Cursive Handwriting Recognition Using CNN with VGG-16

Conference Paper · January 2023

DOI: 10.1109/ICECONF57129.2023.10083561

3 authors, including:

Rajeswari Devarajan
SRM Institute of Science and Technology
70 PUBLICATIONS **426** CITATIONS

SEE PROFILE

2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF) | 979-8-3503-3436-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICECONF57129.2023.10083561

Cursive Handwriting Recognition Using CNN with VGG-16

Anushka Anil Rangari, Swagat Das and Rajeswari D*

*Department of Data Science and Business Systems SRM Institute of Science and Technology Kattankulathur, Tamilnadu, India (*Corresponding Author)*
*E-mail : ar7937@srmist.edu.in sd8074@srmist.edu.in *drajiit@gmail.com*

Abstract- Cursive handwritten text recognition is a process of identifying handwritten text from images. The recognition process is difficult due to person's unique style of writing. Pattern recognition is used to classify the data into various categories. This study used CNN with VGG16 model to identify cursive English alphabets and words in a given scanned text document. The first phase is image acquisition, which involves acquiring the scanned picture, normalizing the image, extracting the features from the image, and applying segmentation. Three different pre-processing techniques like data augmentation, image segmentation and image data generator are implemented in this work. CNN with VGG 16 model is employed for recognition. Three kinds of experiments were done with various combinations of pre-processing techniques combined with CNN model. The experimental results indicates that data augmentation pre-processing technique with CNN produced 98.36% training accuracy and 95.1% testing accuracy which is best than other combinations.

Keywords— CNN, TensorFlow, Deep Learning, Cursive Handwriting Recognition, Data Augmentation.

I. INTRODUCTION

Handwriting recognition is a technique that locates and detects text that is contained in images of different formats. An image's text is transformed into an ASCII or Unicode format that is computer readable it changes a letter's pixel representation into the matching character form. It is an electronic recreation of a reader's interface. The study of word recognition focuses on how robots can keep an eye on their environment, pick out key words from their surroundings, and decide what to do with them [1]. TensorFlow (TF)-based Handwritten Text Recognition (HTR) system developed, trained using IAM off-line HTR dataset. The model produces the identified text after receiving pictures of single words or text lines (many words) as input. Character errors occur at a rate of about 10%, and approximately 3/4 of the words in the validation set are successfully identified. The two decoders included with TF can be replaced with the word beam search decoder. Only words found in a dictionary are allowed, however random non-word character sequences (such as numerals and punctuation signs) can still be recognised. In the example shown in figure 1, word beam search successfully decodes the sample whereas the other decoders are unable to do so.

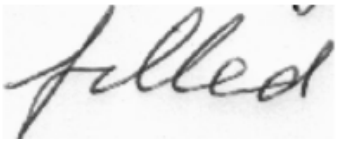

Figure 1: word beam search successfully decodes the sample

Table 1 shows the system that has been created for the thesis, the HTR system, is simplified in the model. The basic minimum is all that is needed to accurately detect text. It has two RNN (LSTM) layers, two CNN layers, and a layer for CTC loss and decoding. Refer to this medium article for further information.

Table 1: Shows the system that has been created for the thesis

World Bean Search	filled"

There are various benefits to turning a printed text into machine-readable text. The information can then be changed, transferred, stored in a compressed format, and searched for using keywords. There are many applications for it. People who are blind or have visual impairments utilize it. To digitalize paper documents, it is used in the financial and legal sectors. In the retail sector, barcode recognition technology is employed. It is frequently used in the fields of finance, education, and automatic number plate recognition [2].

