Effective Handwritten Digit Recognition using Deep Convolution Neural Network

Research *in* International Journal of Advanced Trends in Computer Science and Engineering · April 2020

DOI: 10.30534/ijatcse/2020/66922020

2 authors:

Bharadwaj Yellapragada
K L University

Bhanu PRAKASH Kolla
K L University

ISSN 2278-3091
Volume 9 No.2, March -April 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse66922020.pdf
https://doi.org/10.30534/ijatcse/2020/66922020

Effective Handwritten Digit Recognition using Deep Convolution Neural Network

Yellapragada SS Bharadwaj[1], Rajaram P[2], Sriram V.P[3], Sudhakar S[4], Kolla Bhanu Prakash[5]

[1] Dept. of CSE, Koneru Lakshmaiah Education Foundation, India,
saisrinivasabharadwaj@gmail.com
[2]Dept. of CSE, AKT Memorial College of Engineering and Technology, Kallakurichi, Tamil Nadu, India,
rajaramnov82@gmail.com
[3]Dept. of Management Studies, Acharya Bangalore B School (ABBS), Bengaluru,India
dr.vpsriram@gmail.com
[4]Dept. of CSE, Sree Sakthi Engineering College, Coimbatore, Tamil Nadu India,
sudhasengan@gmail.com
[5]Dept. of CSE, Koneru Lakshmaiah Education Foundation, India,
drkbp@kluniversity.in

ABSTRACT

This paper proposed a simple neural network approach towards handwritten digit recognition using convolution. With machine learning algorithms like KNN,SVM/SOM, recognizing digits is considered as one of the unsolvable tasks due to its distinctiveness in the style of writing. In this paper, Convolution Neural Networks are implemented with an MNIST dataset of 70000 digits with 250 distinct forms of writings. The proposed method achieved 98.51% accuracy for real-world handwritten digit prediction with less than 0.1 % loss on training with 60000 digits while 10000 under validation.

Key words: Convolution neural networks, MNIST dataset, TensorFlow, OCR, Segmentation, Cross-Validation

1. INTRODUCTION

Advancements in the field of computer vision using deep neural networks attract attention; thus, many A.I. practitioners are moving towards it.[1] One of the influencing projects that opted for deep learning is OCR (Object Character Recognition). OCR is a mechanism that converts printed or documented letters into encoded text. Intuitively OCR tool scans a document to extract the information and store it in a digital document. There are two major ways of implementing OCR, one by recognizing the patterns in the characters, and the other one is through segmentation.[1][10] Either way, this problem comes under the radar of machine learning. Handwritten digit recognition (HDR) is a snippet of OCR where instead of taking the whole character's data, HDR detects digits. Comparing to OCR, HDR is light and faster. In fields like medical, banking, student management, and taxation process, HDR possesses great flexibility.[3]

A human brain can easily interpret a sophisticated image and can extract data from it, but for a computer, image is a collection of pixels which are nothing but a list of combinations of numbers ranging from 0 to 255, blends being RGB, B.W., greyscale, etc. Information in an image can be extracted from the pixel values. A human eye contains light-sensitive cells that can segment the image into partitions which help to interpret the information like shape, size, and color which are called features, these features are later sent to the visual cortex that analyses the characteristics and maps it to our memory for identification. Thus our brain is capable of identifying an unseen image.

Neural Networks are a branch of machine learning inspired by the human brain. It adapts layered architecture where the layers represent mathematical functions, and each neutron is a carrier of weights. Layers that lie between input and output layers are called hidden layers, which can be changed based on the task. For image classification CNN (Convolution Neural Networks) performs better, and for speech recognition, Long short-term memory network delivers better.[4] Intuitively convolution is a process of applying various filters on images to highlight the details in it. Each binary classifier called perceptron, many perceptron mapping from each neuron to every other neuron of forwarding layers together forms a neural network.[2] Consider an input greyscale image of size 28x28, and the input layer consists of 784 neurons, each with a value of grey level corresponding to the pixels, and the weight carried by that neuron is called activation. The last layer, aka the output layer, contains neurons equal to the number of target classes; in case of the handwritten digit, recognition number of categories are ten ranging from 0 to 9 with probability values.[3][4] The number of input neurons and output neurons depends on the task, but the hidden layers are arbitrary and are not dependent on the task. That's why they are hidden layers, but the propagations between the hidden layers depend on the activation of the previous layers.[11] [12] Intuitively, the pattern of activations in the presentation layer causes some specific patterns in the next layer; thus, the highest activation neuron is the network's choice of class. In this paper, we are implementing convolution neural network architecture with relu and sigmoid activation functions to predict a real-world handwritten digit by training the network with the MNIST dataset.[10]

