AN EFFICIENT SEMI-SIGMOIDAL NON-LINEAR ACTIVATION FUNCTION APPROACH FOR DEEP NEURAL NETWORKS

CHIENG HOCK HUNG

A thesis submitted in
fulfillment of the requirement for the award of the
Doctor of Philosophy in Information Technology

Faculty of Computer Science and Information Technology
Universiti Tun Hussein Onn Malaysia

JANUARY 2022

DEDICATION

This is dedicated to the field of Artificial Intelligence (AI) and computer science, as well as the academicians who work in these related fields. The advancement in AI has indeed transformed and blessed the world significantly. Without this breakthrough, domains such as healthcare, security, telecommunication, to name a few, are impossible to benefit mankind to a greater extend. I expect that this work will be a blessing to the world. Though the contributions are seemingly small, yet it still taking part in transforming the world. There is a quote saying that,
"...if you have faith as small as a mustard seed, you can say to this mountain, 'Move from here to there', and it will move..."

I also dedicate this work to my beautiful wife Choong Pui Ying, who loves me so much and taking care of me throughout the process of finishing the work.
As well as my daughter Divine Chieng, who always reminds me of the goodness of Abba Father.
Thanks to my Dad, Mum, and sisters, who consistenly encourage me along the journey.

Last but not least, thanks to the Hope family's members, without your supports and prayers, it is impossible for me to endure this journey.

ACKNOWLEDGEMENT

First of all, I would like to give thanks to my Abba Father, for His endless goodness and favor to me. He is the source of my strength, wisdom, and knowledge. He empowered and granted me the strength to soar high on wings like eagles throughout this Ph.D. journey.

Secondly, I would like to thanks my supervisor, Assoc. Prof. Dr. Noorhaniza Wahid. She was a direct contributor to my work. She is like an academically mother to me. Her advice, wisdom, and personal experiences have guided and directed me in the direction toward the fulfillment of this study. Further, she is also a kind person that always enlightens me with her life experiences. Turn out, I was greatly encouraged and inspired to explore beyond the sky.

Thirdly, I would like to appreciate my co-supervisor, Associate. Prof. Dr. Ong Pauline. Her guidance has brought so much impact to my work. She always provided me with another perspective or viewpoint on my research work. These further allow me to gain much deeper insight into my research work. In addition, she also provided me the opportunity to serve as a reviewer for a prominent international journal. In that place, I could learn from other's works, hence my research and writing skills have significantly improved.

Lastly, my special gratitude is due to my brothers and sisters in Hope Batu Pahat. I wish to express my deepest and sincere thanks to them who had accompanied me with prayers, love, concern, support, and advice over the past four years. This research work would not have been completed without many of you.

Thank you!

ABSTRACT

A non-linear activation function is one of the key contributing factors to the success of Deep Learning (DL). Since the revival of DL takes place in 2012, Rectified Linear Unit (ReLU) has been regarded as a de facto standard for many DL models by the community. Despite its popularity, however, ReLU contains several shortcomings that could result in inefficient learning of the DL models. These shortcomings are: 1) the inherent negative cancellation property in ReLU tends to remove all negative inputs and causes massive information lost to the network; 2) the derivative of ReLU potentially causes the occurrence of dead neurons problem to the networks; 3) the mean activation generated by ReLU is highly positive and lead to bias shift effect in the network layers; 4) the inherent multilinear structure of ReLU restricts the non-linear capability of the networks; 5) the predefined nature of ReLU limits the flexibility of the networks. To address these shortcomings, this study proposed a new variant of activation function based on the Semi-sigmoidal (Sig) approach. Based on this approach, three variants of activation functions are introduced, namely, Shifted Semi-sigmoidal (SSig), Adaptive Shifted Semi-sigmoidal (ASSig), and Bi-directional Adaptive Shifted Semi-sigmoidal (BiASSig). The proposed activation functions were tested against the ReLU (baseline) and state-of-the-art methods using eight Deep Neural Networks (DNNs) on seven benchmark image datasets. Further, Adaptive Moment Estimation (ADAM) and Stochastic Gradient Descent (SGD) were selected as optimizers to train the DNNs. The baseline comparison score and mean rank were used to consolidate and analyse the experimental results effectively. The experimental results in terms of the overall baseline comparison score shown that SSig, ASSig, and BiASSig obtained the score of 79, 87, and 86 out of 112, respectively, which achieving outstanding performance than ReLU in more than 70% of the cases. In terms of overall mean rank (OMR), ReLU ranked at tenth (10^{th}), whereas SSig, ASSig, and BiASSig ranked at fifth (5^{th}), first (1^{st}), and second (2^{nd}), showing remarkable performance than ReLU and other comparing methods.

