Recognition And Classification of Indian Scripts in Natural Scene Images

Suryosnata Behera
Dept of Computer Science and Engineering
E-mail - beherasuryosnata@gmail.com

Dr.SatyaRanjan Pattanaik
Dept of Computer Science and Engineering
E-mail - drsatyaranjan@gift.edu.in

ABSTRACT

In the field of computer vision and document analysis, the identification and categorization of Indian scripts in natural scene images pose a difficult yet crucial challenge. The variety of characters and intricate writing styles in Indian scripts require reliable solutions for precise identification under different environmental conditions. This study presents a novel CNN model designed for identifying scripts in Indian multilingual document images captured by cameras. Experimental evaluations of the model's performance were conducted with two regional languages (Odia and Telugu) and one national language (Hindi). The average accuracy in script recognition for the three language combinations is 95.66%, with Odia achieving 99.00%, Hindi 90.33%, and Telugu 98.12%. The model achieved the highest accuracy in recognition. The model achieved the highest accuracy in recognition

Keywords: ***Text Recognition, Image Augmentation, CNN, LSTM, VGG, ResNet, DenseNet, Datasets, Natural Images***

I. Introduction:

Text is regarded as a popular and effective medium for information representation and exchange. Text has always been an essential component for both researchers and people in general [1]. Text can be discovered in the natural world, on paper, even handwritten. Print media comprises books, journals, periodicals, and newspapers; handwritten media comprises notes, letters, and manuscripts; and natural environment media comprises billboards, hoardings, banners, landmarks, and wall advertisements, among other things.

Textual images are images that have text in them. Either scanned or camera-captured photographs can be used as these textual images. While text images that are camera captured are taken using a digital camera or a mobile phone camera, scanned text images are taken with a scanning device. While digital cameras and mobile phone cameras can be used to capture text images from scenes as well as printed documents, scanners can be used to scan printed document images. Digital cameras are a common, effective, and strong picture capture equipment that come installed in practically all portable electronics, such as PDAs, tablets, pens, watches, and cell phones [2].The market for document imaging analysis is shifting due to technological advancements. Sales of built-in cameras, such as those in phones, tablets, and other gadgets, have increased in tandem with a decline in scanner sales. This is because these cameras are frequently used in a variety of application areas. The most difficult aspect of picture analysis is text extraction analysis using digital cameras [3].

II. Literature Review:

A system to identify handwritten characters in the Odia dialect on the internet. The authors had primarily concentrated on classifying distinct classes using strokes [4]. The writers have several character recognition techniques, including feature extraction and input pre-processing, which they describe in great depth. In order to recognize scripts from the camera-captured document images, a new CNN model is

created in this paper. Very little research has been done on script identification from camera-captured document images, scene text, natural scenes, and video scripts in the literature. They have created templates and signatures, applied statistical techniques to extract features, and tackled the script identification from camera-based document images. The distance between the train and test sets is determined by the classification using the hamming distance measure. They may obtain a recognition accuracy of 91.00% from this. Block-wise script identification taken by a camera is demonstrated. The LBP elements that were taken from the English, Hindi, and Kannada scripts have been taken into consideration. The recognition accuracy of KNN and SVM classifiers was 99.70% and 98.00%, respectively.

A convolutional-LSTM network's attention-based strategy for identifying scripts from pictures and video frames of real-world scenes has been documented by certain writers. From the input image, the local and global CNN features are retrieved, and the fusion method is used. With the CNN-LSTM network [5], they were able to obtain the recognition accuracy of 90.23% for ICDAR-17, 96.70% for MLe2e, 96.50% for SIW-13, and 97.75% for CVSI-15, using four standard databases: SIW-13, CVSI2015, ICDAR-17, and MLe2e. The HOG and GLCM features are taken from each word picture and merged to create a new feature vector. Five well-known classifiers were tested with this combined feature vector: SVM, Multi-class, Naïve Bays, MLP, and Multi-class. Out of all the classifiers, the MLP classifier has the highest recognition accuracy, scoring 90.00%. characteristics of an input picture.

