Handwritten character recognition using CNN -review

Abstract

Using a multilayer feed-forward neural network, an attempt is made to recognise handwritten characters for English alphabets. The neural network is trained using the EMNIST dataset, which contains the English alphabet and numbers. The EMNIST balanced dataset has 131,600-character pictures divided into 47 classes. The pixel values are normalised to obtain the feature extraction technique. The intensity of each pixel in the image is represented by pixel values that vary from 0 to 255, and they are normalised to represent values between 0 and 1. The EMNIST dataset is trained using a convolutional neural network as a classifier. The work is expanded by adding new datasets to the EMNIST dataset of Tamil language characters and training the model. The trained classifier provides a forecast for the input image.

1. Introduction

Because of its application in a variety of sectors, handwriting recognition has sparked a lot of interest in the field of pattern recognition and machine learning. A handwriting recognition system is a computer program that recognizes letters and other symbols in natural handwriting. The two types of handwriting recognition are offline and online handwriting recognition. Online handwriting data is inherently dynamic. As a result, online handwriting recognition is known to be influenced by pen trajectory and expressed as a function of time. A scanner frequently records the writing optically and makes it available as an image for offline recognition.

Handwritten character recognition is the gadget in our project that converts the user's handwritten characters or words into a computer-readable format.

HCR performance has increased in recent years, but the wide variety of handwriting styles, the presence of many similar characters, and the large number of character categories make it a difficult task. Convolutional neural networks (CNNs) are a well-known deep learning architecture inspired by the human brain's natural visual perception mechanism. Taking advantage of the recent exponential growth in the volume of annotated data and the rapid increases in the capabilities of graphics processing units, the study of CNN (Convolutional Neural Network) arose quickly and achieved state-of-the-art performance on a variety of tasks, including image classification, text detection, pose estimation, object tracking, action detection, visual saliency detection, scene marking, speech and natural language process.

The core features of CNN architectures are similar, even though there are many variations. Convolutional, pooling, and fully connected layers were among the three types of layers used. It has made significant contributions to computer vision, and it has been implemented in the HCR to obtain good recognition results. As a result, multiple CNN-based models are being used by the researchers to solve HCR difficulties.

Many researchers have worked in the HCR field, however perfect precision is impossible to accomplish. In [1], Kato et al. presented a handwritten approach for recognizing Chinese and Japanese characters. The method utilized to recognize the characters in dataset etl9b, in which each character's directional element attributes were retrieved and the asymmetric mahala Nobis distance was used for fine classification. The document claims a 99.42 percent accuracy rate. Chaudhuri and Adak [6] devised a method for detecting strike-out text in offline handwritten document photographs and removing it. They describe a hybrid SVM classification and shortest path technique for recognizing such texts, with a 94 percent overall output accuracy.

They remove the least important feature maps and their associated filters from the original network topology for channel pruning. Their test was reported to be 95.33 percent accurate. For detecting isolated Bangla HCR, Ferdous and afroge [8] and das et al. [9] used CNN. The accuracy of the two sets of researchers was 95.5 percent and 98.3 percent, respectively. Our goal in this research is to use CNN to create an efficient HCR system. The EMNIST database was used to evaluate the HCR system. We used the dataset EMNIST to test the CNN algorithm's performance and discovered the accuracy of handwritten characters. EMNIST provides photographs of handwritten characters. The photos are divided into two groups: training and testing

This paper has five sections. Section 1 describes about introduction, while section 2 presents the background of the study. Section 3 describes the architecture of CNN, section 4 shows the method of the paper. Finally, the conclusion is described in section 5.

2. Background study

Character recognition is a fundamental yet difficult aspect of pattern recognition, and it has a wide range of applications. It has been a hot topic of research since the dawn of computer science since it is a natural way for computers and humans to connect. Character recognition, more properly, is the process of detecting and recognizing characters in an input image and converting them to ASCII or another machine-editable format.

The visual cortex configuration boosts CNN architecture, which resembles a human brain communication pattern of neurons. By observation and analysis, nerve cells only react to stimuli from visible spectrum. A number of these spectrum fields overlap to fill the entire display area. CNN can capture the spatial relationships inside a picture and relate them to the image's content for recognition. To locate the distinctive elements to be categorized, the CNN design provides a robust match of the original image. During the training step, the weights, parameters, and biases associated with the transformations from the original picture to the feature vector are discovered to better understand the nature of the image.

3. **The architecture of CNN**:

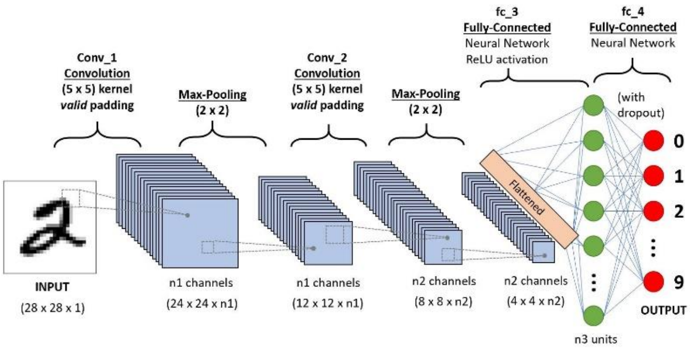

Figure 1:*Architechture of CNN[1]*

Figure 1 shows the process of CNN that consists of the fundamental parts – an input layer, convolution, pooling, fully connected neural network - which is described as follows:

Input layer:

The input layer is the first layer of the first image that contributes to the operation at front of the system architecture. Input for this layer is the image of the character that we must process. The image can be of RGB or grayscale. The dimensions of the image will be in WxHxD where W is for width, H is for height, and D is for depth. Depth for grayscale images is 1 pixel while for RGB depth is 3 pixels.