ABSTRAK

Fungsi pengaktifan tidak linear merupakan salah satu faktor penyumbang utama kepada kejayaan Pembelajaran Mendalam (PM). Sejak berlakunya kebangkitan PM pada tahun 2012, komuniti PM telah menganggap *Rectified Linear Unit* (ReLU) sebagai *standard de facto* untuk model-model PM. Namun begitu, ReLU mempunyai beberapa kelemahan yang boleh mengakibatkan ketidakcekapan pembelajaran model PM. Kelemahan-kelemahannya adalah: 1) sifat pembatalan negatif dalam ReLU membuang kesemua input negatif dan menyebabkan kehilangan maklumat yang berlebihan; 2) terbitan ReLU menyebabkan masalah neuron mati dalam rangkaian; 3) pengaktifan min yang dihasil oleh ReLU sangat positif dan membawa kesan peralihan berat sebelah kepada lapisan-lapisan rangkaian; 4) kewujudan struktur multilinear ReLU menyekat keupayaan tidak linear; 5) sifat ReLU yang tetap menyekat fleksibiliti rangkaian. Untuk mengatasinya, kajian ini mencadangkan varian fungsi pengaktifan baharu berdasarkan pendekatan *Semi-sigmoidal* (Sig). Berdasarkan pendekatan ini, tiga varian fungsi pengaktifan diperkenalkan, iaitu, *Shifted Semi-sigmoidal* (SSig), *Adaptive Shifted Semi-sigmoidal* (ASSig), dan *Bi-directional Adaptive Shifted Semi-sigmoidal* (BiASSig). Fungsi pengaktifan yang dicadangkan diuji terhadap ReLU (garis dasar) dan kaedah *state-of-the-art* dengan menggunakan lapan Rangkaian Neural Dalam (RND) pada tujuh data gambar penanda aras. Selanjutnya, *Adaptive Moment Estimation* (ADAM) dan *Stochastic Gradient Descent* (SGD) dipilih sebagai pengoptimum untuk melatih RND. Skor perbandingan dasar dan kedudukan min digunakan untuk menyatukan dan menganalisis hasil eksperimen dengan berkesan. Hasil eksperimen dari segi keseluruhan skor perbandingan dasar menunjukkan bahawa SSig, ASSig, dan BiASSig masing-masing mencapai skor 79, 87 dan 86 daripada 112, mencapai prestasi cemerlang melebihi 70% kes berbanding ReLU. Dari segi kedudukan min keseluruhan, ReLU berada di kedudukan kesepuluh, sedangkan SSig, ASSig, dan BiASSig berada di kedudukan kelima, pertama, dan kedua, menunjukkan prestasi yang luar biasa daripada ReLU dan kaedah-kaedah lain yang dibandingkan.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

ADAM	-	Adaptive Moment Estimation
AI	-	Artificial Intelligence
ALiSA	-	Adaptive Linearized Sigmoidal Activation
AMR	-	Average of Mean Rank
ANNs	-	Artificial Neural Networks
APL	-	Adaptive Piecewise Linear
ASSig	-	Adaptive Semi-sigmoidal
BiASSig	-	Bi-directional Adaptive Semi-sigmoidal
BN	-	Batch Normalization
CIFAR	-	Canadian Institute for Advanced Research
CNNs	-	Convolutional Neural Networks
DL	-	Deep Learning
DNNs	-	Deep Neural Networks
ELU	-	Exponential Linear Unit
EReLU	-	Elastic Rectified Linear Unit
FN	-	False Negative
FP	-	False Positive
FReLU	-	Flexible Rectified Linear Unit
GANs	-	Generative Adversarial Networks
GELU	-	Gaussian Error Linear Unit
GPU	-	Graphic processing unit
GTSRB	-	German Traffic Sign Recognition Benchmark
LReLU	-	Leaky Rectified Linear Unit
ML	-	Machine Learning
MNIST	-	Modified National Institute of Standards and Technology database
MPELU	-	Multiple Parametric Exponential Linear Unit
NLP	-	Natural Language Processing

NN	-	Neural Network
OMR	-	Overall Mean Rank
PAU	-	Padé Activation Units
PELU	-	Parametric Exponential Linear Unit
PReLU	-	Parametric Rectified Linear Unit
ReLU	-	Rectified Linear Unit
RL	-	Representation Learning
RNNs	-	Recurrent Neural Networks
SCR	-	Standard Competition Ranking
SELU	-	Scaled Exponential Linear Unit
Sig	-	Semi-sigmoidal
SGD	-	Stochastic Gradient Descent
SSig	-	Shifted Semi-sigmoidal
SVHN	-	Street View House Numbers
TP	-	Truly Positive
TN	-	Truly Negative

LIST OF APPENDICES

LIST OF PUBLICATIONS

Chieng, H. H., Wahid, N., Pauline, O., & Perla, S. R. K. (2018). Flatten-T Swish: a thresholded ReLU-Swish-like activation function for deep learning. *International Journal of Advances in Intelligent Informatics*, *4*(2), 76-86.

