TEXT ENCODER USING NEURAL NETWORK

Dr. P. SHANMUGA PRIYA[1], GORUGANTHU NARASIMHA SURYA MUKUND[2],

KOLLI SRI VATSA KARTHIKEYA[3]

[1]*Associate professor, Department of Computer Science and Engineering, SCSVMV, Kanchipuram*
[2]*B.E graduate (IV year), Department of Computer science and engineering, SCSVMV, Kanchipuram*
[3] *B.E graduate (IV year), Department of Computer science and engineering, SCSVMV, Kanchipuram*

ABSTRACT

One of the active and difficult study areas in the world of pattern recognition and image processing has been handwriting recognition.It is an effort to use neural networks to recognise handwritten English alphabets.The neural network is trained using several character data sets. Classification and recognition are performed using the trained network.Character recognition (CR) has undergone substantial research over the past 50 years and has advanced to the point where it can develop technology-driven applications.Designing an expert system for "HCR(English) using Neural Network" is the primary goal of this research. utilizing an Artificial Neural Network technique, is capable of accurately identifying a certain character of type format.Given a specific initial state (data input), neural computers are taught (not programmed) to either classify the input data into one of the many classes or to cause the original data to grow in such a way that a specific desirable attribute is optimized.

Keywords: Character recognition, Neural Network, Trained dataset, Image processing.

1.INTRODUCTION:

The handwriting character identification is one of the most challenging areas of recognition. It is very beneficial in a wide range of real-world situations, such as document analysis, understanding postal addresses, processing bank checks, verifying signatures on papers, and many other things.

Numerous pattern recognition techniques, including statistical methods, structural and syntactic methods, and neural networks, have been used to identify handwriting both online and offline. While some reading programmes focus on recognising strokes, others try to identify individual letters, groups of letters, or full words. In the difficulty of handwritten character recognition, neural networks are frequently used to predict the handwritten characters effectively.

Objective:

The major goal is to identify handwritten characters. The field of handwriting recognition has seen a lot of effort, and there have been several reviews. This method calls for applying deep learning to recognise handwriting.There are numerous approaches to recognise manually entered characters using machine learning that require less human work.

Scope of the Project:

Everyone will have their own writing style, each with their particular writing style. Each author will create the characters in their own unique style. Some are simple to understand, while others are difficult to understand.

Our main goal is to comprehend those characters. We have developed a trained model to recognise that stokes, and we utilise that model in our code to test that character. The output is shown by the user interface by way of a message box.

Existing System and it's drawbacks:

The quality of the input has a significant impact on recognition accuracy in the majority of the current systems. Characters in handwriting frequently touch or overlap. The majority of the segmentation algorithms in use evaluate human writing empirically in order to determine rules. However, there is no assurance that these heuristic criteria will produce the best results in all writing styles. Additionally, handwriting differs between individuals and even within the same individual depending on mood, pace, etc. To extract data, you must use statistical classifiers and artificial neural networks.

Drawbacks :

- High Complexity
- Difficult to analysis
- Time Consumption is more.

II. LITERATURE REVIEW

In his work, Anuj Dutt showed how he could get a very high level of accuracy by using Deep Learning algorithms. He achieved a 98.72% accuracy by using a convolutional neural network using Keras and Theano as the backend. Additionally, CNN execution using TensorFlow produces a strikingly superior outcome of 99.70%.Even if the process and coding seem more complicated when compared to standard machine learning methods, the precision he achieved is becoming more and more visible.

Multilayer Perceptron (MLP) Neural Network was used to implement recognition in a paper by Saeed AL-Mansoori.On a dataset obtained from MNIST, the suggested neural network was trained and put to the test.

In his work, Gaurav Surve expalined the Advantage: When compared to many of the conventional vertical and horizontal approaches, the feature extraction method, such as diagonal and direction techniques, is significantly better at producing high-accuracy outcomes. The additional benefit of utilising a neural network with the best tested layers is that it has a higher tolerance for noise, providing reliable results.The accuracy of the output increases with the size of our training data collection and the quality of our neural network design.

