Nepal Script Text Recognition From Ancient Artifacts: Challenges and Opportunities

Swornim Nakarmi, Sarin Sthapit, Sahil Ratna Tuladhar, Arya Shakya, Bal Krishna Bal, Rajani Chulyadyo
Kathmandu University, Nepal
{sn34021319, ss53021319, st59041720, as48021319}@student.ku.edu.np,
{bal, rajani.chulyadyo}@ku.edu.np

Abstract

Nepal Script, a script of significant linguistic, historical, and cultural importance, can be found in ancient artifacts in Nepal. As this script has faced a decline in use, it is considered among endangered scripts at present. For its revival and preservation, it is important to digitize ancient artifacts written in Nepal Script and create an accessible digital dataset. Among such artifacts are stone inscriptions, and manuscripts, from which we attempt to recognize texts using Artificial Intelligence techniques. This paper presents our approach of preparing a dataset through an extensive data acquisition method, and developing a system that recognizes Nepal Script texts from images. Our system combines the YOLOv8 algorithm with Convolutional Recurrent Neural Network architecture and Connectionist Temporal Classification loss. Our dataset consists of 5,219 text line images from ancient stone inscriptions, manuscripts, and modern handwritten and typed documents. Utilizing an augmented dataset of 41,752 samples, our system achieved 12.61% Character Error Rate. Despite the small training dataset, our model successfully predicted texts in not only new stone inscriptions and manuscripts but also wooden and copper plate inscriptions. We expect our contributions will encourage further research on Nepal Script and other Nepalese scripts.

Keywords: ocr, artifact dataset, artifact digitization, nepal script

1. Introduction

Nepal Script, also known as Prachalit Nepal Script, Newa Script, Newa Akha (Pandey, 2012) is one of the most common scripts for writing Nepal Bhasa, the native language of the Kathmandu Valley in Nepal. It is also used for writing Sanskrit, Pali and Maithili. It is derived from Brahmi Script, also known as Indic Script, and was developed in the 9th century (Nepal Lipi Guthi, 1992). Though several other scripts, such as Ranjana, Bhujimol, Golmol, Litumol, Pachumol, Kwenmol, Hinmol Kunmol were derived from it, Nepal Script is the most widely used one, followed by Ranjana and Bhujimol Script.

Many ancient Hindu and Buddhist manuscripts can be found to be written in the Nepal Script. It can also be seen on stone inscriptions (texts inscribed or carved into stone surfaces such as stone slabs and pillars). This script saw a decline during the Rana regime, the autocratic oligarchic rule of the Rana family in Nepal from 1846 to 1951 (Whelpton, 2005), when Nepal Bhasa and Nepal Script were banned from official use (Shrestha, 1999), suppressing the language and its associated literary and cultural traditions. The lifting of the restrictions after the end of the Rana regime made the study of the Nepal Bhasa possible, consequently reviving Nepal Script along with the other scripts. Today, numerous literature on Nepal Script can be found. Those works cover various topics, from history to grammar to font representation etc. However, only a limited amount of work has been done to automatically recognize texts written in Nepal Script from historically important documents. Such automatic systems can assist not only in the revival of the script, but also in digitization and preservation of ancient artifacts, thereby supporting linguistic, historical and cultural studies.

This work is an extension to our previous work on Nepal Script Text Recognition (Nakarmi et al., 2024). In this paper, we present our approach of developing a system for automatically recognizing texts from the images of ancient artifacts written in Nepal Script. We started with an extensive data acquisition through field visits to identify and photograph stone inscriptions across 85 locations in Kathmandu, supported by secondary data collection from archival records which include stone inscriptions, and religious manuscripts. A team of experts and researchers were involved to ensure that the accuracy and authenticity of the dataset were ensured, through manual verification and a workshop. Next, we performed data preprocessing, and transcription mapping to prepare a dataset for an Optical Character Recognition (OCR) model training, and developed an OCR system capable of recognizing text from images. Our OCR system combines the YOLOv8 algorithm (Redmon et al., 2016) with Convolutional Recurrent Neural Network (CRNN) architecture with Connectionist Temporal Classification (CTC) loss (Nakarmi et al., 2024) for recognizing the texts

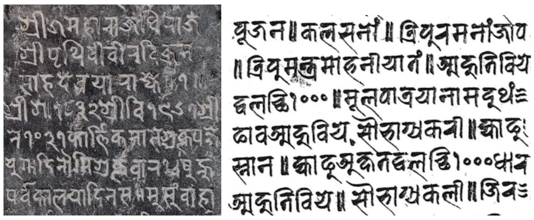
Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 3161–3170
11–16 May 2026. ©ELRA Language Resources Association (ELRA), 2026

at the line level. Despite being trained on a small dataset, our system could achieve a Character Error Rate (CER) of 12.61%. This work could be a significant contribution in the field of Nepal Script OCR, and also assist in heritage preservation, and restoration.