A digital image is modified as part of the processing process to extract data or features (DIP). It raises the bar for machine automation, storage effectiveness, and image quality. In a handwriting recognition system, DIP is crucial. Words that are hand-written are transformed into computer symbols that comply with ASCII through the process of handwriting recognition. The steps in the handwriting recognition process are pre-processing, segmentation, feature extraction, classification, and post-processing. A scanned image in a format like JPEG, BMP, etc. must be used as the input for a recognition system. [3]

II. Literature Survey

Pritam Dhande, Reena Kharat et al. [4] after picture acquisition and segmentation, adjustments such as noise reduction, skew correction, cropping, resizing, normalisation, thinning, binarization, and skeletonization are made. These are used to improve photos for further processing and remove noise from images. On the input scanned picture, morphological processes like dilation and erosion can also be used.

Jamshed and co. [5] in an effort to record the unique characteristics of text written by humans, numerous feature extraction techniques have been tried. Each of these methods starts by extracting local or global components from a picture. Some of the fascinating properties of the Hough transformation include information preservation, robustness against disconnections, noise, and variations in stroke width, as well as the ability to preserve the whole directional information of the image.

Sunanda, Keisham, and Kanchan. Dixit and co. The alphabets are divided using an synthetic neural network, and the text lines are divided using the information energy of each pixel. 92% of the time, the recognition is accurate. The fundamental premise is that the temporal sequence of auditory impulses in speech can be likened to the left-to-right sequence of ink signals in handwriting, according to Salvador Espana-Boquera et al. [7]. Renata F. P. Neves et al. calculated the separation between the two classes to find the separation plane, or "optimal separation plane," that provides the greatest margin of safety.

T. Wakabayashi, F. Kimuratime, et al. are the authors in U.Pal. For the goal of recognising offline handwritten digits in six well-known Indian scripts, a modified quadratic classifier-based technique is suggested. The authors are Namrata Dave and others. This study explores character segmentation methods for interpreting handwritten text. Text-lines, word, and character segmentation are the three observable levels of segmentation. It is investigated why segmentation is important and what factors affect it. Made Iping Supriana, among others, and Edwin Wira Putra [11] the distance between comparable graphs is determined using the Levenshtein distance. In terms of recognition accuracy, this method performed better.

III. Proposed System

A. Dataset Description

For author identification and verification experiments, also for testing and training acknowledging the handwritten text, the IAM Handwriting Database has handwritten English text forms. In 1999, ICDAR initially made this database available. This database was used to develop an HMM-based handwriting recognition system, which was demonstrated at ICPR 2000. [12-13] unrestricted handwritten forms were 300 dpi scanned and exported images in the form of PNGs with 256 shades of grey. The database includes these documents. The following figures provide a sample of the entire form, the text lines, and the extricated words. (in Figure 2). [14]

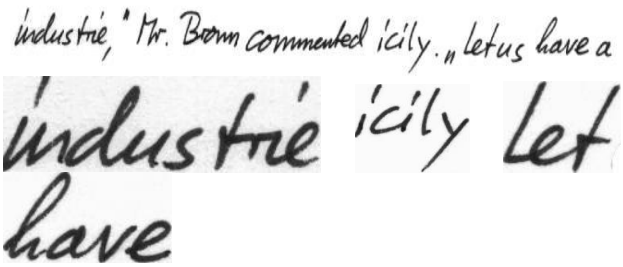

Figure 2: Example of the full form, the lines of text, and the extracted words

B. Propsed Architecture

The proposed method's process flow is shown in Figure 3. The proposed method is constructed as follows: The handwritten character that needs to be recognized is obtained via a scanner. The captured image is not appropriate for the recognition procedure. So that it may be used in later stages of the recognition process, the acquired image needs to be preprocessed. Skew correction and normalizing are carried out during preprocessing. [15]

