Project Report | CSD346 | Seminar

Project done by:
Shruti Bansal(2110110492)

Guided by
Shankar Ghosh
(Professor at Shiv Nadar University)

Date: 5th March, 2024

Abstract

This project report examines the practical applications and advancements of the VGG16 model, a deep convolutional neural network architecture. Through analysis of recent research papers, this report highlights the model's effectiveness in diverse domains. By leveraging transfer learning techniques, VGG16 demonstrates its potential as a powerful tool for image analysis.

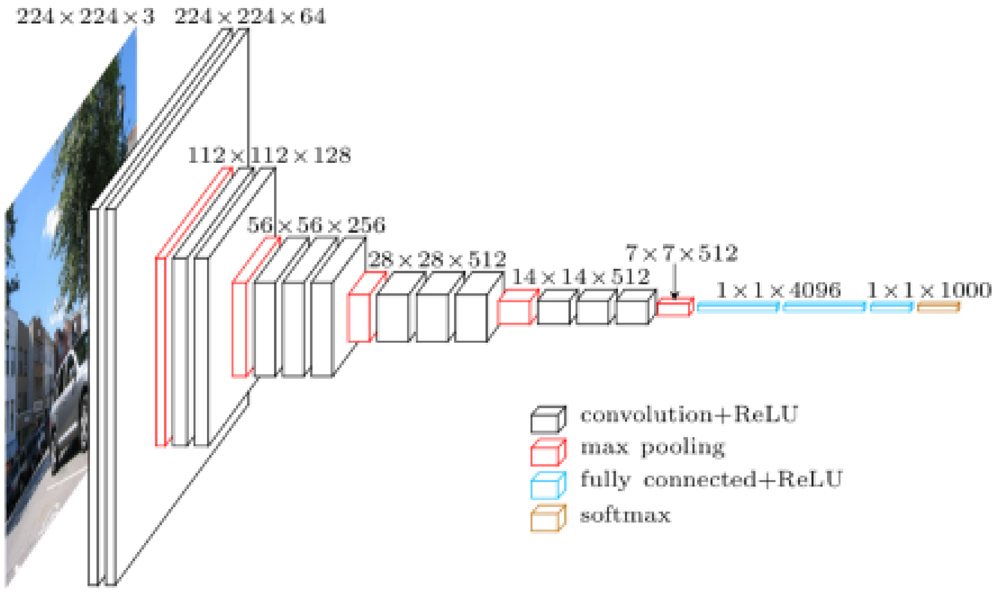
1 Introduction

Developed by the Visual Geometry Group at the University of Oxford, VGG16 stands out for its straightforward architecture, comprising 16 weight layers, hence the nomenclature. These layers include 13 convolutional layers followed by 3 fully connected layers, forming a deep neural network capable of extracting intricate features from input images.

At its core, VGG16 employs a series of small 3x3 convolutional filters with a stride of 1 pixel and a padding of 1 pixel, facilitating the preservation of spatial information throughout the network. By utilizing such small filters, VGG16 achieves a deeper network architecture while maintaining a manageable number of parameters. Additionally, max-pooling layers interspersed between convolutional blocks serve to down-sample feature maps, enhancing computational efficiency and reducing overfitting.

One of the defining characteristics of VGG16 is its homogeneity in architecture, where each convolutional block comprises multiple convolutional layers followed by a max-pooling layer. This uniformity simplifies model design and fosters interpretability, making VGG16 an attractive choice for researchers and practitioners alike. Furthermore, the availability of pre-trained weights on large-scale image datasets such as ImageNet has accelerated its adoption and facilitated transfer learning, enabling the model to be repurposed for various image recognition tasks with minimal data requirements.

In the following discourse, we delve deeper into the architecture and workings of VGG16, exploring its applications across diverse domains ranging from object recognition in natural images to medical image analysis.



