Received: 1 July 2022 | Revised: 31 August 2022 | Accepted: 6 September 2022 | Published online: 6 September 2022

Artificial Intelligence and Applications
2023, Vol. 1(2) 103–111
DOI: 10.47852/bonviewAIA2202293

RESEARCH ARTICLE

Region-Based Convolutional Neural Network for Segmenting Text in Epigraphical Images

Padmaprabha Preethi[1,*] and **Hosahalli Ramappa Mamatha**[1]

[1]*Department of Computer Science and Engineering, PES University, India*

Abstract: Indian history is derived from ancient writings on the inscriptions, palm leaves, copper plates, coins, and many more mediums. Epigraphers read these inscriptions and produce meaningful interpretations. Automating the process of reading is the interest of our study, and in this paper, segmentation to detect text on digitized inscriptional images is dealt in detail. Character segmentation from epigraphical images helps in optical character recognizer in training and recognition of old regional scripts. Epigraphical images are drawn from estampages containing scripts from various periods starting from Brahmi in the 3rd century BC to the medieval period of the 15th century AD. The scripts or characters present in digitized epigraphical images are illegible and have complex noisy background textures. To achieve script/text segmentation, region-based convolutional neural network (CNN) is employed to detect characters in the images. Proposed method uses selective search to identify text regions and forwards them to trained CNN models for drawing feature vectors. These feature vectors are fed to support vector machine classifiers for classification and recognize text by drawing a bounding box based on confidence score. Alexnet, VGG16, Resnet50, and InceptionV3 are used as CNN models for experimentation, and InceptionV3 performed well with good results. A total of 197 images are used for experimentation, out of which 70 samples are of printed denoised epigraphical images, 40 denoised estampage images, and 87 noisy estampage images. The segmentation result of 74.79% for printed denoised epigraphical images, 71.53 % for denoised estampage epigraphical images, and 18.11% for noisy estampage images are recorded by InceptionV3. The segmented characters are used for epigraphical applications like period/era prediction and recognition of characters. FAST and FASTER region-based design approach was also tested and illustrated in this paper.

Keywords: region-based CNN, text segmentation, support vector machine, Resnet50, FAST RCNN, FASTER RCNN

1. Introduction

Evolution of Indian history is derived from the study of ancient inscriptions. The written text on these inscriptions belong to various dynasties ruled. All Indian writings are derived from base script i.e., Brahmi script, dates to third century B.C and gradually scripts have changed to another form over a period. These scripts are perceived on the temple walls, copper plates, palm leaves, coins and on pots. The content on these above mentioned input is extracted and decoded by experienced epigraphers and paleographers. Traditional ways of decoding these inscriptions are by creating replica by stamping methods, interpretation, documenting them for future reference, study, transliteration, and publication. But finding such epigraphers is difficult as they are seldom and extraction of information from inscriptions takes a lot of time. Digitizing and building an automatic recognizer is the requirement for the current generation, as we need to pass this information to next generation. Despite research on inscriptions, reading and understanding has taken sneak peak, era prediction and recognition is still difficult because of the texture of input material, type of the input considered, materials using which the scripts are carved, and absence of experts. These challenges fascinated our research direction, and, in this paper, we are presenting segmentation of scripts by training region based convolutional neural network models to identify complete characters. The complete characters can be used for recognition/transliteration and prediction of the era.

In the recent days, many researchers are working on applying machine learning tools for applications like historical document cleaning, regeneration of washed-out characters, image identification, color regeneration, text reading, scene understanding, etc.

Character segmentation is to extract the scripts from the images and can be used later for identification. we have followed few methodologies such as segmenting the characters through k-nearest black neighbors connected to each other, watershed algorithm and basic machine learning algorithms were applied and connected components worked well for extracting characters from epigraphical/inscriptional images. The outcome of these above said methods had small non character segments such as prominent noise, partially cropped characters & protruding scratches and

*__Corresponding author:__ Padmaprabha Preethi, Department of Computer Science and Engineering, PES University, India. Email: ppreethijain@gmail.com

these are considered as the primary challenges to resolve in this experimentation. Paper describes the process of building classifiers, which are modeled to differentiate input images into –i) Noisy, ii) Partially cut characters as non-text and iii) Useful complete character for recognition.

Survey briefs about various segmentation procedures applied on epigraphical images, proposed method briefs about building input dataset, classification models built in detail. The fourth section describes results obtained and last section concludes segmentation results and at last scope of future work. Figure 1 to 3 shows the challenges associated with the collected images.

