International Journal of Information, Engineering & Technology
Volume 12, Issue 10, PP 156-163, ISSN: 2360-9194, April, 2025
OTL: 4272-1454-381-12112-1
Double Blind Peer Reviewed International Research Journal
Journal Series: Global Academic Research Consortium (garc)
arcnjournals@gmail.com
https://arcnjournals.org

DYNAMIC HYBRID MULTISTAGE GRADIENT DESCENT OPTIMIZATION TECHNIQUES FOR ENHANCING PERFORMANCE OF DEEP NEURAL NETWORK IN HIGH-DIMENSIONAL APPLICATIONS

Usman Ali[1], Fatima Aliyu Shugaba[2], Muhammad Habibullah Abdulfattah[3], Modu Mustapha Tijjani[4]
Department of Computer Engineering, Ramat Polytechnic, Maiduguri, Borno state of Nigeria[1]
Department of Computer Engineering, University of Maiduguri, Borno state of Nigeria[2,3] Department of Elect/Electronics Engineering, Ramat Polytechnic, Maiduguri, Borno state of Nigeria[4]

Abstract: *In high-dimensional applications like natural language processing (NLP) and image identification, training deep neural networks (DNNs) presents several challenges such as convergence, poor generalization, and computational instability. Despite their widespread use, traditional gradient descent (GD) optimization techniques frequently fail to adequately handle these problems throughout training. While optimizers like SGD with Momentum perform well in later training stages but have slow early convergence, advanced optimizers like Adam and RMSprop increase convergence but may result in inferior generalization. The current work presents a unique framework for hybrid multistage gradient descent optimization, which attempts to adapt optimization methods at different stages of training. This approach tries to reduce computing costs while improving convergence speed, model accuracy, and generalization by utilizing the strengths of numerous optimizers, including Adam, SGD with Momentum, and RMSprop. Adaptive learning rate scheduling and real-time optimizer switching based on training dynamics will be integrated into the framework. A series of extensive experiments will be conducted on high-dimensional image recognition (like CIFAR-10 and ImageNet) and NLP (like IMDB and BERT fine-tuning) datasets to validate the framework's performance across different deep-learning architectures. This work will address the difficulties of training DNNs in intricate, high-dimensional environments and provide a valuable theoretical and practical contribution to optimization strategy development.*

Keywords: *Optimization, Convergence, Computation, Dataset, Training.*

INTRODUCTION

Deep learning algorithms such as Deep Neural Networks (DNNs) have been the dominant paradigm recently in the areas of artificial intelligence (AI) (Hayashi, 2023), which has led to notable advances in so many variety of applications such as computer vision, finance, natural language processing (NLP), healthcare, and autonomous systems (Mahadeo & Dhanalakshmi, 2020). DNNs are invaluable tools for solving real-world problems that traditional algorithms have struggled to solve because of their ability to learn complicated patterns from high-dimensional data (Jiang et al., 2022). Furthermore, in the field of natural language processing

(NLP), deep neural networks (DNNs) models have led to significant advancements in machine translation and speech recognition, as well as state-of-the-art performance in tasks involving object detection and facial recognition (Chen *et al.*, 2023).

However, training these complicated models is a very difficult undertaking, especially when datasets get bigger and more complex. When it comes to DNN training, gradient descent (GD) and its variations are the most used optimization methods. By minimizing a loss function through iterative model parameter adjustments, these techniques eventually allow the network to produce precise predictions (Scale et al., 2019). The efficacy of the GD optimization technique used in training has a significant impact on a DNN's performance. Even though GD is widely used, researchers have found that training DNNs presents several significant difficulties, including sluggish convergence, vanishing and exploding gradients, overfitting, and high computing costs (Galv, 2021). Dealing with high-dimensional data, as in many contemporary applications, exacerbates these issues (Guo *et al.*, 2020).

