Ilkogretim Online - Elementary Education Online, 2021; Vol 20 (Issue 3): pp. 3522-3527
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2021.03.369

Recognition Of Handwritten Digits Using Cnn

[1]Ishika Gupta, [2]Manoj Diwakar, [3] Ms Manisha Aeri

[1,2]Graphic Era deemed to be University Dehradun, Uttarakhand, India.

[3]Assistant Professor, Department of Computer Science and Engineering, Graphic Era Hill University, Dehradun.

ABSTRACT

In deep learning a lot of changes have come over the years and one such change is the use of Convolution Neural Network(CNN). CNN is a sub-domain of Artificial Neural Network (ANN), discovered by a postdoctoral researcher Yann LeCunn. In today's time, deep learning is used in many industries with different applications like unmanned cars, news clustering, fraud news identification, processing high-level language, fraud detection, etc. Convolution neural networks are very useful in extracting distinct features of handwritten characters which makes them the natural choice for solving complex problems related to handwritten digit recognition. This paper aims to identify different optimization algorithms that can be used for handwritten recognition, evaluate those optimization techniques, and find the most accurate optimization technique.

Key words: convolution neural network, stochastic gradient descent, adam, Rmsprop (root mean square prop), adadelta, adagrad

1. INTRODUCTION

In today's age, the world is moving towards digitalization and handwritten documents are being stored electronically. But due to different writing styles, we often get confused about distinctly identifying numbers. And this is where deep learning is used, to scan the document, extract the features of the digit, identify the digit, and store them in a standard font and format electronically. This technique can be applied to digit recognition in license plates, postal letter-sorting, cheque validation, and storing documents of historical and archaeological importance electronically. Its use can be extended by using it to store old library books as well. All these areas deal with data in huge volumes and require high recognition accuracy. The novelty of the proposed work is a thorough investigation of optimization techniques used in CNN architecture to deliver the best recognition accuracy for MNIST dataset digit recognition.

In the context of deep learning methodologies, convolutional neural networks are included in the category of deep neural networks. The input is subjected to a convolutional linear process, during which it is multiplied by a set of weights. Convolution neural networks are also known as shift invariant neural networks and space invariant neural networks. The shared weight architecture of the convolution kernels or filters, which slide with respect to the input features and provide translation equivariant responses known as feature maps, is another name for these types of neural networks. CNNs take the input data and use it to apply weights and biases to the objects in the image so that they may be differentiated from one another. The data is modified by utilizing filters inside the various layers of the CNN. This network's net layer contains all the neurons that are connected to the neurons in all the other layers.

The machine learning and pattern recognition technology known as K. Fukushima's neocognitron was first presented to the public in the year 1980. It was the first step toward the artificial intelligence that we have today and was inspired by the work of Hubel and Wiesel. The necognitron was responsible for the implementation of convolution layers and downsampling layers into CNNs. When a convolution layer is used, the receptive fields of the units in that layer cover patches from the layer below it. It is possible for units to share their filters. The receptive fields of the convolution layers are covered by the units that make up the downsampling layers. J. Weng and colleagues came up with a method that they named max-pooling as an alternative to Fukushima's spatial averaging. In this method, a downsampling unit calculates the maximum of the activations of the units in its patch, which ultimately results in the cresceptron. LeNet-5 is a pioneering 7-level convolution network that was built by LeCun et al in 1998. It has recently been utilized by a number of banks in order to recognize hand-written numerals on scanned images of cheques that have 3232 pixels. Convolution neural networks need to be larger and more sophisticated in order to process higher-resolution images because the processing of these images requires more computer resources.