Our main contributions are summarized as follows:

1. We prepared a dataset of ancient artifacts for training the OCR models for recognizing texts in Nepal Script. The dataset consists of 2,371 text line images from stone inscriptions, and 1,758 from religious manuscripts together with their corresponding transcriptions in Nepal Script. Our primary dataset, which we collected ourselves, will be released for public use, whereas our secondary dataset, which we collected through multiple organizations, could not be released due to their license.

2. We developed an OCR system that takes an image of a document (ancient or modern) written or inscribed in Nepal Script, and predicts the texts written or inscribed on it. Our system first detects text lines using the YOLOv8 algorithm, and then predicts the texts using CRNN architecture with CTC loss.



(a) (b)

Figure 1: (a) A stone inscription, (b) A religious manuscript written in Nepal Script

(a) (b) (c)

Figure 2: (a) Variants of some vowels, (b) Variants of some consonants, (c) Variants of some numerals

2. Background

Once a widely used script, the usage of Nepal Script greatly declined due to various reasons, mostly political, and it was replaced for many decades by the Devanagari script, which is used for several South Asian languages such as Sanskrit, Hindi, Nepali, Pali, and Maithili, among others. However, in recent years, due to continuous campaigning and increased awareness, the Nepal script has been experiencing a revival, and the number of users is also increasing. Additionally, numerous tools, applications, and fonts have been developed. The release of its own Unicode code block, Newa (Unicode, Inc., 2023) in 2016 enabled it to be used on many devices.

Nepal Script is closely related to Devanagari Script. Like Devanagari, Nepal Script is written from left to right. It consists of vowels, consonants, numerals, and diacritics, many of which look similar to their Devanagari equivalents. Unlike Devanagari, where a header line, called *shirorekha*, runs along the top of full letters to connect letters in a word, the presence or absence of the header line in Nepal Script determines the type of some vowel modifiers to be used. Besides, ancient artifacts rarely separate words or even sentences. Words may be broken to avoid overflow. Figure 1 shows an example of a stone inscription and a handwritten manuscript, where separation of words cannot be seen. Additionally, there are numerous possible conjuncts, and some letters have variants, and may change their form during conjunct formation and while using modifiers. Figure 2 shows some variants of characters found in ancient manuscripts and inscriptions. Such variations pose challenges in training the text recognition models.

OCR is the process of converting text images into an editable digital format. It involves the acquisition of images, processing them, and applying text recognition algorithms to finally recognize the texts on the images. Developing such an OCR tool demands a collection of images together with the corresponding digitized text (called ground truth). Such ground truth is used for training an AI model so as to enable the model to make predictions on new texts presented to it. To our knowledge, such ground truth is available for some manuscripts (O'Neill and Hill, 2022), and handwritten and printed texts (Nakarmi et al., 2024) in

Nepal Script but not for stone inscriptions. Having ground truth for heterogeneous source could help build a more accurate optical character recogniser. In this paper, we present our attempt for i) developing a dataset or ground truth required for developing an OCR system for Nepal Script, and ii) developing a system that could recognize texts from ancient artifacts written in Nepal Script.

3. Literature Review

Following the lifting of the ban on Nepal Bhasa, Nepal Script witnessed a revival through awareness programs, collaborative training initiatives, cultural preservation efforts, and scholarly research on Nepal Bhasa and related scripts. However, compared to Devanagari, study of Nepal Script is still limited even though they are pretty much similar to each other. Several linguistic studies on Nepal Script can be observed, while only a limited number of research works on the automatic text recognition can be found. Only a handful of datasets are available for OCR on Nepal Script (O'Neill and Hill, 2022; Nakarmi et al., 2024), or Ranjana Script (Bati and Dawadi, 2023), a closely related script. On the other hand, several benchmark datasets (Dutta et al., 2018; Pant et al., 2012; Shaw et al., 2008; Singh et al., 2011) are publicly available for Devanagari Script, facilitating research on the OCR field for this script.

Traditionally, Machine learning techniques, such as Hidden Markov Model (HMM) (Shaw et al., 2008), Support Vector Machine (SVM) (Singh et al., 2011), K-Nearest Neighbors (Singh et al., 2011), Radial Basis Function (RBF) classifier (Pandey et al., 2017) etc. were commonly used for OCR on Nepali and Hindi documents written in Devanagari Script. However, with the advancement of Deep Learning techniques, the OCR community has also adopted these techniques to develop more advanced and more accurate OCR models. Dutta et al. (2018) released a handwritten word dataset, called IIIT-HW-Dev, and benchmarked it using a CNN-RNN hybrid architecture. Dwivedi et al. (2020), and Mondal and Jawahar (2022) used an attention-based encoder-decoder framework for OCR of Indic handwritten documents. Vaidya and Bal (2024) also used the same framework along with CRNN for Nepali handwritten word recognition.

In the context of automatic text recognition for Nepal Script and related scripts, O'Neill and Hill (2022) released ground truth consisting of the images of manuscripts written in Nepal Script and corresponding transcriptions. They trained a model on the Handwritten Text Recognition (HTR) engines on the Transkribus platform. Nakarmi et al. (2024) used Convolutional Recurrent Neural Network (CRNN) architecture with Connectionist Temporal Classification (CTC) loss to detect Nepal Script texts on handwritten and printed documents with a Character Error Rate (CER) of 6.65% and a Word Error Rate (WER) of 13.11%. Bati and Dawadi (2023) introduced a dataset of handwritten characters in Ranjana script, and proposed a CNN architecture that produced 99.73% accuracy.