Chieng, H. H., Wahid, N., & Ong, P. (2020). Parametric Flatten-T Swish: an adaptive nonlinear activation function for deep learning. *Journal of Information and Communication Technology*, *20*(1), 21-39.

CHAPTER 1

INTRODUCTION

1.1 Background of Study

In recent years, Deep Learning (DL) has brought a significant advancement to the Artificial Intelligence (AI) domain. Such revolutionary change is mainly attributed to the improvements made in these four aspects: high-performance graphic processing unit (GPU), accessible large-scale datasets, effective learning algorithms, and powerful Neural Network (NN) structures (Jiang *et al*., 2018; Li *et al*., 2018). Due to these advancements, a remarkable success of DL is achieved in diverse real-world applications, including autonomous driving (Grigorescu *et al*., 2020), image recognition (Xie *et al*., 2020), medical diagnostics (Ma *et al*., 2020), speech recognition (Wang *et al*., 2020), and recommender system (Bobadilla *et al*., 2020).

DL is also known as multi-level Representation Learning (RL), where raw data features are taken directly as input and transformed into higher levels of abstraction to make sense of the machine when performing complex tasks (Nwankpa *et al.*, 2018). Meanwhile, RL belongs to Machine Learning (ML), whereas ML is one of the AI classes (Goodfellow *et al.*, 2016). Figure 1.1 shows the relationship between AI, ML, RL, and DL.

DL models vary in terms of architecture and size, depending on the type or domain of application. Some examples of DL models are as follows: (a) Convolutional Neural Networks (CNNs), also known as ConvNets, consist of multiple layers and are specifically designed for analyzing visual imagery (LeCun *et al.*, 2010); (b) Recurrent Neural Networks (RNNs) are commonly used for Natural Language Processing (NLP) (Graves *et al.*, 2013); (c) Generative Adversarial Networks (GANs) are invented to deal with modern unsupervised learning problems

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Kudlur, M. (2016). Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)* (pp. 265-283).

Aghdam, H. H., & Heravi, E. J. (2017). Convolutional neural networks. In *Guide to convolutional neural networks* (pp. 85-130). Springer, Cham.

Agostinelli, F., Hoffman, M., Sadowski, P., & Baldi, P. (2014). Learning activation functions to improve deep neural networks. *arXiv preprint arXiv:1412.6830*.

Amodei, D., Ananthanarayanan, S., Anubhai, R., Bai, J., Battenberg, E., Case, C., ... & Chen, J. (2016, June). Deep speech 2: End-to-end speech recognition in english and mandarin. In *International conference on machine learning* (pp. 173-182).

Apicella, A., Isgrò, F., & Prevete, R. (2019). A simple and efficient architecture for trainable activation functions. *Neurocomputing*, *370*, 1-15.

Apicella, A., Donnarumma, F., Isgrò, F., & Prevete, R. (2020). A survey on modern trainable activation functions. *arXiv preprint arXiv:2005.00817*.

Bawa, V. S., & Kumar, V. (2019). Linearized sigmoidal activation: A novel activation function with tractable non-linear characteristics to boost representation capability. *Expert Systems with Applications*, *120*, 346-356.

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, *35*(8), 1798-1828.

Bisong, E. (2019). Google Colaboratory. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform* (pp. 59-64). Apress, Berkeley, CA.

Bobadilla, J., Alonso, S., & Hernando, A. (2020). Deep learning architecture for collaborative filtering recommender systems. *Applied Sciences*, *10*(7), 2441.

Brahimi, M., Boukhalfa, K., & Moussaoui, A. (2017). Deep learning for tomato diseases: classification and symptoms visualization. *Applied Artificial Intelligence*, *31*(4), 299-315.

Brock, A., Donahue, J., & Simonyan, K. (2018). Large scale gan training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*.

Bulatov, Y. (2011). Notmnist dataset. *Google (Books/OCR), Tech. Rep.[Online]. Available: http://yaroslavvb. blogspot. it/2011/09/notmnist-dataset. html*.

Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, *106*, 249-259.

Campolucci, P., Capperelli, F., Guarnieri, S., Piazza, F., & Uncini, A. (1996, May). Neural networks with adaptive spline activation function. In *Proceedings of 8th Mediterranean Electrotechnical Conference on Industrial Applications in Power Systems, Computer Science and Telecommunications (MELECON 96)* (Vol. 3, pp. 1442-1445). IEEE.

