Handwritten Digit Recognition

Pooja Lakhane[1], Milind Rane[2]

[1]B Tech Electronics Engineering, Vishwakarma Institute of Technology Pune, India

[2]Electronics Engineering, Vishwakarma Institute of Technology Pune, India

Abstract: *In recent times, with the proliferation of the Artificial Neural Network (ANN), deep learning has brought greater prominence to the field of machine learning by making it more intelligent. Deep learning is surprisingly used in a large range of fields due to a variety of applications such as surveillance, health, medicine, sports, robots, drones, etc. In deep learning, the Convolutional Neural Network (CNN) is at the center of amazing progress that integrates the Artificial Neural Network (ANN) with deep learning strategies to date. They are widely used in pattern recognition, sentence segmentation, speech recognition, facial recognition, text input, text analysis, incident, and handwritten digit recognition. The goal of this paper is to look at the difference in accuracy of CNN to separate handwritten numbers using different numbers of hidden layers and epochs and to make comparisons between accuracy. For this CNN performance evaluation, we performed our experiments using the Modified National Institute of Standards and Technology (MNIST) dataset. In addition, the network is trained using the stochastic gradient origin and the backpropagation algorithm*

Keywords: Handwritten digit recognition, Convolutional Neural Network (CNN), Deep learning, MNIST dataset, Epochs, Hidden Layers, Stochastic Gradient Descent, Backpropagation

1. Introduction

Acceptance identifies or distinguishes an object or person from the past or from learning. Similarly, Digit Recognition is nothing more than seeing or pointing to digits in any document. Digit recognition framework is simply a function of the machine to prepare or interpret digits. Handwritten Digit Rechipition is the computer's ability to translate handwritten numbers from various sources such as messages, bank checks, papers, photos, and so on in various handwriting situations used on PC tablets, identifying mobile number plates, managing bank checks, digits included in any of the forms etc. Machine learning offers a variety of ways in which human effort can be reduced to recognizing handwritten numbers. Deep learning is a machine learning technique that trains computers to make things easier for humans: learning by example. By using deep learning methods, human effort can be reduced to seeing, learning, seeing and learning in many more fields. Using deep learning, the computer learns to perform image classification tasks or content from any text. Deep learning models can achieve position accuracy, in addition to human-level performance. The digital recognition model uses large databases to identify numbers from different sources. Handwriting recognition has been around since the 1980s. Handwritten digit recognition function, using a classifier, has more value and use such as - online digit recognition on PC tablets, zip codes by mail, processing bank exam numbers, numbered sections in handwritten formats (for example - tax forms) and more . There are various challenges he faces while trying to solve this problem. Handwritten numbers are not always the same size, size, or shape and position related to marks. The main purpose was to prove the pattern matching method to see the handwritten digits assigned to the Minist data set of handwritten digits (0-9).

2. Literature Survey

CNN is assuming a significant job in numerous parts like picture handling. It powerfully affects numerous fields.

Indeed, in nano-advancements like assembling semiconductors, CNN is utilized for issue recognition and characterization. Manually written digit acknowledgment has become an issue of enthusiasm among specialists. There are an enormous number of papers and articles are being distributed nowadays about this theme. In examine, it is demonstrated that Deep Learning calculation like multilayer CNN utilizing Keras with Theano and Tensorflow gives the most noteworthy exactness in examination with the most generally utilized AI calculations like SVM, KNN and RFC. On account of its most noteworthy exactness, Convolutional Neural Network (CNN) is being utilized for a huge scope in picture order, video investigation, and so forth. Numerous analysts are attempting to make feeling acknowledgment in a sentence. CNN is being utilized in characteristic language preparing and notion acknowledgment by changing various parameters. It is truly testing to get a decent presentation as more parameters are required for the huge scope neural system. Numerous specialists are attempting to expand the precision with less mistake in CNN. In another examination, they have indicated that profound nets perform better when they are prepared by basic back-proliferation. Their design brings about the most minimal blunder rate on MNIST contrast with NORB and CIFAR10. Scientists are dealing with this issue to decrease the blunder rate however much as could be expected in penmanship acknowledgment. In one research, a mistake pace of 1.19% is accomplished utilizing 3-NN prepared and tried on MNIST. Profound CNN can be customizable with the info picture clamor. Intelligence intermittent convolutional arrange (CRCN) is a multimodal neural engineering. It is being utilized in recuperating sentences in a picture. A few analysts are attempting to think of new methods to maintain a strategic distance from downsides of conventional convolutional layer's. Ncfm (No mix of highlight maps) is a strategy which can be applied for better execution utilizing MNIST datasets. Its exactness is 99.81% and it very well may be applied for huge scope information. New uses of CNN are creating step by step with numerous sorts of research. Analysts are making a decent attempt to limit mistake rates. Utilizing MNIST datasets and CIFAR, mistake rates are being watched. To

