The effect of variation of Batch size on Validation Accuracy and training time for AlexNet deep learning network

Sudhir G Hate[1], Rahul M Pethe[2], Sahebrao N Patil[3], Amalraj Shankar[1]

[1]G H Raisoni University, Amravati, India
[2]G H Raisoni Institute of Engg. and Technology, RTMNU Nagpur, India
[3]JSPM's Bhivarabai Sawant Institute of Technology and Research, SPPU, Pune, India
[1]sudhir.hate@raisoni.net, [2]rahul2480@gmail.com , [3]jspmwagholipatil@gmail.com , [1]shankaramalraj@gmail.com

Abstract:- Deep Neural Networks training is handled with the different parameters like size of batch, rate of learning and time need for training. Network training with the parameters decides speed of training the network and it is expected to be low as much as possible. Use of machine learning and artificial intelligence in agriculture increasing day by day. In this research work, the experimentation of training the AlexNet network is done using MATLAB R2020a tool. The investigation of trained network's training time requires the variable batch size and learning rate consecutively. The behaviour on size of batch on the performance of AlexNet is analysed. The dataset is used in this training of AlexNet is taken from the farming images collected from the Vidarbha region of Maharashtra, a state of India. As we are using more data for analysis so required to use data science techniques for managing more data. The obtained results show the high accuracy always not achieved by keeping high learning rate. The higher learning rate needs more training time to train the network. Not only keeping learning rate low and small batch size will allow training better but it requires more training time additionally. The batch size is affected by another hyper parameters such as learning rate so the combination of these hyper parameters is as important as batch size itself.

Keywords: Weed Classification, Machine learning, deep learning, AlexNet;

DOI: 10.48047/ecb/2022.11.12.88

I. Introduction

CNNs used in primary image classification algorithms from two decades. It is having more advantages from normal neural network like higher scaling, multilayer and higher accuracy [1][2]. There are several pretrained networks such as AlexNet, Google net, VGG net etc for various applications [2]. AlexNet architecture classifying millions of images with with an accuracy of 85% whereas the conventional algorithm gives only 74% of accuracy [3]. This is the reason that the CNN is highly used in image processing area [4]. The natural advantage of CNN is feature extraction task is not done manually. So that it is robust against new datasets. Because of the above advantages CNN has many applications like classification of text, changes detection in climate , speech recognition and image processing domain[5][6]. This research focus on the study of effect of variation on batch size in the training time for AlexNet deep learning network. The dataset is used in this training of AlexNet is taken from the farming images collected from the Vidarbha region of Maharashtra, a state of India. For the classification of images, changes requires in some of the hyper parameters in the training of CNN. The network [7] performance is affected by all the hyper parameters. This research work focus on the tuning of the batch size in the training of the AlexNet pretrained network. The images collected from the agricultural lands are used in every epoch to train. By keeping hyper parameter very high system takes more time for convergence. The CNN architecture used is VGG16[8]. The effects of variation in batch size on validation accuracy and training time is observed. In next Section, literature survey on batch size is shown. In other section aims and task of the research. In Section 4 experimental Results is presented and Section 5 discuss about the conclusion.

Using machine learning recent survey was done in batch. To control the stochastic noise algorithms such as Stochastic Variance Reduced Gradient (SVRG) method and mix SGD-like steps with batch computations to [9][10][11][12]. M. Dereziński et al and M. Takac et al suggested that, the number to be set not more than 64 [13]. The trained networks demonstrate nice general [14]. The maximum value of the batch size is ranges from 64 to 512[15][8]. Y. Chen et al have analysed with higher values of recognition accuracy using number of datasets [16]. The aim is achieved with the completion of following tasks. The first task is investigating the batch size value in the accuracy of image classification. The second task is investigating the batch size value in the training time. The collected datasets is used for training the different batch size and learning rate values.

