International Journal on Engineering Technology (InJET)

www.kec.edu.np/journal
Volume 1, issue no. 1, Nov 2023, Page 166-180
Received Date: Aug 4, 2023
Accepted Date: Oct 8, 2023
ISSN: 3021-940X (print)

Offline Handwritten Text Extraction and Recognition Using CNN-BLSTM-CTC Network

Ranjila Shrestha[1*], Oshin Shrestha[2], Monika Shakya[3], Urja Bajracharya[3], Subash Panday[3]

[1]*National College of Engineering, Talchikhel, Lalitpur, Nepal, ranjilashrestha2057@gmail.com*
[2]*National College of Engineering, Talchikhel, Lalitpur, Nepal, oshinshrestha.080@gmail.com*
[3]*National College of Engineering, Talchikhel, Lalitpur, Nepal, monikashakyaofficial@gmail.com*
[3]*National College of Engineering, Talchikhel, Lalitpur, Nepal, urjabajj2013@gmail.com*
[3]*National College of Engineering, Talchikhel, Lalitpur, Nepal, kavrelysubash@gmail.com*

Abstract

Offline handwriting recognition is a significant research area that aims at tackling problems encountered with handwritten forms in college application and registration processes. The objective of this study is to address the problems of English language offline handwriting recognition via CNN-BLSTM-CTC neural network applied for an NCE Admission form. The system uses OpenCV for image processing, TensorFlow for neural network training and handwritten text recognition, and trains and tests it on the IAM database using image segmentation-based handwriting recognition. With the help of proper image verification, the system allows the users to upload images of the NCE Admission form provided that they strictly comply with the specified format; it denies access to images not conforming to the set standards. Following the successful delivery of a valid image, the form goes through extensive processing that includes text extraction from specific regions of interest (ROIs). The extracted texts are then passed to text recognition block. The recognized texts are then recorded in a CSV file under respective fields. The text recognition model has a CER of approximately 9.33%. The study performed with 15 NCE Admission forms found that the average Character Error Rate (CER) was approximately 12.2% for scanned images and 19.3% for camera-captured images. The results show that accuracy depends on aspects such as the quality and orientation of the image; thus, scanned images are preferred for better performance.

Keywords: handwritten forms, offline handwriting recognition, NCE Admission forms, image segmentation, CNN-BLSTM-CTC

1. Introduction

Handwriting recognition has been a subject of research for several decades. Offline handwriting recognition automatically converts text in an image into letter codes usable within computer and text-processing applications (Sahu, 2013). Despite challenges like varying handwriting styles and image quality, advancements in machine learning have improved the accuracy of recognizing complete words and sentences. Automated handwritten text extraction and recognition offer several advantages. They significantly enhance efficiency by replacing the time-consuming manual transcription of handwritten text with automated processes, reducing the time and resources needed for document processing.

Additionally, electronic data storage is more cost-effective and space-efficient than physical files, eliminating the need for manual sorting and organization. Businesses can leverage data analysis on handwritten text from surveys and forms, facilitating more accessible insights extraction. Preserving fragile historical documents, such as letters and manuscripts, becomes possible through digitization, increasing accessibility and safeguarding cultural legacy. The primary objective of the system "Offline Handwritten Text Extraction and Recognition using CNN-BLSTM-CTC Network" is to create an application capable of taking an image of the NCE Admission Form and saving the

*Corresponding Author

recognized handwriting into a CSV file. The other objective is to analyze the complexities of recognizing handwritten text using CNN and BLSTM networks and calculate the loss using the CTC network. The system employs Optical Character Recognition (OCR) and a mighty Neural Network to extract and recognize text from NCE Admission forms. Optical character recognition is a technique of automatically identifying a different character from a recorded picture. It additionally provides full alphanumeric recognition of printed or handwritten characters, text numerals, letters, and symbols in a computer-processable layout, including ASCII, Unicode, and so forth (Awel & Abidi, 2019). It involves various steps like image preprocessing, feature extraction, and classification.

Traditional feature-based methods have been widely used but often require complex feature engineering suited for simpler handwriting styles. They struggle with diverse handwriting styles and complex scripts, demanding manual feature design. Hidden Markov Models (HMMs) have been used traditionally for handwriting recognition. They model the temporal dependencies in handwriting strokes and can be combined with various feature extraction methods. However, they require handcrafted features and explicit modeling of state transitions, which can be less flexible and efficient.

In contrast, deep learning with Convolutional Neural Networks (CNNs) automatically extracts hierarchical features, adapting better to various handwriting styles, albeit with challenges in recognizing entire sentences. Recurrent Neural Networks (RNNs), like LSTMs, excel at capturing sequence context vital for handwriting recognition but can face vanishing gradient issues in long sequences. CNN-RNN hybrid models, like CNN-BLSTM, strike a balance by combining CNNs for feature extraction and RNNs for sequence modeling, effectively capturing local and global information. The CNN-BLSTM-CTC approach likely stands out due to its excellence in recognizing complete words and sentences, adaptability to diverse styles, and robust handling of variable-length data using the CTC (Connectionist Temporal Classification) loss function, making it a solid choice for offline handwriting recognition.

First, the system uses pytesseract to determine whether the given form is an NCE Admission form, rejecting others. Once an NCE Admission form is recognized, the system crops the text fields and passes them to CNN-BLSTM-CTC Neural Network, designed explicitly for recognizing handwritten text on paper documents. CNN (Convolutional Neural Network) extracts relevant features from the image. BiLSTM (Bidirectional Long-Short-Term Memory) layers consider information from past and future steps to capture temporal dependencies and contextual information within the extracted features. BLSTM layers can better understand sequence patterns and relationships in handwriting data by incorporating this bidirectional perspective. The CTC loss function in handwriting recognition aligns input sequences (images) with target sequences (text) without explicit alignment information. It handles variable-length inputs and outputs, allowing the model to associate input images with the corresponding target text, even with alignment variations and uncertainties.

