to some extent.

3.Dropout

When train the data, the network only randomly chooses 50% of the neurons (the weights to the rest are set to zero), which is similar to traditional average method to MLP: For example, use the same training data to train 5 separate neural network. And after all the NNs are trained, we can use the average values as the final NN. The dropout works similar function here, as for every input, the neural network has a different structure, but the parameters inside share with each other, thus to avoid overfitting

Three weak points of the paper (Yicheng Wang):

1.There is an obvious mistake in the paper that the size of input images in the paper 224*224 can not lead to exactly 55*55 activation map in the second layer, so the proper input should be 227*227
2.Compared with latest model like GoogleNet or ResNet, there are still too many parameters which will lead to a really long training time.
3.The last layer is fully connected layer which has tremendous parameters, which almost cost 90% of total number of parameters, and will also lead to overfitting

Future work brainstorming:
1. The structure of the CNN is still very simple, we can enhance the structure by creating more connections between each layer, then the first convolution layer information will not be vanish.
2. The function of the normalization seems not that important, therefore, we can try the network without this function.

Reference:
[1] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.
[2] Ciregan D, Meier U, Schmidhuber J. Multi-column deep neural networks for image classification[C]//Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012: 3642-3649.
[3] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.