In this paper, our contribution includes:

Presenting model that automatically extracts script on the epigraphical images of ancient Kannada inscriptions. Our paper includes technology behind extraction, comparison results with ground truth, which was prepared manually with the help of epigraphists. Also results of the proposed method for various challenges and conclusions are discussed.

Figure 1
Uneven sized characters on document images

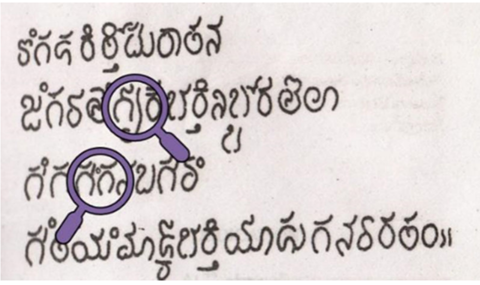
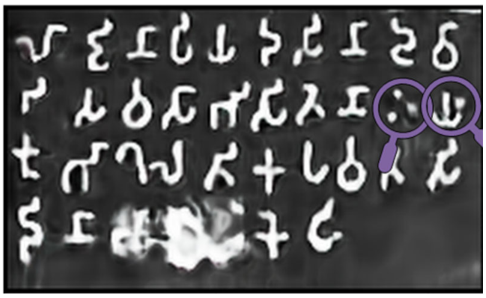


Figure 2
Overlapped characters

Figure 3
Multiple component characters



2. Literature Survey

Content extraction or script cropping plays a major role in building optical character recognizer as they are input. As the input considered in our proposed work is Historical documents, inscriptions, and Estampages, building recognizer is a big challenge due to the lack of ground truth, training and test samples. Many researchers have worked and published papers. Input images considered for their respective research are digital copies of epigraphical images of various centuries belonging to south India-Karnataka, and character spotting from these images for the purpose of era prediction and recognition remains an unresolved problem statement. This portion of the paper will review literature on character segmentation from epigraphical images, challenges considered, and possible techniques applied in the domain of document image segmentation.

2.1. Segmentation using traditional algorithms

As the document images are treated as graph representation and scripts as components, graph based connected component extraction methodologies are explored.

Likforman-sulem et al., (2007) in their paper explains the importance of epigraphy and also details about the research outcome in this field. For the Arabic document, author applied noise removal preprocessing methods and text extraction procedure like projection profile, drawing a bounding box to recognize the area of the text using connected components. The research revolves around 16^{th} century data and has good quality.

Murthy and Kumar (2004) proposed nearest neighboring clustering for extracting text from the Epigraphical images. Input for the proposed method is of printed scripts with varying spaces, skewness in the text writing and touching characters. Authors describe characters, words, and lines segmentation. In Bhat and Balachandra Achar (2016), connected component algorithm was used for segmentation of input image into lines and character segmentation. The segmented text will be later used for Era identification (to recognize the dynasty) and the recognition (to recognize the text and content written.

In Zirari et al., (2013), segmentation of characters uses connected component algorithm and this algorithm proves to work on noiseless images. In Manigandan et al., (2017) also applies connected component algorithm on binary input. In Sowmya and Kumar (2015), works with drop fall and water reservoir approach to segment the text in printed records having varying degradations. Author confirms the working of algorithm on preprocessed images.

Feature detectors are helpful in spotting characters and text content in the images, below review updates the behavior of super pixels and detectors.

Hu et al., (2017) extracts maya codices from noisy historical images using region based segmentation. A binary support vector machine was trained to classify background and foreground pixels of the input image. To strengthen the label consistency - connected conditional random field model in the pipeline was added. The complete model will extract the characters from the maya script images using super pixels.

A novel segmentation algorithm was designed on Greek inscription. simple linear iterative clustering (SLIC) super pixel with region merging was adopted to work on geometry of the surface. Based on the uninscribed surface, strokes and breaks on rock surface points are classified in Sapirstein (2019).

Chen et al. (2016) build support vector machine which is a multiple classifier classifies four classes such as background, text, boundaries and margin using super pixels. Based on the intensity values of the super pixels model discriminate between the classes.

Mohana et al. (2014) propose text removal from inscriptions. A method applied includes cropping character, preprocessing images using morphological operators to highlight script and finding scale invariant feature transform features. The extracted reference image is compared to find structural similar characters in the input image. A reference image is matched based on the number of foreground and background pixels matched. The similarity calculates hit and miss count of the pixels matched while scanning of foreground pixels as it contains text. Threshold values is customized to count matching characters in the stone inscriptions. The method proves to work on clean epigraphical images with few text on it and achieved result recorded by the author in the research paper is 88% accuracy.