Advanced gradient descent optimization techniques like Adam, RMSprop, Adagrad, and SGD with Momentum have been developed to tackle these difficulties (Sherif *et al*., 2021). By introducing mechanisms for adjusting the learning rate during training, these techniques aid in performance improvement and convergence stabilization. However, depending on the training stage and the dataset's properties, each optimizer has advantages and disadvantages that affect how effective it is (Khan & Maity, 2020). For instance, SGD with Momentum is slower but frequently produces superior results in terms of model correctness on test data, whereas Adam is known for fast convergence but may result in poor generalization. Because of this, using just one optimizer for the whole training process may produce less-than-ideal results, particularly when training big and intricate DNNs (L*i et al.*, 2020).

This work aims to provide a hybrid multistage gradient descent optimization framework in light of these difficulties. This framework will dynamically switch between several advanced optimization approaches as it is being trained, combining them in a step-by-step manner. The hybrid multistage approach that has been suggested seeks to handle the difficulties related to high-dimensional data while enhancing convergence speed, accuracy, and generalization. This work intends to further the field of AI and machine learning by providing a more efficient and versatile approach to DNN optimization by examining the relationships between various optimizers and training phases.

To summarize, although tremendous progress has been made in DNN optimization, the problems associated with training big models on high-dimensional data are not entirely addressed by the techniques now in use (Liu *et al.,* 2023). By creating a unique hybrid multistage optimization framework that leverages the advantages of several GD techniques, this research aims to close this gap and eventually improve the state of the art in DNN training.

1.2 Significance of Research

To sum up, this study is of paramount important since it has the potential ability to improve theoretical understanding as well as practical deep learning optimization skills. This study or research tends to improve the performance, efficiency, and generalizability of DNNs across many high-dimensional applications by creating a novel hybrid multistage gradient descent optimization framework or model, which will promote greater adoption of AI technologies in crucial industries.

Review of Related Work

Table 2.1: Summary of Literature Review

Author(s)	Year	Study Focus	Key Contributions	Limitations
Kingma, D. P. & Ba, J.	2015	Adam: A Method for Stochastic Optimization	Presented Adam, which combines the benefits of AdaGrad and RMSprop for adaptive learning rates	It tends to overfit and generalize poorly in later training stages, especially in high-dimensional models
Loshchilov, I. & Hutter, F.	2019	Uncoupled Weight Decay Regularization (AdamW)	Improved Adam by decoupling weight decay from the gradient-based updates, resultant in better generalization	Still trusts on a single optimizer throughout the training process; does not adapt optimizers dynamically
Smith, L. N.	2018	Cyclical Learning Rates (CLR)	Proposed cyclical learning rates that dynamically oscillate to avoid local minima and improve performance	Limited to learning rate adjustments without addressing the use of multiple optimizers in a hybrid procedure.
Loshchilov, I. & Hutter, F.	2017	SGDR: Warm Restarts in SGD	Introduced stochastic gradient descent with warm restarts (SGDR) to advance model convergence by restarting the learning rate	It does not explore combining optimizers; relies on SGD and learning rate restarts alone
You, Y. et al.	2020	Large Batch Optimization with LARS and LAMB	Established LARS and LAMB optimizers for training deep learning models with large batches, improving scalability	Focuses on batch size adjustments, not on the dynamic switch between different optimization algorithms
Zaheer, M. *et al.*	2018	Adaptive Gradient Methods with Dynamic Boundaries	They proposed Yogi, an adaptive gradient method that prevents exponential growth of gradients in Adam	Yogi enhances Adam but remains a single-optimizer approach; lacks dynamic adaptability across different stages
Chen *et al.*	2021	Revisiting Batch Normalization (BN) in Optimizers	Showed that BN can help SGD-based approaches converge faster and	Emphases on batch normalization rather than combining