Vision Transformers (ViTs) (Dosovitskiy, 2020) have emerged as a transformative architecture in computer vision by processing images as sequences of patches and employing self-attention mechanisms to capture global context rather than relying exclusively on convolutional operations. In the HTR-VT approach, this paradigm is adapted for handwritten text recognition by integrating a CNN feature extractor, a span mask strategy, and the Sharpness-Aware Minimization (SAM) optimizer to boost model generalization and efficiency even with limited training data (Li et al., 2025). This methodology can be particularly valuable for developing a text recognition system for the Nepal Script, where annotated data is often scarce and handwriting styles can vary significantly.

4. Methodology

Our work employs a typical OCR pipeline, as shown in Figure 3. The input image first undergoes text line detection using YOLOv8, and the detected regions are extracted as segmented line images. These line images are converted to grayscale and then passed to the text recognition model to produce the final text output.

With the aim of developing an OCR system for recognizing texts written in Nepal Script, we first prepared a dataset consisting of images of ancient stone inscriptions, and manuscripts, and their corresponding transcriptions. After data preprocessing, and data augmentation, we developed an OCR model, and evaluated it. These steps are explained in detail in the following sections.

4.1. Data Acquisition

Primary dataset was prepared through field visits whereas secondary dataset was prepared by collecting archival data from Central Department of Nepal Bhasa, Tribhuvan University, and also by purchasing some data from the National Archives of Nepal.

4.1.1. Primary Data Collection

Various places in Kathmandu Valley, including Kathmandu Metropolitan City, Patan, Banepa, and Thimi, were visited to identify the sites containing stone inscriptions. As most inscriptions are located in publicly accessible areas, photography

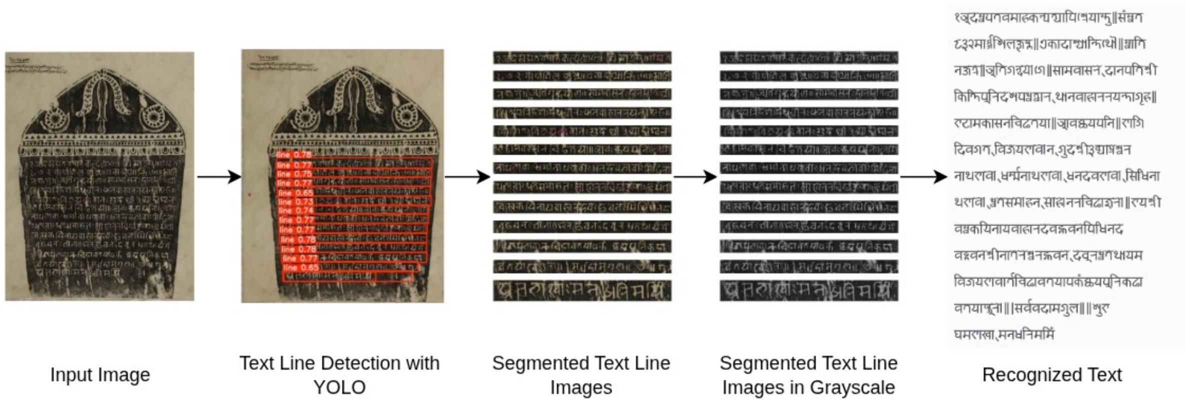


Figure 3: Overview of the OCR Pipeline

was carried out without special permissions, while the relevant local government bodies were consulted where required. A total of 85 sites were identified. Photos were taken after cleaning the stone inscriptions, applying white powder such as limestone powder, chalk powder, over them. A sample of stone inscription after applying white powder is shown in Figure 4a. After taking photographs, the inscriptions were cleaned again and left in the original condition. Some inscriptions did not need white powder as they were in good condition. They could be photographed after basic cleaning.

4.1.2. Secondary Data Collection

As a supplement to our primary dataset, we collected secondary data of two types from multiple sources - image data and transcription data. We purchased images of 95 stone inscriptions (taken using the rubbing technique, as shown in Figure 4b) from the National Archives of Nepal[1]. These purchases were made in accordance with the respective institutional policies of the National Archives of Nepal and the Central Department of Nepal Bhasa. Besides images of the inscriptions, we also need their transcriptions to train the model. So, we purchased Kantipur Shilalekh Suchi (Rajbanshi, 1970), which comes in 4 parts, from the National Archives, containing transcriptions of stone inscriptions in Kathmandu. These books contain texts in Devanagari Script. As we have their hard copy only, we needed to perform OCR of those texts, and convert the Devanagari texts into Nepal Script.

Additionally, we obtained both images and transcriptions in Devanagari Script of some ancient manuscripts from the Central Department of Nepal Bhasa[2] at Patan Multiple Campus, Nepal. Additionally, we also included the original dataset of 1,090 samples, which was previously used to prepare the Nepal Script Text Dataset (NST Dataset) (Nakarmi et al., 2024).