2.2. Segmentation using machine learning models

Machine learning models are widely used due to availability of high-end machines and cloud support in solving real world problems, the below review details about the use of machine learning/deep learning approaches in solving epigraphical domain problems.

Kavitha et al. (2016) proposed a method to segment historical Indus input with degradation. Experimentation includes preprocessing method for the removal of noise using hybrid laplacian and sobel filters followed by skeletonization. Initially nearest neighboring based clustering method was used to extract closed proximal components and then two clusters are formed to identify script and background using features of text. Authors records 90% as the result of this segmentation for the printed historical document belongs to 16th-20th century images.

Chandrakala et al. (2019) considers 11th century digitized kannada epigraphical script for recognition using deep neural networks. Features extracted from deep neural networks are fed to multiclass classification methods such as stochastic gradient decent multiclassifier (SGDM) and SVM for recognition. The author also claims intra class variability of handwritten characters causes bottleneck in recognizing characters to map into modern kannada script.

Alberti et al. (2017) proposed a deep neural network and initializing the weights using linear discriminant analysis. To achieve high performance, these weight values are initialized layer wise with the intension of quick convergence. The transformation matrix are used to separate text and background regions. Method was tested medieval period dataset having 150 pages of scanned images. Author has experimented on narrow layered architecture of CNN.

Badjatiya et al. (2018) proposed a bidirectional long short-term memory an RNN model to segment sentence embedding. The text segments are recognized from noiseless documents.

Jo et al. (2020) have implemented a CNN that uses handwritten samples for training and segments scripts in the document based on the features. The authors' claim is that the segmentation accuracy achieved lies between 75% to 90%. In AI-Rawi et al. (2019), the author describes use of GAN to extract text automatically using unsupervised text segmentation. The Generator model and Discriminator models helps in recognizing text and non-text. The data set considered are of scene images having text with good accuracy and speed. In Reza et al. (2019), Invoices are the input to segmentation, where text content is segmented using conditional generative adversarial network having challenges like table area localization and skewness. The design of the model has Signet architecture for encoder and decoder with skip connections for GAN.

In Li et al. (2017) character segmentation was considered under complex background and model classifies the segmentation into characters and just spaces. In Abtahi et al. (2015) segmentation module was designed based on reinforcement learning. Input the model is projection profiles extracted from input having only two text lines with equally spaced text.

These reviews listed varied from traditional computer vision approach, super pixel based method to machine learning models in segmenting characters in images of text documents with various processing challenges. The challenges addressed are spacing between the characters, skewness, noise present, background, script information etc. In this paper, we propose the Region based convolutional neural network to use as a classifier in classifying the script on the images.

3. Proposed Methodology

Epigraphical images considered for experimentation are printed document and estampages of various periods from 3rd century to 15th century data having south Indian Kannada scripts and are collected from "The Archaeological Survey of India (ASI), Mysore, Karnataka." The images are naturally noisy; hence, preprocessed images are considered for segmentation, and method adopted for cleaning images is discussed detail in paper Preethi et al. (2019).

Epigraphical text is harder to segment into letters due to varying and uneven character dimensions. Connected component algorithm uses graph traversal technique, which works well on image segmentation, but generates image segments having both text and non-text. Non-text includes prominent noise, cuts, and incomplete text. Feeding the output of connected component technique directly to recognition and prediction will yield poor results.

3.1. Generation of dataset for classifier models

Traditional method of applying image processing steps such as converting input image into gray scale followed by edge detection filter and morphological operations with dilation and erosion followed by finding the centroid of the contours would help us in finding the region of interest (ROI) in the given images. This technique would work if the image were complete and of medical images, satellite images, and scenic images. Traditional method fails when document images having characters connected, skewed, and disconnected parts of meaningful characters are fed as input image. Since the interest of the proposed method is to have segmentation tool for epigraphical images, with the help of trained epigraphist, we manually created the ground truth or ROI for the benefit of segmentation.