Currently, the system cannot recognize Nepali handwriting and emails present within the form. We can significantly augment the system's capabilities by integrating additional datasets and tailoring our models to recognize Nepali handwriting and identify emails within the forms. This enhancement will not only ensure more comprehensive results but also bolster the accuracy of our recognition. As we move forward, we envision the development of multilingual recognition systems that can accurately decipher handwritten text across a diverse array of languages. This exciting prospect holds the potential to bridge linguistic divides and empower individuals and organizations with a truly versatile and globally applicable solution.

2. Literature Review

There are five significant steps in text recognition from images: preprocessing, segmentation, feature extraction, classification, and post-processing. Preprocessing converts the color image into a binary image, while segmentation separates characters. Feature extraction captures essential information for recognition. Classification applies defined rules to identify the text, and post-processing reduces errors Adyanthaya (2020).

Zaitoun and Aqel, 2015 provided methods for Image Segmentation. The methods are classified into two main categories: Layer-Based Segmentation Methods and Block-Based Segmentation Methods. Layer-based Segmentation Methods for object detection and image segmentation composite the output of a bank of object detectors to define shape masks and explain the appearance depth ordering that evaluates both class and instance segmentation. Block-Based Segmentation Methods are based on various features found in the image. It might be color information used to create histograms or information about the pixels that indicate edges, boundaries, or texture information.

Abdulateef and Salman, 2021 performed a comprehensive review of Image Segmentation Techniques. The effectiveness of approaches varies according to object arrangement, lighting, shadow, and other factors. However, there is no generic approach for successfully segmenting all images; some approaches have been proven more effective than others. The study's primary goal was to provide a summary of the disadvantages and advantages of each of the reviewed image segmentation approaches which is shown in Table 1.

Table 1. Aggregation of Test Accuracies

Technique	Description	Advantages	Disadvantages
Boundary-based segment	Based on the discontinuity detection.	1- Suitable for simple images. 2- It is beneficial for images with a prominent contrast between objects. 3- Low computation intensive is making the methods more ideal for use.	1-It does not function well with images with many edges or unclear edges. 2-Producing a closed curve or boundary is not. 3-It is not suitable for images to be very noisy.
Region-based segment	Similarity is in terms of a set of predefined criteria.	1- More impervious to noise. 2-When defining similarity criteria is straightforward, this method is applicable.	Expensive according to memory and time.
Hybrid-based segment	Based on combining region and edge.	It depends on the combination of techniques.	It depends on the combination of techniques.

Krishna et al., 2020 aimed to improve the accuracy and effectiveness of perceiving and predicting handwritten digits ranging from 0 to 9. To achieve this, they utilized the renowned MNIST dataset. They employed the Histogram of Oriented Gradients (HOG) algorithm and the Support Vector Classifier (SVC) to analyze and fit the data. The study assessed parameters such as recognition rate, error rate, misclassified image rate, and computing time to evaluate accuracy and performance. Based on their analysis, the Handwritten Digit Recognition system utilizing the HOG algorithm achieved an impressive accuracy of 90%.

Geetha et al. (2021) proposed a hybrid handwritten text recognition (H2TR) model, combining deep neural networks with the sequence-to-sequence (Seq2Seq) approach. This model leveraged the strengths of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) with long short-term memory (LSTM). The CNN was employed to extract salient features from handwritten images. These extracted features were then utilized in a sequence-to-sequence framework, where an RNN-LSTM network encoded the visual features and decoded the sequence of letters present in the handwritten image. The proposed model was evaluated using the IAM and RIMES handwritten databases, demonstrating competitive letter and word accuracy.

According to Zhang (2021), recognizing text in natural scenes had low accuracy. In order to address this issue, an end-to-end deep learning text recognition algorithm called CRNN was employed for the OCR of images. The CRNN algorithm mainly comprises CNN, RNN, and CTC (Connectionist Temporal Classification) algorithms. Among them, CNN uses the improved VGG model to extract the sequence feature of the text line. In order to eliminate the gradient dispersion problem in the training process of RNN and strengthen the semantic information of the context, BLSTM was utilized to replace the RNN model for label prediction. Then, the CTC algorithm was used to complete the transcription and output of the final recognition result. The experimental results showed that the improved CRNN text recognition algorithm has an accuracy rate of 96.6%, which is 1% higher than the basic CRNN text recognition algorithm, and this design of the end-to-end network structure also significantly reduced the text recognition time.

Chandio et al. (2022) proposed a segmentation-free method for recognizing cursive Urdu text in natural scenes. Their approach eliminated the need for character segmentation and focused on cropped Urdu word images. The framework consisted of feature extraction, sequence labeling, and text transcription. Performance evaluations on a newly developed dataset showed that their model achieved a CRR of 95.75%, WRR of 87.13%, and WRR1F of 94.21%. Despite language complexities, the proposed method outperformed existing Arabic text recognition techniques. Future work involves incorporating language models and applying attention and transformer techniques to improve Urdu text recognition accuracy.

A fine-tuned AlexNet model with data augmentation achieved superior recognition performance, surpassing state-of-the-art methods. Three different settings of data were used for evaluation, namely, digits only, characters only, and hybrid, comprising both digits and characters. The accuracy for digit recognition was up to 98.21%, character recognition was 97.08%, and for the hybrid setting, it was 94.92%. The study suggests using deep neural networks like ResNet, GoogleNet, and other machine learning algorithms for future research (Rashees et al., 2022).