[1] http://narchives.gov.np/

[2] https://cdnb.tu.edu.np/

(a) (b)

Figure 4: (a) Image of a stone inscription taken after applying white powder, (b) Rubbing of a stone inscription

4.1.3. Image to Transcription Mapping

Training OCR models require a collection of images together with the corresponding digitized text. To prepare such a collection, transcriptions of the acquired images were extracted from Kantipur Shilalekh Suchi (Rajbanshi, 1970) and Madhya Kalka Abhilekha (Vajracharya, 1999). A mapping of the images to the transcriptions was then constructed. Transcriptions available in these books are in Devanagari Script, and, therefore, needed to be converted to Nepal Script. Some stone inscriptions were manually transcribed by our team. Though we obtained both images and transcriptions from the Central Department of Nepal Bhasa, we needed to map the manuscript images to their corresponding transcriptions. Since the transcriptions were in Devanagari Script and in the image format, they had to be converted to digital format by first performing OCR and then apply-

Table 1: Summary of our dataset

Type	No. of Line Segments
Stone inscriptions (Primary)	1,097
Stone inscriptions (Secondary/Rubbings)	1,274
Manuscripts (Secondary)	1,758
Handwritten and printed (Secondary)	1,090

ing Devanagari to Nepal Script Converter[3]. Next, the manuscript images were segmented and their corresponding transcriptions were checked and mapped.

4.1.4. Data Verification

Data was first manually verified by the authors. Further verification was conducted by a team of Nepal Script experts during a workshop organized by the authors on Data Verification and Annotation of Stone Inscriptions in Nepal Script, held on January 11, 2025, at Durbar High School, Kathmandu. During the workshop, several inconsistencies arising from multiple data collection sources and the manual transcription process were identified. To address these inconsistencies, the transcriptions were cleaned and standardized: the hash symbol (#) was used for illegible characters, continuous spaces were removed, and visually similar characters such as 𑑑 (Nepal Script numeral equivalent to "1") and 𑑇 (the Siddhi character), were carefully differentiated.

4.1.5. Dataset Summary

After the comprehensive data acquisition process and data preprocessing, we prepared a collection of 2,371 text line images of stone inscriptions, and 1,758 of manuscripts. Table 1 presents the summary of the dataset.

4.2. Data Preprocessing

Data preprocessing workflow consisting of several key steps was performed to improve the quality of images. These steps are explained in the following sections.

4.2.1. Line Segmentation

Initially, manual segmentation was carried out using Adobe Photoshop on a limited dataset to create a prototype. Later, automatic text line segmentation was conducted using the YOLOv8 algorithm, trained on 115 annotated images, to detect and extract text lines from the images. This resulted in a collection of 5,219 line segments and each of the image segments were then mapped to their corresponding transcriptions. These image segments were further augmented to increase the dataset size, as described in Section 4.3.

4.2.2. Image Gray-Scale Conversion

All images were converted from color to grayscale, as the color information did not significantly contribute to model performance. Moreover, using RGB images requires three channels, which increases computational resources compared to the single channel used for grayscale images.

4.2.3. Image Resizing

All images were standardized to a fixed dimension of 508×64 pixels (W×H) to ensure uniformity. Images with varying sizes were adjusted by adding black padding pixels. Subsequently, these images were transposed to a dimension of 64×508 pixels (W×H) to match the input size of our model.

4.3. Dataset Augmentation

Deep Learning problems like text recognition systems require a large amount of dataset to achieve better results. As our dataset was constrained by limited resources, we applied dataset augmentation techniques to produce variations in our original dataset. Dataset augmentation is a method to increase the existing dataset by applying random transformations. It helps improve model generalization and robustness.

Various data augmentation techniques, shown in Table 2, were applied in that order to our dataset using Albumentations library with a fold of 8. After applying these techniques, we obtained 41,752 text line images.

4.4. Model Development

Our work employs the YOLOv8 algorithm for detecting text lines, and CRNN-CTC architecture for text recognition. The input to our model is an artifact image, which is first passed through a text line detection model based on the YOLOv8 algorithm to detect and segment text lines in the image, as depicted in Figure 3. Each text line image is then passed to the CRNN network, which accepts grayscale input images of size 64×508 pixels (W×H). As shown in Table 3, it comprises five convolutional layers with ReLU activation, batch normalization, and L2 Regularization, followed by two max-pooling layers, each with a kernel size of 2×2. The convolutional layers utilize a kernel size

[3] https://www.nepalbhasa.org/nepal-lipi-converter

Table 2: Transformations performed for data augmentation

Transformation	Parameters
Affine Transformation	rotate=(-1, 1) shear=(-2, 2) scale=(0.98, 1.02) p=1.0
Perspective Transformation	scale=(0.02, 0.04) p=0.35
Elastic Transformation	alpha=1 sigma=50 p=0.7
Gaussian Noise Injection	std_range=(0.2, 0.3) p=0.5
Random Brightness and Contrast	brightness_limit=0.2 contrast_limit=0.4 p=0.6
Grayscale Conversion	p=0.2

of 2×2 with progressively increasing filter sizes of 32, 32, 64, 64, and 128. The resulting feature maps are subsequently reshaped into dimensions of 127×2048 for sequence modeling. Temporal dependencies are captured through three stacked Bi-LSTM layers with 256, 128, and 128 hidden units, each incorporating a dropout rate of 0.3. The sequential output is further processed by a time-distributed dense layer with 128 units and L2 Regularization, followed by a softmax layer with 105 output classes, representing 104 character classes and one blank label for the CTC decoder. Finally, the CTC decoder converts the per-time-step predictions into the final text transcription.