2. Related Work

[1] A novel deep learning architecture called DIGINET was introduced by Huseyin Kusetogullari et al that uses a large digit dataset known as DIDA, which is in red, green, and blue color spaces to detect digits in historical documents dating back to the 19th century, written by different priests in different handwriting.
[2]In 2017, Mahmoud M. Abu Ghosh et al published by keeping Deep Neural Network(DNN), Deep belief network(DBN) and convolution neural network(CNN) as the main focus. In this paper, the authors have used both standard and random datasets to prove that among the three neural network approaches, deep neural network is the most accurate.
[3]In a paper published by Caiyun Ma et al, a new and effective method of fusing images into a deep neural network was implemented for extracting important features of data and the results showed that the performance of this technique of varied feature extraction and data attributes is better than traditional methods, which make use of basic visual features. Moreover, this feature can be used for a variety of images in different datasets.
[4] Sonia Flora et al proposed a convolution neural network in which the CNN model gives less average error than the artificial neural network model.
[5]Archana N. Vyas et al proposed the recognition of Gujrati handwritten numeral using three techniques, the first being spatial domain in which a modified chain code method is used which reduces local noise, the second method in which 85 dimensional Fourier

descriptors are computed, and the third method which uses DCT coefficient(which are used as feature vectors). The result proves that using handwritten recognition in areas where speed matters like postal mail sorting, modified chain code method with some improvement can be used.

[6]T. Y. Zhang et al proposed an algorithm that performs multiple operations in a given time for Thinning digital patterns and sub-iterations and aims at removing the pixels which form the exoskeleton, finally providing just a skeleton of the pattern.

[7]G. G. Rajput et al described a method for recognition of Marathi handwritten numerals which makes use of Fourier descriptors, to differentiate between different numerals on the basis of their shape. In this method, the SVM classifier proves to be more accurate in classifying the digits correctly with respect to other classifiers.

[8]Nafiz Arica et al discussed and compared current character recognition techniques and proposed directions for future research.

[9]Vijay Laxmi Sahu et al discussed the relevance of extraction of different data attributes as the most important factor resulting in higher accuracy in character recognition.

[10]Simone Marinai et al discussed document analysis and recognition, and the principal themes in DAR research.

[11]Lawrence O'Gorman et al reviewed specific techniques used in Optical Character Recognition(OCR)

3. Methods:

There are several stages involved in the process of handwritten digit recognition which are shown in figure 1.

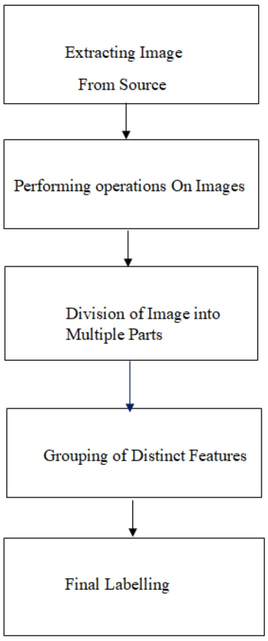

Figure 1: Framework for handwritten digit recognition

- Extraction Of Image From Source: It is the step in which a handwritten digit is stored electronically with the help of a camera with proper resolution.
- Performing Operation on Images: In this step, an operation on the image pixel is performed to remove noise from the image.

- Division Of Images into Multiple Parts: In this step, all images are made of uniform size
- Grouping of Distinct Features: In this step, we extract the features of the image like edges and horizontal and vertical lines.
- Final Labelling: In this step, the image is classified with the help of its features.
-