clean obscure pictures CNN is being utilized. For this reason, another model was proposed utilizing MNIST dataset. This methodology arrives at an exactness of 98% and misfortune run 0.1% to 8.5%. In Germany, a traffic sign acknowledgment model of CNN is proposed. It proposed a quicker presentation with 99.65% precision. Misfortune work was structured, which is relevant for light-weighted 1D and 2D CNN. For this situation, the correctness's were 93% and 91% individually.

3. Existing System

Nowadays, a developing number of individuals are utilizing pictures to move information. It is additionally the primary stream to isolate delicate information from pictures. Picture Recognition is a significant research territory for its usually utilized applications. By and large, the field of example acknowledgment, one of the most troublesome errands is exact PC produced acknowledgment of human penmanship. Indeed, this is a troublesome issue on the grounds that there are wide varieties of penmanship from individual to individual. In spite of the way that, these distinctions don't mess up individuals, be that as it may, it is by the by turning out to be increasingly more hard to educate PCs to interpret standard original copy. With regards to picture acknowledgment, for instance, in transcribed areas, it is essential to turn out how the subtleties are shown in the photographs.

Manual acknowledgment from MNIST information is notable among researchers for utilizing various varieties of different parameters, blunder rate has diminished, for instance, from a classar classifier (1-layer NN) with 12% to 0.23% by 35 neural choices. The proportion of this is to utilize a manually written note acknowledgment framework and consider various differentiators and various procedures with an attention on the best way to accomplish the nearest to one's exhibition. The way toward naming various digits (0-9) for various individuals the most well-known issue can be for the request for digits and the closeness between digits, for example, 1 and 7, 5 and 6, 3 and 8, 9 and 8, etc. Moreover, individuals make a similar degree in various kinds of representations, varieties and penmanship in various human penmanship comparatively adds to the turn of events and nearness of digits.

3.1 MINST Dataset

Adjusted National Institute of Standards and Technology (MNIST) is a huge arrangement of PC vision dataset which is broadly utilized for preparing and testing various frameworks. It was made from the two extraordinary datasets of National Institute of Standards and Technology (NIST) which holds paired pictures of transcribed digits. The preparation set contains written by hand digits from 250 individuals, among them half preparing dataset was representatives from the Census Bureau and its remainder was from secondary school underststudies. Be that as it may, it is frequently ascribed as the first datasets among different datasets to demonstrate the viability of the neural systems.

The database contains 60,000 pictures used for getting ready similarly as very few of them can be used for cross-endorsement purposes and 10,000 pictures used for testing. All the digits are grayscale and arranged in a fixed size where the power lies at the point of convergence of the image with 28×28 pixels. Since all the photos are 28×28 pixels, it shapes a display which can be fixed into 28*28=784 dimensional vector. Each section of the vector is a twofold worth which depicts the power of the pixel.

3.2 Convolution Neural Network

To perceive the transcribed digits, a seven-layered convolutional neural system with one information layer followed by five shrouded layers and one yield layer is planned.
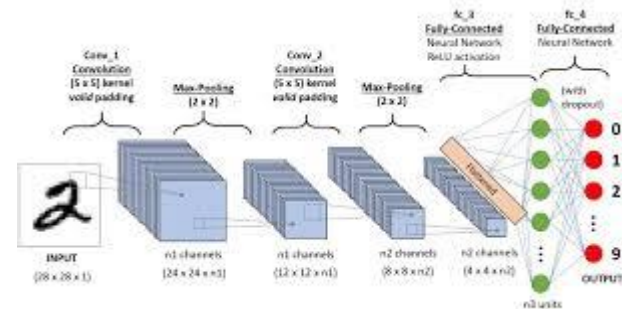
The info layer comprises of 28 by 28 pixel pictures which imply that the system contains 784 neurons as information. The info pixels are grayscale with a worth 0 for a white pixel and 1 for a dark pixel. Here, this model of CNN has five concealed layers. The main concealed layer is the convolution layer 1 which is liable for highlight extraction from an information. This layer performs convolution activity to little limited zones by convolving a channel with the past layer. Likewise, it comprises of various element maps with learnable bits and corrected direct units (ReLU). The piece size decides the region of the channels. ReLU is utilized as an enactment work toward the finish of every convolution layer just as a completely associated layer to upgrade the presentation of the model. The following shrouded layer is the pooling layer 1. It lessens the yield data from the convolution layer and diminishes the quantity of parameters and computational multifaceted nature of the model. The various sorts of pooling are max pooling, min pooling, normal pooling, and L2 pooling. Here, max pooling is utilized to subsample the component of each element map. Convolution layer 2 and pooling layer 2 which has a similar capacity as convolution layer 1 and pooling layer 1 and works similarly with the exception of