4.5. Model Evaluation

For evaluating the model, the prepared dataset was divided into train, validation, and test sets with a split ratio of 70:15:15. The training, validation, and test sets consisted of 29,226, 6,263, 6,263 samples respectively. Character Error Rate (CER) was calculated separately for stone inscription, stone inscription rubbing, manuscript, and regular handwritten and printed text images. Word Error Rate (WER) is not reported because in stone inscriptions and manuscripts, the characters are equally spaced without distinction between words. Additionally, in some cases, words are broken down into two lines to prevent overflow, i.e. line segments may contain incomplete words.

5. Results and Discussions

Our model was trained on a Kaggle kernel using a P100 GPU. With 250 epochs, the Adam optimizer (learning rate = 0.001), and a batch size of 32, the training took approximately 10 hours and 55 minutes to complete. Additionally, a learning rate

Table 3: Text recognition model implementation summary; f, k, p, n, and d stand for filter size, kernel size, padding type, number of hidden units, and dropout rate respectively.

Layer	Configurations
Input	Gray-scale image of dimension 64×508 pixels (W×H)
Convolution	f=32, k=2×2, p=same with L2 Regularization
Batch Normalization	-
Convolution	f=32, k=2×2, p=same with L2 Regularization
Batch Normalization	-
Convolution	f=64, k=2×2, p=same with L2 Regularization
Batch Normalization	-
Max Pooling	k=2×2
Convolution	f=64, k=2×2, p=same with L2 Regularization
Batch Normalization	-
Convolution	f=128, k=2×2, p=same with L2 Regularization
Batch Normalization	-
Max Pooling	k=2×2
Map-to-sequence	-
Dense	n=64 with L2 Regularization
Batch Normalization	-
Dropout	d=0.3
Bidirectional-LSTM	n=256, d=0.3
Bidirectional-LSTM	n=128, d=0.3
Bidirectional-LSTM	n=128, d=0.3
Time Distributed Dense	n=128 with L2 Regularization
Batch Normalization	-
Dropout	d=0.3
Time Distributed Dense	n=104 (size of character set) + 1 (blank character) with Softmax
CTC Decoder	-
Transcription	-

scheduler (particularly, `ReduceLROnPlateau`, a scheduler which reduces learning rate when a metric has stopped improving) with a factor of 0.5, and a patience of 5 was employed to adjust the learning rate dynamically. Figure 5 shows the training and validation CTC loss curves.

As shown in Table 4, the model achieved CERs of 3.82%, 25.21%, 18.31%, and 3.09% for regular handwritten and printed texts, stone inscriptions, stone inscription rubbings, and manuscripts respectively, with an average CER of 12.61%. The model performed best on regular and manuscript texts, as the characters in these categories are

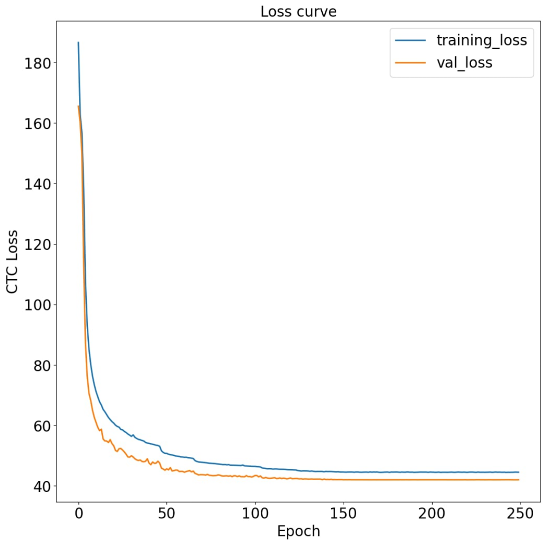


Figure 5: Training and validation CTC loss curves

Table 4: CERs by artifact type

Artifact Type	CER (%)
Regular handwritten and printed text	3.82
Stone inscription	25.21
Stone inscription rubbing	18.31
Manuscript	3.09

more easily distinguishable from the background. In the case of inscriptions and rubbings, the model showed moderate performance due to the presence of smudges, stone surface textures, and numerous illegible characters. Some sample predictions are illustrated in Figure 6, which also includes prediction of texts inscribed on some copper plates, and wooden surfaces.

