Journal of Ubiquitous Computing and Communication Technologies (UCCT) (2021)
Vol.03/ No.01
Pages: 10-22
https://www.irojournals.com/jucct/
DOI: https://doi.org/10.36548/jucct.2021.1.002

Overview of Configuring Adaptive Activation Functions for Deep Neural Networks - A Comparative Study

Dr. Wang Haoxiang,
Director and lead executive faculty member,
GoPerception Laboratory,
NY, USA
Email id: hw496@goperception.com.

Dr. S. Smys,
Professor,
Department of CSE,
RVS Technical Campus,
Coimbatore, India.
Email id: smys375@gmail.com

Abstract- Recently, the deep neural networks (DNN) have demonstrated many performances in the pattern recognition paradigm. The research studies on DNN include depth layer networks, filters, training and testing datasets. Deep neural network is providing many solutions for nonlinear partial differential equations (PDE). This research article comprises of many activation functions for each neuron. Besides, these activation networks are allowing many neurons within the neuron networks. In this network, the multitude of the functions will be selected between node by node to minimize the classification error. This is the reason for selecting the adaptive activation function for deep neural networks. Therefore, the activation functions are adapted with every neuron on the network, which is used to reduce the classification error during the process. This research article discusses the scaling factor for activation function that provides better optimization for the process in the dynamic changes of procedure. The proposed adaptive activation function has better learning capability than fixed activation function in any neural network. The research articles compare the convergence rate, early training function, and accuracy between existing methods. Besides, this research work provides improvements in debt ideas of the learning process of various neural networks. This learning process works and tests

ISSN: 2582-337X
Submitted:20.02.2021
Revised: 22.03.2021
Accepted: 17.04.2021
Published: 1.05.2021

the solution available in the domain of various frequency bands. In addition to that, both forward and inverse problems of the parameters in the overriding equation will be identified. The proposed method is very simple architecture and efficiency, robustness, and accuracy will be high when considering the nonlinear function. The overall classification performance will be improved in the resulting networks, which have been trained with common datasets. The proposed work is compared with the recent findings in neuroscience research and proved better performance.

Keywords: Deep networks, Adaptive activation function

1. INTRODUCTION

The Artificial Neural Network (ANN) method is simulated around 100 trillion links in the human brain. Also, this network consists of billions of computational units. The inputs of a neuron are based on output frequency, which is passed from one to another unit [1]. The neurons are connected in a network of around 100 trillion links. The hierarchical setup of neurons is a more efficient way in the neural networks. Figure 1 shows various notations of artificial intelligence (AI).

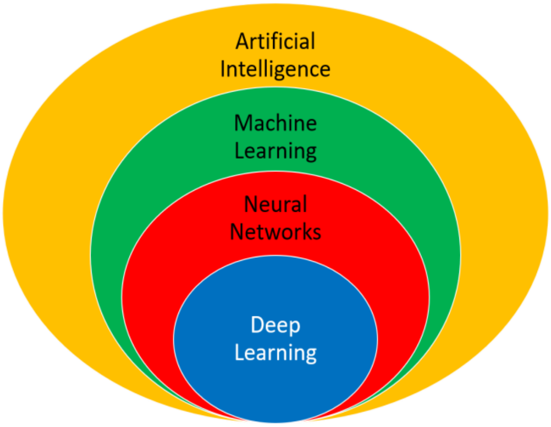

Figure 1 Various Notation of Artificial Intelligence

The study of node function in each layer, regularization techniques, which are used to minimize the ill-posed problems, network structure, and bounce forward connection, which is a great deal [2]. Generally, the activation function will be enabled in each node of a network to

ISSN: 2582-337X
Submitted:20.02.2021
Revised: 22.03.2021
Accepted: 17.04.2021
Published: 1.05.2021

learn a complex nonlinear behavior by ANN [3]. This method procedure restricts the output range of the neurons by using mathematical functions [4]. One of the efficient activations is rectified linear units (ReLU), which is the most successful procedure among all other functions [5]. The difference between ML and DL is shown in figure 2 as a block diagram representation.

