EVALUATING CNN WITH OSCILLATORY ACTIVATION FUNCTION

Jeevanshi Sharma
Department of Electrical Engineering
Aligarh Muslim University, Aligarh
sharma.jivanshi@gmail.com

ABSTRACT

The reason behind CNN’s capability to learn high-dimensional complex features from the images is the non-linearity introduced by the activation function. Several advanced activation functions have been discovered to improve the training process of neural networks, as choosing an activation function is a crucial step in the modeling. Recent research has proposed using an oscillating activation function to solve classification problems inspired by the human brain cortex. This paper explores the performance of one of the CNN architecture AlexNet on MNIST and CIFAR10 dataset using oscillatory activation function (GCU) and some other commonly used activation functions like ReLu, PReLu, Mish.

Keywords Activation Function · Convolutional Neural Networks · Growing Cosine Unit · XOR Function

1 Introduction

The convolutional neural network is a type of feed-forward multi-layer neural network, which is generally used for image datasets. It is composed of a convolution layer (Conv), a down-sampling layer (or pooling layer), and a full-connection layer (FC). The convolution layer is responsible for the extraction of the different features from the input images, which gets convoluted by several filters on the convolution layer. Then the feature is blurred by the down-sampling layer. Finally, a set of eigenvectors is acquired through a full connection layer. Activation functions are an important part of CNN’s architecture and its performance. CNN’s have been successfully in practically various fields. The general architecture of a convolutional neural network is represented in Figure 1.

The practical implementation of a CNN model still demands improvement. Many studies have been done on different image classification methods, design of adaptive learning rate, design of drop-out layer, and at the core research on better activation functions for mapping data better. Neural networks are non-linear. It is their non-linearity that makes them powerful. As natural images have highly co-related pixel values, this attribute can only be better exploited with the use of non-linear functions. Linear activation functions make no use of the multi-layer CNNs.

Over the years, there has been a lot of significant development in CNN architectures to mitigate problems regarding computational efficiency, error rates, and other improvements. Architectures like LeNet-5, AlexNet, Inception, VCGNet, ResNet, etc. were put forward to deal with issues of overfitting, issues with high-resolution images, etc. Past studies have shown to achieve better results when these architectures are optimized with different activation functions and optimization algorithms. Along with this, many variants of activation functions like ReLu and different combinations of activation functions have been proposed to achieve better results with these architectures. In particular, the use of the ReLU like activation functions can result in faster training compared to saturating sigmoidal type activation functions because these activation functions do not saturate for a wider range of inputs and avoid the vanishing gradient problem. The classification of activation functions is primarily based on whether they can modify their shape during training. Oscillatory activation functions which are inspired by biological comprises of multiple hyperplanes in their decision boundaries, which enables neurons to make more complex decisions. In the author suggests that deep neural networks with oscillatory activation functions might partially bridge the
gap present in biological and artificial neural networks. One such function is GCU (Growing Cosine Unit). In contrast to the common perception of a single plane decision boundary, the GCU has infinite evenly spaced solutions resulting in infinite evenly spaced hyperplanes, which solves the XOR problem in neural networks.

This study talks about the development of trainable activation function and their performance on CNN architectures, AlexNet and ResNet.

2 Activation Functions

In a neural network, neurons map the weighted sum of their input to a single output with the help of Activation Functions. The main purpose served by the activation function is to introduce non-linearity in the network. The important properties commonly shared by activation functions are non-linearity, differentiability, continuous, bounded, and zero-centering. Non-linear functions are universal approximates and using a linear activation function in a multi-layer network is the same as using a single-layer network. Continuously differentiable activation functions are desired when using gradient-based optimization algorithms. Activation functions are preferred to be centered at zero so that the gradients do not shift in a particular direction for gradients to be unbiased the output of the activation function must be symmetrical at zero so that the gradients do not shift in a particular direction.

