Deep Residual Learning for Image Recognition (ResNet Explained)

ALGORITHM | COMPUTER VISION | DEEP LEARNING | IMAGE ANALYSIS | INTERMEDIATE | RESEARCH & TECHNOLOGY

Introduction

Deep learning has revolutionized computer vision and paved the way for numerous breakthroughs in the last few years. One of the key breakthroughs in deep learning is the ResNet architecture, introduced in 2015 by Microsoft Research. In this article, we will discuss the ResNet architecture and its significance in the field of computer vision.

Learning Objectives

- Address the problem of vanishing gradients in deep neural networks by introducing residual connections.
- Propose a new architecture for deep residual networks (ResNet) that uses residual connections.
- Features of ResNet
- Evaluate the performance of ResNet on the ImageNet dataset and compare it with traditional deep neural networks.
- Learn the effectiveness of the residual connections in addressing the vanishing gradient problem and enabling the training of very deep networks with up to 152 layers.

This article was published as a part of the Data Science Blogathon.

Table of Contents

What is ResNet, and Why was it Proposed?

ResNet stands for Residual Neural Network and is a type of convolutional neural network (CNN). It was designed to tackle the issue of **vanishing gradients** in deep networks, which was a major hindrance in developing deep neural networks. The ResNet architecture enables the network to learn multiple layers of features without getting stuck in local minima, a common issue with deep networks.

What are the Features of ResNet?

Here are the key features of the ResNet (Residual Network) architecture:

- **Residual Connections:** ResNet incorporates residual connections, which allow for training very deep neural networks and alleviate the vanishing gradient problem.
- **Identity Mapping**: ResNet uses identity mapping as the residual function, which makes the training process easier by learning the residual mapping rather than the actual mapping.
- **Depth**: ResNet enables the creation of very deep neural networks, which can improve performance on image recognition tasks.
- **Fewer Parameters**: ResNet achieves better results with fewer parameters, making it computationally more efficient.
- **State-of-the-art Results:** ResNet has achieved state-of-the-art results on various image recognition tasks and has become a widely used benchmark for image recognition tasks.
- **General and Effective Approach:** The authors conclude that residual connections are a general and effective approach for enabling deeper networks.

How ResNet Works?

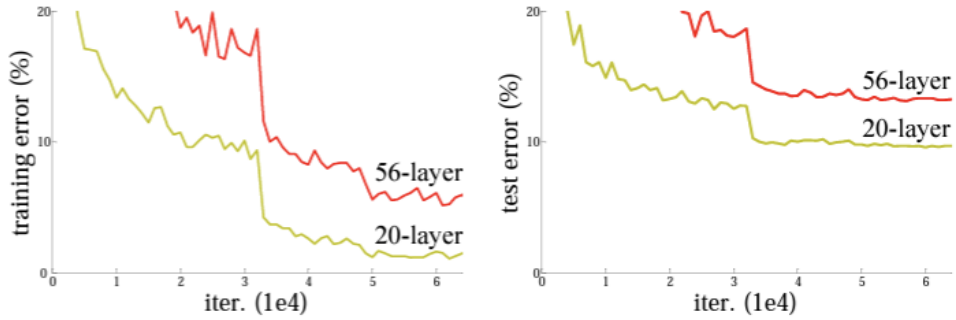
- ResNet works by adding residual connections to the network, which helps to maintain the information flow throughout the network and prevents the gradients from vanishing.
- The residual connection is a shortcut that allows the information to bypass one or more layers in the network and reach the output directly.
- The residual connection allows the network to learn the residual function and make small updates to the parameters, which enables the network to converge faster and achieve better performance.
- This enables the network to learn residual functions and helps the network to converge faster and achieve better performance.
- The residual connection is based on the idea that instead of trying to learn the complex mapping between the inputs and the outputs, it is easier to learn the residual function, which maps the inputs to the desired outputs.

Paper Analysis: Deep Residual Learning for Image Recognition

Paper link: https://arxiv.org/abs/1512.03385

Problem Statement

Deep Neural Networks provide more accuracy as the number of layers increases. But, when we go deeper into the network, the accuracy of the network decreases instead of increasing.



Source: arxiv.org

An increase in the depth of the network increases the training error, which ultimately increases the test error. Here, the training error on a 20-layer network is less than that of 56 layer network. Because of this, the network cannot generalize well for new data, which becomes inefficient. This degradation indicates that the increase in the model layer does not aid the model's performance.

- When more layers are piled, there occurs a problem of vanishing gradient. The paper mentions that vanishing gradients have been addressed by adding the intermediate normalization layers.
- Also, the degradation is not caused by overfitting.

What Might be the Reason for Degradation, and How to Resolve it?

Adding more layers to a suitably deep model leads to higher training errors. The paper presents how architectural changes like residual learning tackle this degradation problem using residual networks.

Residual Network adds an identity mapping between the layers. Applying identity mapping to the input will give the output the same as the input. The skip connections directly pass the input to the output, effectively allowing the network to learn an identity function.

The paper presents a deep convolutional neural network architecture that solves the vanishing gradients problem and enables the training of deep networks. It showed that deep residual networks could be trained effectively, achieving improved accuracy on several benchmark datasets compared to previous state-of-the-art models.