2. RELATED WORK

Many techniques have been developed to recognize handwritten digits; most of the A.I. practitioners use this to test their model's performance. In the past decades, a segmentation-based approach was used to solve this problem later with the advancements in machine learning a segmentation less approach was introduced. Even though the implementation changes, the issue still remains the same and open for anyone to solve.

Bailing Zhang [5] utilizes the ASSOM technique to provide numerical stability that can predict precise digits. Their main idea was to use the SOM clustering algorithm with autoencoder neural networks in a nonlinear approach. The modularity of SOM helped in extracting several features in a digit. For each digit, individual ASSOM was constructed and compared with ten several construction-related errors to minimize the misclassification. Their model shows promising results, even with small training samples.

Saleh Aly [3] proposed a technique for handwritten numerical strings of arbitrary length recognition using SVM and PCA, addressing the major challenge in word detection, which is overlapping characters. Their method uses hybrid PCA called PCANet for segmentation and SVM for segmentation classification together called PCA-SVMNet. Their experiment shows high efficiency in recognizing unknown handwritten number classification without any segmentation method applied [15]

Yue Yin; Wei Zhang [1] have concluded that out of all neural network implementations CNN method is valid for OCR based image classification systems. They claimed that OCR had become a preliminary technique in the field of computer vision; they state that if an image classification model performs well in OCR, then it can be used for any image classification systems.

Mahmoud M. Abu Ghosh [8] compared CNN, DNN, DBN approaches to determine which neural network is useful in the field of computer vision, specifically in image classification like OCR. Their analysis claims that DNN's(Deep Belief Network) accuracy beats other neural networks with 98.08% accuracy but falls behind CNN in terms of execution time. they concluded that any algorithm would only have 1 or 2 percent error rate towards the similarity of digits [13] [14]

A.K. Jain [9] have presented an approximation based KNN classification algorithm for handprinted digit recognition. He performed matching on two character's deforming edges and dissimilarities. Their proposed work is on patterns in low-dimensional space where the scaling is 2000 times lesser results 99.25% accuracy.

R.Alhajj [10] has applied a completely new approach called the agent-oriented approach. In a sentence, they appoint agents to each character where their only purpose is to identify hills and valleys, which are nothing but blacks and whites. The ability of agents to socialize with each other is highlighting features compared with any other image classification techniques. Overlapping digits are identified based on the cut-points, which is nothing but the intersection of agent paths. The results of their methods are surprisingly higher when compared to many other ANN-based approaches at about 97% accuracy.

3. METHODOLOGY

In this paper, we used MNIST as a primary dataset to train the model, and it consists of 70,000 handwritten raster images from 250 different sources out of which 60,000 are used for training, and the rest are used for training validation. MNIST data is represented in the IDX file format and are look like in figure 1. Our proposed method mainly separated into stages, as shown in Figure 2, pre-processing, Data Encoding, Model Construction, Training & Validation, Model Evaluation & Prediction. Since the loading dataset is necessary for any process, all the steps come after it.

Figure 1: Sample MNIST data

3.1 Pre-Processing

After loading the data, we separated the data into X and y where X is the image, and y is the label corresponding to X. As shown in figure 3, the first layer/input layer for our model is convolution. Convolution takes each pixel as a neuron, so we need to reshape the images such that each pixel value is in its own space, thus converting a 28x28 matrix of greyscale values into 28x28x1 tensor. With the right dimensions for all the images, we can split the images into train and test for further steps [19][20][21].

Figure 2: Flowchart

3.2 Data Encoding

This is an optional step since we are using the cross-categorical entropy as loss function; we have to specify the network that the given labels are categorical in nature.