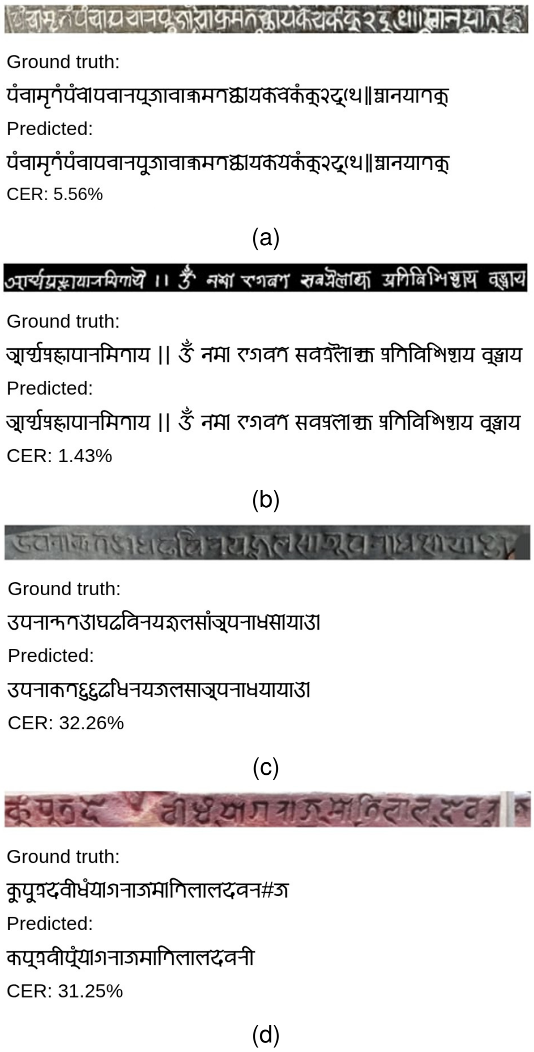


Figure 6: Some text recognition results on (a) a stone inscription, (b) a regular handwritten document, (c) a copper plate inscription, and (d) a wooden inscription

5.1. Challenges and Opportunities

Text recognition using ancient artifacts as a data source leads to a distinct set of challenges which is significantly different from OCR tasks involving printed or modern digital text. Firstly, character shapes often vary according to historical period, regional style, or the individual technique employed while inscribing. Also, the absence of a standardized form of the script across time further complicates the development of generalized recognition models. In many cases, even human experts find it difficult to interpret certain inscriptions, which highlights the difficulty of training machines to do the same.

Another major challenge is the lack of annotated training data. In particular, large, high-quality datasets for Nepal Script OCR are extremely rare, especially for inscriptions carved into stone. Consequently, most research efforts lie in collecting and annotating data manually, which is extremely time-consuming. In addition, a very limited number of people possess the expertise needed to transcribe historical Nepal Script accurately. Overall, the lack of properly standardized practices makes it very inconvenient to compare methods or measure progress across studies.

The current condition of the artifacts is also a major cause of complications because many artifacts are centuries old and may be weathered, eroded, dirty, or partially missing. Moreover, image acquisition also introduces noise through uneven lighting, glare from stone surfaces, shadows, and perspective distortion due to photographing angles. Consequently, custom image processing might be required. Similarly, decorative elements carved into the stone and irregular spacing between characters or lines further increase the difficulty of layout analysis.

Importantly, the current OCR tools offer limited to no support for Nepal Script. There is a need to design custom architectures and preprocessing techniques from scratch, as there are no reliably pretrained models or well-defined pipelines specific to this script, which leads to increased development time, computational cost, and model instability. Thus, Nepal Script OCR remains a very underexplored and technically challenging area within the broader research community.

Regardless of these challenges, this field presents many opportunities. One of the most impactful directions is the development of open, standardized datasets, and benchmarks. In particular, well-annotated and publicly available datasets would ease reproducible research, proper model comparison, and faster innovation. Meanwhile, interdisciplinary collaboration also holds significant promise. Particularly, historians, epigraphers, linguists, and computer scientists each can contribute crucial expertise, making it possible to interpret inscriptions accurately while designing effective recognition systems grounded in historical and linguistic context.

Furthermore, advancements in Nepal Script OCR with the emergence of newer AI technologies could make it possible to create annotated digital archives which linguists, researchers, educators, heritage-enthusiasts, and the general public could actually use. These digital archives would help heritage preservation and promotion, support academic study, and provide easier access to historical materials. Additionally, OCR could also play instrumental role in restoration work by allowing virtual reconstruction of inscriptions that are in poor conditions using the applications of Augmented Reality and/or Virtual Reality. On top of that, interactive tools and academic platforms could be useful in reviving the interest in studying or learning ancient Nepal Script, helping to promote the language and culture.

5.2. Current Initiatives

Several organizations, along with local and governmental institutions, have recently conducted notable initiatives to preserve and promote Nepal Bhasa and its scripts. Several youth led organizations like Callijatra Foundation have been actively involved in conducting workshops, live demonstrations, online tutorials, and in-person classes to teach how to write Nepal Bhasa scripts, engaging learners across different age groups.