Current Approach

High-level Contributions of the Current Approach

The authors proposed a method to approximate the residual function and add that to the input. A residual block consists of two or more convolutional layers, where the output of the block is obtained by adding the input of the block to the output of the last layer in the block. This allows the network to learn residual representations, meaning that it learns the difference between the input and the desired output instead of trying to approximate the output directly.

Source: arxiv.org

Mathematical Formulation

H(x) = F(x) + x

Where,

X is the input to the set of layers
F(x) is the residual function
H(x) is the mapping function from input to output

The authors speculated that optimizing the residual mapping is easier than optimizing the original, unreferenced mapping.

The main aim would be to learn the residual function F(x), which would increase the overall accuracy of the network. The authors introduced the concept of zero padding and linear projection to tackle cases when there is a dimension mismatch.

Zero Padding:

- If the dimension of the output is greater than the input, then the extra zero is padded.
- If the dimension of the input is greater than the output, then striding is performed on the input.

Linear Projection:

Instead of H(x) = F(x) + x, a linear projection(LP) is applied to input to obtain **H(x) = F(x) + W(x)*x**

Architcture Diagram:

The paper used the baseline model of VGGNet as a plain network with mostly 3×3 filters with two design rules:

- The layers have the same filters for the same feature map size of the output.
- If the size of the feature map is halved, the number of filters is doubled. This is done to preserve the time complexity per layer.

Source: arxiv.org

The network comprises a global average pooling layer and a 1000-way fully connected layer with a softmax at the end. The dotted lines indicate a change in the size of the image from one residual block to another and are Linear projections that can be accomplished using 1×1 kernels.

Source: arxiv.org

Experimental Setup

Dataset Used: Imagenet

Resnet architecture was evaluated on the ImageNet 2012 classification dataset consisting of:

- 1000 classes
- 1.28 million training images
- 50k validation images
- 100k images

Identity vs. Projection Shortcuts

The author compared:

- A: Zero padding shortcuts for increasing dimensions
- B: Projection shortcuts for increasing dimension, other shortcuts for identity
- C: All shortcuts are projections

The accuracy increased from A to B to C. This concludes that projection is not important for addressing the degradation problem.

CIFAR-10 Dataset

Resnet architecture was evaluated on the CIFAR-10 dataset consisting of the following:

- 10 classes
- 50k training images
- 10k testing images

110 layers on the residual network have the same order of parameters as the previous network, only 19 layers. With a deeper network, it outperformed the previous network. However, with a 1202-layer network, the error increased.

Results

Imagenet Dataset results

Source: arxiv.org

- Training and validation error is higher in deeper networks, 34-layered, than in 18-layered.
- With the introduction of residual networks, 34-layer with residual connection has much lower training and validation error than 18-layer.
- Resnet 34 performed better than the Resnet 18 network with the same number of layers as in the plain network.
- Resnet-34 performed well in generalizing validation data and showed less error.

Training of 50, 101, and 152 layer ResNets on ImageNet dataset

- The depth of ResNet for best accuracy is over four times deeper than previous deep networks.
- Achieved 3.57% top 5 error rate on the test set with 152 layer ResNet on ensemble model.

CIFAR-10 Dataset results

Source: arxiv.org

On the CIFAR-10 dataset with 1202 layers, validation error increased on increasing the network layer, but training error is the same for both 110 and 1202 layers, near zero. Here the authors concluded that it was overfitting with a lot of layers.

Comparison of Current Approach with Previous

Source: arxiv.org

The table shows the training error with different layers.

- The 34-layer plain network has a training error more than an 18-layer plain network, which was caused by degradation as we go deeper into the plain network.
- With the use of ResNet, when a shortcut connection was added, the error decreased for more layers and performed well in generalizing the validation data. This resolved the problem of degradation.

Residual connection to 18 layers is marginally better. However, introducing a residual connection to the 34 layers is much better.

Conclusion

ResNet architecture, which incorporates residual connections, significantly outperforms prior state-of-the-art models on image recognition tasks such as ImageNet. The authors demonstrate that residual connections help alleviate the vanishing gradient problem and enable much deeper networks to be trained effectively. The ResNet architecture achieves better results with fewer parameters, making it computationally more efficient. Residual connections are a general and effective approach for enabling deeper networks, and ResNet architecture will become a new benchmark for image recognition tasks.

Key Takeaways

- The addition of identity mapping in the residual network does not introduce any extra parameters. Hence, the computational complexity of the network does not increase.
- The accuracy gains are higher in ResNet as the depth increases. This produces results considerably better than other previous networks, such as VGG net.
- Residual connection avoids exploding or vanishing gradient problems by having shallow networks in the ‘ensembles.’

Questions Unanswered

- How can the residual block be combined with other techniques, such as transfer learning or attention mechanisms?

- What are the optimal hyperparameters for training residual networks, such as the number of residual blocks, the number of filters, and the size of the shortcuts?

Article Url - https://www.analyticsvidhya.com/blog/2023/02/deep-residual-learning-for-image-recognition-resnet-explained/

Babina Banjara