The Effect of Optimizers on Siamese Neural Network Performance

Conference Paper · August 2022

1 author:

Farah F. Alkhalid
Uuniversity of Technology-Iraq

The Effect of Optimizers on Siamese Neural Network Performance

Farah F. Alkhalid
Control and Systems Engineering
University of Technology-Iraq
Baghdad, Iraq
farah.f.alkhalid@uotechnology.edu.iq

Abstract

Optimizers are approaches or algorithms dependent to enhance the characteristics of the Neural Network (NN) like weights and learning rate in order to decrease the loss rate, On the other hand, Siamese Neural Network (SNN) are two identical sub-networks, they work in parallel and they are sharing parameters and weight, SNN uses for indicate similarity. In this research, we study the effect of optimizers Siamese Neural Network, using Digits handwritten (MINST) dataset, the effects is studied for Adam, Nadam, Adadelta and SGD optimizers with respect to process time and accuracy, the accuracy is 97%, 97%, 79% and 92%.

Keywords

Deep learning, **Siamese Network, Image classification, CNN, Adam, Adadelta, SGD,Nadam.**

1. Introduction

Deep Convolutional Neural Networks (CNNs) are becoming the highest development model for image classification and recognition procedures (Sun et al., 2020)(Kalash et al., 2018). One of the major bounds is that CNNs need a bulk classified data (data with labels) (Farah Flayyeh Alkhalid et al., 2022). In several cases, gathering this amount of data is occasionally not possible. One Shot Learning goals to resolve this big problem (Jadon & Srinivasan, 2021).

for typical classification output, the input images are applied into a sequence of convolved layers, then lastly, the output may create a probability spreading over all the named classes. In another word, if the input image classifies as dog or cat or elephant or horse, then the model generates probability value for each class (Algan & Ulusoy, 2021)(F. Alkhalid, 2020). Couple significant facts have to be attended. Firstly, large number of images are required during learning and training process. Secondly, if the neural network is trained on the above four classes, then the network cannot predict any other image from other class i.e. “Bear”. If the class “Bear” must be classified, it is very important to train the model on many “Bear images” (M et al., 2019). There are approaches may use in case no enough images are available for class or when the entire number classes is giant and dynamically varying. Thus, the cost of data collection and periodical re-training is too high (M & Alkhalid, 2019). Cannot always trust on gathering extra data, to resolve this problem, a novel kind of neural network architecture named Siamese Networks. Siamese Network depends rare images for training to get improved estimates. The skill to learn using very few data made Siamese networks high recommended to depend.

In this paper, we demonstrate deep learning using Siamese network with MNIST dataset and how it effects with changing optimizers.

2. Literatutre survey

Robin M. and *et al* in (Schmidt et al., 2020) recognized a meaningly compact subgroup of exact optimizers and parameters selections that mostly prime to competitive results in their results, ADAM stays a robust challenger, with fresher approaches failing to significantly and reliably outclass it.

Less W. and *et al* in (Wright & Demeure, 2021) introduced Ranger21, an innovative optimizer which combines AdamW with 8 components, they concluded that the final optimizer offers pointedly enhanced validation accuracy and training speed.

Shaziya and *et al* in (Shaziya & Zaheer, 2021) suggested model, which was measured well however the batch magnitude of the architecture was being a smaller amount than the testing group of database's magnitude. the suggested optimizer deported the other optimizer like Stochastic Gradient Descent (SGD). The suggested optimizer seemed better than the Adadelta optimizer, the Adam optimizer, the AdaGrad optimizer, and AdaMax optimizer also.

Xueliang W. and *et al in* (Wang et al., 2022) proposed Smish, a new nonlinear activation function, which might overcome other activation functions through well properties. results showed that Smish tendef to operate more professionally than Logish, Mish, and other activation functions on models with open datasets.

Wendyam E. and *et al* in (Ilboudo et al., 2022) proposed AdaTerm which is predictable to ignore the calculated gradients for changes, and support the strength of the next changes.