Initiatives have also been conducted on the digital front to promote Nepal Bhasa. One such effort was the launch of Nepal Bhasa application in June 2024, which was led by Dabuli[4] in partnership with Nepal Lipi Guthi, Kathmandu Metropolitan City, and Nepal Bhasa Academy. It provides a 41,000-word trilingual (English-Nepali-Nepalbhasa) dictionary, script tutorials, quizzes, and proverbs to promote learning and engagement in Nepal Bhasa.

Furthermore, Kathmandu Metropolitan City has mandated and included Nepal Bhasa education in the curriculum from Grade I to Grade VIII in all public and private schools within the metropolis. This initiative makes sure that future generations gain knowledge and are aware about Nepal Bhasa and its scripts regardless of their ethnic background (Himalayan News Service, March 14, 2021). Together, these initiatives reflect a coordinated effort by various organizations and institutions to safeguard, preserve, and promote Nepal Bhasa and its associated heritage.

6. Conclusion

Ancient artifacts such as stone inscriptions, manuscripts, books, coinage, etc., are invaluable sources of information for cultural, historical, and linguistic studies. Digitization of such artifacts is an important step in their preservation, restoration, and wider accessibility. In this study, we concentrated on the digitization of stone inscriptions and manuscripts written in Nepal Script (aka Prachalit Nepal Script). This script is traditionally used for writing Nepal Bhasa (or Newari), a Tibeto-Burman language spoken by the Newar community of Nepal. However, despite its historical significance, the once abundantly used script is at risk of endangerment.

In an effort to support and revitalize Nepal Script, we have developed an OCR model specifically designed for recognizing Nepal Script texts. Our model is trained on our primary dataset, prepared through an extensive data acquisition method and supplemented with a secondary dataset. Dataset preparation involved expert consultations, field visits, photography, manual annotations for text segmentation, images-to-transcriptions mapping, and data verification by experts. The resulting dataset comprises 5,219 text line images, which were augmented to 41,752 images using various data augmentation techniques. Trained on this augmented dataset, our model achieved an average Character Error Rate (CER) of 12.61%. Its performance on regular handwritten and printed texts, and manuscripts (CER of 3.82% and 3.09% respectively) surpasses the 6.65% CER on handwritten and printed texts reported by Nakarmi et al. (2024) but approaches the 2.6% CER on manuscripts with consistent characters reported by O'Neill and Hill (2022) (they obtained higher CER for cruder form of the script, though the value is not explicitly specified). The highest CER (25.21%) was obtained on photographed images of stone inscrip-

[4]https://www.dabuli.org

tions, likely due to smudges, stone surface textures, and unclear/damaged characters. However, a smaller CER (18.31%) was achieved on the images of rubbing of stone inscriptions, where character outlines are clearer than in photographed images. Some possible directions for improving the OCR model could be to extend the current dataset, and use linguistic models to predict the eroded or unclear characters.

Despite being trained on images of stone inscriptions, manuscripts, and recent handwritten and printed texts, our model effectively recognized texts from copper plate and wooden inscriptions. This indicates generalization capability of our model across diverse artifact types, and highlights its potential contribution to heritage preservation and restoration.

Even though numerous significant challenges (e.g. lack of large, high-quality dataset, character variations across historical timeline, scarcity of experts, eroded artifacts, high cost of collecting data, etc.) exist, the potential cultural, academic, and technological benefits cannot be discounted. This combination of difficulty and impact shows why Nepal Script OCR is such a vital area for research and collaboration, and is worth pursuing.

7. Acknowledgements

This work was conducted under the Kathmandu Metropolitan Research Fellowship Program, funded by Kathmandu Metropolitan City, Nepal. We sincerely thank Ms. Rajju Hada from the National Archives of Nepal, Mr. Sarad Kasa from Aasha Safoo Kuthi, Mr. Ananda Kumar Maharjan, and Ms. Nitu Dangol from Callijatra Foundation, Mr. Pabitra Bir Singh Kansakar, Mr. Raja Bajracharya, and Mr. Anil Sthapit from Nepal Lipi Guthi, Dr. Rajan Lal Joshi, Mr. Jeevan Kumar Maharjan, and Prof. Omkareshwar Shrestha from Central Department of Nepal Bhasa, Tribhuvan University, as well as Mr. Alok Siddhi Tuladhar, Mr. Aashish Mishra, Mr. Rishi Govinda Amatya, Mr. Saurav Thapa Shrestha, and Mr. Shree Ram Khaitu for their valuable insights and resources, and also supporting us during the data refinement process. We extend our acknowledgement to the volunteers Anish Manandhar, Bigyan Byanju Shrestha, Jessica Thapa, Prayash Shakya, Shreela Sapkota, Shubhashish Karki, and Subani Malla for their time and effort in preparing the dataset.

8. Ethical Considerations

The primary data collection was conducted by the authors with voluntary participation from seven undergraduate students from the Department of Computer Science and Engineering at Kathmandu University, Nepal. The team visited various sites in and around the Kathmandu Valley to collect photographs of stone inscriptions and process the data required for training the text recognition model. For each site visit, the team obtained formal and oral permission from relevant authorities, including local committees, site caretakers, and community representatives. Even when inscriptions were located in publicly accessible areas, the team followed ethical data collection practices to preserve their integrity.