3.3 Model Construction

After data encoding, the images and labels are ready to be fitted into our model [22] [23]. Summary of the model can be seen in Figure 4

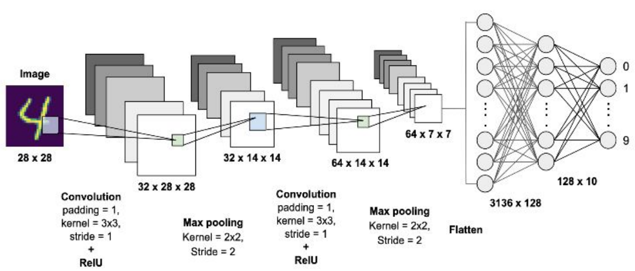

Figure 3: Proposed model

Our model is composed of feature extraction with convolution and binary classification. Convolution and max-pooling are carried out to extract the features in the image, and a 32 3x3 convolution filters are applied to a 28x28 image followed by a max-pooling layer of 2x2 pooling size followed by another convolution layer with 64 3x3 filters. In the end, we obtain 7x7 images to flatten. Flatten layer will flatten the 7x7 images into a series of 128 values that will be mapped to a dense layer of 128 neurons that are connected to the categorical output layer of 10 neurons.

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_1 (Conv2D)            (None, 26, 26, 32)        320
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 13, 13, 32)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 11, 11, 64)        18496
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 5, 5, 64)          0
_________________________________________________________________
flatten_1 (Flatten)          (None, 1600)              0
_________________________________________________________________
dense_2 (Dense)              (None, 128)               204928
_________________________________________________________________
dense_3 (Dense)              (None, 10)                1290
=================================================================
Total params: 225,034
Trainable params: 225,034
Non-trainable params: 0
_________________________________________________________________
```

Figure 4: Model Summary

3.4 Training & Validation

After building the model [24], we compiled a model with adam optimizer and particular cross-entropy loss function, which are standard in making a convolution neural network. Once the model is successfully assembled, then we can train the model with training data for 100 iterations, but as the number of iteration increases, there is a chance for overfitting. Therefore we limit the training up to 98% accuracy, as we are using real-world data for prediction, test data was used to validate the model [16].

3.5 Model Evaluation & Prediction

For real-world image classification prediction, we need to do a little image pre-processing on the real-world images as model training was done with greyscale raster images. The steps of image pre-processing are,

1. Loading image
2. Convert the image to greyscale
3. Resize the image to 28x28
4. Converting the image into a matrix form
5. Reshape the matrix into 28x28x1

After pre-processing, we predict the label of the image by passing the pre-processed image through the neural network. The output we get is a list of 10 activation values 0 to 9, respectively. The position having the highest value is the predicted label for the image [18].

4. RESULTS AND DISCUSSION

Our model is built to work on real-world data, and real-world images are not even close to MNIST raster images, a lot of pre-processing was done to make a real image to look like a raster image.

4.1 Accuracy score

Our model stopped training at the 2nd epoch as it reached 98.21% training accuracy and 98.51% validation accuracy with 5% training loss and 4% validation loss. The progression of accuracy and loss are represented in Figure 5.

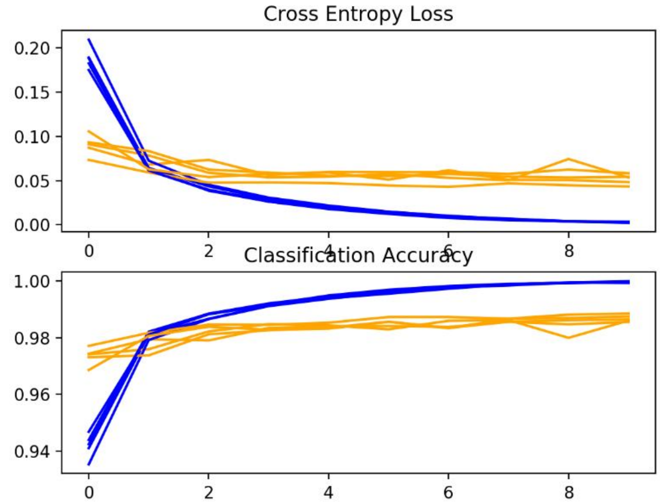

Figure 5: Loss and Accuracy Learning Curves

