Assignment 2

A review of the scientific paper "Deep Learning and computer vision"

Bendik Nordeng (Student ID: 478128, user name: bendikno)
Peder Møyner Lund (Student ID: 478109, user name: pederml)

Summary of the paper

Deep Residual Learning for Image Recognition, written by Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, discusses the use of deep residual neural networks for image recognition. More specifically, it presents a residual learning framework to ease the training of "very deep" neural networks.

The paper starts with an explanation of convolutional neural networks as a tool for image processing. Furthermore, the authors specify that the network depth "is of crucial importance" (He et al., 2016) for the results of the image classification. The paper then moves on to discuss whether or not "*learning better networks is as easy as stacking more layers*"(He et al., 2016). Based on other sources and own scientific results, the authors quickly note that learning better networks for image classification is far more difficult than just stacking layers in the network. The paper then goes on to discuss a phenomenon referred to as the *degradation problem*. The degradation problem is a case with classical image recognition where the training accuracy of neural networks rapidly worsens with increased depth. The problem is rather counterintuitive, where you would expect a more complex model to overfit and increase only the test error instead.

The authors then delve more into how different network architectures deal with the degradation problem, explaining how the residual networks with identity mapping may improve performance dealing with the degradation problem. This is followed by a comprehensive comparison between residual networks and plain networks – which simply stack layers. Furthermore, the relevant networks are used in ensemble models, combining them with other approaches referenced in the paper. Experiments are then performed to underline the differences between the models. Lastly, the paper looks at test results using other datasets to investigate the general performance of the model. The results show that the proposed residual network outperform all other state of the art models.

Research goals

Addressing the degradation problem, the paper introduces a deep residual learning framework, and tries to prove that the models that takes advantage of this approach greatly outperform plain neural networks (stacked layers with no further considerations) in terms of classification accuracy. The authors hypothesize that “it is easier to optimize the residual mapping than to optimize the original, unreferenced mapping” (He et al., 2016). The model takes advantage of *identity mapping* through shortcut connections in the network, which neither adds extra parameter nor computational complexity. The authors then try to show that the deep residual networks will obtain far greater accuracy and decreased errors with substantially deeper levels.

Research methodology

The paper makes use of theoretical, experimental and analytic methodology. Using an extensive evaluation of the different network architectures, the authors theoretically explain how their approach to a residual model will learn the dataset with greater ease. Through referencing experiments later in the paper, they provide comprehensive empirical evidence that the extremely deep residual nets are both easy to optimize, and that they can easily enjoy accuracy gains from greatly increased depth. Using a variety of network depths, the authors want to give an overview of how the different architectures perform with different dimensions. Graphs and tables are used to visualize the results of the experiments. Using the results from the experiments, the authors further analytically argue and substantiate their claim that the residual learning model is superior.

Results

Using the ImageNet 2012 classification dataset, consisting of 1000 classes, the models explained in the paper was trained on 1.28 million training images and evaluated on 50 000 validation images. The test set consisted of 100 000 test images. For the plain networks, evaluating on a wide range of depths, the degradation problem was verified. The results showed greater error through the whole training procedure for the deeper networks. With 18 layers, the plain model had a test error of 27.94%, while it had a test error of 28.54% with 34 layers. The residual networks, however, showed great promise in the experiments. The situation was reversed, where the 18-layer network had a test error of 27.88% and the 34-layer network had a test error of 25.05%, which is an improvement of 2.83%. With accuracy gains with increased depths, this is an indication that the degradation problem has been well addressed. With 152 layers, the residual network obtained a top-1 (meaning that the target label is the model's top prediction) error of 19.38% and a top-5 (target label among the model's top five predictions) error of 4.49%.

Combining the residual networks with other models referenced in the paper, the ensemble models were able to further improve the error rates. Obtaining a top-5 test error of 3.57%, this model won the 1st place in ILSVRC 2015.

Furthermore, the models were tested on the CIFAR-10 dataset. This dataset contains 50 000 training images and 10 000 test images in 10 classes. Also with this dataset, the residual networks managed to overcome the optimization difficulty and demonstrate accuracy gains when depth was increased. Testing with layer sizes 20, 32, 44, 56, 110 and 1202, results were visualized with an even greater range than with the ImageNet dataset. Also on this dataset, the degradation problem affected test results for the plain model, with test error rates higher than 60% when only using 110 layers. With the residual networks, performance improved for models with up to 110 layers. With the 1202-layer model, error seemed again to increase. However, the authors mention that there were no optimization difficulty, and that it was able to achieve training errors below 0.1%. This was also the case with the 110-layer model.

Evaluation of the results

The results are evaluated and analyzed throughout the paper. Firstly, the optimization difficulty causing degradation with the plain network is questioned to be caused by vanishing gradients. Following this is a discussion arguing that this is not the case, showing that neither forward nor backward signals can vanish. This makes for a greater degree of credibility in the results obtained. Different aspects with each architecture are discussed in depth, e.g. looking at which components that are and are not needed to address the degradation problem, and also complexity issues. The results are compared. Furthermore, the text continuously makes references to the results found in the experiments to substantiate the analysis of the different networks. Comparing the results to current "State-of-the-art"-methods, the authors strengthen their claims about their own model.

Using the excessively deep 1202-layered models for the CIFAR-10 dataset, the authors are testing the boundaries of their models. They further argue that the lack of performance with these models are caused by overfitting, and also that they may be unnecessarily large for CIFAR-10, being a fairly small dataset. This is a reasonable approach, where the goal is to check all aspects of the model, both strengths and weaknesses.

Personal opinion of the paper

In this paper, the authors thoroughly analyze the cause and effects of vanishing gradients in neural networks and demonstrate an effective solution to the problem with ultra-deep networks. There are several interesting aspects of this paper. Firstly, the topic of this paper is very relevant today and a solution to the problem can have significant impact on future research and the performance on image processing. A second positive thing with this paper is the originality of solution proposed by the authors, using past approaches to compensate for information loss. The solution could also be applied to other neural networks with different depths. A third aspect of the paper is its structure and use of illustrations to convey the research and its findings. This makes it easy for the reader to understand the topic.

Despite being an interesting scientific paper, discussing a relevant topic, there are some aspects that could be improved. Firstly, the focus of the authors was, as they state, on “extremely deep” networks. It would be interesting to evaluate if their solution could be applied to shallower networks as well, which could further strengthen the idea of using residual blocks. Secondly, the paper could address several ways of improving the computational cost due to deep-layered networks.

Reference

[He et al., 2016] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), *pages* 770–778.