2.1 ReLU

ReLU was primarily used to overcome the vanishing gradient problem. ReLU is the most common activation function used for classification problems.
2.2  PReLU

The disadvantage of ReLU is that it cannot adapt to sudden changes and zero the negative inputs (dying ReLU problem). PReLU generalises the traditional with a slope for negative inputs using a learnable parameter that leads to adaptive constants which can increase accuracy. [30]

Figure 3: PReLU

2.3  Mish

Mish is an unbounded in positive direction and a bounded negative domain. Mish has properties like non-monotonicity and smooth profile. [31]

Figure 4: Mish

2.4  GCU

Growing Cosine Unit was introduced in 2021 [32]. It is an oscillatory activation function whose amplitude increases with increasing values. GCU has infinite evenly spaced solutions resulting in infinite evenly spaced hyper-planes, instead of a single plane decision boundary which resolves the long-rooted XOR problem as for separating classes in XOR datasets two hyper-planes are needed which required the use activation function with multiple zeros. [33]

3  Dataset and Methodology

3.1  Experimental Setup

In this following, a comparison of the AlexNet on CIFAR10 and MNIST with GCU and other activation functions is presented. The SGD and Adam optimizers are used on CIFAR and MNIST datasets respectively with sparse categorical
Evaluating CNN with Oscillatory Activation Function

Figure 5: GCU

crossentropy loss function (softmax classification head). For both the datasets, we experiment with the choice of activation functions at two locations in AlexNet: one for all of the convolutional layer and the other for the dense layers. The results for CIFAR 10 are reported on 50 epochs and 40 epochs for MNIST.

4 Results

The performance of GCU with various activation functions are visualized and presented in Table 1 and Table 2 for comparing the oscillating activation functions with other prominently used non-linear activation functions (Mish, Prelu).

| Architecture | Convolutional Layer | Activation Dense Layer | Validation Accuracy | Test Accuracy | Val Loss | Loss |
|--------------|---------------------|------------------------|---------------------|--------------|----------|------|
| AlexNet      | ReLU                | ReLU                   | 0.9940              | 0.9974       | 0.0850   | 0.0145|
| AlexNet      | GCU                 | ReLU                   | 0.9830              | 0.9791       | 0.0791   | 0.9830|
| AlexNet      | PReLU               | ReLU                   | 0.9945              | 0.9904       | 0.0469   | 0.0435|
| AlexNet      | Mish                | ReLU                   | 0.9940              | 0.9928       | 0.1190   | 0.0625|

Table 1: Performance on MNIST

| Architecture | Convolutional Layer | Activation Dense Layer | Validation Accuracy | Test Accuracy | Val Loss | Loss |
|--------------|---------------------|------------------------|---------------------|--------------|----------|------|
| AlexNet      | ReLU                | ReLU                   | 0.7011              | 0.9777       | 1.6989   | 0.0666|
| AlexNet      | GCU                 | ReLU                   | 0.6036              | 0.8256       | 1.3953   | 0.4913|
| AlexNet      | PReLU               | ReLU                   | 0.7015              | 0.9765       | 1.6207   | 0.0722|
| AlexNet      | Mish                | ReLU                   | 0.7143              | 0.9802       | 1.6899   | 0.0591|

Table 2: Performance on CIFAR 10

5 Conclusion

The activation function is an important part of a Convolutional Neural Network. Many research have proposed multiple methods which can improve the performance of CNNs. From the view of activation function, this paper explores and experiment the use of oscillatory activation function (GCU) in AlexNet, and compares it with PReLU and Mish. As Mish is more computationally heavy than ReLU, we have only placed these function in the convolution layers and ReLU for the deep connected layers. GCU have shown comparable performance on both MNIST and CIFAR 10 dataset.
References

[1] Albert Gidon, Timothy Adam Zolnik, Pawel Fidzinski, Felix Bolduan, Athanasia Papoutsi, Panayiota Poirazi, Martin Holtkamp, Imre Vida, and Matthew Evan Larkum. Dendritic action potentials and computation in human layer 2/3 cortical neurons. Science, 367(6473):83–87, 2020.

[2] Ronan Collobert and Jason Weston. A unified architecture for natural language processing: Deep neural networks with multitask learning. In Proceedings of the 25th International Conference on Machine Learning, ICML ’08, page 160–167, New York, NY, USA, 2008. Association for Computing Machinery.