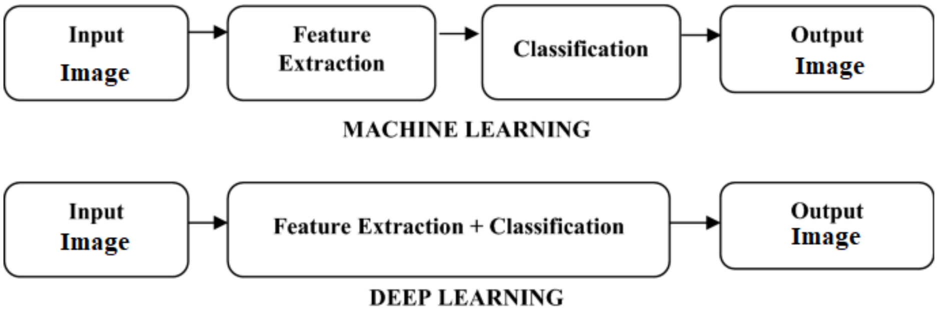

Figure 2 Differences of ML and DL

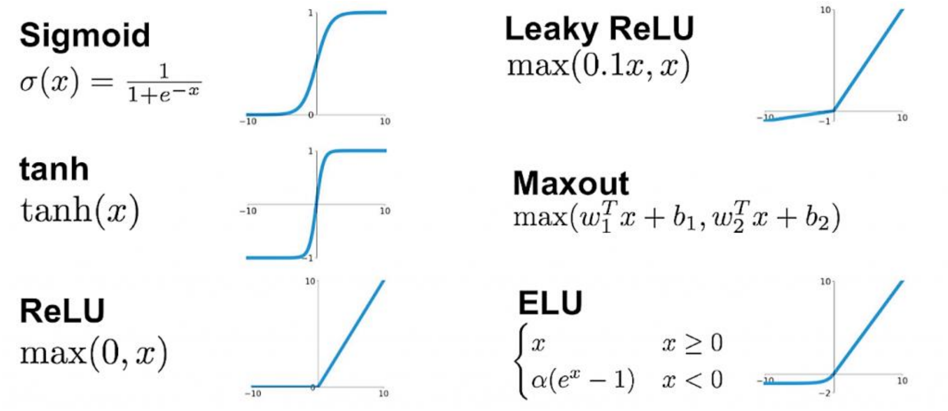

Figure 3 Various Activation Function in Neural Network

This transfer of incoming data is roaming from one unit to another through this activation function with a firing rate [6]. Figure 3 shows some graphical representations of various activation functions in neural networks. The activation function may be odd or even with inhibitory or excitatory function. The concentration of units on the network's surface is based on the concentration of the units within the units [7]. The paper has proposed a mathematical model of the feed-forward network for any continuous function of approximations [8]. A theorem has been proposed for a neural network, where inherent approximation has been activated by itself. The simplified activation approach is showing in figure 4.

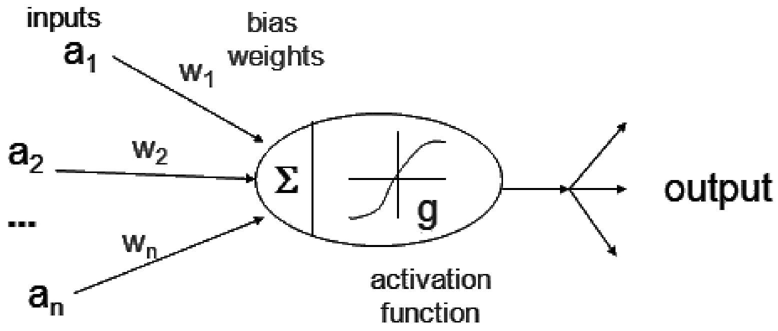

Figure 4 Simplified block diagram of activation function

The inherent architecture of the feed-forward contains a lot of hidden layers or units, which is required to approximate the function present in the account. This possible training impact on any activation function such as Rectified linear unit (ReLU), Leaky ReLU, PReLU, Exponential Linear Unit (ELU), Softmax, and soft plus with training time and memory requirement unit [9] [10][11]. The deep CNN-based architecture contains the rectified linear units, which will gain more accuracy based on its property. Besides, nonlinear function trains faster than other activation function [11][12]. Each convolutional layer is scaling the intensities within the networks with positive values of the output of the units to grow unbounded will not change the decision capabilities of any neural networks [13]. This invariant of inputs shows a good impact on any classical problems. Our research work analysis the adaptive activation function for larger deep networks with neurons in the networks. It can be configured to transform any changes in the activation function. However, this switching “ON” and “OFF” of activation function is called adaptive function inactivation procedure used for training the dataset [14]. This new framework incorporates the adaptive function inactivation procedure of neural networks, which are used to minimize the classification errors [15].

