Note on “Deep Residual Learning for Image Recognition”

Created on 2024-1-23
Author Di Yu (yudi.0211@foxmail.com)

Introduction

Deep convolutional neural networks are a powerful tool for image classification tasks. Deep networks are capable of capturing high-level features and can approximate complicated mappings with higher accuracy. For these reasons, deep neural networks may exhibit better performance than shallow ones in image recognition problems. However, training such deep learning models is difficult because the training accuracy usually degrades as the network depth increases (see Fig. 1). This problem of training accuracy hinders the demonstration of the full potential of deep neural networks, which could solve image recognition problems with higher accuracy.

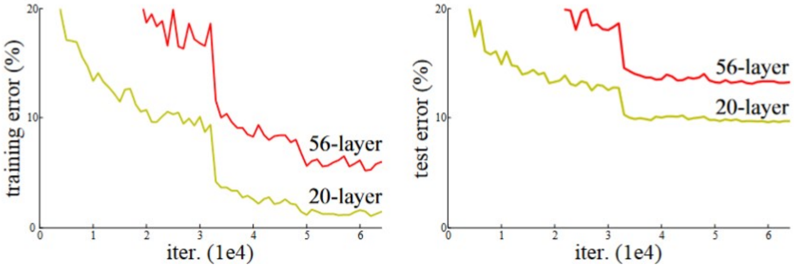

Figure 1: Training accuracy reduces with network depth [1]

In 2016, Kaiming He published a paper titled “Deep Residual Learning for Image Recognition” [1] with his colleagues at Microsoft. This paper was selected as the best paper at CVPR in that year. In this paper, Kaiming He proposed the residual learning framework (known as ‘ResNet’), a novel building block for deep neural networks that features significantly improved training accuracy. The basic idea is to introduce a shortcut connection (illustrated in Fig. 2), allowing information to flow and skip multiple nonlinear layers. This shortcut connection makes it easier for multiple stacked layers to approximate an identity mapping, which is key to training deep learning models.

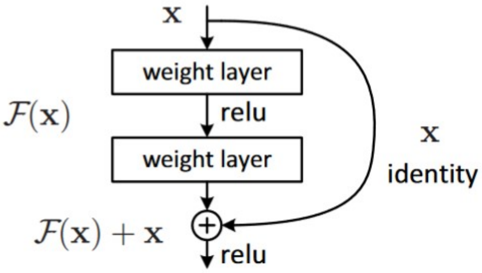
x
weight layer
$\mathcal{F}(\mathbf{x})$
relu
weight layer
x
identity
$\mathcal{F}(\mathbf{x}) + \mathbf{x}$
relu

Figure 2: Schematic for ResNet [1]

ResNet has had a significant impact on machine learning research. The ResNet paper has garnered over 200,000 citations as of 2024, establishing itself as a fundamental building block in modern deep learning models. This includes its incorporation into large language models like GPT.

Observations, Comments, and Interesting Facts

- The idea of utilizing shortcut connections to reduce training error was inspired by the observation that nonlinear layers in deep neural networks perform worse than identity mapping. This results in a training error that accumulates as network depth increases, making deep models perform worse than shallow ones.
- While previous leading works utilized neural network models with a depth of 16 to 30, the ResNet architecture enables a depth of up to 152 layers.
- The extremely deep residual network performs surprisingly well in experiments; it won the 1st prize in multiple important competitions in 2015, including ILSVRC classification competition, ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation, demonstrating the effectiveness and generality of the ResNet architecture.
- Kaiming He realized that the solver is responsible for the reduction in training accuracy as network depth increases, instead of the model itself. Remarkably, he chose to modify the model to ease the optimization process by the solver, rather than altering the solver configuration.
- This choice might be motivated by the high cost of any modification to the solver, which has by that time become a cornerstone of common deep learning frameworks. In contrast, a transition between plain networks without shortcut connections to ResNet might be easier since ResNet can be optimized with a traditional solver using backpropagation, without entailing additional computational cost.
- One important reason for ResNet is that for deep neural network models, the optimal solution for each block is approximately an identity mapping, for which a ResNet can easily approximate by simply nullifying all parameters.
- When multiple ResNet blocks are stacked together, the shortcut connections form a series connection, allowing information to flow directly from input to output. It turns out that such long-range connections play an important role in finding the optimal solution when training a complicated deep learning model.

Reference

- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).