Sundaresan and Lin focused on developing a complete pipeline for text recognition in any input image. Their approach involves two main steps: image segmentation to extract individual characters and classification of these characters into their corresponding labels. They built upon Lecun et al.'s work on digit recognition and extended it to character recognition for the entire alphanumeric range (a-z, A-Z, 0-9). They found that the choice of architecture plays a crucial role in achieving high performance, even when working with similar datasets. For example, they observed that using the LeNet architecture on the Chars74K dataset resulted in subpar accuracy compared to other approaches. They came with conclusions, which are summarized in Table 2.

Table 2. Aggregation of Test Accuracies

	MNIST	Chars74K
KNN	0.9328	0.3547
Linear Classifier	0.8869	0.3015
LeNet	0.9905	0.4536
AlexNet	~	0.6338
Our Convnet	0.9812	0.7169
DeCampos	~	0.5526

Patel and Thakkar (2015) presented a comprehensive review of Handwritten Character Recognition (HCR) in English. They stated that immense work and research has been done in the handwritten separate character recognition. However, 100% accuracy has not been achieved so far, which gives scope for further work in this direction. Separate characters give good accuracy, but word recognition is affected by different writing styles. The holistic method eliminates complicated segmentation, but they use a limited vocabulary. Segmentation-based method, due to its complexity, acquires less accuracy. Good accuracy is observed in the classifier where the scope of words is limited to fixed numbers as it must deal with a limited number of variations.

2. Methodology

The block diagram of the system is shown in Figure 1. The system takes the input image from the user. The image is preprocessed and validated whether it is the desired form. The next step is ROI segmentation, which results in clipped images of form fields with their contents. The cropped images are then passed to the text recognition block, which is recognized using neural network layers. These recognized texts are stored in a CSV file with their associated fields. Finally, the user can download the CSV file.

The entire system is divided into four blocks. They are:

- Image preprocessing block
- Form validation block
- ROI segmentation block
- Text recognition block

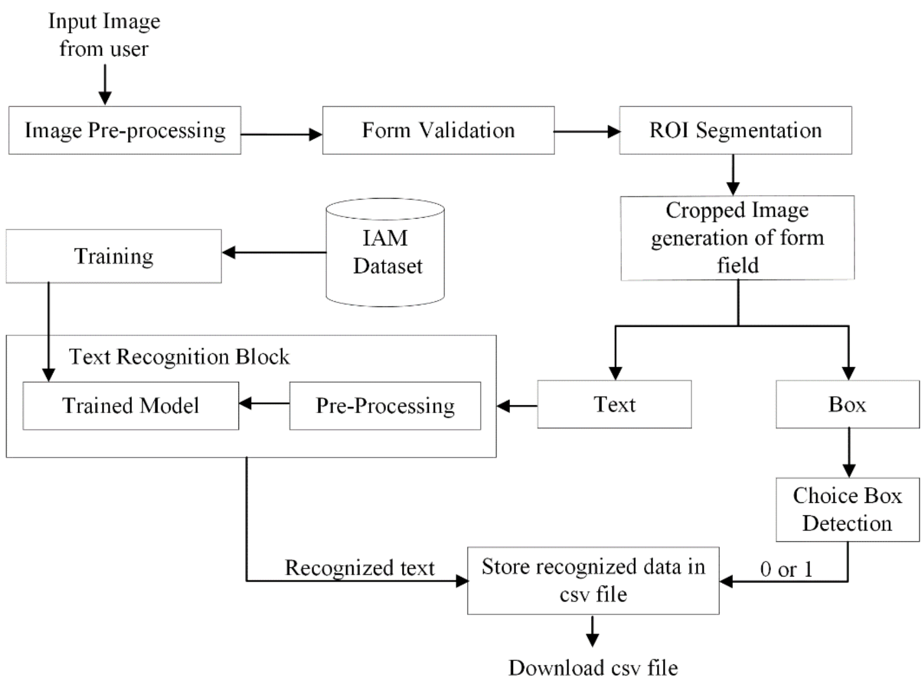


Figure 1. System block diagram

3.1. Image preprocessing block

The first step in form image preprocessing is to resize them to a fixed size of 4819 x 6873 pixels. It ensures that processing and analysis are consistent. The ORB (Oriented FAST and Rotated BRIEF) detection algorithm is then applied to the images to identify key points, as shown in Figure 2. The algorithm uses the Harries corner measure to find the top N key points based on distinct regions, such as intensity variations. These key points are then transformed into binary descriptors, representing feature vectors of 0s and 1s. Following that, a brute-force matching technique is used to compare the descriptors of features in the first and second images using hamming distance calculation, and the closest matches are returned, as shown in Figure 3. For improving the accuracy of perspective transformation, a homography matrix is employed. However, to account for outliers caused by noise or image artifacts, the RANSAC (Random Sample Consensus) technique is utilized to estimate the homography matrix. RANSAC helps filter out incorrect feature points, more accurately representing the geometric relationship between the images and facilitating perspective-warping transformation. Figure 4 shows the NCE Admission Form after perspective transformation.