Submitted:20.02.2021
Revised: 22.03.2021
Accepted: 17.04.2021
Published: 1.05.2021

2. ORGANIZATION OF THE RESEARCH

The organization of the research is that, section 3 shows the relative works of an effective deep learning process. Methodology for configuring adaptive activation function in deep neural networks is discussed in section 4. The obtained results are categorized and discussed in section 5. Finally, Section 6 delivers a conclusion and further extension of the research.

3. PRELIMINARIES

DNN deploys various application-based networks based on shapes and size to evaluate both the accuracy and efficiency. The input for DNN is a simple set of values, which is used to represent the data analysis. The pixel values of an image are available with numerical values along with the state of systems. There are many group-based convolution deep learning architectures deployed for efficient networks. Designing an efficient deep neural network is based on depth-wise convolution layers [16]. The point-wise grouping of the layer for channel streaming is designed in this paper [17] to deliver an efficient approach. The learned group convolution architecture is one among them. All the aforementioned methods are effective and they are constructed based on redundant weights for obtaining a better accuracy. Recently, the deep learning approach takes on the direction towards adaptive learning for providing an efficient exploration of many applications [18][19]. These adaptive inference purposes in deep learning help to achieve efficiency in any system process. Generally, this adaptive approach gains high performance in terms of efficiency, flexibility, and accuracy metrics. Bolukbasi et al introduces the ensemble model of multiple-stage deep networks to compute the decision function of the variable size of networks [20]. Huang et al propose a multi-scale convolutional network with adaptive inferences during the computation stage [21]. Veit et al and Wang et al designed a Resnet architecture with variable-based choosing layers for adaptive inference [22][23]. Figurenov et al propose the entry-level adaptive spatial location method for many applications [24]. Enabling adaptive activation function in pixel tasks for every attention gating are introduced in RNN architecture is used for many dynamic processes according to the variable

ISSN: 2582-337X
Submitted:20.02.2021
Revised: 22.03.2021
Accepted: 17.04.2021
Published: 1.05.2021

coefficient in the neural network framework [21]. This RNN architecture is accelerating for many visual tracking systems with adaptive functions [15].

Most of the prior works is focusing on designing a framework with adaptive inferences for efficient works. This research article is comparing convergence rate and early training function between the existing methods. Besides the choosing, the adaptive activation function is a challenging task for various applications. For example, the sigmoid activation function is used for consisting of 2 output categories. More than 2 output categories should be used softmax activation function. Our model applied various adaptive activation function with multiple intermediated classifiers [14][2][16]. Lan et al discuss about one stage online network distillation architecture with the multi-level network for enhancing the final network [25]. Their approach is containing a huge number of networks and retaining higher accuracy. In multi-level classifiers, every single network is comprised of adaptive inference is used to provide good accuracy and efficiency.

4. METHODOLOGIES

4.1 Training dataset

The datasets are splitting into two named training and testing datasets. The training datasets are very essential to train the network in any supervised learning. We select a training dataset with normal distribution which will depend on large gradient regions. The spectral method is performing here to get a high numerical solution [26]. Besides, training data is acquired from performing experiments with low fidelity data sets. Figure 5 shows our proposed adaptive activation function for a single layer network.

ISSN: 2582-337X
Submitted:20.02.2021
Revised: 22.03.2021
Accepted: 17.04.2021
Published: 1.05.2021

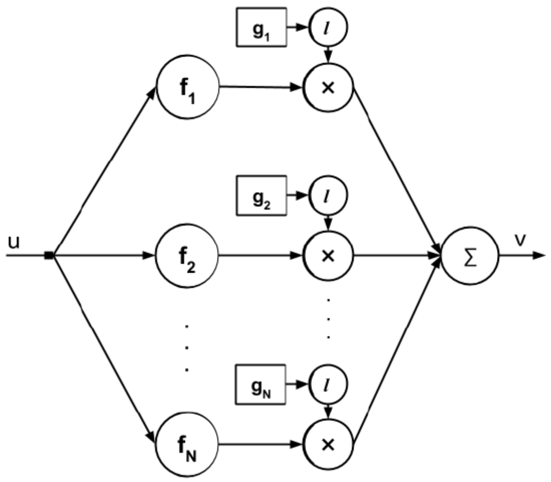

Figure 5 Adaptive Activation Function Blocks for Single Layer

4.2 Proposed Adaptive Activation Algorithm

Step 1:

The gradient descent approach is noted here;

$$w^{*} = arg \min_{w \in \Theta}(J(W))$$

$$b^{*} = arg \min_{b \in \Theta}(J(b))$$

Step 2:

The hyper-parameter “a” in the adaptive activation function,

$$\sigma\left(a\mathfrak{L}_k(x^{k-1})\right)$$

Which will be optimized, also resulting optimization provides the minimum of loss function based on weights and bias conditions,

Step 3:

The optimization issues provide to find the loss function with the help of optimized parameters is defined here;

$$a^{*} = arg \min_{a \in \mathbb{R}+\setminus\{0\}}\left(J(a)\right)$$

This parameter can be updated as,

$$a^{m+1} = a^{m} - \eta_l \nabla_a J^{m}(a)$$

Step 4:

The loss function has minimized with optimal weights is defined as,

$$J(\Theta) = MSE_R + MSE_B$$

Where, the mean square error (MSE) is given by,

$$MSE_R = \frac{1}{N_R}\sum_{i=1}^{N_R}\left|R\left(x_R^i, y_R^i, t_R^i\right)\right|^2 \;\&\; MSE_B = \frac{1}{N_B}\sum_{i=1}^{N_R}\left|B\left(x_B^i, y_B^i, t_B^i\right)\right|^2$$

Where, MSE_R denotes the residual training sets and MSE_B denotes boundary training sets.

Step 5:

Sigmoid activation function is defined as;

Sigmoid =$\frac{1}{1+e^{-ax}}$

ReLU =max(a, ax)

Leaky ReLU = max$(0, ax) - Vmax(0, -ax)$

4.3 Adaptive approach

Many attempts are created for the adaptive activation functions per layer in the deep learning algorithm. The layer optimization includes a complex non-linear function that is used for effective conditions on modern datasets. The network architectures are appended with adaptive capabilities for the function which will reduce the training times and provides better accuracy. This integration of the adaptive model with neural network architecture is used to explore the optimization concepts [27].

5. RESULTS & DISCUSSION

Various activation functions are playing an important role in the final output blocks of the neural network. This activation function will be linear transformation which is used to solve the complex problem with the absence of weights and bias limited capacity. The model graphs are showing in figure 6 for activation functions.

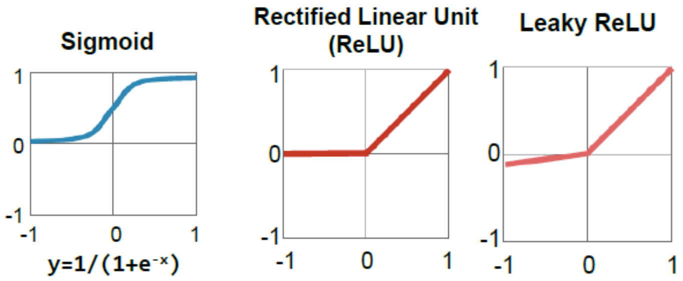

Figure 6 Model Units for Activation Function

The back propagation algorithm is evaluating the gradient of the loss function with gradient parameters. These possible gradients are supplied to the update and weights for any linear function without a differentiable equation which is not possible [28].

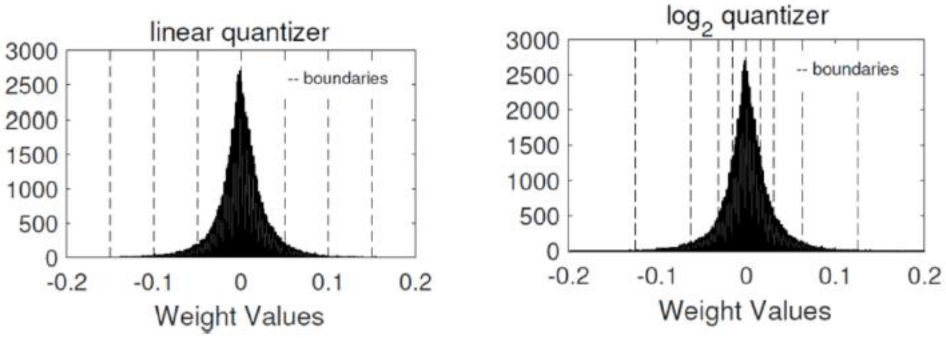

Figure 7 Results of Linear and Log Quantization

The proposed work shows better results in activation function at the tail of DNN. The correlated function in a deep network is used to solve the complex problem of training sets. Figure 7 shows the results of linear and log quantization of activation function. Mostly, the

ISSN: 2582-337X
Submitted:20.02.2021
Revised: 22.03.2021
Accepted: 17.04.2021
Published: 1.05.2021

suitable architecture is applied for adaptive based on the classification conditions. Based on this condition, the classification error can be minimized compared to the existing algorithm. The graphs show in figure 8 that comparison charts perform between the activation function with two outputs verification.