Chang, Y., Wu, X., & Zhang, S. (2018, August). Piecewise Linear Units for Fast Self-Normalizing Neural Networks. In *2018 24th International Conference on Pattern Recognition (ICPR)* (pp. 429-434). IEEE.

Cheng, Q., Li, H., Wu, Q., Ma, L., & King, N. N. (2020). Parametric Deformable Exponential Linear Units for deep neural networks. *Neural Networks*.

Chieng, H. H., Wahid, N., Pauline, O., & Perla, S. R. K. (2018). Flatten-T Swish: a thresholded ReLU-Swish-like activation function for deep learning. *International Journal of Advances in Intelligent Informatics*, *4*(2), 76-86.

Chieng, H. H., Wahid, N., & Ong, P. (2021). Parametric flatten-t swish: an adaptive nonlinear activation function for deep learning. *Journal of Information and Communication Technology (JICT)*, *20*(1), 21-39.

Ciuparu, A., Nagy-Dăbâcan, A., & Mureşan, R. C. (2020). Soft++, a multi-parametric non-saturating non-linearity that improves convergence in deep neural architectures. *Neurocomputing*, *384*, 376-388.

Clanuwat, T., Bober-Irizar, M., Kitamoto, A., Lamb, A., Yamamoto, K., & Ha, D. (2018). Deep learning for classical japanese literature. *arXiv preprint arXiv:1812.01718*.

Clevert, D. A., Unterthiner, T., & Hochreiter, S. (2016). Fast and accurate deep network learning by exponential linear units (elus). In *Proceedings of the International Conference on Learning Representations (ICLR)*.

Corless, R. M., Gonnet, G. H., Hare, D. E., Jeffrey, D. J., & Knuth, D. E. (1996). On the LambertW function. *Advances in Computational mathematics*, *5*(1), 329-359.

Dede, G., & Sazlı, M. H. (2010). Speech recognition with artificial neural networks. *Digital Signal Processing*, *20*(3), 763-768.

Deng, L. (2011). An overview of deep-structured learning for information processing. In *in Proc. Asian-Pacific Signal & Information Processing Annual Summit & Conference*.

Dramsch, J. S., & Lüthje, M. (2018). Deep-learning seismic facies on state-of-the-art CNN architectures. In *Seg technical program expanded abstracts 2018* (pp. 2036-2040). Society of Exploration Geophysicists.

Dubey, S. R., & Chakraborty, S. (2018). Average biased ReLU based CNN descriptor for improved face retrieval. *arXiv preprint arXiv:1804.02051*.

Dugas, C., Bengio, Y., Bélisle, F., Nadeau, C., & Garcia, R. (2001). Incorporating second-order functional knowledge for better option pricing. In *Advances in neural information processing systems* (pp. 472-478).

Erhan, D., Bengio, Y., Courville, A., & Vincent, P. (2009). Visualizing higher-layer features of a deep network. *University of Montreal*, *1341*(3), 1.

Fiori, S. (2000). Blind signal processing by the adaptive activation function neurons. *Neural Networks*, *13*(6), 597-611.

Flennerhag, S., Yin, H., Keane, J., & Elliot, M. (2018). Breaking the activation function bottleneck through adaptive parameterization. In *Advances in Neural Information Processing Systems* (pp. 7739-7750).

Gao, H., Cai, L., & Ji, S. (2020). Adaptive Convolutional ReLUs. In *Thirty-Fourth AAAI Conference on Artificial Intelligence*.

Gao, R., Huo, Y., Bao, S., Tang, Y., Antic, S. L., Epstein, E. S., ... & Landman, B. A. (2020). Multi-path xD recurrent neural networks for collaborative image classification. *Neurocomputing*.

Garg, A., Upadhyaya, S., & Kwiat, K. (2013). A user behavior monitoring and profiling scheme for masquerade detection. In *Handbook of Statistics* (Vol. 31, pp. 353-379). Elsevier.

Gibson, E., Li, W., Sudre, C., Fidon, L., Shakir, D. I., Wang, G., ... & Whyntie, T. (2018). NiftyNet: a deep-learning platform for medical imaging. *Computer methods and programs in biomedicine*, *158*, 113-122.

Glorot, X., & Bengio, Y. (2010, March). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249-256).

Glorot, X., Bordes, A., & Bengio, Y. (2011, June). Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics* (pp. 315-323).

Godin, F., Degrave, J., Dambre, J., & De Neve, W. (2018). Dual rectified linear units (DReLUs): A replacement for tanh activation functions in quasi-recurrent neural networks. *Pattern Recognition Letters*, *116*, 8-14.