Convolution:

As seen in fig. 1 the conv1 area consists of channels of learning capacities and those channels are named as components of the layer. Each channel is known as a filter, which is a square matrix that has spatial width and length in pixels with a depth

A convolutional layer applies sliding filters vertically and horizontally to the input layer. This layer learns the features of all the regions of input image while scanning. It computes a scalar product of values of the filter with the values of image regions and adds a bias for each region. A rectified linear unit applies element wise activation function, viz., max (0, x), tanh, sigmoid: 1/ (1 + e−x) to the output of this layer for thresholding.

Pooling:

In the CNN architecture, the pooling layer is situated in between the convolutional and fully connected layers. The pooling layer is used after one or more convolutional layers to shrink the volume of the data to make the network computationally faster. It restricts the networks from overfitting as well as supplies the network with translation invariance. To keep the critical data local pooling called sub testing is done. MAX pooling and average pooling are used to implement pooling. It applies sliding filters vertically and horizontally through the input image to get the max value or average value for each region of the input data. MAX pooling is the most used pooling. It has a pooling layer with channels of size 2x2 utilized with a stage of 2 down-inspecting local pooling.

Fully connected neural network:

This is the last layer of the CNN architecture. The 2-dimensional image matrix that was received from the pooling layer will be converted to a 1-dimensional vector form and then applied to the network system.

4. Methods

HCR is one of the most significant topics in machine learning and computer vision. The method that has been used to recognize characters from an image to a document is mentioned in this section.

The figure below shows us the four stages used during the classification and detection i.e.: Pre-Processing, Feature Extraction, Classification and Recognition.

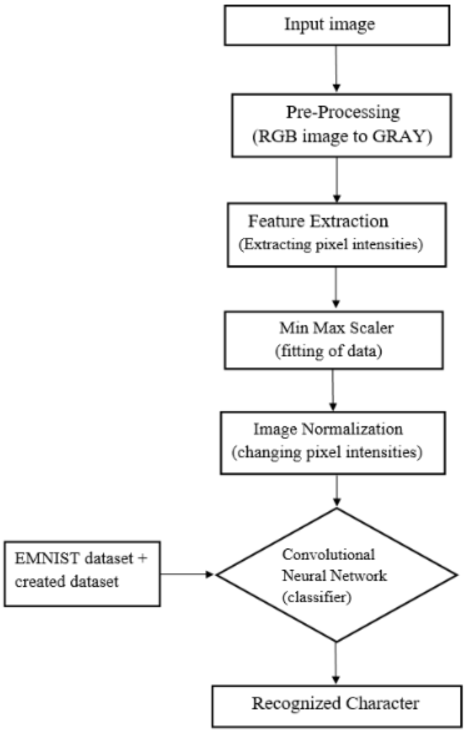

Fig. 2: proposed system architecture

1. Pre-processing
   1.1. In this step, the input image is cleaned by performing cleaning tasks. Image is also enhanced by removing noise from it. Image is also converted to grayscale or binary format in this stage.
2. Segmentation
   2.1. After images are pre-processed, they are segmented in this step using segmentation techniques. Characters are then stored in a sequence of images. Borders are also eliminated if it is present. After which images are scaled to a specific size.
3. Feature extraction
   3.1. Feature extraction is made on the segmented characters. In our case, the features are extracted using CNN with the ReLU activation function as shown in figure 1. CNN works on each character image to form a matrix of reduced size using convolution and pooling. Finally, the reduced matrix is compacted to a vector form using the ReLU function. This vector is regarded as the feature vector
4. Classification and recognition

4.1. The feature vector received from the last step is used as individual input to formulate the corresponding class. During the training phase, the parameter, biases, and weights are calculated which are used in the testing phase of classification and recognition purpose

5. Conclusion

5.1. The paper discusses in detail about handwritten character recognition. The most correct solution provided in this area depends upon the quality as well as the nature of the material to be read. From the study done so far, it is analysed that the selection of the classification as well as the feature extraction techniques needs to be proper to reach a fast rate in recognizing the character. Studies in the paper reveal that there is still scope for enhancing the algorithms as well as enhancing the rate of recognition of characters

6. References

[1], I. Khandokar, M. Hasan, F. Ernawan, S. Islam, και M. N. Kabir, 'Handwritten character recognition using convolutional neural network', Journal of Physics: Conference Series, τ. 1918, τχ. 4, σ. 042152, Ιουνίου 2021.

[2] Kai Ding, Zhibin Liu, Lianwen Jin, Xinghua Zhu, "A Comparative study of GABOR feature and gradient feature for handwritten chinese character recognition", International Conference on Wavelet Analysis and Pattern Recognition, pp. 1182-1186, Beijing, China, 2-4 Nov. 2007.

[3] Pranob K Charles, V.Harish, M.Swathi, CH. Deepthi, "A Review on the Various Techniques used for Optical Character Recognition", International Journal of Engineering Research and Applications, Vol. 2, Issue 1, pp. 659-662, Jan-Feb 2012.

[4] Bhatia Neetu, "Optical Character Recognition Techniques", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 5, May 2014.

[5] Liana M. Lorigo and Venu Govindaraju, "Offline Arabic Handwriting Recognition: A Survey", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 28 Issue 5, May 2006.

[6] K. Gaurav and Bhatia P. K., "Analytical Review of Preprocessing Techniques for Offline Handwritten Character Recognition", 2nd International Conference on Emerging Trends in Engineering & Management, ICETEM, 2013.

[7] B. Koyuncu and H. Koyuncu , "Handwritten Character Recognition by using Convolutional Deep Neural Network; Review", International Journal of Engineering Technologies IJET, vol. 5, no. 1, pp. 1-5, Mar. 2019, doi:10.19072/ijet.528775