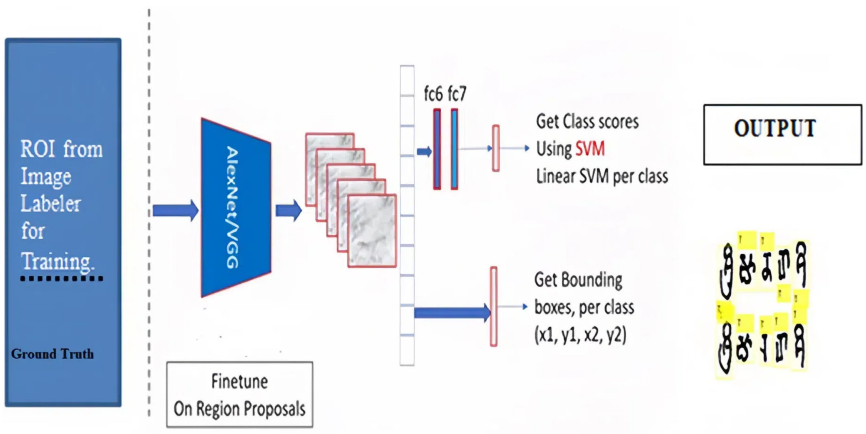
A rectangular ROI having the two top left and bottom right coordinates is extracted for all training images. The training images are grouped as printed, denoised inscribed, and noisy inscribed epigraphical images. Among 70 printed epigraphical images, 40 images are used for training and 30 images are used for testing. Among 40 denoised inscribed epigraphical images, 25 are used for training and 15 images are used for testing. Out of 87 noisy epigraphical estampage images, 50 images are used for training and 37 images are used for testing. For each training sample, a rectangular kernel is drawn around every character in the image and labeled as a character class. CNN must have a proportional training sample to perform well and to avoid underfitting and overfitting, augmentation methods are applied. Augmentations such as flipping the image, adding some skewness, varying size, changing illumination, and rotating to some angles are applied to training samples, and CNN is invariant to all these augments.

3.2. Region-based CNN

In our early study, a CNN was implemented to segment characters on the input images. The network was trained on images having complete text and non-text cropped from sliding window algorithm (of the varying kernel) from various epigraphical images. Kernel slides over input image from top left boundary to bottom right boundary to capture regions of the image. Test image generates a huge set of input images from sliding window procedure, and model will classify cropped regions into text and non-text. The classified texts are used for recognition and era prediction. A total of 150 input images were considered for the experimentation with various challenges. The outcome was appreciable with some drawbacks. The model took a lot of time to process all the generated thousands of kernels, and outcome had multiple copies of the same-segmented characters.

To overcome the above-said model, we adopted region-based CNN in our study. First, model generates proposed regions using selective search method. Secondly, CNN extracts feature vectors, and finally, design of a linear support vector machine will classify into text. Figure 4 depicts segmentation model used in the paper.

Figure 4
Region-based CNN model for text detection on epigraphical images



3.2.1. Selective search

Selective search proposes approximately 2000 regions per test image to use in feature extraction as regions describe more features than individual pixels. Identify initial regions using graph-based method as described in Susmaga (2004), which provides high-quality starting points. The two regions are combined using a greedy approach based on the similarity measures of the neighboring regions. This process repeats for each neighboring regions in the image being similar. The similarity measure is calculated based on the color, texture, size, and fillings. These measures are in the range of [0,1]. Paper (Felzenszwalb & Huttenlocher, 2004) describes selective search method and its similarity estimation in detail.

3.2.2. Region-based CNN

Characters in input images are segments/objects expected, and every input image has variable number of such characters of varying size, and hence, standard convolutional network cannot directly used for segment detection.

Fully connected CNNs are trained on training images with corresponding ground truth for extracting features from input samples. CNN extracts feature vectors for all proposed regions obtained by selective search method. Experimentation used implementation of Alexnet described by Krizhevsky et al. (2012) and VGG16 Simonyan & Zisserman (2015); however, we have also experimented using Resnet50 described by Szegedy et al. (2016) and Inception V3 presented in Szegedy et al. (2017).

Input size for Alexnet is 227×227, VGG16, and Resnet50 use input size as 224×224, and Inceptionv3 has an input size of 299×299. All proposed regions are wrapped along with a bounding box to required input size of the network used. Wrapped regions are forward propagated from first layer to the penultimate layer before the softmax classifier to extract feature vectors. Alexnet and VGG16 have feature vectors of dimension 4096, and Resnet50 and InceptionV3 are of size 2048.

Proposed methods use stochastic gradient descent training with two classes' text and background for classification. All region proposals with ≥ 0.7 intersection over union (IOU) overlap with ground truth are true positives, and the rest all are considered as negatives. The learning rate is initialized to 1e-6 with minibatch size as 32 positive windows and maximum epochs set as 10 for fine-tuning. Experimentation was conducted on "Google Colab" environment with 70, 40, and 87 images from denoised printed epigraphical images, denoised epigraphical images, and noisy estampage images, respectively.

