**Let’s keep it simple,**​

**Using simple architectures to outperform deeper and more complex architectures.**

**paper reading report**​

​Jinhe Fan UFID:41328609

Department of Electrical and Computer Engineering, University of Florida

Gainesville, FL

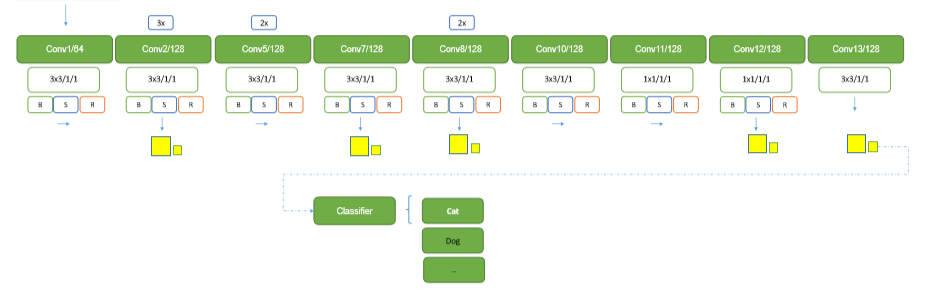
1. CONTEXT

Convolutional Neural Networks(CNNs) is a mapping from input to output in essence, it can study the mapping relationship between input and output , without precise mathematical expression between any input and output, as long as the known pattern of convolutional network training, network has the mapping ability between the input and output.

Convolutional Neural Network is a kind of advanced neural network and its have powerful function, it has achieved great success in the identification and classification, image NLP classification and so on, it has been widely used in industry, for example, it will be used for Google image search, Amazon and it will be used for product recommendation. CNNs mainly have three shortcomings. Shortcoming one: the realization is more complex. Shortcoming two: training takes a long time. Shortcoming three: Not a identical algorithm, different tasks need to be trained individually.

1. SUMMARY

This paper present a well-crafted simple architecture, which called SimpleNet. SimpleNet have mainly several advantages. Advantage one: Its with minimum reliance on new features. Advantage two: SimpleNet can be used in the embedded devices. Advantage three: Its occupy the memory consumption is small because it can be further compressed using technique. Advantage four: It can boost the performance and has a good compatibility. Advanced five: With fewer parameters, it can perform either superior to deeper architectures or on par. The simple convolutional network is provided with 13 layers. The network using 3×3 kernels for convolutional layer expect 11th and 12 th layers (1 × 1 kernel) and 2×2 kernels for pooling operations. In the problem of vanishing gradient, they used batch-normalization before any ReLU non-linearity. They also used the CIFAR10/100 as datasets.



1. RELATED WORKS

This paper consider several principles to design their architecture.

1. This paper recommend a deeper network with thinner layers, which tends to perform better than the same network being much shallower with wider layers.
2. The simple convolutional network is provided with 13 layers. The network using 3×3 kernels for convolutional layer expect 11th and 12 th layers (1 × 1 kernel) and 2×2 kernels for pooling operations. In the problem of vanishing gradient, they used batch-normalization before any ReLU non-linearity. They also used the CIFAR10/100 as datasets.
3. The paper suggests to replace 1×1 kernels with 2×2 kernels. Using 2×2 kernels both help to reduce the number of parameters and also to retain neighborhood information.
4. Utilize as much information as it is made available to a network can avoid rapid down sampling especially in first layer.
5. 3 × 3 kernels harness capture a larger area of neighborhood in the input, using them can filter noises and thus capturing better features.

.

IV ADVANTAGES

1. This paper a better valuable academic point and explain the main contents and points in the design briefly and emphatically.
2. This paper present a new opinion, which promoting academic exchanges, promotion of achievements and development of science and technology.
3. This paper not only write the principle of design but also describe the solution of latent problems.

V DISADVANTAGES

1. This paper present some new opinions which are different from previous opinions. But some opinions just present, which are not proved by experiment.
2. They use caffe for training their architecture, which may produce unpredictable result in the future. Because the Caffe's code structure is highly hierarchical, some research and applications are difficult to implement.
3. Some opinions lack of experiment analysis. For example, They present ‘using 2×2 kernels both help to reduce the number of parameters and also to retain neighborhood information’, but there is no experiment analysis to support their opinion.
   1. BRAINSTORMS

This paper presented a simple convolution architecture that takes advantage of the simplicity in its design. If an architecture harness deeper convolutional layer and combine to SimpleNet, we can get a new architecture with accuracy and occupy the memory consumption is small .This kind of architecture can be used in specific area that needs high accuracy and big memory consumption.

Links: https://docs.google.com/document/d/1c7HCV2Xebov03drkorMJqFJ7W\_nJCVIOk2yzDgbd7OM/edit