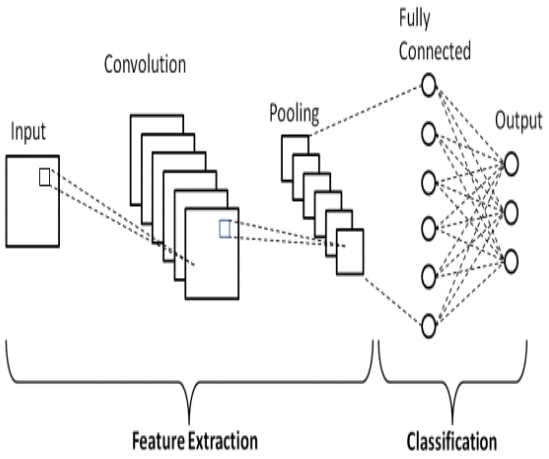


Fig.1 Flow Diagram of Purposed System

III. EXISTING SYSTEM:

Identifying and categorising Indian scripts in photographs of natural scenes is a difficult but doable challenge using current technologies[6]. A strong framework for solving this issue is provided by combining deep learning architectures, machine learning models, and sophisticated image processing approaches. Such systems' accuracy and efficiency will be further improved by ongoing research and development.

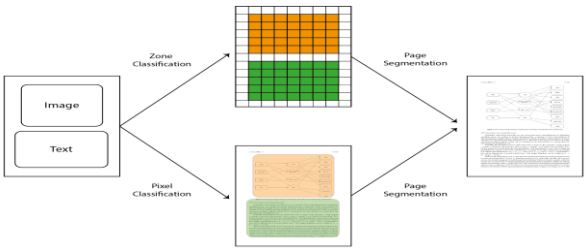


Fig.2. Image Extraction process

This is a high-level summary of the normal functioning of such a system, including information on the current components and methodologies:

i. **Image Acquisition:** Pictures are taken from a variety of places, such as digital libraries, security cameras, and mobile cameras. Better recognition accuracy is achieved with higher-resolution photos.

ii. **Preprocessing:**To eliminate noise, methods like median filtering and Gaussian blurring are employed.adjusting the contrast and brightness to make sure that every image is the same. separating text areas from the backdrop with techniques such as edge detection, linked component analysis, or thresholding

iii. **Script Identification:**Identifying and extracting script-specific features. This may comprise: Local Binary Patterns (LBP) or Gabor filters are utilised. examining the geometry and brushwork.To identify the script, capturing the gradient structure and training classifiers like SVMs, Random Forests, or neural networks on the extracted characteristics

iv. **Character Recognition:**Making use of OCR engines (such as Tesseract) modified to identify Indian scripts.Accuracy can be increased over regular OCR by using Convolutional Neural Networks (CNNs) trained specifically for Indian scripts..

v. **Postprocessing:**Correcting recognition mistakes by using dictionaries or language models.using contextual data to enhance identification, particularly in situations involving many scripts or languages.

Challenges

i. **Diversity of Scripts**: India has multiple scripts (e.g., Devanagari, Tamil, Telugu, Kannada), each with unique characteristics.
ii. **Natural Scene Variability**: Lighting conditions, angles, occlusions, and backgrounds can vary widely.
iii. **Font Variability**: Different fonts and handwritten styles add to the complexity.
iv. **Multilingual Contexts**: Scenes may have text in several languages or scripts, so correctly identifying the script is necessary before recognition.
v. **Limited Data Availability**It might be challenging to obtain sizable, annotated datasets for machine learning model training, particularly for scripts that are not as widely used. It can be difficult to make sure that a balanced dataset sufficiently covers all scripts and variations...

IV. Purposed Methodology

The CNN-LSTM framework is used in this paper to provide a unique approach for script identification that involves extracting local and global features and dynamically weighting them. Before feeding the images into a CNN-LSTM architecture, we first transform them into patches. Using the softmax layer after the LSTM, attention-based patch weights are computed. Then, in order to get local features, we multiply these weights patch-wise using the matching CNN. From LSTM's final cell state, global features are also retrieved. For each patch, we use a fusion technique that dynamically weights the local and global features. Trials have been conducted using the SIW-13, CVSI2015, ICDAR-17, and MLe2e public script identification datasets. By comparison with traditional approaches, the suggested framework produces better results.