2 Research on VGG16 Convolutional Neural Network Feature Classification Algorithm Based on Transfer Learning

2.1 Authors: J. Tao, Y. Gu, J. Sun, Y. Bie, and H. Wang

2.2 Abstract

The paper introduces a feature classification model utilizing the VGG16 neural network for remote sensing applications, achieving accurate identification of water, farmland, buildings, roads, and trees in light and small SAR images. Results indicate that the VGG16 network achieves an average accuracy of 75%, which improves to 81% after parameter optimization. Additionally, integrating a pre-trained model further enhances classification accuracy, reaching an average of 87.5%.

2.3 Methodology

Data Set Construction:

Satellite images containing of 5 features namely roads, buildings, water bodies, trees and farmlands are used. These images are divided into train and test set.

VGG16 Network Construction:

VGG16 has 3x3 kernels with stride 1 and 2x2 max-pooling layers with stride 2. There are three fully connection layers, the first two layers have 4096 channels, and the third layer has 1000 channels.The input image size is 224x224x3, gradually reducing to 7x7x512 before classification through three fully connected layers, producing 1000 output values. By inputting these 1000 values into the softmax activation function, they can be normalized to the [0,1) interval, and their sum is 1. In this way, the 1000 output values can well represent the probability that the image belongs to each category.

VGG16 based on transfer learning:

VGG16 utilizes transfer learning by retaining the convolutional layers and part of the pooling layers while removing the fully connected and softmax layers. It preserves the input image format of 224x224 and weights, incorporating global mean pooling after the fourth pooling layer and adding a 4096-neuron fully connected layer followed by a softmax layer. By freezing the structure up to the fourth pooling layer during training, it optimizes training efficiency and reduces system failures, making it suitable for environments with limited CPU and memory resources.

2.4 Results

The experiment involves preprocessing data by converting grayscale images to pseudo-color images, which are standardized to 224x224x3 pixels, with the dataset split into 80% training, 10% validation, and 10% testing sets. The research results show that the test result accuracy of vgg16 network is 75%. After parameter optimization, the accuracy of ground feature classification can reach 81.25%. The effect of vgg16 network after migration learning is better, and the accuracy can reach 87.5% in the case of small samples.

3 A Novel Parameterized Activation Function in Visual Geometry Group

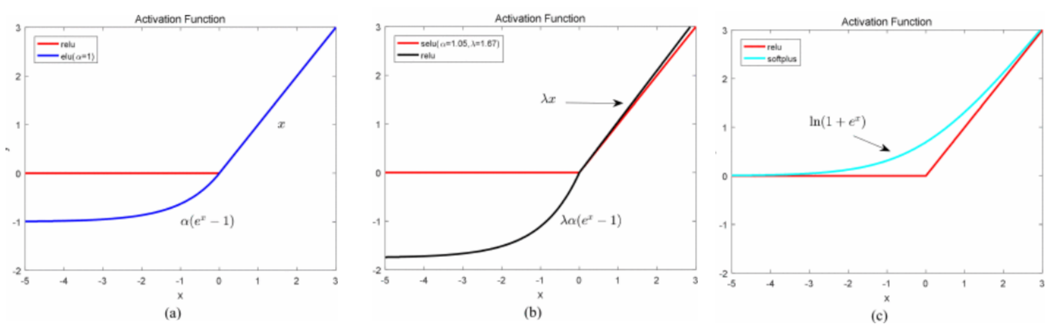
3.1 Authors: Chen Xu, Jie Huang, Sheng-peng Wang and An-qing Hu

3.2 Abstract

This paper investigates the performance of four activation functions—standard rectified linear unit (ReLU), exponential linear unit (ELU), scaled exponential linear unit (SELU), and Softplus—within the context of the Visual Geometry Group (VGG) network, using recognition accuracy and runtime as evaluation metrics. A newly proposed parameterized activation function, combining features of ReLU and Softplus, is compared and tested within the VGG network using a specific dataset. Results demonstrate that the novel activation function significantly enhances recognition accuracy within the VGG framework, particularly when the parameter k exceeds 1.37, improving accuracy by approximately 3.0