Author(s)	Year	Study Focus	Key Contributions	Limitations
			generalize better in image classification	optimizers; lacks exploration of multistage techniques
Shankar *et al.*	2024	AI-Based Intrusion Detection with Optimized Learning	Applied novel optimization techniques to advance intrusion detection systems' accuracy and computational efficiency	Focuses on cybersecurity applications but lacks broader pertinence to other high-dimensional data problems
Wang *et al*.	2023	Adaptive Gradient Descent with Momentum	Proposed hybrid optimizers that combine adaptive learning rates with momentum-based techniques	While effective, the study lacks a complete multistage framework for varying training phases dynamically
Singh & Zubair	2024	Multistage Optimization in Deep Learning	Developed a multistage framework for combining optimizers such as Adam and SGD in different training stages	Provides a foundation for hybrid approaches but limited experimental data on generalization in high-dimensional tasks

Summary

In conclusion, the literature demonstrates a clear need for more flexible and adaptive optimization techniques that can enhance the performance of DNNs in high-dimensional applications. Multistage optimization techniques, which apply different optimizers at various stages of training, hold significant promise in addressing the limitations of single-optimizer approaches. However, more research is needed to develop a formalized framework that can dynamically adjust optimization strategies based on the training phase and model behavior. This research aims to fill these gaps by developing a hybrid multistage gradient descent optimization framework that improves convergence, generalization, and computational efficiency across diverse applications in image recognition and NLP.

PROBLEM DEFINITION

This research thesis seeks to address the above problem by exploring the following sub-problems:

a. Which gradient descent optimizer combinations work best at certain stages of DNN training?
b. How can we respond to real-time training parameters like validation performance, gradient magnitude, and loss stagnation by dynamically switching between optimizers during training?
c. In what ways may the hybrid optimization framework be modified to efficiently manage high-dimensional data in fields like natural language processing and picture recognition?

d. In various multistage optimization applications, what are the trade-offs between convergence speed, generalization, and computational efficiency?

The aim of this study is to develop a flexible and adaptive hybrid optimization framework that would leverages the strengths of different optimizers at different stages of training, ultimately overcoming the inefficiencies and generalization issues associated with single-optimizer methods and approaches. The framework if developed will be validated through extensive experimentation on high-dimensional datasets across diverse application domains/areas, providing a comprehensive solution to the problem of DNN optimization.

Aim of the Study

In high-dimensional applications like image recognition and natural language processing, the goal of this research is to create a hybrid multistage gradient descent optimization framework that dynamically modifies optimization strategies during the training phase to enhance the convergence, generalization, and computational efficiency of deep neural networks (DNNs).

4.2 Objectives of the Study

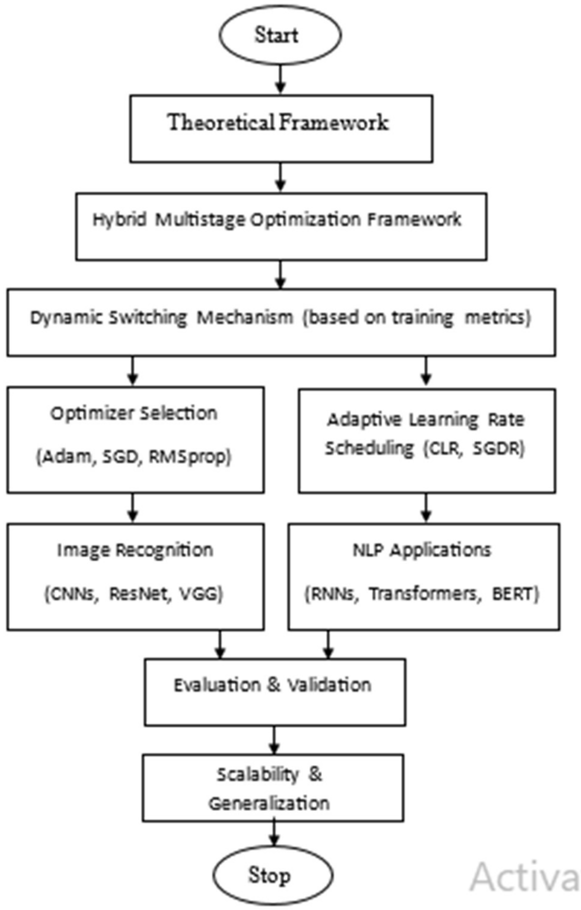
To achieve the above aim, the following specific objectives will guide the proposed study:

- **a.** To design and Develop a Hybrid Multistage Optimization Framework
- **b.** To optimize Hyperparameters Dynamically
- **c.** To apply the Framework to High-Dimensional Applications
- **d.** To provide Open-Source Implementation and Documentation

METHODOLOGY

Design of the Hybrid Multistage Optimization Framework

The major innovation of this proposed study lies in the development of the **hybrid multistage optimization framework model.** This framework if developed will be designed to automatically, effortlessly switch between different GD optimizers at various stages of training process based on the evolving training dynamics. To achieve this, the following phases will be involved:



CONCLUSION

In conclusion, by offering a flexible and scalable optimization strategy that meets the expanding demands of modern artificial intelligence applications, this work seeks to significantly advance the field of deep learning. The results will eventually raise the bar for DNN training techniques and have applications in the fields of deep learning technologies-dependent industries like healthcare, autonomous systems, and natural language processing.

REFERENCES

Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. Neural Networks: Tricks of the Trade, 437-478. https://doi.org/10.1007/978-3-642-35289-8_26

Chen, T., Xu, B., Zhang, C., & Guestrin, C. (2021). Revisiting Batch Normalization in Optimizers. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/CVPR42600.2021.01017

Chen, C., Lai, J., Chang, Y., Lai, C., & Pai, P. (2023). *A Study of Optimization in Deep Neural Networks for Regression*. 1–17.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.

Galv, E. (2021). *Neuroevolution in Deep Neural Networks : Current Trends and Future Challenges*. *4581*(c), 1–21. https://doi.org/10.1109/TAI.2021.3067574

Guo, D., Wang, X., Gao, K., Jin, Y., Ding, J., & Member, S. (2020). *Evolutionary Optimization of High-Dimensional Multiobjective and Many-Objective Expensive Problems Assisted by a Dropout Neural Network*. 1–14.

Hinton, G. (2012). Neural Networks for Machine Learning. Lecture 6e: RMSprop. Coursera Lecture.

Hayashi, T. (2023). A Deep-Learning Method for Path Loss Prediction Using Geospatial Information and Path Profiles. *IEEE Transactions on Antennas and Propagation*, *71*(9), 7523–7537. https://doi.org/10.1109/TAP.2023.3295890

Jiang, Y. A. N., Li, X., & Liu, S. (2022). A Hybrid Three-Staged , Short-Term Wind-Power Prediction Method Based on SDAE-SVR Deep Learning and BA Optimization. *IEEE Access*, *10*(October), 123595–123604. https://doi.org/10.1109/ACCESS.2022.3223435

Khan, M. I., & Maity, R. (2020). Hybrid Deep Learning Approach for Multi-Step-Ahead Daily Rainfall Prediction Using GCM Simulations. *IEEE Access*, *8*(Ml), 52774–52784. https://doi.org/10.1109/ACCESS.2020.2980977

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *International Conference on Learning Representations (ICLR)*. https://arxiv.org/abs/1412.6980

Li, C., Tang, G., Xue, X., Chen, X., & Wang, R. (2020). The short-term interval prediction of wind power using the deep learning model with gradient descend optimization. *Renewable Energy*, *155*, 197–211. https://doi.org/10.1016/j.renene.2020.03.098

Liu, J., Zhan, C., Wang, H., Zhang, X., Liang, X., & Zheng, S. (2023). *Developing a Hybrid Algorithm Based on an Equilibrium Optimizer and an Improved Backpropagation Neural Network for Fault Warning*.