Figure 2. ORB detection

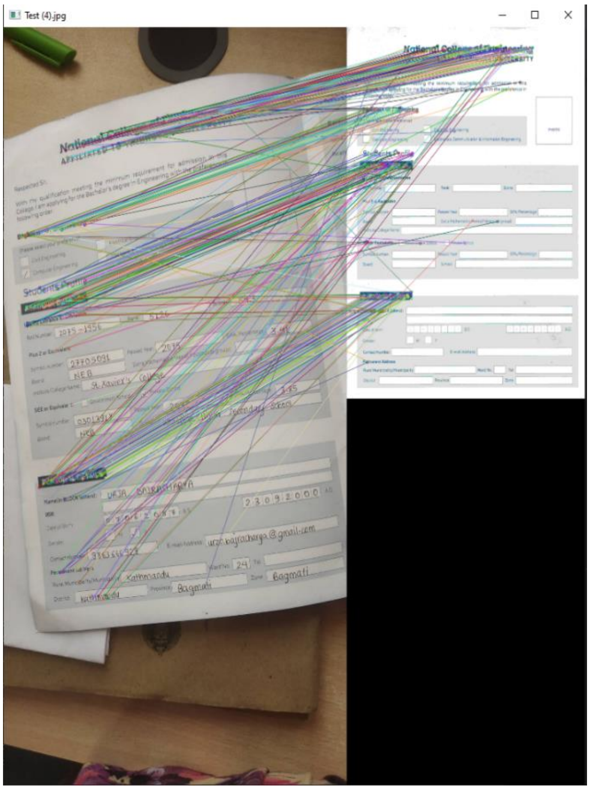
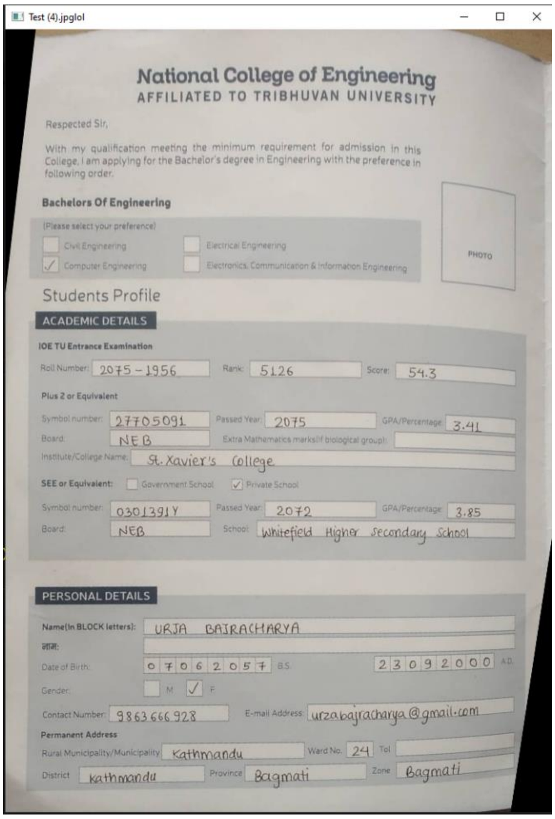


Figure 3. Bruteforce matching


Test (4).jpglol

National College of Engineering
AFFILIATED TO TRIBHUVAN UNIVERSITY

Respected Sir,

With my qualification meeting the minimum requirement for admission in this College, I am applying for the Bachelor's degree in Engineering with the preference in following order.

Bachelors Of Engineering

(Please select your preference)

☐ Civil Engineering ☐ Electrical Engineering
☑ Computer Engineering ☐ Electronics, Communication & Information Engineering

PHOTO

Students Profile

ACADEMIC DETAILS

IOE TU Entrance Examination

Roll Number: 2075 - 1956 Rank: 5126 Score: 54.3

Plus 2 or Equivalent

Symbol number: 27705091 Passed Year: 2075 GPA/Percentage: 3.41
Board: NEB Extra Mathematics marks(if biological group):
Institute/College Name: St. Xavier's College

SEE or Equivalent: ☐ Government School ☑ Private School

Symbol number: 0301391Y Passed Year: 2072 GPA/Percentage: 3.85
Board: NEB School: Whitefield Higher Secondary School

PERSONAL DETAILS

Name(In BLOCK letters): URJA BAJRACHARYA
नाम:
Date of Birth: 07062057 B.S. 23092000 A.D.
Gender: ☐ M ☑ F
Contact Number: 9863666928 E-mail Address: urzabajracharya@gmail.com

Permanent Address

Rural Municipality/Municipality: Kathmandu Ward No. 24 Tol
District: Kathmandu Province: Bagmati Zone: Bagmati

Figure 4. Perspective transformation

3.2. Form validation block

The image the user provides is validated in this procedure to determine whether it is an NCE Admission form. It is accomplished through the use of the Python library pytesseract. The key areas of the form carrying specific context are cropped and, using pytesseract, converted into text. The transformed text is compared to the original text using a sequence matcher. SequenceMatcher is a class in the difflib module of the Python standard library that allows us to compare sequences of elements. The SequenceMatcher class compares two sequences and returns a similarity ratio that indicates how closely the sequences match. The similarity ratio ranges from 0.0 (no similarity) to 1.0 (identical sequences). The form is accepted if the similarity ratio of all the matched contexts is above 0.5. Otherwise, the form is rejected.

3.3. ROI segmentation block

Initially, an empty form or a training image was provided as input. The required coordinates of the data fields on the form are selected using the mouse pointer. The field data type (a text field or a box) and the field name are specified and then organized in a list of regions of interest (ROIs).
As a result, when a completed form is submitted to the system, it returns the desired fields of the form as cropped images, which are then sent to the text recognition block for further processing.