II. Research Methodology

This section describes briefly about the experiment for training AlexNet Network using MATLAB R2020a tools. The collected dataset of soybean weeds from different region of Vidarbha used for image classification. The datasets is trained using AlexNet pretrained Network. The BS and Base LR are varied and observed its effects on training time to train AlexNet pretrained network. The batch size varied from 5 to 100 and at the same time the learning rate also changes from 0.001 to 0.003. By doing variations in the above parameter finding out the training time required to train the network. The graphs showed in next session shows the behaviour of variations of all the above parameters.

III. Results and Discussion

The BS and basic LR are varied and observed its effects on training time to train AlexNet pretrained network. Initially the batch size kept 5 and changes it up to 100 and at the same time the learning rate also changes from 0.001 to 0.003.The following table shows the variation for training time and validation for different batch sizes. The values we obtained using MATLABR2020a tools got in different tables we taken only final value from each table and summarized it in the following tables shown. From the tables as we increases batch size slowly in steps of 5 we observed variations in training time and validation accuracy. Increase in batch size slowly step by step increases the training time but validation accuracy values doesn’t show any such kind of behaviour.

Table 1:- The Batch size, Training Time and validation Accuracy variation for Learning Rate 0.001

Batch Size / Epoch	Training Time (seconds)	Validation Accuracy
5	100.1073	20
10	75.07307	46.67
15	84.3194	66.67
20	100.3387	60
25	78.40708	80
30	84.11556	66.67
35	124.2753	73.33
40	144.3766	80
45	208.7786	66.67
50	238.0637	53.33
55	261.1855	73.33
60	295.8676	73.33
65	328.1584	73.33
70	258.9673	53.33
75	302.8866	53.33
80	319.6313	60
85	327.0904	73.33
90	353.2499	60
95	440.923	53.33
100	452.7012	66.67

Table 2:- The Batch size, Training Time and validation Accuracy variation for Learning Rate 0.002

Batch Size / Epoch	Training Time (seconds)	Validation Accuracy
5	100.2798	20
10	122.2598	20
15	91.04437	20
20	98.54309	46.67
25	73.15162	40
30	100.9877	60
35	116.2652	66.67
40	142.2121	60
45	177.4968	60
50	186.7443	66.67
55	225.2796	73.33
60	315.4687	60
65	307.4649	66.67
70	322.741	46.67
75	358.9309	66.67
80	418.7633	60
85	339.7318	53.33
90	366.2687	73.33
95	389.1206	86.67
100	383.2288	80

Table 3:- The Batch size, Training Time and validation Accuracy variation for Learning Rate 0.003

Batch Size / Epoch	Training Time (seconds)	Validation Accuracy
5	68.31122	1
10	76.80462	20
15	82.4896	20
20	101.106	20
25	74.688	20
30	98.51981	33.33
35	125.7031	60
40	145.5937	73.33
45	202.3968	46.67
50	231.7433	53.33
55	283.7978	60
60	304.5629	66.67
65	273.441	53.33
70	291.8199	66.67

75	288.3749	46.67
80	286.6732	80
85	330.5512	53.33
90	371.1787	66.67
95	414.8518	53.33
100	458.0658	66.67

Fig 1, 2 and 3 shows variations in classification accuracy for different batch size and different learning rate. Fig 4 shows comparison of batch size, learning rate, accuracy and training time. From figure 4 we can observe that when accuracy is good more training time required and when training time is less accuracy is not good. We can't say that particular batch size gives best performance.