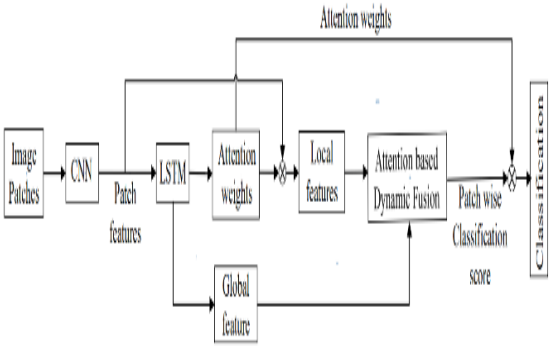


Fig.3. CNN-LSTM framework

Advantages:

i. Improved OCR Systems
ii. Advanced Machine Learning
iii. Business Applications
iv. Advertising and Marketing
v. Cultural Preservation
vi. Education and Accessibility
vii. Real-Time Applications
viii. Innovation and Exploration

V. SOFTWAREREQUIREMENTS:

i. Operatingsystem :Windows7&aboveversions
ii. Coding Language:Python
iii. 5.2HARDWAREREQUIREMENTS:
iv. Processor :Inteli3 andabove.
v. RAM:4GBand higher.
vi. HardDisk :500GBminimum.

VI. Implementation:

1.1. Implementing a CNN in Python for Image Classification:

Now that we have a basic understanding of CNNs, let's see how we can implement a CNN in Python for image classification using the popular Keras library.

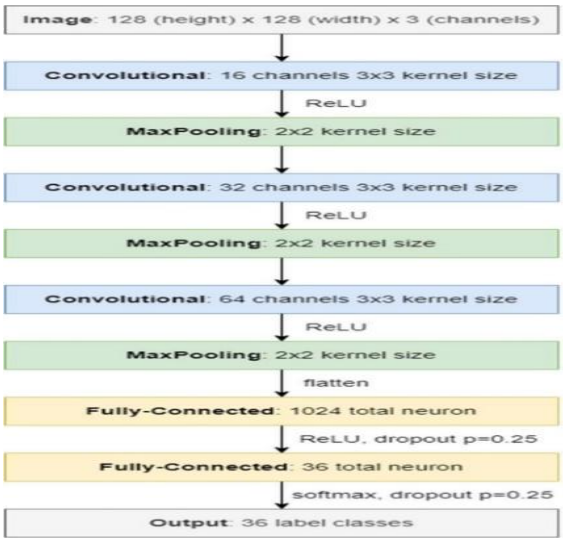
The first thing we need to do is install the required libraries. We will be using the following libraries −

- **NumPy** − A library for working with arrays and matrices
- **Matplotlib** − A library for creating plots and charts
- **Keras** − A high-level library for building and training neural networks

1.2.Convolutionalneuralnetwork:

CNN or Convolutional Neural Network is a type of deep learning neural network algorithm that is often used in the image classification process. Recently, deep learning algorithms have had great success in various computer vision problems . In its application, CNN implements a convolutional layer which is used to carryoutthe convolution processby creating akernelbased on thedata section of theinputdata.The output of the convolutional layer is

then processed at the pooling layer which is used to reduce the amount of data without losing important features by taking data that has been determined by the formula on certain pieces of data, such as the average value or the maximum value. This process is



then repeated with anamount as the total of convolutional and pooling layers. Then, the data is generally flattened to convert the data into one long dimension. And for the last step, the data is then going into the fully-connected layer for the classification process to get the output of the correct label from the CNN process. In this study, thevanilla CNN architecture used for the experiment is illustrated in Figure 4.

Figure4.IllustrationofCNNarchitecture

1.2.1.ReLUfunction

The ReLU (Rectified Linear Unit) activation function is anon-linear activation function that is used to eliminate neurons that have negative output results, and are replaced with a value of 0. Thus, the ReLU activation function is computationally efficient since it will deactivate the neuron with values below 0. The **formula** of ReLU function is expressed in Equation 1.

$$(x) = ma(0,x) \quad (1)$$

Where$(x)$is ReLU result and$x$is neuron value. suggested CNN model is tested with three distinct script combinations—Odia, Hindi, and Telugu—in order to evaluate its performance. The three script combinations yielded values of 99.00%, 90.00%, and 98.00%, respectively. These are very encouraging [10]. It ought to be necessary to increase recognition accuracy, especially for multi-script combinations.

1.2.2 SGDoptimization

Optimization in deep learning is a method of updating theweights by evaluating the performance of the model against the loss function. Stochastic Gradient Descent (SGD) is an optimization algorithm for updating the weights by getting the gradient loss value against the layer weight value by considering the amount of learning rate.

This paper presents a new CNN model for script identification from camera captured Indian multilingual document images. To evaluate the performance of the proposed model taken 2 regional languages(Odia,Telugu),one national language(Hindi).