3.2.3. Text classifier using support vector machine

Training a binary classifier to detect text and background is the motive of using SVM classifier. SVM classifies text and non-text by taking input from selective search and feature vectors from CNNs. Segment of image having text within the bounding box is a positive region, and background is a negative region. If region proposal has partially cut scripts/texts or even noise, we use IOU to define negative regions.

Overlap threshold is defined, and in our study, we have tuned it to 0.7, and this threshold is carefully selected by trial-and-error method. Positive examples are grouped as equal to ground truth images. SVM classifiers refine all the region proposals obtained using selective search and training data from CNNs.

3.2.4. Evaluation metric

While training, IOU as specified in the ground truth for character marked bounding box is checked against predicted bounding box. If IOU is exactly equal to 1 indicates, bounding boxes are exactly matching. IOU is defined below in the equation 1.

$$IOU = \frac{Common\ area\ of\ two\ bounding\ box}{Combined\ area\ of\ the\ bounding\ box} \tag{1}$$

Threshold value is set to classify the predicted boxes into true positive and false positives. In our study, a threshold of 0.7 was selected by trial-and-error method, and hence, ≥ 0.7 are true positives, anything < 0.7 are wrongly detected false positives. Model is unable to identify script/text defined in the ground truth while training is false negative. For object detection/identification of text/script, true negatives are not required as it points to background in the given image.

Confidence score is the probability indicating the bounding box containing script/text, and classifier contributes to this confidence score. The same threshold of 0.7 is used to display segmented text or identified text from input testing sample.

3.2.5. Variants of RCNN

To improve segmentation rate, other variants of RCNN were designed and experimented, FAST RCNN and FASTER RCNN. In FAST RCNN, epigraphical script image is directly fed to CNN models to generate feature map than feeding region proposals. These feature maps help in identifying region of proposals. By ROI pooling layer, we feed it to fully connected layer after reshaping. With the help of ROI feature vector, softmax layer is designed to predict the ROI and to draw bounding box. This method is faster than RCNN as image is parsed only once to generate the feature map.

RCNN and FAST RCNN used selective search as a time-consuming approach (a greedy method). To overcome that, FASTER RCNN was designed. Input to FASTER RCNN is an epigraphical script image to fetch convolutional feature map, and a separate network model is designed to predict the region proposals. The predicted regions are used to classify and predict values of bounding boxes.

As a network model is designed to fetch only the ROI, the data samples for classification model are comparatively less in number than RCNN and FAST RCNN and hence faster in computation.

4. Results and Discussion

In this section, we discuss about model performance for variants of epigraphical images.

Figure 4 depicts the proposed system and its outcome for printed epigraphical scripts. Training input consists of ROI drawn for every character on the epigraphical images. ROI is a rectangular bounding box with coordinates of two extreme corners per one character built with the help of image labeler. ROIs are stored in a datastore, and a sample copy is depicted in Figure 5. CNN feature vectors are fed to SVM classifiers for confidence score. Based on the confidence score and bounding box regression, a bounding box is drawn around character detected. The yellow rectangular box with caption "T" indicates an identified text in the image.

Figure 5
Sample region of interest as ground truth

Variables - T

Prercnn | score | T

	1 Filename	2 ROI
1	'D:\PHD\Segmentation\RCNN\Document_images\forRCNN\b33.jpg'	[35,50,58,75]
2	'D:\PHD\Segmentation\RCNN\Document_images\forRCNN\b33.jpg'	[94,33,60,87]
3	'D:\PHD\Segmentation\RCNN\Document_images\forRCNN\b33.jpg'	[174,45,50,81]
4	'D:\PHD\Segmentation\RCNN\Document_images\forRCNN\b33.jpg'	[246,47,55,81]
5	'D:\PHD\Segmentation\RCNN\Document_images\forRCNN\b33.jpg'	[326,34,57,89]
6	'D:\PHD\Segmentation\RCNN\Document_images\forRCNN\b33.jpg'	[392,29,60,101]
7	'D:\PHD\Segmentation\RCNN\Document_images\forRCNN\b33.jpg'	[465,25,46,98]
8	'D:\PHD\Segmentation\RCNN\Document_images\forRCNN\b33.jpg'	[522,34,71,86]
9	'D:\PHD\Segmentation\RCNN\Document_images\forRCNN\b33.jpg'	[608,37,45,87]
10	'D:\PHD\Segmentation\RCNN\Document_images\forRCNN\b33.jpg'	[661,35,49,93]
11	'D:\PHD\Segmentation\RCNN\Document_images\forRCNN\b33.jpg'	[704,13,48,81]
12	'D:\PHD\Segmentation\RCNN\Document_images\forRCNN\b33.jpg'	[40,164,49,72]
13	'D:\PHD\Segmentation\RCNN\Document_images\forRCNN\b33.jpg'	[98,157,46,113]

Segmentation model outperformed when InceptionV3 model was used as CNN model. All the outputs depicted here are of InceptionV3 model. Figure 6 shows output of printed epigraphical scripts, Figure 7 shows output of denoised estampage images, and Figure 8 shows output of noisy estampage images. Figure 9 demonstrates the challenges addressed and future scope of work in segmentation of epigraphical images.