From the above curve, we can observe that the loss and accuracy are cooperatively changed at every fold during k-fold cross-validation. Before two folds, efficiency almost reached 98%, and that's why the number of iterations stopped at the 2nd epoch. Stability inaccuracy score can be observed from 2nd iteration.

4.2 Prediction

Our model is able to recognize computer-generated digits as well as handwritten digits. Computer-generated digit prediction is more accurate compared to real-world digit prediction, which can be observed in Figure 6.

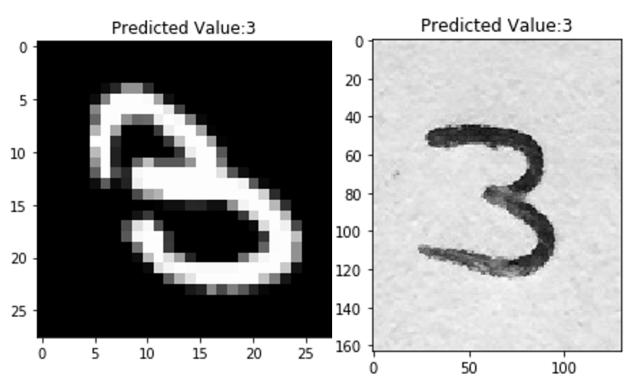

Figure 6: Raster vs. Real image prediction

5. CONCLUSION

The performance of CNN for handwritten recognition performed significantly. The proposed method obtained 98% accuracy and is able to identify real-world images as well; the loss percentage in both training and evaluation is less than 0.1, which is negligible. The only challenging part is the noise present in the real-world image, which needs to look after. The learning rate of the model is much dependent on the number of dense neurons and the cross-validation measure

REFERENCES

1. Y. Yin, W. Zhang, S. Hong, J. Yang, J. Xiong, and G. Gui, "**Deep Learning-Aided OCR Techniques for Chinese Uppercase Characters in the Application of Internet of Things**," in IEEE Access, vol. 7, pp. 47043-47049, 2019. https://doi.org/10.1109/ACCESS.2019.2909401
2. N. B. Venkateswarlu, "**New raster, adaptive document binarization technique,**" in Electronics Letters, vol. 30, no. 25, pp. 2114-2115, 8 Dec. 1994. https://doi.org/10.1049/el:19941439
3. S. Aly and A. Mohamed, "**Unknown-Length Handwritten Numeral String Recognition Using Cascade of PCA-SVMNet Classifiers**," in IEEE Access, vol. 7, pp. 52024-52034, 2019.
4. Seong-Whan Lee and Sang-Yup Kim, "**Integrated segmentation and recognition of handwritten numerals with cascade neural network**," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 29, no. 2, pp. 285-290, May 1999. https://doi.org/10.1109/5326.760572
5. Bailing Zhang, Minyue Fu, Hong Yan, and M. A. Jabri, "**Handwritten digit recognition by adaptive-subspace self-organizing map (ASSOM)**," in IEEE Transactions on Neural Networks, vol. 10, no. 4, pp. 939-945, July 1999.
6. L. Deng, "**The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web],**" in IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 141-142, Nov. 2012. https://doi.org/10.1109/MSP.2012.2211477
7. S.Sudhakar, V.Vijayakumar, C.SathiyaKumar, V.Priya, LogeshRavi, V.Subramaniyaswamy, **Unmanned Aerial Vehicle (UAV) based Forest Fire Detection and monitoring for reducing false alarms in forest-fires**, Elsevier- Computer Communications 149 (2020) 1–16, https://doi.org/10.1016/j.comcom.2019.10.007
8. M. M. A. Ghosh and A. Y. Maghari, "**A Comparative Study on Handwriting Digit Recognition Using Neural Networks,**" 2017 International Conference on Promising Electronic Technologies (ICPET), Deir El-Balah, 2017, pp. 77-81.
9. A. K. Jain and D. Zongker, "**Representation and recognition of handwritten digits using deformable**

templates," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 12, pp. 1386-1390, Dec. 1997.

10. S.Sudhakar, V.Vijayakumar, C.SathiyaKumar, V.Priya, LogeshRavi, V.Subramaniyaswamy, **Unmanned Aerial Vehicle (UAV) based Forest Fire Detection and monitoring for reducing false alarms in forest-fires**, *Elsevier- Computer Communications* 149 (2020) 1–16, https://doi.org/10.1016/j.comcom.2019.10.007
11. R. Alhajj and A. Elnagar, "**Multiagents to separating handwritten connected digits,**" in IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, vol. 35, no. 5, pp. 593-602, Sept. 2005. https://doi.org/10.1109/TSMCA.2005.843389