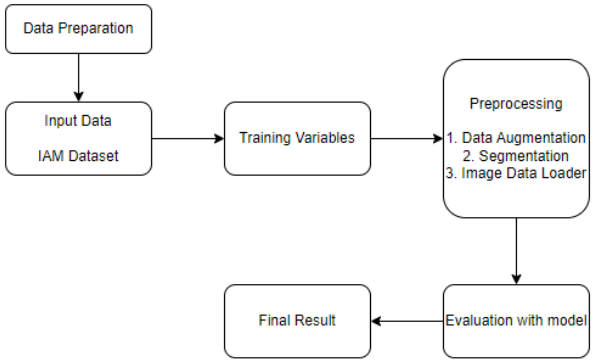

Figure 3: proposed architecture

Unprocessed raw data are converted into numerical features that can be handled while maintaining the details of the original data set through the process of extracting numerical features from unprocessed raw data. Compared to directly applying machine learning to raw data, it yields better outcomes. Feature extraction increases the precision of learnt models by extracting features from the input data.. In order to lower the dimensionality of the data, the general framework's elimination step eliminates irrelevant data. For locating picture line segments, the Hough transform method was developed. This method states that the normal form of a line is represented by:

$$p = x\cos e + y\ sine \qquad (1)$$

A sinusoid for a particular point in parameter space is represented by Equation 1. (x, y). Therefore, the routes travelling to the collinear points will come together in one place (p, e). Finding these crossing points makes it easier to recognize lines in a picture.

Segmentation is the process of separating characters within a word and is the most challenging aspect of cursive handwritten recognition. Here, the segmentation regions are determined by the vertical projection profile peaks. On a white surface, the vertical projection of a binary image resembles a series of hills. After the segmentation regions are extracted, characters are segmented [17].

A recognition system's decision-making process, known as recognition, uses the information that has already been retrieved. A feed forward back propagation neural network for classification with layers—two of which are hidden—is shown in Figure 4. The output layer is a competitive layer since one of the letters needs to be recognized. The buried layers, on the other hand, employ the log sigmoid activation function.

C. Pre Processing Techniques

Preprocessing is the process of carrying out some preliminary steps before the final rendering. Image preprocessing is a very necessary step because that helps us generate the data from the images in a more precise manner. Three preprocessing techniques are used in this study: data augmentation, image segmentation and image data generator.

In data augmentation, photos are created intentionally or images that will result in a more reliable training model are created. Images that have been enhanced can mimic various image capture methods. Randomized geometric transformations, such as rotations and translations that imitate changes in camera orientation with respect to the scene are common forms of picture improvement procedures. Cropping at random simulates changes in scene composition. Artificial noise mimics distortions brought about by processes for upstream data processing or image collection. Scaling improves the useful amount of training data and aids in making the network resistant to biases and common changes in the data. [16-17]

A common method in digital image processing and analysis is image segmentation, which divides a picture into a number of sections or areas according to the characteristics of the pixels in the image. The process of segmenting an image entails dividing the foreground from the background or grouping areas of pixels based on their similarity in color or shape. Over time, a number of picture segmentation algorithms and methods have been created to efficiently address the segmentation issue in this specific application domain employing domain-specific information. These include computer vision, autonomous driving, video surveillance, and medical imaging [18]. The Keras ImageDataGenerator is used to take an input of original data, perform random transformations on this data as well, and return an output containing only the newly transformed data. No data is added. The Keras image data generator class is also used to perform data augmentation aimed at gaining overall gains in model generalization. Operations such as rotation, translation, shearing, scaling, and horizontal flipping are performed randomly on data augmentation using an image data generator [19].

D. Convolutional Neural Networks

CNNs are Synthetic neural systems' are typically used for classification, segmentation, image processing, and other auto-correlated data.In computer vision, examples of CNN include face recognition and image classification. It is comparable to the basic neural network. Similar to neural networks, CNN has learnable parameters like weights, biases, etc. A convolution neural network may extract information from an image with the aid of several hidden layers. The Convolution layer, ReLU layer, Pooling layer, and Fully linked layer are the four most crucial CNN layers [20-21].

Convolution Layer

The process of extracting helpful image features starts with this stage. A convolution layer has a number of filters that conduct the convolution process. Imagine that a matrix of pixel values represents each image.

ReLU Layer

ReLU is the abbreviation for the rectified linear unit. After being recovered, the feature maps must be added to a ReLU layer. The process of extracting helpful image features starts with this stage. The convolution process is carried out by a number of filters within a convolution layer. Imagine that each image is a matrix of pixel values.