J.Pradeep-Classification and recognition tasks are carried out using an artificial neural network as the backend. Neural networks have become the quick and dependable instruments for categorization in the off-line identification system to achieve high recognition accuracy .Since the 1990s, classification methods have been used to recognise handwritten characters. These techniques include artificial neural networks (ANNs), support vector machines (SVM), multiple classifier combinations, and statistical techniques based on the Bayes decision rule.

Preeta S- Using CNN is abbreviated form of Convolution Neural Network. Convolution refers to twisted or coiled. Any neural network is similar to human brain. Neural networks are designed by taking inspiration from brain. CNN is mainly used for Image classification. CNN consists of many layers depending on the requirements. Ahmed Mahdi Obaid and his colleagues proposed an effective handwritten text recognition system using two different learning algorithms. Considering similar configuration, Scaled Conjugate Gradient algorithm proved to perform well in terms of accuracy and training time when compared Resilient Back-propagation algorithm and his colleagues discussed three main layers of CNN are 1. Convolution layer 2. Pooling layer 3. Fully connected layer.

III. PROBLEM STATEMENT:

Massive individual variation and uncertainty in strokes in an individual's handwriting style also changes and fluctuates over time. a source image or document that has degraded over time in quality.While humans do not have to consistently write the same letter on white paper, characters in printed texts sit in a straight line. Character separation and identification are difficult with cursive writing.The system must be able to determine the type of character from a handwritten character.In other words, if we can write the character "A," the system will assume that it is either the real "A" or that the input character is somewhat similar to it. The goal of this project is to train the neural network successfully using the Resnet50 method to recognise the character by taking the handwritten characters as input and processing them.

Proposed System :

The performance of automatic pattern recognition depends critically on the selection of a representative training set and an appropriate set of features. These two elements work hand in hand and in a way that promotes positive recognition outcomes. to get a good training set for the recognition of a certain character in handwriting. to give users a simple-to-use interface so they can recognise the characters In the face of noise, the system should be able to accurately predict the handwritten letters. This character classification is incredibly simple, accurate, and user-friendly.

Module Description:

Primary module(Training of datasets):

In this module , we will train the model using the visual studio code and store the model.h5 file for the further use to identify the randomly choosen character.

Secondary module(Choosing an character randomly):

In this module , from the set of the given data of character we will choose one character in random to identify it.

Third module(GUI):

In this module, it will detect the character that is randomly given by us.

Then the result is displayed using the user interface.

IV. ARCHITECTURE:

The main objective is to detect the handwritten character(Alphabet).

- First we will train the images
- They will undergo the image acquisition , preprocess , Feature Extraction and then the model get trained.
- By using that model in the code we will identify the hand written character uploaded by using the User interface .

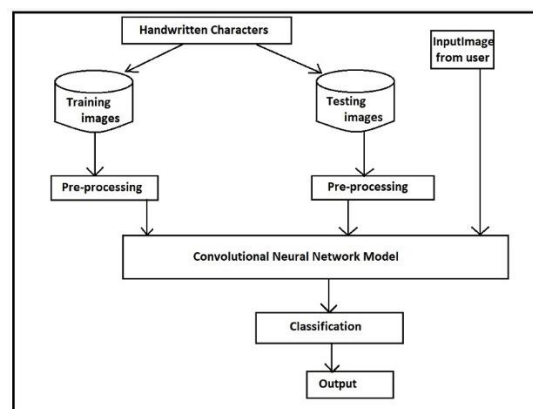DEEP LEARNING- CNN – DATASETS –VISUALSTUDIOCODE - PYTHON - CHARACTER IDENTIFICATION

Figure 01: System architecture

V. PROCESS:

STEP 1:- Import the libraries and load the dataset.

STEP 2:- First, we are going to import all the modules that we are going to need for training

our model.

STEP 3:- Preprocess the data.

STEP 4:- Create the model.

STEP 5:- Train the model.

STEP 6:- Evaluate the model.

STEP 7:- Use the model in the code.

STEP 8:- Test the Character.

VI. METHODOLOGY:

A convolutional neural network (CNN) is a type of artificial neural network used primarily for image recognition and processing; A CNN is a powerful tool but requires millions of labelled data points for training. CNNs must be trained GPU .