Prabhu T. and *et al* in (Prabhu et al., 2020) discussed that a reasonable valuation of optimizers' performance must income the computational charge of hyperparameter tuning into account.
Farah F. in (Farah F. Alkhalid, 2020) studied the effects of two optimizers (Adadelta and Adam) on fingerprint recognition using Deeplearning. the experimental outputs showed that the precision of test was 85.61% for Adadelta Optimizer, and 91.73%. for Adam.

3. Overview of Siamese Networks

Twins who are not completely separated where are still partially united are called Conjoined twins are widely identified as Siamese twins Chang and Eng (1811-1874), the famous conjoined Chinese twins born in Siam (Thailand) figure1 shows Siamese twin.

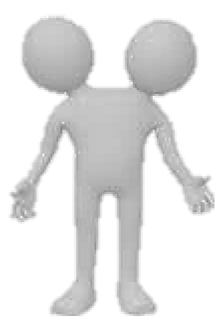


Figure 1: Siamese network

In neural network there is a class called Siamese Neural Network which consists from two or more matching subnetworks, the two subnetworks have the same parameters, weights and layers, along the training process, the weight are updated in both subnetworks and mirrored along. These couple of networks are used to check the similarity of inputs, which may utilize in different applications like Signature matching, Handwriting matching, fingerprint recognition and so on. They used to discover the likeness of the inputs by comparing their feature parameters (Dey et al., 2017), Deep learning architectures such as FaceNet, VGGFace, and dlib's ResNet face recognition model are all cases of Siamese Networks. Figure 2 denotes block diagram of Siamese Network.

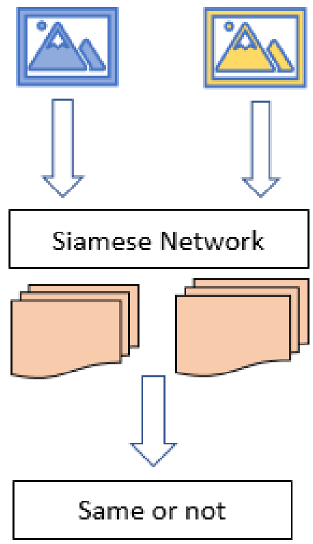


Figure 2: Block diagram of Siamese Neural Network

4. Optimizers

Deep learning training is complex architecture, which does many calculations for obtaining high accuracy. Optimization algorithms are used to eliminate the loss rate and increase accuracy rate. In optimization process, in other words, the aim of optimizers is to decrease the trains' error rate. The main achievement of deep learning statistics is to decrease the overall errors. (Aston Zhang, Zachary C. Lipton, Mu Li, 2021).
Keras is DL Library written in Python, it runs over Machine Learning platform Tensorflow, Keras is established as part of the research effort of project Open ended Neuro Electronic Intelligent Robot Operating System (ONEIROS) by François Chollet (Chollet, n.d.), below are some optimizers work under Keras:

4.1 SGD

Stochastic gradient descent achieves parameters changed for every training input $x^{(i)}$ and class $y^{(i)}$, Next the route of vertical parentage is definite to chief to development, providing that the impartial curve is flat:

$$\theta = \theta - \eta \cdot \nabla_\theta J(\theta; x(i); y(i))$$

4.2 ADAM

The term Adam is resulting from Adaptive Moment approximation. One of the main mechanisms of Adam is that it calculates exponential weighted moving to gain an estimation of the momentum and the second moment of the gradient. That is, it uses the state variables (Peel & Moon, 2020):

$$V_t \leftarrow \beta_1 V_{t-1} + (1- \beta_1)g_t$$
$$S_t \leftarrow \beta_2 S_{t-1} + (1- \beta_2)g_t^{\,2}$$

Here $\beta 1$and $\beta 2$are nonnegative weighting parameters. Common choices for them are $\beta 1=0.9$ and $\beta 2=0.999$.