3.4. Text recognition block

The text recognition block is shown in Figure 5. It consists of the following modules:-

- Preprocessing
- Trained Model
- Decoding

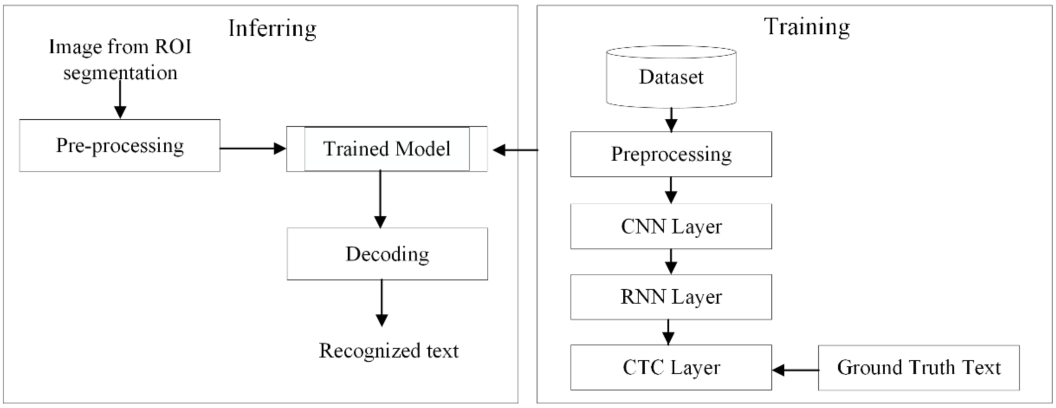


Figure 5. Text Recognition Block

3.4.1. Preprocessing

The image size is set to (256, 32) during the training process. The dataset words are concatenated with spaces to form text lines. Random words are selected and combined into lines. To enhance the dataset, photometric and geometric data augmentation techniques are applied. During inference, a target white image is created based on the width and height of the given image. The warpAffine function is then used to map the input image to the target image. The maximum text length is determined by dividing the image's width by 4. These steps help prepare the images for training and inference, enabling effective data augmentation and accurate processing of text lines.

3.4.2. Trained model

The trained model contains CNN, RNN, and CTC layers.

- CNN
  The input image is passed through the CNN layers, which are responsible for extracting relevant features from the image. Each layer in the CNN consists of several operations. First, a convolution operation is applied using filter kernels of different sizes (5x5 in the first two layers and 3x3 in the last three). This convolution operation convolves the filter kernel with the input image, generating feature maps. Each feature map represents a specific aspect of the image, such as edges or textures. After the convolution operation, a non-linear activation function is applied. The Rectified Linear Unit (ReLU) function is commonly used in CNNs. It introduces non-linearity by thresholding the output of the convolution operation, setting negative values to zero. It helps capture more complex features in the input data. Batch normalization is applied to normalize the activations of the previous layer at each batch during training. Pooling operations are then applied to downsize the feature maps. This pooling layer summarizes image regions and produces a downsized input version. In this architecture, the image height is reduced by a factor of 2 in each layer while the number of feature maps (channels) is increased. This results in an output feature map or sequence with a size of 64x256.

- RNN with BiLSTM
  The feature sequence consists of 256 features per time step, and an RNN architecture is employed to propagate relevant information through this sequence. Specifically, two RNN layers with 256 units each are created and stacked together to form a Bidirectional Long-Short-Term Memory (BLSTM) network. This BLSTM network processes the input sequence in both directions, from front to back and back to forth. Consequently, two output sequences, forward and backward, of size 64×256 are obtained. These two sequences are concatenated along the feature axis, resulting in a sequence of size 64×512. Finally, this sequence is mapped to the output sequence (or matrix) of size 64×80 and subsequently fed into the CTC layer.

- CTC
  During NN training, the CTC layer receives the transposed RNN output matrix and the ground truth text to compute the loss. The CTC loss function penalizes the difference between the network output and the ground truth labels. The loss function calculates the CTC loss between the two inputs, considering the sequence length and whether repeated characters should be merged. The resulting loss is a sparse tensor, with one value for each batch element. Then, another function calculates the mean loss across the batch, which the network tries to minimize during training.

3.4.3. Decoding

The best path decoding algorithm transforms the output sequence from the RNN layers into the final text during the decoding process. This algorithm follows a greedy search approach, selecting the most probable label at each time step. It starts by calculating the best path, considering the highest probability character for each time step. After obtaining the best path, the algorithm undos the encoding process. First, it removes duplicate characters to eliminate repeated labels in the sequence. Then, it removes all blank characters, typically used as placeholders or separators in the recognition process. The remaining characters in the path represent the recognized text.

Table 3. Summary of model

Type	Description
Input	Gray-value line-image (256×32)
Conv+Pool+BN	#Map 32 kernel 5×5, pool 2×2
Conv+Pool+BN	#Map 64 kernel 5×5, pool 2×2
Conv+Pool+BN	#Map 128 kernel 3×3, pool 1×2
Conv+Pool+BN	#Map 128 kernel 3×3, pool 1×2
Conv+Pool+BN	#Map 256 kernel 3×3, pool 1×2
Collapse	Remove dimension
Forward LSTM	256 hidden unit
Backward LSTM	256 hidden unit
Project	Project into 80 classes
CTC	Decode or loss

Table 3 summarizes the model used for the system, which contains 5 CNN layers where each layer is trained to extract relevant features from the image. The LSTM implementation of RNN is used as it can propagate information through longer distances and provides more robust training characteristics. Then, CTC is used for decoding and calculating loss value.