All data collection activities were conducted in accordance with the guidelines of the concerned authorities, using non-invasive methods to ensure the physical condition of the inscriptions remained unharmed. The process prioritized respectful documentation, balancing the need for high-quality imagery with the protection and sanctity of historical sites. The inscriptions primarily contain historical and religious records such as donor names, family members, rulers of specific periods, and texts dedicated to particular deities or sites. A review of the content confirmed that no sensitive personal, political, or religious information is present. The inscriptions were transcribed using the Unicode standard for Nepal Script, Newa (Unicode, Inc., 2023) to accurately reflect their cultural and linguistic context.

9. Data and Code Availability

To support the preservation of Nepal Script and facilitate further research, we make the stone inscription dataset used in this work publicly available along with the source code. The dataset is available on Kaggle as the Nepal Script Stone Inscription Dataset (kaggle.com/dsv/15121632) under CC BY-NC-SA 4.0 license. The source code for our implementation is available at github.com/svarnimn/nepal-script-text-recognition-from-ancient-artifacts under GPLv3 license.

10. Bibliographical References

Jen Bati and Pankaj Raj Dawadi. 2023. Ranjana script handwritten character recognition using cnn. *JOIV: International Journal on Informatics Visualization*, 7(3):984–990.

Alexey Dosovitskiy. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Kartik Dutta, Praveen Krishnan, Minesh Mathew, and CV Jawahar. 2018. Offline handwriting recognition on devanagari using a new benchmark dataset. In *2018 13th IAPR international workshop on document analysis systems (DAS)*, pages 25–30. IEEE.

Agam Dwivedi, Rohit Saluja, and Ravi Kiran Sarvadevabhatla. 2020. An ocr for classical indic documents containing arbitrarily long words. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 560–561.

Himalayan News Service. March 14, 2021. Kmc makes nepal bhasa mandatory in schools. *The Himalayan Times*. Accessed on: October 14, 2025.

Yuting Li, Dexiong Chen, Tinglong Tang, and Xi Shen. 2025. Htr-vt: Handwritten text recognition with vision transformer. *Pattern Recognition*, 158:110967.

Ajoy Mondal and CV Jawahar. 2022. Enhancing indic handwritten text recognition using global semantic information. In *International Conference on Frontiers in Handwriting Recognition*, pages 360–374. Springer.

Swornim Nakarmi, Sarin Sthapit, Arya Shakya, Rajani Chulyadyo, and Bal Krishna Bal. 2024. Nepal script text recognition using crnn ctc architecture. In *Proceedings of the 3rd Annual Meeting of the Special Interest Group on Under-resourced Languages@ LREC-COLING 2024*, pages 244–251.

Nepal Lipi Guthi. 1992. *Pracalit Nepāl Lipiyā Varṇamālā*. Nepal Lipi Guthi.

Alexander James O'Neill and Nathan Hill. 2022. Text recognition for nepalese manuscripts in pracalit script. *Journal of Open Humanities Data*, 8.

Anshuman Pandey. 2012. Proposal to encode the newar script in iso/iec 10646.

Ram Chandra Pandey, Babu Ram Dawadi, Suman Sharma, and Abinash Basnet. 2017. Dictionary based nepali word recognition using neural network. *Int. J. Sci. Eng. Res*, pages 473–479.

Ashok Kumar Pant, Sanjeeb Prasad Panday, and Shashidhar Ram Joshi. 2012. Off-line nepali handwritten character recognition using multilayer perceptron and radial basis function neural networks. In *2012 Third Asian Himalayas International Conference on Internet*, pages 1–5. IEEE.

Shankarman Rajbanshi. 1970. Kantipur shilalekh suchi [kathmandu stone inscriptions list].

Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788.

Bikash Shaw, Swapan Kumar Parui, and Malayappan Shridhar. 2008. Offline handwritten devanagari word recognition: A holistic approach based on directional chain code feature and hmm. In *2008 International Conference on Information Technology*, pages 203–208. IEEE.

Bal Gopal Shrestha. 1999. The newars: The indigenous population of the kathmandu valley in the modern state of nepal. *The Journal of Newar Studies*, 2:1–16.

Brijmohan Singh, Ankush Mittal, MA Ansari, and Debashis Ghosh. 2011. Handwritten devanagari word recognition: a curvelet transform based approach. *International Journal on Computer Science and Engineering*, 3(4):1658–1665.

Unicode, Inc. 2023. Newa Range: 11400-1147F. Accessed on: February 24, 2024.

Regent Nanda Vaidya and Bal Krishna Bal. 2024. Deep learning approach for offline nepali handwritten word recognition. In *International Conference on Analysis of Images, Social Networks and Texts*, pages 138–150. Springer.

Dhanavajra Vajracharya. 1999. Madhya kalka abhilekha [inscriptions of the medieval period].

John Whelpton. 2005. *A history of Nepal*. Cambridge University Press.