Review of "Very Deep Convolutional Networks for Large-Scale Image Recognition"

Riyad Bin Rafiq

1. Paper summary

Convolutional Neural Network (CNN) has shown great success in image classification tasks in recent years. The paper [1] implemented a deep CNN architecture including upto 19 weight layers that demonstrated a great performance in ImageNet Challenge 2014. This research used very small (3x3 and 1x1) convolutional layers in the network and showed the importance of depth in visual representations.

2. Contribution

2.1 Importance of depth

The authors represented the depth of the network that ranged from 16 to 19 weight layers. They also showed how the performance of the classification increased as depth increased. Several attempts were made to improve the performance of the original AlexNet but in this research, the authors proposed new architecture and experimented with different depths using very small convolutional layers. This research contributed to understanding the importance of depth in CNN.

2.2 Power of small receptive field

In the research, stacked small convolutional layers were used instead of big ones. For example, the authors applied three 3x3 convolutional layers instead of one 7x7 convolutional layer as both have the same effect in feature mapping. Also, using small receptive fields reduces the number of parameters in the network. This helped in memory consumption and computation time.

2.3 Experimentation with scaled image

In the research, the authors experimented with different scales of image and demonstrated how it performed in classification. In training, they followed two approaches: scaling S is fixed where values are 256 and 384; scaling S varied in the range between 256 and 512. The evaluation was also conducted in both single scale and multi-scale in the test set.

2.4 Comparison and generalization

Another important contribution is that the authors compared their proposed model with other concurrent networks and noted the performance discrepancy. Moreover, they showed how their model generalized well to a wide range of tasks and datasets. For example, the model was

applied to image classification on Caltech-101 and Caltech-256 dataset and action classification on VOC-2012 dataset.

3. Critique

3.1 Focused more on results rather than science

The paper focused more on representing their results rather than explaining the science behind the network. The authors stated that small receptive fields had been used in the network. But they didn't explain much about how it helped to capture the features and created feature maps further.

3.2 Choice of hyperparameter

The authors didn't mention how they ended up choosing the batch size, momentum, and dropout. Were there any other values that didn't work well? If so, why didn't the values work well? Any explanation regarding the network architecture?

3.3 ConvNet configurations

The depth of configurations included 11, 13, 16 and 19 weight layers. In the research, the authors stated that the depth increased performance accuracy and mentioned the importance of the depth in visual representations. So a question emerges as to why they didn't mention increasing the depth beyond 19 layers. How did it perform then adding more layers to 19 layers? Didn't it classify well? If so, what are the reasons behind it? Will it be considered in the future study? A lot of questions remain unexplained.

Reference

1. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv [cs.CV]. 2014. Available: http://arxiv.org/abs/1409.1556