Pooling Layer

To make feature maps less dimensional, layers are pooled. As a result, the network needs to execute fewer computations and learn fewer parameters.

Fully Connected Layer

Convolutional neural networks (CNNs) have fully linked layers, which have shown to be particularly effective in picture recognition and classification for computer vision.

E. VGG-16

A convolutional neural network with 16 layers is referred to as VGG-16. The ImageNet database contains a pre-trained version of the network that has been trained on a million images [15]. The pre-trained network can classify photos into 1000 different item categories, including a variety of animals, a keyboard, a mouse, and a pencil. The network currently provides comprehensive feature representations for a variety of photos as a result. The network can handle images with a 224 by 224 resolution. The creators of this model demonstrated that it greatly outperformed existing setups by analyzing the network and boosting depth with a comparatively modest (33) convolutional filter architecture [22]. The VGG-16's architecture is pictured in Figure 4.

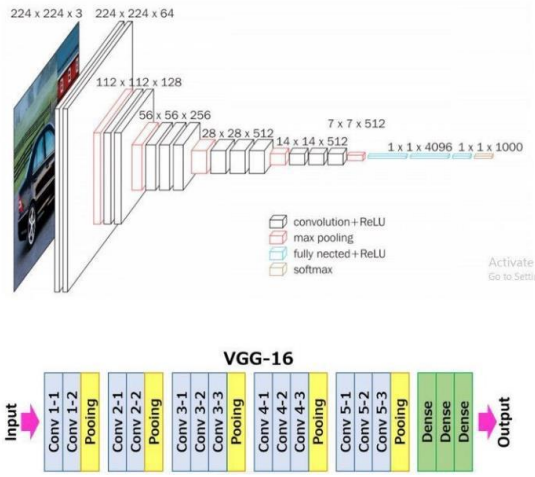

Figure 4: VGG-16 system.

VGG16's 16th digit denotes its 16 weighted layers. Only 16 of the total 21 layers in VGG16 are weight layers. It has 5 max-pooling layers, 3 thick layers, and 13 convolution layers. H.

Layers that have programmable parameters. VGG16 employs 3 RGB channels and accepts input tensor sizes of 224 and 244. This work's concentration on convolutional layers with 3x3 filters in step 1 and constant usage of padding and maxpool layers with 2x2 filters in step 2 instead of a huge number of hyper parameters is what makes VGG16 special. Max pool and convolutional layers are located regularly in the entire design. Conv-2 Layer of 128 filters, Conv-1 Layer consists of 64 filters, Conv-3 Layer of 256 filters, Conv-4 Layer of 512 filters, and Conv-5 Layer of 512 filters. Three Fully-Connected (FC) layers are added after a stack of convolutional layers, the third of which performs 1000-way ILSVRC classification and has 1000 channels. Each of the top two FC levels has 4096 channels (one for each class). The soft-max layer is the final layer.

IV. RESULTS

13,353 photos of handwritten lines of text by 657 writers are available in the IAM database. These authors have transcribed British English texts from the Lancaster Oslo/Bergen Corpus. 657 contributors contributed to a total of 1,539 handwritten pages

with 115,320 words, which are categorized as being a part of the Modern Collection. The database has tags at the word, line, and sentence levels. The algorithm is carried out by Python in PyCharm Community Edition 2022.2.1 in a system with Windows 11 operating system with CPU of 8 GB RAM, i7 8th Gen processor and GPU of NVIDIA GEFORCE GTX1050Ti.

The table represents three different experiments that were conducted in this study. Since three different preprocessing were used, the first experiment includes the results of each preprocessing technique used with CNN model. The second experiment shows the results when CNN model was used with 2 preprocessing techniques. Finally, the third experiment shows the result of all three preprocessing techniques used with CNN model.

The outcomes of the study are tabulated in table 1 as follows.