Godfrey, L. B. (2019, October). An Evaluation of Parametric Activation Functions for Deep Learning. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)* (pp. 3006-3011). IEEE.

Goh, G. B., Hodas, N. O., & Vishnu, A. (2017). Deep learning for computational chemistry. *Journal of computational chemistry*, *38*(16), 1291-1307.

Gomes Mantovani, R., Debiaso Rossi, A. L., Alcobaça, E., Castro Gertrudes, J., Barbon Junior, S., & Ferreira de Carvalho, A. C. P. D. L. (2020). Rethinking Default Values: a Low Cost and Efficient Strategy to Define Hyperparameters. *arXiv e-prints*, arXiv-2008.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672-2680).

Graves, A., Mohamed, A. R., & Hinton, G. (2013, May). Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 6645-6649). IEEE.

Grigorescu, S., Trasnea, B., Cocias, T., & Macesanu, G. (2020). A survey of deep learning techniques for autonomous driving. *Journal of Field Robotics*, *37*(3), 362-386.

Hahnloser, R. H., Sarpeshkar, R., Mahowald, M. A., Douglas, R. J., & Seung, H. S. (2000). Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, *405*(6789), 947-951.

Han, Z., Hong, M., & Wang, D. (2017). *Signal processing and networking for big data applications*. Cambridge University Press.

Hayou, S., Doucet, A., & Rousseau, J. (2019). On the impact of the activation function on deep neural networks training. *arXiv preprint arXiv:1902.06853*.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision* (pp. 1026-1034).

Hendrycks, D., & Gimpel, K. (2016). Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*.

Ioffe, S., & Szegedy, C. (2015, July). Batch normalization: accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning-Volume 37* (pp. 448-456). JMLR. org.

Jagtap, A. D., Kawaguchi, K., & Karniadakis, G. E. (2020). Adaptive activation functions accelerate convergence in deep and physics-informed neural networks. *Journal of Computational Physics*, *404*, 109136.

Jiang, X., Pang, Y., Li, X., Pan, J., & Xie, Y. (2018). Deep neural networks with elastic rectified linear units for object recognition. *Neurocomputing*, *275*, 1132-1139.

Jinsakul, N., Tsai, C. F., Tsai, C. E., & Wu, P. (2019). Enhancement of Deep Learning in Image Classification Performance Using Xception with the Swish Activation Function for Colorectal Polyp Preliminary Screening. *Mathematics*, *7*(12), 1170.

Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, *6*(1), 27.

Kiliçarslan, S., & Celik, M. (2021). RSigELU: A nonlinear activation function for deep neural networks. *Expert Systems with Applications*, *174*, 114805.

Kim, J., Sangjun, O., Kim, Y., & Lee, M. (2016). Convolutional neural network with biologically inspired retinal structure. *Procedia Computer Science*, *88*, 145-154.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Klambauer, G., Unterthiner, T., Mayr, A., & Hochreiter, S. (2017). Self-normalizing neural networks. In *Advances in Neural Information Processing Systems* (pp. 971-980).

Krizhevsky, A., & Hinton, G. (2009). Learning multiple layers of features from tiny images.

Krizhevsky, A., & Hinton, G. (2010). Convolutional deep belief networks on cifar-10. *Unpublished manuscript*, *40*(7), 1-9.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

Krizhevsky, A., Nair, V., & Hinton, G. (2009). *The CIFAR-10 dataset*. Retrieved on March 30, 2020, from https://www.cs.toronto.edu/~kriz/cifar.html

Krueger, D., Maharaj, T., Kramár, J., Pezeshki, M., Ballas, N., Ke, N. R., ... & Pal, C. (2016). Zoneout: Regularizing rnns by randomly preserving hidden activations. *arXiv preprint arXiv:1606.01305*.

Laurent, T., & von Brecht, J. (2017). The multilinear structure of ReLU networks. *arXiv preprint arXiv:1712.10132*.

LeCun, Y. A., Bottou, L., Orr, G. B., & Müller, K. R. (2012). Efficient backprop. In *Neural networks: Tricks of the trade* (pp. 9-48). Springer, Berlin, Heidelberg.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436.

LeCun, Y., Cortes, C., & Burges, C. J. (2010). MNIST handwritten digit database. *AT&T Labs [Online]. Available: http://yann. lecun. com/exdb/mnist*, *2*.

LeCun, Y., Kavukcuoglu, K., & Farabet, C. (2010, May). Convolutional networks and applications in vision. In *Proceedings of 2010 IEEE International Symposium on Circuits and Systems* (pp. 253-256). IEEE.