4. Results

We performed the training with the IAM dataset, which contained 115,320 isolated labeled words. The model was trained using Adam Optimizer and by concatenating the random words into lines. The default hyperparameters used for the Adam optimizer are as follows:

- learning_rate: 0.001
- beta1: 0.9
- beta2: 0.999
- epsilon: 1e-08

The CER obtained for each epoch after concatenating into lines is shown below:

Table 4. CER in each epoch

Epoch	CER
1	0.13465840710578195
2	0.12080715303568396
3	0.11658384955588863
4	0.11662076654434837
5	0.11329823758297093
…	…
31	0.09469207539925723
32	**0.0933113800308626**
33	0.10209762328428296
…	…
42	0.09750514991989014

Table 4 shows the CER for each epoch, which ran till 42 epoch. We have used an early stopping of 10 based on no improvement in CER, resulting in a model being saved at epoch 32 with a CER of **0.0933113800308626**.

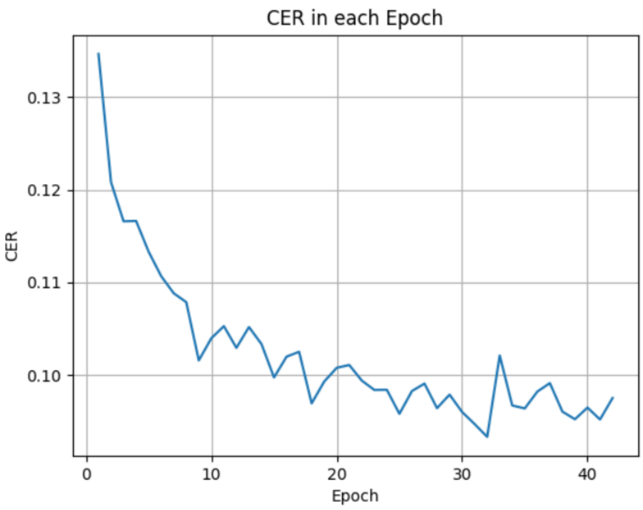
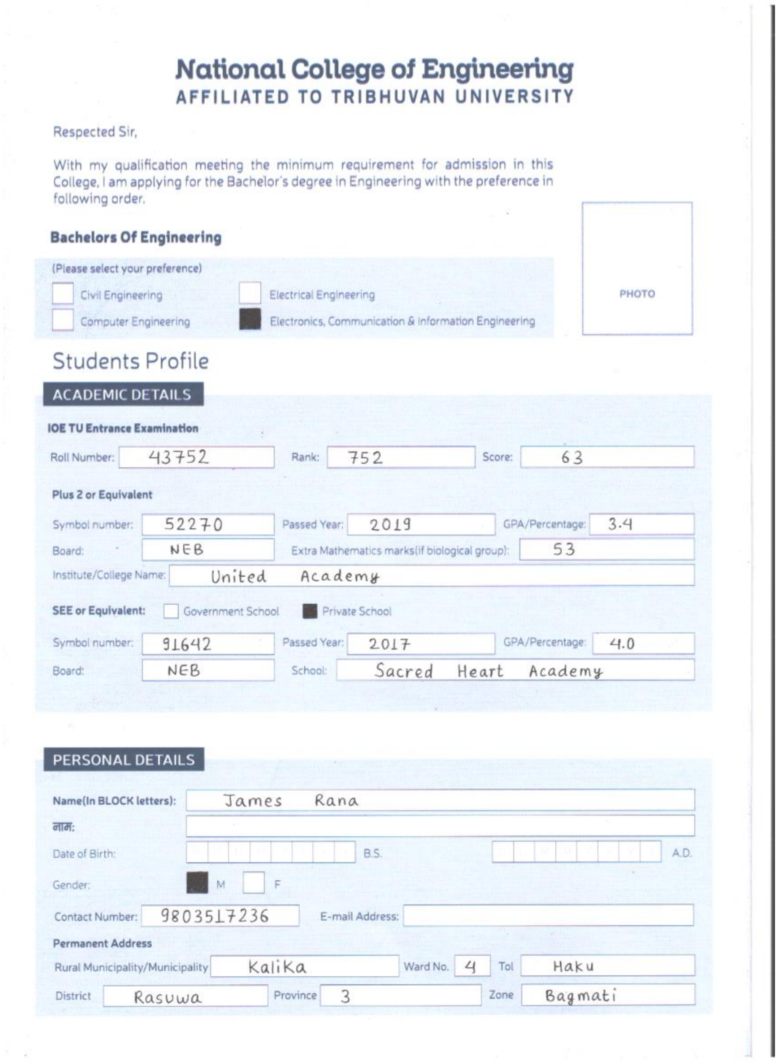


Figure 6. CER vs Epoch for training with words after concatenating into lines

The graph shown in Fig.6 depicts decrement in CER till epoch 32, which later increases on further epochs.

The input to the system and the results obtained are shown below:-



National College of Engineering
AFFILIATED TO TRIBHUVAN UNIVERSITY

Respected Sir,

With my qualification meeting the minimum requirement for admission in this College, I am applying for the Bachelor's degree in Engineering with the preference in following order.