Table 2. Proposed model training and validation accuracy

Expt. No	Deep Learning method used	Training accuracy	Validation accuracy
1	CNN with VGG16 and data augmentation	98.36%	95.1%
	CNN with VGG16 and image segmentation	95.25%	91.8%
	CNN with VGG16 and image data generator	92.47%	89.36%
2	CNN with VGG16 and data augmentation & image segmentation	96.02%	92.1%
	CNN with VGG16 and image segmentation & data image generator	94.75%	90.95%
	CNN with VGG16 and data augmentation & image data generator	93.27%	90.67%
3	CNN with VGG16 and data augmentation, image segmentation & image data generator	97.85%	94.58%

Figure 5, depicts the pictorial representation of the various preprocessing technique with CNN model. It clearly indicates that data augmentation with CNN produced better accuracy in both training and testing.

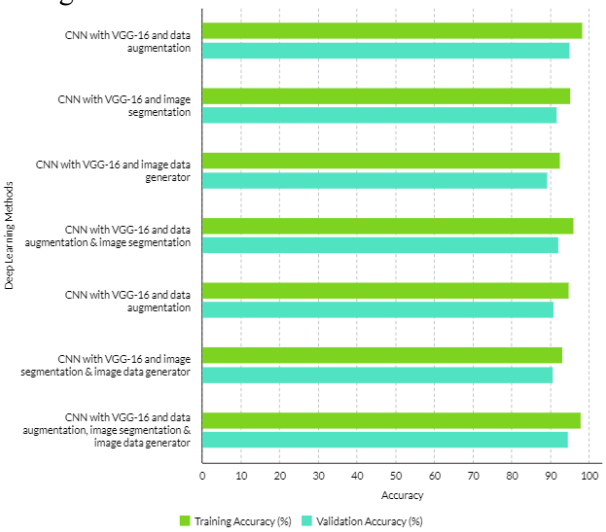

Figure 5: Pictorial representation of training and validation accuracy percentages.

V. Conclusion

The CNN in VGG-16 was used in this study to demonstrate and test a method for reading cursive characters and/or words. CNN is an artificial neural network and are commonly used for classification, segmentation, image processing, and other autocorrelation data. In computer vision, examples of CNNs are face recognition and image classification. This is similar to basic neural networks. CNNs have learnable parameters similar to neural networks like, weight, distortion, etc. On the other hand, binary and/or grayscale photos can be employed with the feature extraction method. This paper used three types of pre-processing techniques which are data augmentation, image segmentation and image data generator. The result for the first pre – processing technique, data

augmentation is 98.36% for training accuracy and 95.1% for validation accuracy. On the other hand, the image segmentation produces training accuracy for 96.25% and has validation accuracy of 91.8%. The last technique, Image data generator generates training accuracy of 93.47% and validation accuracy of 89.36%. In future different preprocessing techniques and some new deep learning model can be used in this study.

REFERENCES

[1] Batuhan Balci, Dan Saadati, & Dan Shiferaw: Handwritten Text Recognition using Deep Learning. In International Research Journal of Engineering and Technology Conference, 2017, pages 1-8.

[2] Noemí Bresó-Grancha, María José Jorques-Infante and Carmen Moret-Tatay: Reading digital- versus print-easy texts: a study with university students who prefer digital sources. In Psicologia: Refexão e Crítica, 2022, pages 1-9.

[3] Mandeep Kour, Vimal Kishore Yadav, Vikas Maheshwari, and Deepak Prashar: Image processing review, Volume 4, Issue 1, 2013, in International Journal of Electronic Communication and Computer Engineering; pages 1-6.

[4] Pritam Dhande and Reena Kharat: Recognizing Handwritten Cursive English Characters, International Conference on Computing, Communication, Control and Automation (ICCUBEA), 2017, pages 1-5

[5] MUEEN UDDIN, JAMSHED MEMON, and MAIRA AMI: Optical Character Recognition (OCR) of handwriting, in International Journal of Advance Research and Innovative Idead in Education, 2022, pages 1-10

[6] Kanchan Keisham and Sunanda Dixit: Recognition of Handwritten English Text Using Energy Minimization Information Systems Design and Intelligent Applications in International Journal of Engineering Reseasch and Technology, 2015, pages 1-4.