Lee, J., Shridhar, K., Hayashi, H., Iwana, B. K., Kang, S., & Uchida, S. (2019). ProbAct: A Probabilistic Activation Function for Deep Neural Networks. *arXiv preprint arXiv:1905.10761*.

Li, B., Tang, S., & Yu, H. (2019). Better approximations of high dimensional smooth functions by deep neural networks with rectified power units. *arXiv preprint arXiv:1903.05858*.

Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D. D., & Chen, M. (2014, December). Medical image classification with convolutional neural network. In *2014 13th* *international conference on control automation robotics & vision (ICARCV)* (pp. 844-848). IEEE.

Li, Y., Fan, C., Li, Y., Wu, Q., & Ming, Y. (2018). Improving deep neural network with multiple parametric exponential linear units. *Neurocomputing*, *301*, 11-24.

Liang, H., Sun, X., Sun, Y., & Gao, Y. (2017). Text feature extraction based on deep learning: a review. *EURASIP journal on wireless communications and networking*, *2017*(1), 1-12.

Lin, G., & Shen, W. (2018). Research on convolutional neural network based on improved Relu piecewise activation function. *Procedia computer science*, *131*, 977-984.

Liu, B., & Liang, Y. (2019). Optimal Function Approximation with Relu Neural Networks. *arXiv preprint arXiv:1909.03731*.

Liu, B., & Liang, Y. (2021). Optimal function approximation with ReLU neural networks. *Neurocomputing*, *435*, 216-227.

Liu, Y., Ji, L., Gao, C., Zhang, J., & Qu, J. (2019). Natural-Logarithm-Rectified Activation Function in Convolutional Neural Networks. *arXiv preprint arXiv:1908.03682*.

Lu, L., Shin, Y., Su, Y., & Karniadakis, G. E. (2019). Dying relu and initialization: Theory and numerical examples. *arXiv preprint arXiv:1903.06733*.

Ma, X. (2020). *Banknote Serial Number Recognition Using Deep Learning.* Auckland University of Technology: Ph.D. Thesis.

Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013, June). Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml* (Vol. 30, No. 1, p. 3).

Madasu, A., & Rao, V. A. (2019). Effectiveness of Self Normalizing Neural Networks for Text Classification. *arXiv preprint arXiv:1905.01338*.

Manessi, F., & Rozza, A. (2018, August). Learning combinations of activation functions. In *2018 24th International Conference on Pattern Recognition (ICPR)* (pp. 61-66). IEEE.

Min, K., Kim, D., Park, J., & Huh, K. (2019). RNN-Based Path Prediction of Obstacle Vehicles With Deep Ensemble. *IEEE Transactions on Vehicular Technology*, *68*(10), 10252-10256.

Mishra, K., Sathish, R., & Sheet, D. (2016, December). Tracking of retinal microsurgery tools using late fusion of responses from convolutional neural network over pyramidally decomposed frames. In *International Conference on Computer Vision, Graphics, and Image processing* (pp. 358-366). Springer, Cham.

Misra, D. (2019). Mish: A Self Regularized Non-Monotonic Neural Activation Function. *arXiv preprint arXiv:1908.08681*.

Mohamed, A. R., Dahl, G. E., & Hinton, G. (2011). Acoustic modeling using deep belief networks. *IEEE transactions on audio, speech, and language processing*, *20*(1), 14-22.

Molina, A., Schramowski, P., & Kersting, K. (2019). Pad\'e Activation Units: End-to-end Learning of Flexible Activation Functions in Deep Networks. *arXiv preprint arXiv:1907.06732*.

Moore, R. E., & Cloud, M. J. (2007). *Computational functional analysis*. Elsevier.

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)* (pp. 807-814).

Naug, A., Ahmed, I., & Biswas, G. (2019, June). Online Energy Management in Commercial Buildings using Deep Reinforcement Learning. In *2019 IEEE International Conference on Smart Computing (SMARTCOMP)* (pp. 249-257). IEEE.

Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., & Ng, A. Y. (2011, December). Reading digits in natural images with unsupervised feature learning. In *NIPS workshop on deep learning and unsupervised feature learning* (Vol. 2011, No. 2, p. 5).

Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. (2018). Activation functions: Comparison of trends in practice and research for deep learning. *arXiv preprint arXiv:1811.03378*.

Ohn, I., & Kim, Y. (2019). Smooth function approximation by deep neural networks with general activation functions. *Entropy*, *21*(7), 627.

Oostwal, E., Straat, M., & Biehl, M. (2021). Hidden unit specialization in layered neural networks: ReLU vs. sigmoidal activation. *Physica A: Statistical Mechanics and its Applications*, *564*, 125517.