4.3 NADAM

Nesterov-accelerated Adaptive Moment Estimation, associations Adam and NAG. In order to join NAG into Adam, it needs to update its momentum term m_t. First, recollection the momentum update rule using the current notation:

$$g_t = \nabla_\theta J(\theta_t)$$
$$mt = \gamma mt\text{-}1 + \eta gt$$
$$\theta_{t+1} = \theta_t - m_t$$

4.5 ADADELTA

In this optimizer, couple important parameters are used, the first is *St,* which is hold a faulty middling for the next moment of the gradient, the second is *Δxt* which is used to hold a faulty middling of the next moment of the updated of parameters in the same architecture. However, it used the novel system:

$$St = \rho\ St\text{-}1 + (1\text{-}\ \rho)\ gt\ 2$$

5. Database

Modified National Institute of Standards and Technology database (MNIST), which is consists of hundred thousand of images with size 64x64, of handwriting digits (from 0 to 9), this dataset is generally used in order to make training on numerous applications of image classification. This dataset is too extensively used to train and test the model of deep learning as one of the standard datasets. It is about 240,000 images for trainings , and about 40,000 images for tastings (LeCun et al., n.d.)

6. Model Architecture and Results

The dataset that is used in this research is MNIST dataset for digit handwriting. A Siamese Neural Network is a class of neural network architectures that contain many layers are running in 20 epochs as demonstrates in model summary below in Figure 3:

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 28, 28, 1)]	0
conv2d (Conv2D)	(None, 24, 24, 4)	104
average_pooling2d (AveragePo	(None, 12, 12, 4)	0
conv2d_1 (Conv2D)	(None, 8, 8, 16)	1616
average_pooling2d_1 (Average	(None, 4, 4, 16)	0
flatten (Flatten)	(None, 256)	0
dense (Dense)	(None, 10)	2570

Total params: 4,290
Trainable params: 4,290
Non-trainable params: 0

Figure3: Model summery

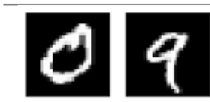
Train on 108400 samples, validate on 17820 samples, Total parameters are 4290, the Siamese model give high accura cy in training and test, reaches to 97% for training dataset, and 97% for test dataset, and best time for processing, the a ccuracy differs by differing optimizers that used in model, and the best are Adam and Nadam, where Table 1 below sh ows the accuracy with respect to optimizers:

Table 1: optimizers effects

Optimizer	Training accuracy	Test accuracy	Process Time (msec)
Adam	97%	97%	1575.4982
Adadelta	79%	82%	1866.0196
SGD	92%	93%	1976.3390
Nadam	97%	97%	1647.6546

Table 2 below, are samples of prediction using Siamese pairs for 0 handwritten digit with different digits:

Table 2: Samples of prediction for 0 digit

Pairs	Siamese Prediction
	0.2215
	0.1169
	1.0764

7. Conclusion

Siamese networks are popular model that used for check similarity, the state of the art of this work is studying the effect of optimizers on accuracy process, as denoted in the results, the best outcome is recorded when using Adam optimizer and Nadam (97%), because of these two algorithms are close to each other, while the worse one is Adadelta (79%), this results are recorded for Siamese network, it is not need to be a rule for other deeplearning models, so, the future work recommended is to compare among optimizers with different models.

References

Algan, G., & Ulusoy, I. (2021). Image classification with deep learning in the presence of noisy labels: A survey. *Knowledge-Based Systems*, *215*(89). https://doi.org/10.1016/j.knosys.2021.106771

Alkhalid, F. (2020). Online Preprocessing of Gesture Signs Using Background Substructure and Edge Detection Algorithms. *International Journal of Simulation Systems Science & Technology*, 1–6. https://doi.org/10.5013/ijssst.a.21.02.02

Alkhalid, Farah F. (2020). The effect of optimizers in fingerprint classification model utilizing deep learning. *Indonesian Journal of Electrical Engineering and Computer Science*, *20*(2), 1098–1102. https://doi.org/10.11591/ijeecs.v20.i2.pp1098-1102

Alkhalid, Farah Flayyeh, Albayati, A. Q., & Alhammad, A. A. (2022). Expansion dataset COVID-19 chest X-ray using data augmentation and histogram equalization. *International Journal of Electrical and Computer Engineering*, *12*(2), 1904–1909. https://doi.org/10.11591/ijece.v12i2.pp1904-1909

Aston Zhang, Zachary C. Lipton, Mu Li, and A. J. S. (2021). *Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola - Dive into Deep Learning.pdf*.