their component maps and piece size changes. A Flatten layer is utilized after the pooling layer which changes over the 2D highlighted map grid to a 1D include vector and permits the yield to get took care of by the completely associated layers. A completely associated layer is another shrouded layer otherwise called the thick layer. It is like the concealed layer of Artificial Neural Networks (ANNs) yet here it is completely associated and interfaces each neuron from the past layer to the following layer. So as to decrease overfitting, dropout regularization strategy is utilized at completely associated layer 1. It haphazardly turns off certain neurons during preparing to improve the exhibition of the system by making it progressively powerful. This makes the system become prepared to do better speculation and less convincing to overfit the preparation information. The yield layer of the system comprises of ten neurons and decides the digits numbered from 0 to 9. Since the yield layer utilizes an enactment capacity, for example, softmax, which is utilized to improve the presentation of the model, groups the yield digit from 0 through 9 which has the most noteworthy actuation esteem.

The MNIST written by hand digits database is utilized for the analysis. Out of 70,000 checked pictures of written by hand digits from the MNIST database, 60,000 filtered pictures of digits are utilized for preparing the system and 10,000 examined pictures of digits are utilized to test the system. The pictures that are utilized for preparing and testing the system all are the grayscale picture with a size of 28×28 pixels. Character x is utilized to speak to a preparation input where x is a 784-dimensional vector as the contribution of x is viewed as 28×28 pixels. The comparable wanted yield is communicated by y(x), where y is a 10-dimensional vector. The system points is to locate the advantageous loads and inclinations with the goal that the yield of the system approximates y(x) for all preparation inputs x as it totally relies upon weight esteems and predisposition esteems.

4. Results

5. Conclusion

In this paper, the varieties of correctnesses for manually written digit were watched for 15 ages by fluctuating the shrouded layers. The exactness bends were produced for the six cases for the distinctive parameter utilizing CNN MNIST digit dataset. The six cases perform contrastingly due to the different mixes of shrouded layers. The layers were taken arbitrarily in an intermittent succession so each case carries on diversely during the investigation. The most extreme and least correctnesses were watched for various concealed layers variety with a bunch size of 100. Among all the perception, the greatest exactness in the presentation was discovered 99.21% for 15 ages in the event that 2 (Conv1, pool1, Conv2, pool2 with 2 dropouts). In digit acknowledgment, this kind of higher precision will coordinate to accelerate the exhibition of the machine all the more satisfactorily. In any case, the base exactness among all perception in the exhibition was discovered 97.07% in the event that 6 (Conv1, pool1, Conv2, pool2 with 1 dropout). In addition, among all the cases, the all out most elevated test misfortune is around 0.049449 found on the off chance that 3 without dropout and the all out least test misfortune is roughly 0.026303 found in the event that 2 with dropout. This low misfortune will give CNN better execution to accomplish better picture goals and commotion handling. Later on, we intend to watch the variety in the general characterization precision by shifting the quantity of shrouded layers and group size.

References

[1] Y. LeCun et al., "Backpropagation applied to written by hand postal district acknowledgment," Neural calculation, vol. 1, no. 4, pp. 541-551, 1989.
[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet arrangement with profound convolutional neural systems," in Advances in neural data handling

frameworks, 2012, pp. 1097-1105.

[3] D. Hubel and T. Wiesel, "Variant visual projections in the Siamese feline," The Journal of physiology, vol. 218, no. 1, pp. 33-62, 1971.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Profound learning," nature, vol. 521, no. 7553, p. 436, 2015.

[5] D. Cireşan, U. Meier, and J. Schmidhuber, "Multi-section profound neural systems for picture arrangement," arXiv preprint arXiv:1202.2745, 2012.

[6] K. Fukushima and S. Miyake, "Neocognitron: A self-sorting out neural system model for a component of visual example acknowledgment," in Competition and collaboration in neural nets: Springer, 1982, pp. 267-285.

[7] Y. LeCun et al., "Written by hand digit acknowledgment with a back-engendering system," in Advances in neural data handling frameworks, 1990, pp. 396-404.

[8] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Slope based learning applied to record acknowledgment," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, 1998.