Ouyang, J., Chen, K. T., Gong, E., Pauly, J., & Zaharchuk, G. (2019). Ultra-low-dose PET reconstruction using generative adversarial network with feature matching and task-specific perceptual loss. *Medical physics*, *46*(8), 3555-3564.

Özbay, Y., & Tezel, G. (2010). A new method for classification of ECG arrhythmias using neural network with adaptive activation function. *Digital Signal Processing*, *20*(4), 1040-1049.

Park, Y., & Yang, H. S. (2019). Convolutional neural network based on an extreme learning machine for image classification. *Neurocomputing*, *339*, 66-76.

Pascanu, R., Mikolov, T., & Bengio, Y. (2012). Understanding the exploding gradient problem. *CoRR, abs/1211.5063*, *2*, 417.

Pascanu, R., Mikolov, T., & Bengio, Y. (2013, February). On the difficulty of training recurrent neural networks. In *International conference on machine learning* (pp. 1310-1318).

Patwardhan, N., Ingalhalikar, M., & Walambe, R. (2018). ARiA: Utilizing Richard's Curve for Controlling the Non-monotonicity of the Activation Function in Deep Neural Nets. *arXiv preprint arXiv:1805.08878*.

Pedamonti, D. (2018). Comparison of non-linear activation functions for deep neural networks on MNIST classification task. *arXiv preprint arXiv:1804.02763*.

Pereyra, G., Tucker, G., Chorowski, J., Kaiser, Ł., & Hinton, G. (2017). Regularizing neural networks by penalizing confident output distributions. *arXiv preprint arXiv:1701.06548*.

Pinto, M. M., Libonati, R., Trigo, R. M., Trigo, I. F., & DaCamara, C. C. (2020). A deep learning approach for mapping and dating burned areas using temporal sequences of satellite images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *160*, 260-274.

Qian, S., Liu, H., Liu, C., Wu, S., & San Wong, H. (2018). Adaptive activation functions in convolutional neural networks. *Neurocomputing*, *272*, 204-212.

Qiu, S., Xu, X., & Cai, B. (2018, August). FReLU: Flexible Rectified Linear Units for Improving Convolutional Neural Networks. In *2018 24th International Conference on Pattern Recognition (ICPR)* (pp. 1223-1228). IEEE.

Ramachandran, P., Zoph, B., & Le, Q. V. (2017). Searching for activation functions. *arXiv preprint arXiv:1710.05941*.

Robbins, H., & Monro, S. (1951). A stochastic approximation method. *The annals of mathematical statistics*, 400-407.

Scardapane, S., Comminiello, D., Hussain, A., & Uncini, A. (2017). Group sparse regularization for deep neural networks. *Neurocomputing*, *241*, 81-89.

Senior, A., & Lei, X. (2014, May). Fine context, low-rank, softplus deep neural networks for mobile speech recognition. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 7644-7648). IEEE.

Shin, Y., & Karniadakis, G. E. (2019). Trainability and Data-dependent Initialization of Over-parameterized ReLU Neural Networks. *arXiv preprint arXiv:1907.09696*.

Si, J., Harris, S. L., & Yfantis, E. (2018, November). A dynamic ReLU on neural network. In *2018 IEEE 13th Dallas Circuits and Systems Conference (DCAS)* (pp. 1-6). IEEE.

Stallkamp, J., Schlipsing, M., Salmen, J., & Igel, C. (2012). Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. *Neural networks*, *32*, 323-332.

Sun, K., Yu, J., Zhang, L., & Dong, Z. (2019, June). A Convolutional Neural Network Model Based on Improved Softplus Activation Function. In *International Conference on Applications and Techniques in Cyber Security and Intelligence* (pp. 1326-1335). Springer, Cham.

Sun, Y. F., Yang, X. D., Hu, Y. L., & Wang, P. (2015). ELM algorithm based on Softplus activation function and improved fisher discrimination. *Journal of Beijing University of Technology*, *41*(9), 1341-1348.

Sütfeld, L. R., Brieger, F., Finger, H., Füllhase, S., & Pipa, G. (2018). Adaptive blending units: Trainable activation functions for deep neural networks. *arXiv preprint arXiv:1806.10064*.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, *15*(1), 1929-1958.

Sussillo, D., & Abbott, L. F. (2014). Random walk initialization for training very deep feedforward networks. *arXiv preprint arXiv:1412.6558*.

Talukder, A., & Casasent, D. (2001, July). Adaptive activation function neural net for face recognition. In *IJCNN'01. International Joint Conference on Neural Networks. Proceedings (Cat. No. 01CH37222)* (Vol. 1, pp. 549-552). IEEE.