Figure 2 shows the framework of CNN for handwritten digit recognition

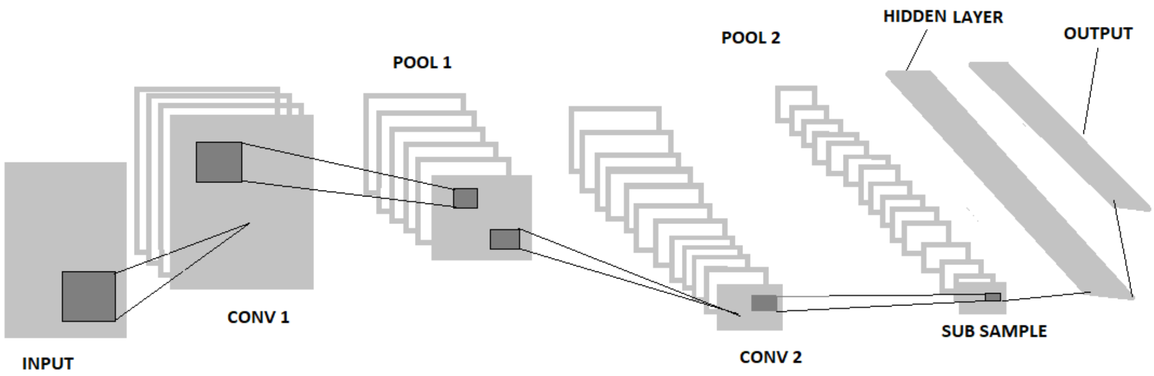

Figure 2: CNN framework

Optimizers:

1)Stochastic Gradient Descent: In this optimization technique we change the cost function and try to minimize the error between the predicted vs actual value for just one training data point at a time. The overall cost function here becomes the average of individual cost functions. It can be expressed as:

$\mathrm{O} = \mathrm{O} - \mathrm{N} \cdot \nabla \mathrm{OT}(\mathrm{O}; \mathrm{A}(\mathrm{i}); \mathrm{B}(\mathrm{i}))$

Batch gradient descent is time efficient when working with large datasets because it recompiles gradients for related samples prior to each parameter update. This saves time over individual gradient updates. By simply executing one update at a time, SGD removes the necessity for doing a numerous number of updates. Because of this, it is typically quite a bit quicker, and it can also be utilized for the purpose of online learning.

2)Adagrad: Adagrad was originally used for convex objectives that had empirical loss form. It is a modified form of stochastic gradient descent. It can be expressed as:

$$\mathrm{OK} + 1, \mathrm{i} = \mathrm{OK}, \mathrm{i} - \mathrm{N}\sqrt{\mathrm{MK}}, \mathrm{ii} + \mathrm{D} \cdot \mathrm{ZK}, \mathrm{iO} + 1,$$
$$\mathrm{i} = \mathrm{Ok}, \mathrm{i} - \mathrm{NMk}, \mathrm{ii} + \mathrm{D} \cdot \mathrm{Dk}, \mathrm{i}.$$

3)Adadelta: It was the goal of Adadelta to soften the blow of Adagrad's relentlessly accelerating decline in learning rate. Adadelta is the tool you need to use if you want to view every squared gradient that has ever been compiled into a database. A recursion of decaying means of prior squared gradients is employed so that there is no need for keeping track of w squared gradients inefficiently. When y is given as a percentage, the running average P[Z2]k at time step t is solely dependent on the most recent average as well as the gradient that is now being experienced. It can be expressed as:

$$P[Z2]k = yP[Z2]k - 1 + (1 - y)Z2kP[Z = yP[Z2]k - 1 + (1 - y)Zk2.$$

4)Adam: Adaptive Moment Estimation (Adam) is an alternative method that can be utilized to compute adaptive learning rates (Adam) for each parameter. The concepts of Momentum and RMSProp have been merged into a single concept. In the same way that momentum is retained, Adam keeps track of an exponentially decaying mean of prior squared gradients vlk in addition to the previously mentioned vlk squared gradient average. The decaying average gradients of the past and the past squared gradients are computed using this method:

$$mlt = B1mlt - 1 + (1 - B1)$$

$$Zkvlk = B2vlk - 1 + (1 - B2)Z2k$$

where

$$ml' = (ml^{k})/(1-B^{k}{}_{1})$$

$$vl_{k}'=(vl^{k})/(1-B_{2}{}^{k})$$

3. Result and Discussion:

Dataset: In the field of computer vision, the data collection known as MNIST (which stands for "Modified National Institute of Standards and Technology") is the most popular option. Since they were first shown to the public in 1999, these pictures have served as the gold standard baseline for all classification systems. In this collection, the digits 0 through 9 are each represented by a grayscale image that is 28 pixels by 28 pixels. This dataset includes 60,000 images for training purposes, as well as 10,000 test images. There are total 70,000 pictures included in the collection, and they have been organised using 10 different categories.