Loshchilov, I., & Hutter, F. (2017). SGDR: Stochastic Gradient Descent with Warm Restarts. *International Conference on Learning Representations (ICLR)*. https://arxiv.org/abs/1608.03983

Loshchilov, I., & Hutter, F. (2019). Decoupled Weight Decay Regularization. *International Conference on Learning Representations (ICLR)*. https://arxiv.org/abs/1711.05101

Mahadeo, U., & Dhanalakshmi, K. R. (2020). High - dimensional microarray dataset classification using an improved adam optimizer (iAdam). *Journal of Ambient Intelligence and Humanized Computing*, *0123456789*. https://doi.org/10.1007/s12652-020-01832-3

Matsumoto, K., & Kovacs, T. (n.d.). *XCSR Based on Compressed Input by Deep Neural Network for High Dimensional Data*. 1418–1425.

Qian, N. (1999). On the momentum term in gradient descent learning algorithms. Neural Networks, 12(1), 145-151. https://doi.org/10.1016/S0893-6080(98)00116-6

Rao, K. N., Rao, K. V., & D, P. R. P. V. G. (2021). A hybrid Intrusion Detection System based on Sparse autoencoder and Deep Neural Network. *Computer Communications*, *180*(April), 77–88. https://doi.org/10.1016/j.comcom.2021.08.026

Review, A. (2021). *Artificial Neural Networks Based Optimization Techniques : Ml*.

Roy, B., Singh, M. P., Kaloop, M. R., Kumar, D., Hu, J., Kumar, R., & Hwang, W. (2021). *applied sciences Data-Driven Approach for Rainfall-Runoff Modelling Using Equilibrium Optimizer Coupled Extreme Learning Machine and Deep Neural Network*.

Sarker, I. H. (2021). Deep Learning : A Comprehensive Overview on Techniques , Taxonomy , Applications and Research Directions. *SN Computer Science*, 1–20. https://doi.org/10.1007/s42979-021-00815-1

Scale, L., Turbine, W., Using, D., Learning, D., & Analysis, C. (2019). *A New Hybrid Approach to Forecast Wind Power for Large Scale Wind Turbine Data Using Deep Learning*.

Shankar, S., Bui, T., Chakrabarti, P., Chakrabarti, T., & Parasa, G. (2024). A Novel Optimization-Based Deep Learning Approach for Intrusion Detection Systems. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-11885-4

Singh, A. K., & Zubair, S. (2024). Combination of Optimization Methods in a Multistage Approach for Deep Neural Networks. *International Journal of Information Technology*, 16(3), 1855–1861. https://doi.org/10.1007/s41870-023-01568-1

Smith, L. N. (2018). Cyclical Learning Rates for Training Neural Networks. *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, 464-472. https://doi.org/10.1109/WACV.2017.58

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2818-2826). https://doi.org/10.1109/CVPR.2016.308

Sherif, K., Hassan, M., Feng, Y., Najah, A., Hoon, C., & El-shafie, A. (2021). A review of the hybrid artificial intelligence and optimization modelling of hydrological streamflow forecasting. *Alexandria Engineering Journal*. https://doi.org/10.1016/j.aej.2021.04.100

Tan, Y., Zhan, C., Pi, Y., Zhang, C., Song, J., Chen, Y., & Golmohammadi, A. (2023). *A Hybrid Algorithm Based on Social Engineering and Artificial Neural Network for Fault Warning Detection in Hydraulic Turbines*.

Wang, Y., Liang, P., & Chang, J. (2023). Adaptive Gradient Descent with Momentum for Neural Networks. Journal of Machine Learning Research, 24(1), 1-25. https://doi.org/10.5555/3566068.3566079

Yang, S., Tian, Y., He, C., Zhang, X., Member, S., Tan, K. C., & Jin, Y. (2021). *A Gradient-Guided Evolutionary Approach to*. 1–15.

Zaheer, M., Reddi, S. J., Kale, S., Kumar, S., & McMahan, H. B. (2018). Adaptive Methods for Nonconvex Optimization. NeurIPS. https://arxiv.org/abs/1804.04272.