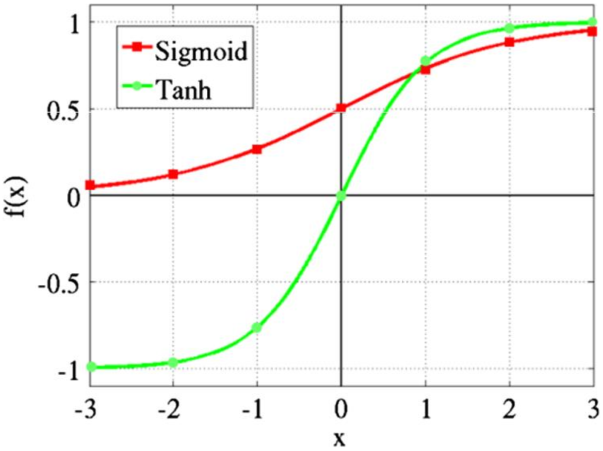

Figure 8 Comparison of Two Activation Function with 2 Outputs

This validation method is mostly a trial and error-based technique. The fine-tuning of the network provides a good performance. Then, the optimization techniques ensembles for each activation function and it is applied in the neural network for every single neural network.

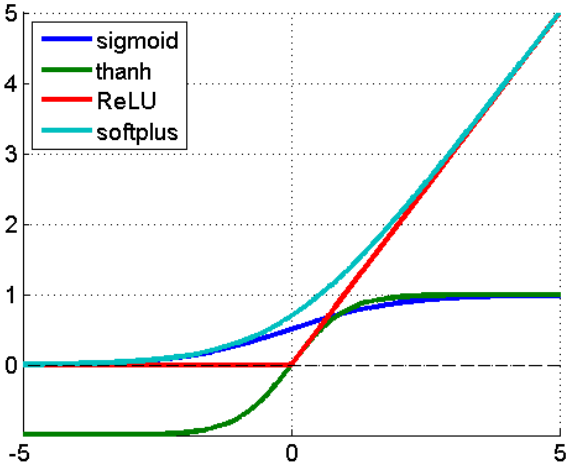

Figure 9 Comparison of Activation Function with more than 2 Outputs

ISSN: 2582-337X
Submitted:20.02.2021
Revised: 22.03.2021
Accepted: 17.04.2021
Published: 1.05.2021

This operation is applied in the process, which should be optimized in more than two outputs entity of neural network performance. The individual activation function can be performed for obtaining two types of outputs in neural networks. Thus, it shows the soft plus activation function performs better results in both the conditions shown in figure 9. The convolutional neural network intimates every node representation with the pixel of the images obtained from one layer to the next layer for individual activation function, which is optimized for dynamic conditions.

6. CONCLUSION

The proposed research paper provides an overview of the adaptive activation function that configures with a mathematical model. Based on the loss function, the network is optimized for every separated layer. This makes the classification accuracy higher and reduces the errors. The DNN consist of a large number of layers and variety of activation function to be tested by using adaptive techniques, which includes the negative impact minimization. In addition to that, the future research is extending the negative impact minimization for the activation function.

REFERENCES

[1] H. Owhadi, Bayesian numerical homogenization, Multiscale Model. Simul. 13, 812-828, 2015.

[2] E. J. Parish, K. Duraisamy, A paradigm for data-driven predictive modeling using field inversion and machine learning, J. Comput. Phys. 305, 758-774, 2016.

[3] S. Qian, et al, Adaptive activation functions in convolutional neural networks, Neurocomputing Volume 272, 10 January 2018, Pages 204-212.

[4] N. Rahaman, et al., On the spectral bias of deep neural networks, arXiv preprint arXiv:1806.08734, 2018.

[5] M. Raissi, G.E. Karniadakis, Hidden physics models: machine learning of nonlinear partial differential equations. J. Comput. Phys., 357, 125-141, 2018.