In this paper, major optimization techniques like Stochastic Gradient Descent, Adam, Adagrad, and Adadelta are used and compared. The result can be clearly seen in table 1.

Table 1: Comparative study

OPTIMIZER	MEAN ACCURACY	LEARNING RATE
SGD	98.46	0.01
ADAM	97.82	0.01
ADADELTA	92.90	0.01
ADAGRAD	97.83	0.01

By looking at the table we can clearly conclude that Stochastic Gradient Descent is the best optimization technique for getting good accuracy but at the same time we cannot rule out the possibility of using other techniques which can also provide good accuracy and better performance than Stochastic Gradient Descent.

5. Conclusion:

After considering the accuracy provided by stochastic Gradient Descent we can conclude that it is the best optimization technique for getting the most accurate results and is the first choice for getting the best results in handwritten digit recognition using CNN.

6. References:

[1] Huseyin Kusetogullari, Amir Yavariabdi, Johan Hall, Niklas Lavesson,DIGITNET: A Deep Handwritten Digit Detection and Recognition Method Using a NewHistoricalHandwrittenDigitDataset,BigDataResearch,Volume23,2021,100182,ISSN22145796,https://doi.org/10.1016/j.bdr.2020.100182.(https://www.sciencedirect.com/science/article/pii/S2214579620300502)

[2] M. M. Abu Ghosh and A. Y. Maghari, "A Comparative Study on Handwriting Digit Recognition Using Neural Networks," 2017 International Conference on Promising Electronic Technologies (ICPET), 2017, pp. 77-81, doi: 10.1109/ICPET.2017.20.

[3] Caiyun Ma and Hong Zhang, "Effective handwritten digit recognition based on multi-feature extraction and deep analysis," 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 2015, pp. 297-301, doi: 10.1109/FSKD.2015.7381957.

[4] Sonia Flora, Parth Goel, Anju Kakkad, "A Survey on Feature Extraction Methods & Classifiers for Handwritten Gurmukhi Character Recognition," International Journal of Computer Sciences and Engineering, Vol.7, Issue.2, pp.313-320, 2019.

[5] Vyas, Archana & Switalmac04@gmail.com,. (2015). Classification of offline gujarati handwritten characters. 10.1109/ICACCI.2015.7275831.

[6] T. Y. Zhang and C. Y. Suen. 1984. A fast parallel algorithm for thinning digital patterns. Commun. ACM 27, 3 (March 1984), 236–239. DOI:https://doi.org/10.1145/357994.358023

[7] Ganapatsingh G Rajput and Shankar Mali from https://www.researchgate.net/publication/45718432_Marathi_Handwritten_Numeral_Recognition_using_Fourier_Descriptors_and_Normalized_Chain_Code.

[8] Arica, Nafiz & Yarman Vural, Fatos. (2001). An overview of character recognition focused on off-line handwriting. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on. 31. 216 - 233. 10.1109/5326.941845.

[9] Vijay Laxmi Sahu, Babita Kubde, "Offline Handwritten Character Recognition Techniques using Neural Network: A Review", International Journal of Science and Research (IJSR), https://www.ijsr.net/search_index_results_paperid.php?id=IJSR13010129, Volume 2 Issue 1, January 2013, 87 – 94.

[10] Marinai S. (2008) Introduction to Document Analysis and Recognition. In: Marinai S., Fujisawa H. (eds) Machine Learning in Document Analysis and Recognition. Studies in Computational Intelligence, vol 90. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-76280-5_1.

[11] Kasturi, R., O'Gorman, L. & Govindaraju, V. Document image analysis: A primer. Sadhana **27,** 3–22 (2002). https://doi.org/10.1007/BF02703309