Figure 6
2[a, b, c]: Output of segmentation on printed epigraphical scripts (a) Brahmi script, (b) Hoysala script, and (c) Kadamba script
(Image courtesy: The ASI, MYSORE)

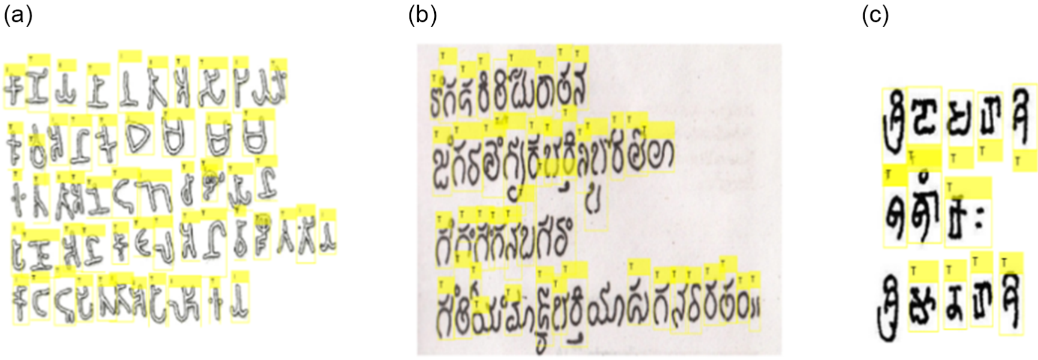


Figure 7
3[a, b, c]: Output of epigraphical scripts from (a) Brahmi, (b) Elongated Brahmi, and (c) Rashtrakuta Dynasty
(Image courtesy: The ASI, MYSORE)

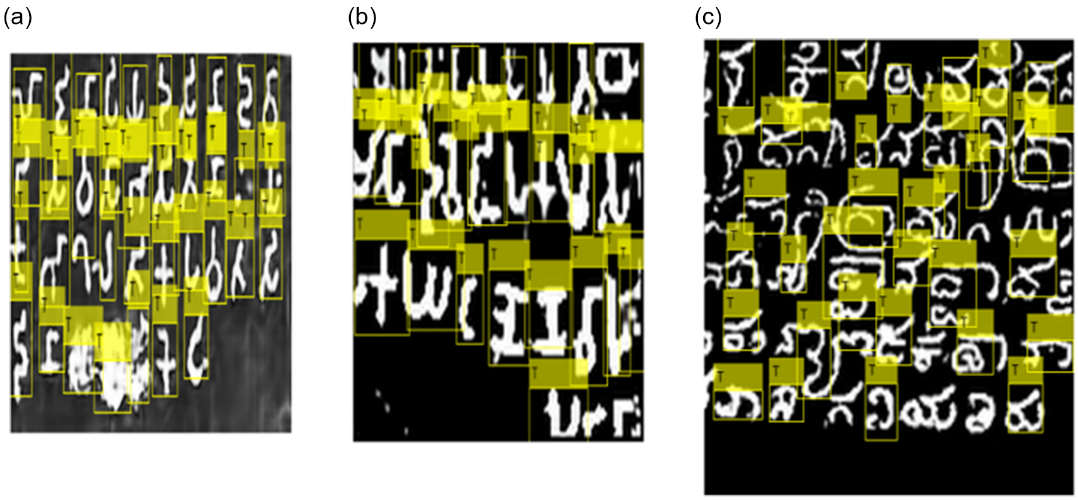


Figure 8
Red circle shows identified complete text having multiple components, and blue circle shows prominent noise also addressed as text
(Image courtesy: The ASI, MYSORE)