[7] Salvador Espan a-Boquera, Francisco Zamora-Martinez, Maria Jose Castro-Bleda, and Jorge Gorbe-Moya: Hybrid HMM/ANN Models for Improving Offline Handwritten Text Recognition, IEEE 2011, pages 1-13.

[8] Alberto N. G. Lopes Filho, Carlos A.B. Mello, Cleber Zanchettin, and Renata F. P. Neves: A SVM-Based Offline Handwritten Digit Recognizer in IEEE International Conference on Systems, Man, and Cybernetics, 2011, DOI: 10.1109/ICSMC.2011.6083734, pages 1-6.

[9] U. Pal, T. Wakabayashi, and F. Kimura: Handwritten Numeral Recognition of Six Popular Scripts," Ninth International Conference on Document Analysis and Recognition ICDAR 07, Vol. 2, pp. 749–753, 2007

[10] Namrata Dave: Segmentation Methods for Hand Written Character Recognition, in International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 8, no. 4, 2015, pp. 155–164

[11] U. Marti and H. Bunke. A full English sentence database for off-line handwriting recognition. In Proc. of the 5th Int. Conf. on Document Analysis and Recognition, pages 705 - 708, 1999.

[12] U. Marti and H. Bunke. Handwritten Sentence Recognition. In Proc. of the 15th Int. Conf. on Pattern Recognition, Volume 3, pages 467 - 470, 2000.

[13] ImageNet. http://www.image-net.org

[14] Karen Simonyan and Andrew Zisserman: Very Deep Convolutional Neural Networks for Large Scale Image Recognition, at International Conference on Learning Representations, 2015, pages 1-14.

[15] Dr. Harvendra Kumar, Tarun Patel, Shivansh Tiwari, Vaibhav Dubey, and Vaibhav Singh: A Study Comparing Python and Convolutional Neural Networks for Handwriting Digit Recognition, International Journal of Science and Research, 2022, pages 1480-1485

[16] Agnieszka Mikołajczyk; Michał Grochowski: Data Augmentation for improving deep learning in image classification problem, 2018 International Interdisciplinary PhD Workshop (IIPhDW), pages 1-5

[17] Jia Shije, Wang Ping, Jia Peiyi, Hu Siping: Research on Data Augmentation for image classification based on convolutional neural netwroks, at Chinese Automation Congress (CAC), 2017, pages 1-5

[18] Priyanka Shivhare, Vinay Gupta: Review of Image Segmentation Techniques Including Pre & Post Processing Operations, International Journal of Engineering and Advanced Technology (IJEAT), 2015, pages 1-5

[19] Ruiming Li, Jinxin Wang, Wenlev Zhou, Xi Wu, Chaojun Dong, Yikui Zhai, Yebin Chen: Actual Image Data Generation and Application via Generative Adversarial Network, at 2nd International Conference on Applied Machine Learning (ICAML), 2020, pages 1-6

[20] M. Yadav, R. Goel and D. Rajeswari, "A Deep Learning Based Diabetic Retinopathy Detection from Retinal Images," 2021 International Conference on Intelligent Technologies (CONIT), 2021, pp. 1-5, doi: 10.1109/CONIT51480.2021.9498502.

[21] O. T. Khan and D. Rajeswari, "Brain Tumor detection Using Machine Learning and Deep Learning Approaches," 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), 2022, pp. 1-7, doi: 10.1109/ACCAI53970.2022.9752502.

[22] R. Mathur, T. Chintala and D. Rajeswari, "Identification of Illicit Activities & Scream Detection using Computer Vision & Deep Learning," 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), 2022, pp. 1243-1250, doi: 10.1109/ICICCS53718.2022.9787991