Tan, M., & Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946*.

Tang, Z., Luo, L., Peng, H., & Li, S. (2018). A joint residual network with paired ReLUs activation for image super-resolution. *Neurocomputing*, *273*, 37-46.

Tarnowski, W., Warchoł, P., Jastrzębski, S., Tabor, J., & Nowak, M. A. (2018). Dynamical Isometry is Achieved in Residual Networks in a Universal Way f or any Activation Function. *arXiv preprint arXiv:1809.08848*.

Tou, J. Y., Lau, P. Y., & Tay, Y. H. (2007, January). Computer vision-based wood recognition system. In *Proceedings of International workshop on advanced image technology* (pp. 197-202).

Tripathi, G. C., Rawat, M., & Rawat, K. (2019, October). Swish Activation Based Deep Neural Network Predistorter for RF-PA. In *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)* (pp. 1239-1242). IEEE.

Trottier, L., Gigu, P., & Chaib-draa, B. (2017, December). Parametric exponential linear unit for deep convolutional neural networks. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 207-214). IEEE.

Tsai, C. W., & Fang, Z. Y. (2021). An effective hyperparameter optimization algorithm for DNN to predict passengers at a metro station. *ACM Transactions on Internet Technology (TOIT)*, *21*(2), 1-24.

Van Rossum, G. (2007, June). Python Programming Language. In *USENIX annual technical conference* (Vol. 41, p. 36).

Vecci, L., Piazza, F., & Uncini, A. (1998). Learning and approximation capabilities of adaptive spline activation function neural networks. *Neural Networks*, *11*(2), 259-270.

Villalba, J., Chen, N., Snyder, D., Garcia-Romero, D., McCree, A., Sell, G., ... & Torres-Carrasquillo, P. A. (2020). State-of-the-art speaker recognition with neural network embeddings in nist sre18 and speakers in the wild evaluations. *Computer Speech & Language*, *60*, 101026.

Wan, S., Gong, C., Zhong, P., Du, B., Zhang, L., & Yang, J. (2019). Multiscale dynamic graph convolutional network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(5), 3162-3177.

Wang, H., & Raj, B. (2017). On the origin of deep learning. *arXiv preprint arXiv:1702.07800*.

Wang, L., Li, Q., & Guo, H. (2019). A Research on Deep Learning Model for Face Emotion Recognition Based on Swish Activation Function. *Journal of Image and Signal Processing*, *8*(3), 110-120.

Wang, W., Wang, L., Ge, X., Li, J., & Yin, B. (2020). Pedestrian Detection Based on Two-Stream UDN. *Applied Sciences*, *10*(5), 1866.

Wang, X., Qin, Y., Wang, Y., Xiang, S., & Chen, H. (2019). ReLTanh: An activation function with vanishing gradient resistance for SAE-based DNNs and its application to rotating machinery fault diagnosis. *Neurocomputing*, *363*, 88-98.

Wang, Y., Li, Y., Song, Y., & Rong, X. (2020). The Influence of the Activation Function in a Convolution Neural Network Model of Facial Expression Recognition. *Applied Sciences*, *10*(5), 1897.

Xiao, H., Rasul, K., & Vollgraf, R. (2017). Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*.

Xie, C., Tan, M., Gong, B., Wang, J., Yuille, A. L., & Le, Q. V. (2020). Adversarial examples improve image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 819-828).

Yang, X., Chen, Y., & Liang, H. (2018, July). Square Root Based Activation Function in Neural Networks. In *2018 International Conference on Audio, Language and Image Processing (ICALIP)* (pp. 84-89). IEEE.

Yarotsky, D. (2018). Optimal approximation of continuous functions by very deep relu networks. *arXiv preprint arXiv:1802.03620*.

Zaheer, R., & Shaziya, H. (2018, January). GPU-based empirical evaluation of activation functions in convolutional neural networks. In *2018 2nd International Conference on Inventive Systems and Control (ICISC)* (pp. 769-773). IEEE.

Zaremba, W., Sutskever, I., & Vinyals, O. (2014). Recurrent neural network regularization. *arXiv preprint arXiv:1409.2329*.

Zhang, Y. D., Pan, C., Sun, J., & Tang, C. (2018). Multiple sclerosis identification by convolutional neural network with dropout and parametric ReLU. *Journal of computational science*, *28*, 1-10.

Zheng, H., Yang, Z., Liu, W., Liang, J., & Li, Y. (2015, July). Improving deep neural networks using softplus units. In *2015 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-4). IEEE.

Zhou, Z. H., & Feng, J. (2017, August). Deep forest: towards an alternative to deep neural networks. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence* (pp. 3553-3559).