12. H. I. Avi-Itzhak, T. A. Diep and H. Garland, "**High accuracy optical character recognition using neural networks with centroid dithering,**" in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 17, no. 2, pp. 218-224, Feb. 1995. https://doi.org/10.1109/34.368165
13. Naimin Zhang, Wei Wu, and Gaofeng Zheng, "**Convergence of gradient method with momentum for two-Layer feedforward neural networks,**" in IEEE Transactions on Neural Networks, vol. 17, no. 2, pp. 522-525, March 2006.
14. R.Vasanthi, R.Jayavadivel, K.Prasadh, J.Vellingiri, G.A kilarasu, S.Sudhakar, P.M.Balasubramaniam, **A novel user interaction middleware component system for ubiquitous soft computing environment by using the fuzzy agent computing system**, Journal of Ambient Intelligence and Humanized Computing (2020), Springer, doi.org/10.1007/s12652-020-01893-4.
15. Jagadeesh Gopal, Vellingiri J, Gitanjali J, Arivuselvan K, Sudhakar S, **An Improved Trusted On-Demand Multicast Routing with QoS for Wireless Networks**, International Journal of Advanced Trends in Computer Science and Engineering, Vol.9,No.1, January – February 2020, P.P.:261-265, https://doi.org/10.30534/ijatcse/2020/39912020.
16. Satheesh N, Sudha D, Suganthi D, Sudhakar S, Dhanaraj S, Sriram V.P., Priya V, "**Certain improvements to Location aided packet marking and DDoS attacks in internet**," Journal of Engineering Science and Technology Vol. 15, No. 1 (2020), pp: 94 - 107, School of Engineering, Taylor's University.
17. Sathiya Kumar C, Priya V, Sriram V P, Sankar Ganesh K, Murugan G, Devi Mani, Sudhakar S, "**An Efficient Algorithm for Quantum Key distribution with Secure Communication**, Journal of Engineering Science and Technology Vol. 15, No. 1 (2020), pp:77-93, School of Engineering, Taylor's University.
18. S.Sudhakar, N.Satheesh, S.Balu, Amireddy Srinish Reddy, G.Murugan,2019,"**Optimizing Joins in a Map-Reduce for Data Storage and Retrieval Performance Analysis of Query Processing in HDFS for Big Data,"** International Journal of Advanced Trends in Computer Science and Engineering, (IJATCSE), Vol.8 No.5,pp:2062-2067, DOI:10.30534/ijatcse/2019/33852019.
19. K.B., Prakash. **"Information extraction in current Indian web documents."** International Journal of and Technology(UAE), 2018: 68-71. https://doi.org/10.14419/ijet.v7i2.8.10332
20. Kolla B.P., Dorairangaswamy M.A., Rajaraman A. **"A neuron model for documents containing multilingual Indian texts."** 2010 International Conference on Computer and Communication Technology, ICCCT-2010,2010: 451-454. https://doi.org/10.1109/ICCCT.2010.5640489
21. Kolla B.P., Raman A.R. **"Data Engineered Content Extraction Studies for Indian Web Pages."** Advances in Intelligent Systems and Computing, 2019: 505-512. https://doi.org/10.1007/978-981-10-8055-5_45
22. Prakash K.B., Dorai Rangaswamy M.A. **"Content extraction studies using neural network and attributegeneration."** Indian Journal of Science and Technology, 2016: 1-10.
23. Prakash K.B., Dorai Rangaswamy M.A., Raman A.R. **"Text studies towards multi-lingual content mining for web communication."** Proceedings of the 2nd International Conference on Trendz in Information Sciences and Computing, TISC-2010, 2010: 28-31. https://doi.org/10.1109/TISC.2010.5714601
24. Prakash K.B., Rangaswamy M.A.D. **"Content extraction of biological datasets using soft computing techniques.",** Journal of Medical Imaging and Health Informatics, 2016: 932-936. https://doi.org/10.1166/jmihi.2016.1931