Bachelors Of Engineering

(Please select your preference)

Civil Engineering | Electrical Engineering
Computer Engineering | Electronics, Communication & Information Engineering

PHOTO

Students Profile

ACADEMIC DETAILS

IOE TU Entrance Examination

Roll Number: 43752 Rank: 752 Score: 63

Plus 2 or Equivalent

Symbol number: 52270 Passed Year: 2019 GPA/Percentage: 3.4
Board: NEB Extra Mathematics marks(if biological group): 53
Institute/College Name: United Academy

SEE or Equivalent: Government School Private School

Symbol number: 91642 Passed Year: 2017 GPA/Percentage: 4.0
Board: NEB School: Sacred Heart Academy

PERSONAL DETAILS

Name(In BLOCK letters): James Rana
नाम:
Date of Birth: B.S. A.D.
Gender: M F
Contact Number: 9803517236 E-mail Address:

Permanent Address

Rural Municipality/Municipality KaliKa Ward No. 4 Tol Haku
District Rasuwa Province 3 Zone Bagmati

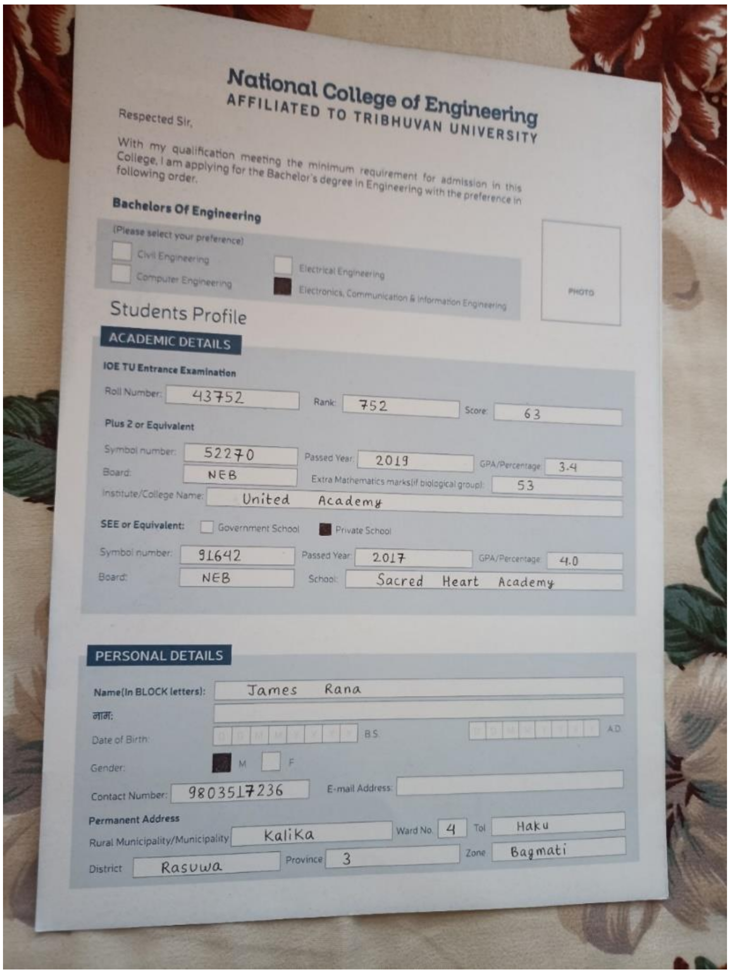
Figure 7. A scanned image of the NCE admission form

Table 5. Original data and recognized data of each field for Figure 7

Field	From image	Recognized
Roll no.	43752	43752
Rank	752	7752
Score	63	63
Symbol no. 1	52270	32270
Passed Year 1	2019	2019 .
GPA 1	3.4	3.4
Board 1	NEB	NEB
Extra Math	53	53
Institute/College	United Academy	United academy
Government School	0	0
Private School	1	1
Symbol no. 2	91642	91642
Passed Year 2	2017	2017
GPA/PER 2	4.0	4
Board 2	NEB	NeB

School	Sacred Heart Academy	Sacred Heart academy
Name	James Rana	James Rana
Male	1	1
Female	0	0
Contact	9803517236	8803527236
Municipality	Kalika	Kalika
Ward no.	4	4
Tole	Haku	MakU
District	Rasuwa	Rasuwa
Province	3	3
Zone	Bagmati	Baymati
Civil	0	0
Electrical	0	0
Computer	0	0
Electronics	1	1

When providing a scanned image of the NCE Admission form into the system, as shown in Figure 7, we obtain the results for each field, as shown in Table 5. A mean CER of 0.113 is obtained after calculation.


National College of Engineering
AFFILIATED TO TRIBHUVAN UNIVERSITY

Respected Sir,

With my qualification meeting the minimum requirement for admission in this College, I am applying for the Bachelor's degree in Engineering with the preference in following order.

Bachelors Of Engineering

(Please select your preference)

Civil Engineering
Computer Engineering
Electrical Engineering
Electronics, Communication & Information Engineering

PHOTO

Students Profile

ACADEMIC DETAILS

IOE TU Entrance Examination

Roll Number: 43752 Rank: 752 Score: 63

Plus 2 or Equivalent

Symbol number: 52270 Passed Year: 2019 GPA/Percentage: 3.4
Board: NEB Extra Mathematics marks(if biological group): 53
Institute/College Name: United Academy

SEE or Equivalent: Government School Private School

Symbol number: 91642 Passed Year: 2017 GPA/Percentage: 4.0
Board: NEB School: Sacred Heart Academy

PERSONAL DETAILS

Name(In BLOCK letters): James Rana
नाम:
Date of Birth B.S. A.D.
Gender: M F
Contact Number: 9803517236 E-mail Address:

Permanent Address
Rural Municipality/Municipality: KaliKa Ward No. 4 Tol: Haku
District: Rasuwa Province: 3 Zone: Bagmati