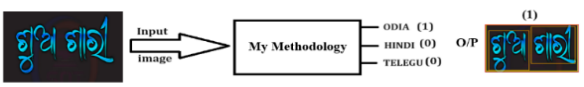


Fig.5 Proposed Model

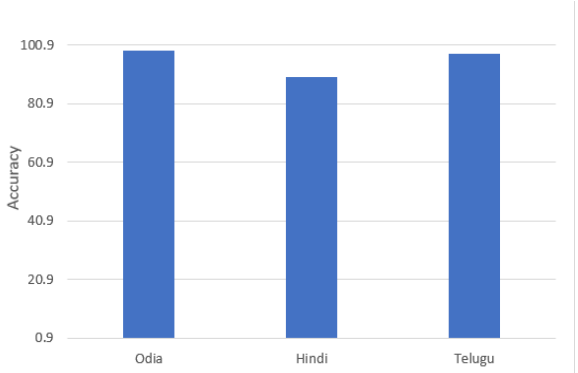


Fig.6. Confusion Matrix

VII.CONCLUSIONANDFUTURESCOPE

Through the development of a novel CNN model, this paper has given a camera-based script identification. The ability to recognize scripts in Odia, Hindi, and Telugu has been consolidated into a single model[9]. The different standard and own datasets were used to evaluate the performance of the proposed CNN model. The Text from natural photographs can be extracted and used for a variety of real-world applications, such as industry automation, automated robot navigation, portable aids, and text-to-speech generators for the blind. The future scope of recognition and classification of Indian scripts in natural scene images is vast and promising. Advancements in this field can

lead to significant improvements and new applications across various sectors. Here are some potential areas of future development and impact:

1. **Enhanced Machine Learning Models**: Continued development of more sophisticated neural network architectures, such as transformers and attention-based models, can further improve accuracy and efficiency.
2. **Multimodal Systems**: Combining visual recognition with data from other sensors (e.g., GPS, accelerometers) to improve context understanding and recognition accuracy.
3. **Real-time Translation and Interpretation**: Enhancing mobile applications to provide instant translation and interpretation of scripts in real-time, useful for tourists and multilingual communities.
4. **Automated Public Services**: Implementing recognition systems in public spaces to automate services like wayfinding, information kiosks, and public transport navigation.
5. **Surveillance and Security**: Using text recognition in surveillance systems to detect and interpret signage and notices for enhanced security monitoring.
6. **Personalized Shopping**: Enabling personalized shopping experiences by recognizing and interpreting product labels and advertisements in various script
7. **Digital Marketing**: Developing targeted digital marketing strategies based on text recognition from natural scenes to better engage with diverse linguistic groups.
8. **Language Learning Tools**: Creating interactive educational tools that help users learn Indian scripts by recognizing and translating text from their surroundings.
9. **Multilingual Support**: Enhancing language learning applications to support multiple Indian scripts, helping learners understand script differences and usage contexts.

References

[1] Rikiya Yamashita, Mizuho Nishio, Richard Kinh, Gian Do and Kaori Togashi, "Convolutional Neural Networks: An Overview and Application in Radiology", Insights Imaging, Vol. 9, pp. 611-629, 2018.

[2] S. Albawi, T.A. Mohammed and S. Al-Zawi, "Understanding of a Convolutional Neural Network", Proceedings of International Conference on Engineering and Technology, pp. 1-6, 2017.

[3] X. Lu and TzyyChyang, "CNN Convolutional Layer Optimisation based on Quantum Evolutionary Algorithm", Connection Science, Vol. 33, No. 3, pp. 482-494, 2021.

[4] A.K. Bhunia, A.K. Bhunia, P. Banerjee, A. Konwer, A. Bhowmick, P.P. Roy, U. Pal, Word Level Font-to-Font Image Translation using Convolutional Recurrent Generative Adversarial Networks, In Pattern Recognition (ICPR), 2018 24th International Conference on IEEE, (2018).

[5] A. Konwer, A.K. Bhunia, A. Bhowmick, A.K. Bhunia, P. Banerjee, P.P. Roy, U. Pal, Staff line Removal using Generative Adversarial Networks, In Pattern Recognition (ICPR), 2018 24th International Conference on IEEE, (2018)

[6] B. V. Dhandra, Satishkumar Mallappa, Gururaj Mukarambi (2020) "Script Identification at Line-level using SFTA and LBP Features from Bi-lingual and Tri-lingual documents Captured from the Camera

Cite this article as:

AnkanKumarBhunia,AishikKonwer,AyanKumarBhunia,AbirBhowmick,ParthaP.Roy,UmapadaPal "Script Identification in Natural Scene and Video Frame using Attention based Convolutional LSTM Network"