[6] M. Raissi, P. Perdikaris, G.E. Karniadakis, Numerical Gaussian processes for time-dependent and nonlinear partial differential equations. SIAM J. Sci. Comput. 40, A172-A198, 2018.

ISSN: 2582-337X
Submitted:20.02.2021
Revised: 22.03.2021
Accepted: 17.04.2021
Published: 1.05.2021

[7] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural network: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comput. Phys., 378, 686-707, 2019.

[8] M. Raissi, P. Perdikaris, G.E. Karniadakis, Inferring solutions of differential equations using noisy multi-fidelity data, J. Comput. Phys. 335 (2017) 736-746.

[9] M. Raissi, P. Perdikaris, G.E. Karniadakis, Machine learning of linear differential equations using Gaussian processes, J. Comput. Phys., 348, 683-693, 2017.

[10] M. Raissi, Z. Wang, M.S. Triantafyllou, G.E. Karniadakis, Deep learning of vortex-induced vibrations, J. Fluid Mech. (2019), vol. 861, pp. 119-137.

[11] S. Ruder, An overview of gradient descent optimization algorithms, arXiv: 1609.04747v2, 2017.

[12] S.H. Rudy, et al., Data-driven discovery of partial differential equations, Sci. Adv. 3(4), 2017.

[13] A.G. Baydin, B.A. Pearlmutter, A.A. Radul, J.M. Siskind, Automatic differentiation in machine learning: a survey, Journal of Machine Learning Research, 18 (2018) 1-43.

[14] J. Berg, K. Nystrom , Data-driven discovery of PDEs in complex datasets, J. Comput. Phys. 384 (2019) 239-252.

[15] K. Duraisamy, Z.J. Zhang, A.P. Singh, New approaches in turbulence and transition modeling using data-driven techniques, AIAA paper 2015-1284, 2015.

[16] D. P. Kingma, J. L. Ba, ADAM: A method for stochastic optimization, arXiv:1412.6980v9, 2017.

[17] Merkel, Cory, Dhireesha Kudithipudi, and Nick Sereni. ”Periodic activation functions in memristor-based analog neural networks.” Neural Networks (IJCNN), The 2013 International Joint Conference on. IEEE, 2013.

[18] A.-R. Mohamed, G. E. Dahl, and G. Hinton, ”Acoustic modelling using deep belief networks,” IEEE Transactions on Audio, Speech and Language Processing, Vol. 20, pp. 14-22, 2012.

[19] Krizhevsky, Alex and Sutskever, Ilya and Hinton, Geoffrey E. “Image net classification with deep convolutional neural networks.” Advances in neural information processing systems, pp. 1097-1105, 2012.

[20] Tolga Bolukbasi, Joseph Wang, Ofer Dekel, and Venkatesh Saligrama. Adaptive neural networks for fast test-time prediction. In *ICML*, 2017.

[21] Gao Huang, Danlu Chen, Tianhong Li, Felix Wu, Laurens van der Maaten, and Kilian QWeinberger. Multi-scale dense networks for resource efficient image classification. In *ICLR*, 2018.

[22] Andreas Veit and Serge Belongie. Convolutional networks with adaptive inference graphs. In *ECCV*, 2018.

[23] Xin Wang, Fisher Yu, Zi-Yi Dou, Trevor Darrell, and Joseph E Gonzalez. Skipnet: Learning dynamic routing in convolutional networks. In *ECCV*, 2018.

[24] Michael Figurnov, Maxwell D Collins, Yukun Zhu, Li Zhang, Jonathan Huang, Dmitry Vetrov, and Ruslan Salakhutdinov. Spatially adaptive computation time for residual networks. *arXiv preprint arXiv:1612.02297*, 2016.

[25] Xu Lan, Xiatian Zhu, and Shaogang Gong. Knowledge distillation by on-the-fly native ensemble. *arXiv preprint arXiv:1806.04606*, 2018.

[26] Karen Simonyan and Andrew Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition” CoRR, abs/1409.1556, 2014.

[27] Sergey Ioffe and Christian Szegedy. “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift.” CoRR, abs/1502.03167, 2015.

[28] He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian. ”Delving deep into rectifiers: Surpassing human-level performance on imagenet classification.” arXiv:1502.01852, 2015.

ISSN: 2582-337X
Submitted:20.02.2021
Revised: 22.03.2021
Accepted: 17.04.2021
Published: 1.05.2021