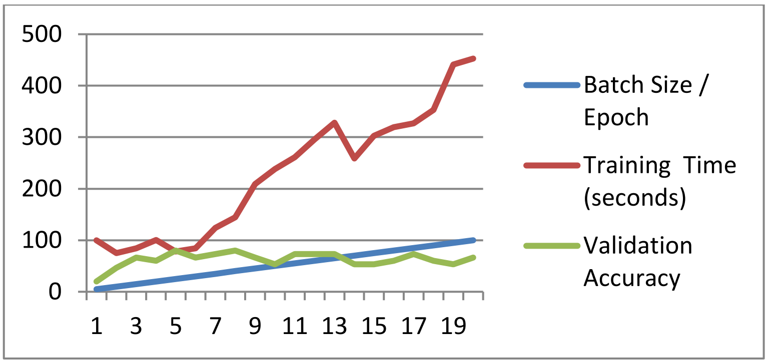

Fig 1:- For Base Learning Rate 0.001 Variations in Batch size / Epoch, Training Time & Validation Accuracy

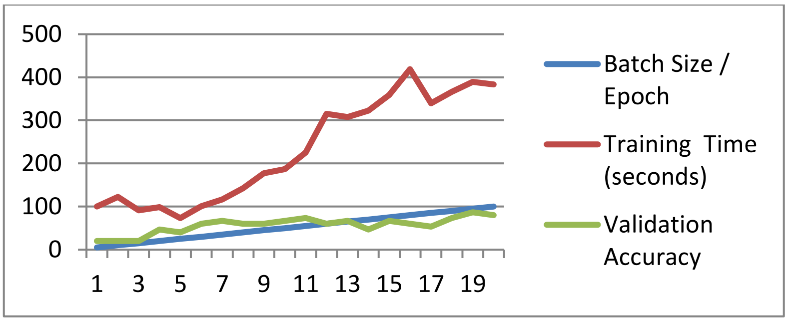

Fig 2:- For Base Learning Rate 0.002 Variations in Batch size / Epoch, Training Time & Validation Accuracy

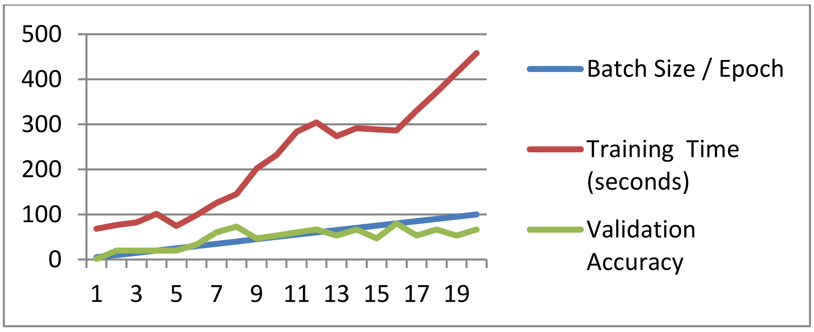

Fig 3:- For Base Learning Rate 0.003 Variations in Batch size / Epoch, Training Time & Validation Accuracy

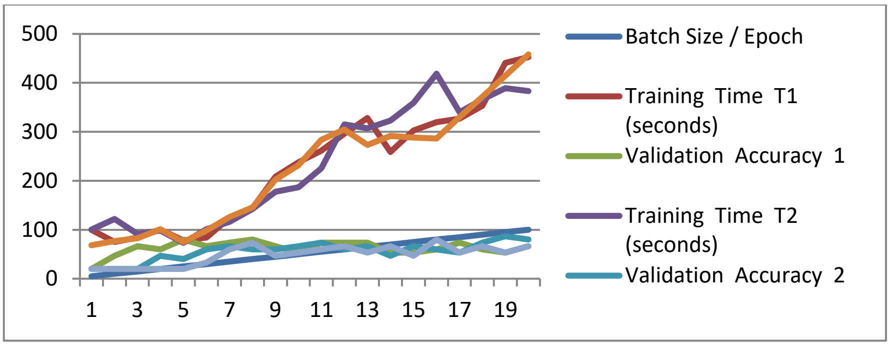

Fig 4:- Comparison of variations in Base Learning Rate, Batch size / Epoch, Training Time & Validation Accuracy

IV. Conclusion

The classification accuracy of AlexNet pretrained network by changing batch sizes and learning rates observed. As per observations, the learning rate and the batch size has a major role on the performance. When the learning rates keeps maximum ; the batch size is more then it gives better results than its small value. The greater the parameter value, the higher the image recognition accuracy. On the other hand, the large batch size value leads to more computational costs.