3.3 Methodology

Comparative Studies Between Activation Functions



Comparation functions ((a) relu & elu; (b) relu & selu; (c)relu & softplus)

The non-negative interval gradient of the ReLU activation function equals to 0, which fully reveals that the gradient vanishing case is totally avoided when the ReLU activation function is applied. For Softplus, it's similar with that of the ReLU since its curve is close to that of the ReLU, it should be noted that the running time increases along with non-zero interval gradient value and it easily leads to a gradient vanishing case when compared with the ReLU. While, for ELU and SELU activation functions, both leads a burden of calculation and a low training efficiency when compared with ReLU activation function.

Although, the ReLU performs well when compared with the other three activation functions, however, once the size of the training data is huge or the learning rate of the network is too large, the training processes in the network will become unstable, and the recognition accuracy will decrease, even the ReLU activation function applied.

Improvements and General Expressions

Through the analysis of the characteristics of the abovementioned activation functions, a novel activation function is proposed in this subsection by combining the advantages of the two activation functions, namely, ReLU and Softplus, for their better performance and high accuracy in recognition training.

In a special case, the proposed novel activation function is expressed as follows

f(x)={ $\ln\left(1+e^{x}\right) \quad for \quad x<0$
$x+\ln 2 \quad for \quad x \geq 0$

Based on this, a parameter k (Slope) is introduced in the special expressions above, aiming at finding a better activation function for general cases. Thus, the expressions of the novel parameterized activation function for general cases is expressed as

f(x)={ $ln(1+e^{x}) \quad for \quad x<0$
$kx+\ln 2 \quad for \quad x \geq 0$

3.4 Results

The study introduces a new activation function that merges the strengths of both ReLU and Softplus functions. Experimental findings indicate that higher values of parameter k in the proposed function notably enhance network recognition accuracy compared to individual activation functions. Future investigations will explore its performance across diverse networks and datasets, delving deeper into additional attributes of the novel activation function.

4 Summary and Conclusion

The project report investigates two key research papers to explore advancements in the field of convolutional neural networks (CNNs) and activation functions.

1. "Research on VGG16 Convolutional Neural Network Feature Classification Algorithm Based on Transfer Learning" by J. Tao et al.:

This paper explores the application of transfer learning techniques in training VGG16 CNNs for feature classification tasks, such as identifying water bodies, roads, buildings, and other features in satellite images.

2. "A Novel Parameterized Activation Function in Visual Geometry Group" by C. Xu et al.:

This paper introduces a parameterized activation function that combines the strengths of ReLU and Softplus functions. The proposed function enhances network recognition accuracy with larger values of the parameter k. The study sheds light on novel approaches to improve the performance of CNNs, particularly in the context of the Visual Geometry Group (VGG) network.

The project report delves into the methodologies, experimental findings, and implications of these papers, highlighting advancements in CNN architectures and activation functions for various applications, including image classification and feature recognition. Additionally, the report underscores the significance of transfer learning techniques in optimizing CNN performance for specific tasks and datasets.

5 References

1. C. Xu, J. Huang, S. -p. Wang and A. -q. Hu, "A Novel Parameterized Activation Function in Visual Geometry Group," 2018 2nd International Conference on Data Science and Business Analytics (ICDSBA), Changsha, China, 2018, pp. 386-389, doi: 10.1109/ICDSBA.2018.00079. keywords: Training;Standards;Visualization;Convolution;Neural networks;Geometry;Textiles;VGG Network;ReLU;Softplus;Activation Function,
2. J. Tao, Y. Gu, J. Sun, Y. Bie and H. Wang, "Research on vgg16 convolutional neural network feature classification algorithm based on Transfer Learning," 2021 2nd China International SAR Symposium (CISS), Shanghai, China, 2021, pp. 1-3, doi: 10.23919/CISS51089.2021.9652277. keywords: Training;Water;Convolution;Roads;Neural networks;Buildings;Transfer learning;Classification of features;VGG16;Deep neural network;transfer learning,