The input to a CNN is a tensor with the following dimensions: (number of inputs)* (input height)* (input width) *(input channels). The image is abstracted to a feature map, also known as an activation map, with the following dimensions: (number of inputs)* (feature map height)* (feature map width)* (feature map channels).

ResNet-50 is a convolutional neural network with 50 layers .The ImageNet database contains a pretrained version of the network that has been trained on more than a million images . ResNet-50 is a 50-layer convolutional neural network (48 convolutional layers, one MaxPool layer, and one average pool layer).The pretrained network can categorize photos into 1000 different object categories. Image resolution accepted by the network is 224 by 224.

Functions()	Description
latten()	For resizing the image accordingly.
Plt .show()	To show the image.
Plt .save()	To save the image.
Prdicttt()	To predict the image.
Upload()	To upload the image.
Image.open()	To open the particular image.
load_model()	To load the model.
message.showinfo()	To display the final output.

Table 01: methods involved in the project

VII. IMPLEMENTATION:

This describes the working of the system. We will load all the libraries.First we will train the model. Then we will save the model. Then we will test .We will give any character in random. Then it will display the result using GUI.

Advantages-Less Complicated,Easy to process,Accuracy is more.

Required specifications:
- Visual studio code
- User interface
- Python
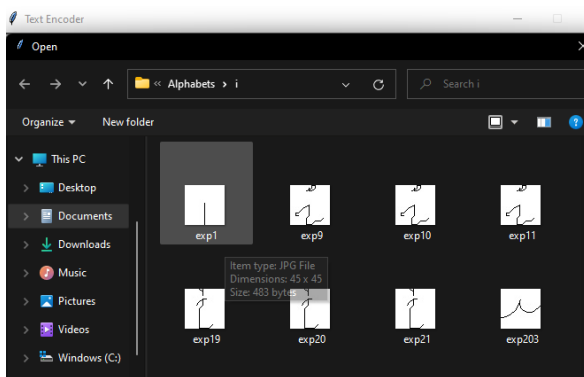- Keras
- TensorFlow
- ResNet50
- NumPy
- Glob

Results:

Input-1:
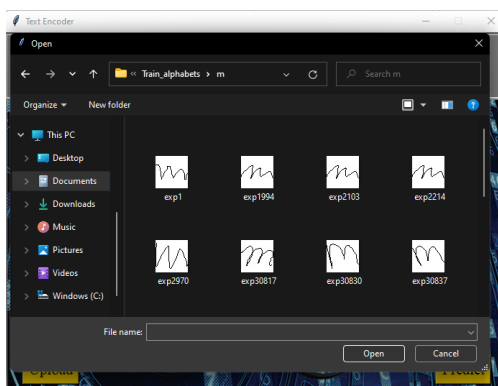

Figure 02(a): Input of the Character

Output-1:

Figure 02(b): Output of the Character

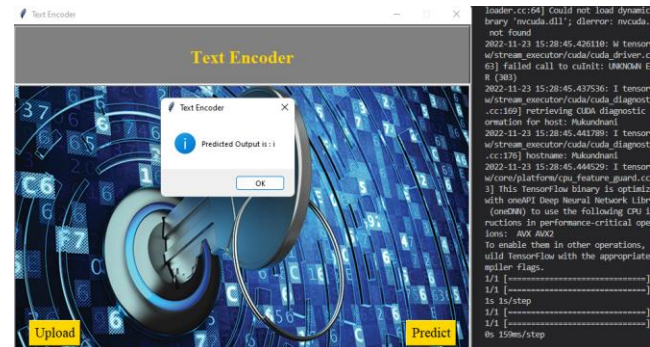**Input-2:**

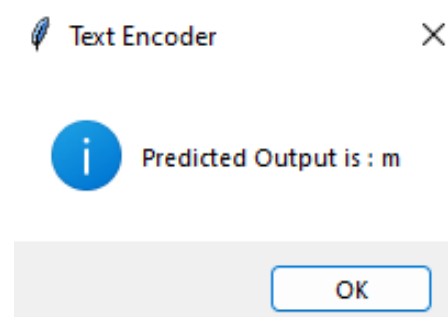**Figure 02(c): Input of the Character**

Output-2:

Figure 02(d): Output of the Character

IX. CONCLUSION:

Picture recognition is a crucial stage in the processing of images. Image feature extraction is subject to a number of limitations, including differences in lighting and the angle at which the image was shot. Since everyone has a different handwriting style, it is more challenging to identify images in handwriting. As a result, handwriting will be more difficult to recognise than computer-generated writings, which already have a distinct standard shape.