V. References

[1] J. Yu, A. W. Schumann, Z. Cao, S. M. Sharpe, and N. S. Boyd (2019) Weed Detection in Perennial Ryegrass With Deep Learning Convolutional Neural Network Front. Plant Sci., vol. 10, no. October, pp. 1–9, doi: 10.3389/fpls.2019.01422.

[2] R. Chauhan, K. K. Ghanshala, and R. C. Joshi (2018) Convolutional Neural Network (CNN) for Image Detection and Recognition ICSCCC 2018 - 1st Int. Conf. Secur. Cyber Comput. Commun., no. December, pp. 278–282, 2018, doi: 10.1109/ICSCCC.2018.8703316.

[3] O. Russakovsky et al. (2015) ImageNet Large Scale Visual Recognition Challenge Int. J. Comput. Vis., vol. 115, no. 3, pp. 211–252, doi: 10.1007/s11263-015-0816-y.

[4] T. F. Gonzalez (2007) Handbook of approximation algorithms and metaheuristics Handb. Approx. Algorithms Metaheuristics, pp. 1–1432, doi: 10.1201/9781420010749.

[5] A. Fesseha, S. Xiong, E. D. Emiru, M. Diallo, and A. Dahou (2021) Text classification based on convolutional neural networks and word embedding for low-resource languages: Tigrinya Inf.,vol.12,no.2,pp.1–17,doi: 10.3390/info12020052.

[6] M. Ech-Chouyyekh, H. Omara, and M. Lazaar (2019) Scientific paper classification using Convolutional Neural Networks ACM Int. Conf. Proceeding Ser., no. October doi: 10.1145/3372938.3372951.

[7] N. M. Aszemi and P. D. D. Dominic (2019) Hyperparameter optimization in convolutional neural network using genetic Algorithms Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 6, pp. 269–278, doi: 10.14569/ijacsa.2019.0100638.

[8] K. Simonyan and A. Zisserman (2015) Very deep convolutional networks for large-scale image recognition 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., pp. 1–14.

[9] L. Wang, Y. Yang, R. Min, and S. Chakradhar (2017) Accelerating deep neural network training with inconsistent stochastic gradient descent Neural Networks, vol. 93, pp. 219–229, doi: 10.1016/j.neunet.2017.06.003.

[10]M. Derezinski, D. Mahajan, S. S. Keerthi, S. V. N. Vishwanathan, and M. Weimer (2018) Batch-expansion training: An efficient optimization framework Int. Conf. Artif. Intell. Stat. AISTATS, vol. 84, pp. 736–744, 2018.

[11] P. M. Radiuk (2018) Impact of Training Set Batch Size on the Performance of Convolutional Neural Networks for Diverse Datasets, Inf. Technol. Manag. Sci., vol. 20, no. 1, pp. 20–24, doi: 10.1515/itms-2017-0003.

[12] D. R. Wilson and T. R. Martinez (2003) The general inefficiency of batch training for gradient descent learning Neural Networks, vol. 16, no. 10, pp. 1429–1451, doi: 10.1016/S0893-6080(03)00138-2.

[13] M. Takáč, A. Bijral, P. Richtárik, and N. Srebro (2013) Mini-batch primal and dual methods for SVMs, 30th Int. Conf. Mach. Learn. ICML, vol. 28, no. PART 3, pp. 2059–2067, 2013.

[14] P. Goyal et al. (2017) Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour.

[15] K. He, X. Zhang, S. Ren, and J. Sun (2016) Deep residual learning for image recognition Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-December, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[16] M. Li, T. Zhang, Y. Chen, and A. J. Smola (2014) Efficient mini-batch training for stochastic optimization Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., pp. 661–670, doi: 10.1145/2623330.2623612.