Chollet, F. (n.d.). *Keras*. Retrieved March 2, 2022, from https://keras.io/

Dey, S., Dutta, A., Toledo, J. I., Ghosh, S. K., Llados, J., & Pal, U. (2017). *SigNet: Convolutional Siamese Network for Writer Independent Offline Signature Verification*. *1*, 1–7. http://arxiv.org/abs/1707.02131

Ilboudo, W. E. L., Kobayashi, T., & Sugimoto, K. (2022). *AdaTerm: Adaptive T-Distribution Estimated Robust Moments towards Noise-Robust Stochastic Gradient Optimizer*. http://arxiv.org/abs/2201.06714

Jadon, S., & Srinivasan, A. A. (2021). Improving Siamese Networks for One-Shot Learning Using Kernel-Based Activation Functions. *Advances in Intelligent Systems and Computing*, *1175*, 353–367. https://doi.org/10.1007/978-981-15-5619-7_25

Kalash, M., Rochan, M., Mohammed, N., Bruce, N. D. B., Wang, Y., & Iqbal, F. (2018). Malware Classification with Deep Convolutional Neural Networks. *2018 9th IFIP International Conference on New Technologies, Mobility and Security, NTMS 2018 - Proceedings*, *2018-Janua*, 1–5. https://doi.org/10.1109/NTMS.2018.8328749

LeCun, Y., Cortes, C., & Burges, C. J. C. (n.d.). *THE MNIST DATABASE of handwritten digits*. Retrieved July 18, 2021, from http://yann.lecun.com/exdb/mnist/

M, A. M., & Alkhalid, F. F. (2019). *Recognize Arabic Handwritten using CNN Model*. *27*, 359–367.

M, A. M., Alkhalid, F. F., & Oleiwi, B. K. (2019). *Online Blind Assistive System using Object Recognition*. *3*(12), 47–51.

Peel, C., & Moon, T. K. (2020). Algorithms for optimization. In *IEEE Control Systems* (Vol. 40, Issue 2). https://doi.org/10.1109/MCS.2019.2961589

Prabhu, T. S., Mai, F., Vogels, T., Jaggi, M., & Fleuret, F. (2020). Optimizer benchmarking needs to account for hyperparameter tuning. *37th International Conference on Machine Learning, ICML 2020*, *PartF16814*, 8983–8992.

Schmidt, R. M., Schneider, F., & Hennig, P. (2020). *Descending through a Crowded Valley - Benchmarking Deep Learning Optimizers*. http://arxiv.org/abs/2007.01547

Shaziya, H., & Zaheer, R. (2021). Impact of Hyperparameters on Model Development in Deep Learning. *Lecture Notes on Data Engineering and Communications Technologies*, *56*, 57–67. https://doi.org/10.1007/978-981-15-8767-2_5

Sun, Y., Xue, B., Zhang, M., & Yen, G. G. (2020). Evolving Deep Convolutional Neural Networks for Image Classification. *IEEE Transactions on Evolutionary Computation*, *24*(2), 394–407. https://doi.org/10.1109/TEVC.2019.2916183

Wang, X., Ren, H., & Wang, A. (2022). *Smish : A Novel Activation Function for Deep Learning Methods*.

Wright, L., & Demeure, N. (2021). *Ranger21: a synergistic deep learning optimizer*. http://arxiv.org/abs/2106.13731

Biography

Farah Alkhalid: MSc. In Computer Engineering Department, Al-Nahrain University in Iraq, I work as lecturer at University of technology, Iraq in Control and systems Engineering Department, I have 15 published papers in journals and two published papers in conferences, I supervised 5 graduated projects, I awarded in two competitions (7th Robotics and Automation exhaustion, Explorer Robots exhaustion), I worked previously in Computer Center at the University of Technology, I worked in programming database systems and I have 3 database systems depended in the university. I presented courses in (Database management systems, SQL Server, Windows server, Arduino, AutoCAD and Visual Studio.NET).

- Scopus Link: https://www.scopus.com/authid/detail.uri?authorId=57204283244