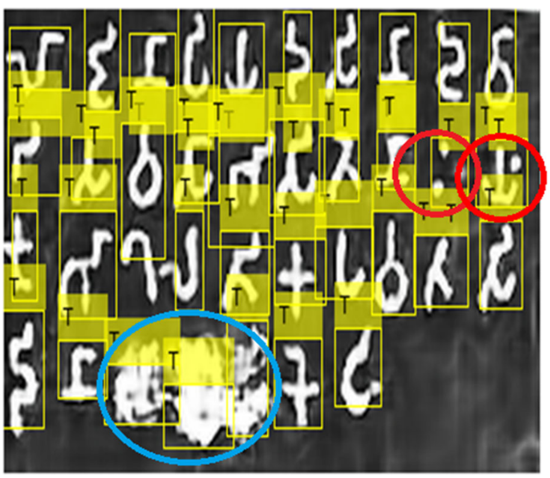
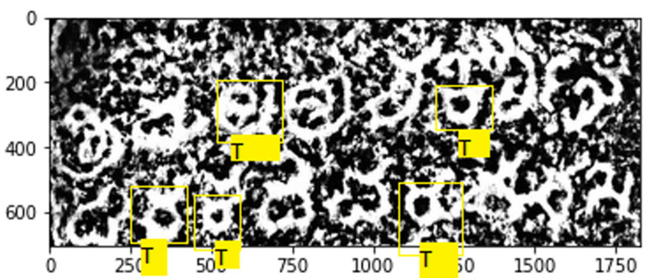


Figure 9
Character identification of estampage images having prominent texture noise
(Image courtesy: The ASI, MYSORE)



Variants of RCNN model like FAST and FASTER RCNN were also tried and tested on epigraphical script images. The outcome of these networks showed less segmentation rate than RCNN as depicted in Table 2.

RCNN performed well on denoised epigraphical scripts compared to FAST and FASTER RCNN. As per the analysis,

Table 1
Comparison of Alexnet, VGG16, and Resnet50 on various epigraphical images

MODEL	Segmentation rate			
	Alexnet	VGG 16	Resnet 50	Inception V3
Printed epigraphical images	68.2%	70.9%	74.56%	74.79%
Denoised epigraphical images	38.72%	47.3%	71.53%	71.53%
Noisy epigraphical images	9.16%	14.21%	17.8%	18.11%

Table 2
Segmentation rate recorded by variants of RCNN on printed epigraphical script images

Models	Accuracy				Comments
	Alexnet	VGG 16	Resnet 50	Inception V3	
RCNN	58.2%	70.9%	74.56%	74.79%	Slow and uses custom regions Prediction time: ~49 s
FASTRCNN	54.31%	57.2%	53.1%	62.41%	Uses custom regions Prediction time: ~4 s
FASTER RCNN	51.04%	54%	49.43%	57.92%	Fast but has its own anchor boxes Prediction time: ~1 s

Table 3
Comparative study of existing methods

Author	Segmentation method	Input type	Size of the dataset	Segmentation rate (%)
Murthy and Kumar (2004)	Partial eight direction-based line segmentation and nearest neighbor clustering	Epigraphical documents	Not specified	Not specified
Sowmya and Kumar (2015)	Drop fall and water reservoir techniques	Enhanced document historical images	150 samples	85–90% segments sample characters.
Sridevi and Subhashini (2012)	Connected components with nearest neighbor	Enhanced ancient Tamil inscriptions	Not specified	82%
Proposed method	Segmentation using Region-based CNN (InceptionV3)	Printed document epigraphical script	43	74.8%
		Printed epigraphical estampage images	47	71%
		Raw estampage images	144	18.79%

RCNN searches for characters exhaustively for each possible region in an image, whereas FAST and FASTER RCNN process input image randomly once in search of region proposed.

The probability of identifying the characters on the image will be less in FAST and FASTER CNNs compared to the native region-based CNN method. Hence, time taken by the region-based CNN is more than the other two methods.

A total of 197 images are considered for the experimentation collected from "The ASI, Mysore" and images from "Indian Council for Historical Records, Bangalore." The CNN Architectures used are Alexnet, VGG 16 Resnet 50, and InceptionV3. Model segmentation rate is calculated based on 60:40 input split, and all the models ran for 100 epochs. The results are as depicted in Tables 1 and 2.

InceptionV3 model performed well on denoised epigraphical images irrespective of background. On noisy estampage images, all the networks showed poor results.

We compare our proposed method with the existing methods for character segmentation and show as Table 3. Type of dataset used for the experimentation are document noiseless images and the models records an accuracy of 85%.

Kannada characters are evolved from 3rd century BC to 15th century AD to its present form of usage. The information written using the scripts of 3rd century BC needs expert hands to decipher, and there is no similarity between the characters of 3rd century scripts to 5th, 6th, 9th, 11th, and 15th century AD script. Before assigning epigraphers to read/decipher specific inscription, it is important to find first to which century inscription belongs to. A classifier model is designed to identify century for which input must be the characters present on the inscriptional image. To retrieve these characters, segmentation module is designed. Segmented characters from RCNN are used to feed multiclass classifier designed to predict the era of these scripts. Based on the period prediction, inscriptions are further processed for decipherment. A recognition module trained with all possible characters of the period also requires segmented characters for identification and transliteration. These requirements demand the need of segmentation module.