[3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. Communications of the ACM, 60:84 – 90, 2012.

[4] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[5] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 4700–4708, 2017.

[6] Wartini Ng, Budiman Minasny, Maryam Montazerolghaem, Jose Padarian, Richard Ferguson, Scarlett Bailey, and Alex B McBratney. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. Geoderma, 352:251–267, 2019.

[7] Huizhen Zhao, Fuxian Liu, Han Zhang, and Zhibing Liang. Research on a learning rate with energy index in deep learning. Neural Networks, 110:225–231, 2019.

[8] Shaofeng Cai, Yao Shu, Gang Chen, Beng Chin Ooi, Wei Wang, and Meihui Zhang. Effective and efficient dropout for deep convolutional neural networks. arXiv preprint arXiv:1904.03392, 2019.

[9] Gianluca Maguolo, Loris Nanni, and Stefano Ghidoni. Ensemble of convolutional neural networks trained with different activation functions. Expert Systems with Applications, 166:114048, 2021.

[10] Arun Kumar Dubey and Vanita Jain. Comparative study of convolution neural network’s relu and leaky-relu activation functions. In Applications of Computing, Automation and Wireless Systems in Electrical Engineering, pages 873–880. Springer, 2019.

[11] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. Neural networks, 2(5):359–366, 1989.

[12] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324, 1998.

[13] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. Communications of the ACM, 60(6):84–90, 2017.

[14] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1–9, 2015.

[15] Mainak Bandyopadhyay. Multi-stack hybrid cnn with non-monotonic activation functions for hyperspectral satellite image classification. Neural Computing and Applications, 33(21):14809–14822, 2021.
Evaluating CNN with Oscillatory Activation Function

[22] Tongtong Jiang and Jinyong Cheng. Target recognition based on cnn with leakyrelu and prelu activation functions. In 2019 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC), pages 718–722. IEEE, 2019.

[23] Zhiguan Huang, Xiaohao Du, Liangming Chen, Yuhe Li, Mei Liu, Yao Chou, and Long Jin. Convolutional neural network based on complex networks for brain tumor image classification with a modified activation function. IEEE Access, 8:89281–89290, 2020.

[24] Andrea Apicella, Francesco Donnarumma, Francesco Isgrò, and Roberto Prevete. A survey on modern trainable activation functions. Neural Networks, 138:14–32, 2021.

[25] Matthew Mithra Noel, Shubham Bharadwaj, Venkataraman Muthiah-Nakarajan, Praneet Dutta, and Geraldine Bessie Amali. Biologically inspired oscillating activation functions can bridge the performance gap between biological and artificial neurons. arXiv preprint arXiv:2111.04020, 2021.

[26] Mathew Mithra Noel, Advait Trivedi, Praneet Dutta, et al. Growing cosine unit: A novel oscillatory activation function that can speedup training and reduce parameters in convolutional neural networks. arXiv preprint arXiv:2108.12943, 2021.

[27] Leonid Datta. A survey on activation functions and their relation with xavier and he normal initialization. arXiv preprint arXiv:2004.06632, 2020.

[28] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. Neural networks, 2(5):359–366, 1989.

[29] Abien Fred Agarap. Deep learning using rectified linear units (relu). arXiv preprint arXiv:1803.08375, 2018.

[30] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE international conference on computer vision, pages 1026–1034, 2015.

[31] Diganta Misra. Mish: A self regularized non-monotonic activation function. arXiv preprint arXiv:1908.08681, 2019.

[32] Mathew Mithra Noel, Advait Trivedi, Praneet Dutta, et al. Growing cosine unit: A novel oscillatory activation function that can speedup training and reduce parameters in convolutional neural networks. arXiv preprint arXiv:2108.12943, 2021.

[33] Ehsan Lotfi and M-R Akbarzadeh-T. A novel single neuron perceptron with universal approximation and xor computation properties. Computational intelligence and neuroscience, 2014, 2014.