Figure 8. An image of the NCE admission form captured from Smart Phone

Table 6. Original data and recognized data of each field for Figure 8

Field	From image	Recognized
Roll no.	43752	43752
Rank	752	75 2
Score	63	63
Symbol no. 1	52270	227o
Passed Year 1	2019	2019 .
GPA 1	3.4	34
Board 1	NEB	NEB
Extra Math	53	53
Institute/College	United Academy	United cademy
Government School	0	0
Private School	1	0
Symbol no. 2	91642	91642
Passed Year 2	2017	2017
GPA/PER 2	4.0	4.0
Board 2	NEB	NEB
School	Sacred Heart Academy	Sacred Heartcademy
Name	James Rana	Tames Rana
Male	1	1
Female	0	0
Contact	9803517236	19803517236
Municipality	Kalika	KKalika
Ward no.	4	1
Tole	Haku	MakU
District	Rasuwa	RRasuwa
Province	3	3
Zone	Bagmati	Bagmatid
Civil	0	0
Electrical	0	0
Computer	0	0
Electronics	1	1

When providing an image of the NCE Admission form captured from a phone into the system, as shown in Figure 8, we obtain the results for each field, as shown in Table 6. A mean CER of 0.18 is obtained after calculation.

5. Discussion

We conducted a study involving a total of 15 forms, which were scanned and photographed. After analysis, we found that the average Character Error Rate (CER) for the scanned images was approximately 12.2%, while for the images captured from a camera, it was approximately 19.3%.
The insights gained from analyzing Character Error Rates (CERs) for various forms highlight a crucial aspect: the system's effectiveness is significantly influenced by the form's position and orientation. Ensuring accurate alignment is essential for precise recognition. When a form is not properly oriented according to the system's expectations, it can lead to misinterpretations of characters and words. Additionally, the form's placement within the scanning or imaging area is equally critical. If a form is not correctly centered or positioned, it can result in characters being cut off or distorted at the image's edges, causing recognition errors as the system may not access complete characters. Similarly, maintaining accurate spacing between characters and words is essential to prevent further misinterpretations, including errors in recognizing entire words, thus contributing to higher error rates. It's

also vital for all characters to stay within the provided boxes, as only text within these cropped boxes is sent to the system for recognition. Any handwritten text outside of these boxes will not be recognized.
Regarding potential areas for future improvement and development, there are various avenues to explore. These include integrating additional datasets and training models explicitly designed for recognizing email addresses and dates within forms. Another avenue involves advancing image processing techniques and employing more sophisticated machine learning algorithms capable of recognizing handwritten text without the need for region-of-interest (ROI) segmentation, thereby allowing text written outside of the box to be included.

6. Conclusion

In this project, we have demonstrated the potential of deep learning algorithms, precisely the CNN-BLSTM-CTC model, in recognizing handwritten data and extracting relevant information from NCE admission forms. The application we have built, which uses different programming languages and frameworks, can extract and recognize handwritten data and store it in a CSV file that can be easily downloaded and manipulated.
Recognizing handwritten data is challenging because handwriting styles can vary significantly between individuals. However, leveraging the IAM dataset achieved a 9.33% CER (Character Error Rate), which is a significant accomplishment. We used the scanned images and photographed images of the form for experiment. The analysis conducted on 15 NCE Admission forms revealed that the mean Character Error Rate (CER) stood at around 12.2% for scanned images and 19.3% for images captured by a camera. These findings underscore the significance of image quality and orientation as influential factors in accuracy. As a result, the preference leans towards scanned images for enhanced performance.
Overall, this project has demonstrated the power and potential of deep learning algorithms in solving real-world problems, particularly in data extraction and recognition. With further research and development, this technology can potentially transform how we handle and process handwritten data.

References

Sahu, V. L. (2013). Offline Handwritten Character Recognition Techniques using Neural Network: A Review. *International Journal of Science and Research*, 2(1).

Awel, M. A., & Abidi, A. I. (2019). Review on Optical Character Recognition. *International Journal of Engineering Research and Technology*, 6(6)**.**

Adyanthaaya, S. K. (2020). Text Recognition from Images: A Study. *International Journal of Engineering Research and Technology*, 7(8), 13-117.

Zaitoun, N. M., & Aqel, M. J. (2015). Survey on Image Segmentation Techniques. *Procedia Computer Science,* 65, 797-806

Abdulateef, S. K., & Salman M. D. (2021). A Comprehensive Review of Image Segmentation Techniques. *Iraqi Journal for Electrical and Electronic Engineering*, 17, 166-175.

Krishna, D. et al. (2020). Handwritten Digit Recognition and Text Conversion using MNIST Dataset. Vignan's Institute of Engineering for Women, Visakhapatnam, India.

Geetha, R. et al. (2021). Effective offline handwritten text recognition model based on a sequence-to-sequence approach with CNN-RNN networks. Neural Computing and Applications, 33, pp. 10923-10934. doi: 10.1007/s00521-020-05556-5.

Zhang, Y. (2021). Research of Natural Scene Text Recognition Algorithm Based on OCR. Sch J Eng Tech, 9(10), pp. 164-167.

Chandio, A. A. et al. (2022). Cursive Text Recognition in Natural Scene Images Using Deep Convolutional Recurrent Neural Network. IEEE Access, 10. doi: 10.1109/ACCESS.2022.3144844.

Rasheed, A. et al. (2022). Handwritten Urdu Characters and Digits Recognition Using Transfer Learning and Augmentation With AlexNet. *IEEE Access*, 10, pp. 102629-102645.

Sundaresan, V., & Lin, J. Recognizing Handwritten Digits and Characters.

Patel, M., & Thakkar, S. P. (2015). Handwritten Character Recognition in English: A Survey. *International Journal of Advanced Research in Computer and Communication Engineering*, 4(2).