The Convolutional Neural Network Method has the highest accuracy and productivity.

We obtained the training model's accuracy, which let us forecast the handwritten character very easy. In the presence of noise, neural networks offer an effective method for character recognition. This work develops and tests a neural network for the recognition of handwritten characters in noisy environments.

In this study, handwritten character identification is accomplished using CNN, a deep learning approach. The primary goal of this research is to determine whether CNN is capable of accurately identifying the characters in a dataset. Utilizing deep learning techniques, handwritten digit recognition has been put into practise.

In order to compare the classifiers, the most popular machine learning algorithms, KNN, SVM, RFC, and CNN, were trained and tested on the same dataset. It is possible to get a high level of accuracy by using these deep learning approaches. By increasing the accuracy of classification models by more than 98%, this approach to study focuses on which classifier performs better than others.

Using Keras as backend and TensorFlow as the software, a CNN model is able to give accuracy. This has satisfied the objective that I had stated above.

X. FUTURE SCOPE:

We have developed a trained model to recognise that stokes, and we utilise that model in our code to test that character. The output is shown by the user interface by way of a message box.The accuracy is more and it is very usefull to identify the characters very easily and in short span of time.

XI. REFERENCES:

1) Handwriting Recognition using Artificial Intelligence Neural Network and Image Processing in (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 7, 2021.

2) Literature Survey for Applications of Artificial Neural Networks, pp. 669-682,2022 ,pp. 669-682,Pooja Deepakbhai Pancholi ,Sonal Jayantilal Patel.

3) U. Bhattacharya and B. B. Chaudhuri, "Handwritten numeral databases of Indian scripts and multistage recognition of mixed numerals", IEEE Transaction on Pattern analysis and machine intelligence, vol. 31, no. 3, pp. 444-457, 2021.

4) Jagan Mohan Reddy D, A Vishnuvardhan Reddy "Recognition of Handwritten Characters using Deep Convolutional Neural Network",Nov 25 2021.

5) Polaiah Bojja., Naga Sai Satya Teja Velpuri, Gautham Kumar Pandala,S D Lalitha Rao Sharma Polavarapu "Handwritten Text Recognition using Machine Learning Techniques in Application of NLP",Dec 2021.

6) Hamid, Norhidayu Abdul, and NilamNur Amir Sjarif,Handwritten recognition using SVM, KNN and neural network, arXiv preprint arXiv:1702.00723 (2022).

7) Verdha Chaudhary, Mahak Bansal and Gaurav Raj(2021). Handwritten Character Recognition using Convolutional Neural Networks (CNN).International Journal of Research in Advent Technology(IJRAT).

8) A. Dutta and A. Dutta, Handwritten digit recognition using deep learning, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), vol. 8, Dec 2021.

9) Megha Agarwal, Shalika, Vinam Tomar, Priyanka Gupta(2022).Handwritten Character Recognition using Neural Network and TensorFlow.International Journal of Innovative Technology and Exploring Engineering (IJITEE).

10) Shetal Takkar and Monica Patel. "Handwritten Character Recognition:A Survey."IJARCCE-2022

XII. AUTHOR'S PROFILE:

1. Dr. P. Shanmuga Priya,Associate Professor, Department of Computer Science and Engineering at SCSVMV, Enathur, Kanchipuram, India.

2. Mr. Goruganthu Narasimha Surya Mukund, Graduate Student,B.E, Department of Computer Science and Engineering at SCSVMV, Enathur, Kanchipuram, India.

3. Mr. Kolli Srivatsa Karthikeya, Graduate Student, B.E, Department of Computer Science and Engineering at SCSVMV, Enathur, Kanchipuram, India.