5. Conclusion and Future Work

Epigraphical study reveals about the ancient culture, life, findings, prosperity, and many more. In the world of digitization, archeological department follows old procedure to manually read, store, and publish information on the inscriptions, which takes a lot of time and resources. Many researchers in India and abroad have taken up this challenge and trying to digitize monuments, sculptures, inscriptions, coins, and they are successful too. In our research study, we are working toward the automatic transliteration of ancient Kannada language written on epigraphical images into modern Kannada language. In building such a model, segmentation plays a major role, and, in this paper, we have used object detectors to detect script/text on the input image, used for recognition and prediction of the era. From the experimentation, region-based CNN can be used to classify/detect text from images. The images considered are of three types, and they are printed denoised epigraphical images, images of denoised estampage, and noisy inscriptional estampage images. As discussed in the result section, InceptionV3 model used in classification modeling has outperformed than Alexnet, VGG16, and Resnet50. The computational time for finding the text on the images was around 52 s. A lot of time is consumed in training the network in Inception V3 model.

Future work is to improve segmentation on noisy epigraphical images and to train a new model to work on all kinds of images. The efficiency of segmentation contributes to predicting the period (era) of the inscriptions and in transliteration to readable script. Future work is to collect scripts from different centuries of Kannada language and to recognize them on the newly found inscription within minimal time.

Acknowledgement

We thank Archaeological Survey of India, Mysore, for providing dataset for the project which helped us to present our findings.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

References

Abtahi, F., Zhu, Z., & Burry, A. M. (2015). A deep reinforcement learning approach to character segmentation of license plate images. In *Proceedings of the International Conference on Machine Vision Applications*, 539–542.

Al-Rawi, M., Bazazian, D., & Valveny, E. (2019). Can generative adversarial networks teach themselves text segmentation. In *IEEE/CVF International Conference on Computer Vision Workshop*, 3342–3350.

Alberti, M., Seuret, M., Pondenkandath, V., Ingold, R., & Liwicki, M. (2017). Historical document image segmentation with LDA-initialized deep neural networks. In *Proceedings of the International Workshop on Historical Document Imaging and Processing*, 95–100.

Badjatiya, P., Kurisinkel, L. J., Gupta, M., & Varma, V. (2018). Attention-based neural text segmentation. In *Proceedings of the European Conference on Information Retrieval*, 180–193.

Bhat, S. S., & Balachandra Achar, H. V. (2016). Character recognition and Period prediction of ancient Kannada Epigraphical scripts. *In National Conference on Recent trends in Technology*, *3*.

Chandrakala, H. T., & Thippeswamy, G. (2019). Deep convolutional neural networks for recognition of historical handwritten kannada characters. In *Frontiers in Intelligent Computing: Theory and Applications*. Advances in Intelligent Systems and Computing, 1014.

Chen, K., Liu, C., Seuret, M., Liwicki, M., Hennebert, J., & Ingold, R. (2016). Page segmentation for historical document images based on superpixel classification with unsupervised feature learning. In *Proceedings of the IAPR Workshop on Document Analysis Systems,* 299–304.

Felzenszwalb, P. F., & Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. International Journal of Computer Vision*,* 59, 167–181.

Hu, R., Odobez, J.-M., & Gatica-Perez, D. (2017). Extracting maya glyphs from degraded ancient documents via image segmentation. *ACM Journal of Computational. Cultural. Heritage*, *10*, 71–93. https://doi.org/10.1145/2996859

Jo, J., Koo, H. I., Soh, J. W., *et al.* (2020). Handwritten text segmentation via end-to-end learning of convolutional neural networks. *Multimedia Tools Applications*, *79*, 32137–32150. https://doi.org/10.1007/s11042-020-09624-9

Kavitha, A. S., Shivakumara, P., Kumar, G. H., & Lu, T. (2016). Text segmentation in degraded historical document images. *Egyptian Informatics Journal*, *17*, 189–197. https://doi.org/10.1016/j.eij.2015.11.003

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, *6*, 1097–1105.

Li, X., Zhang, X., Yang, B., & Xia, S. (2017). Character segmentation in text line via convolutional neural network. In *Proceedings of the International Conference on Systems and Informatics*, 1175–1180.

Likforman-Sulem, L., Zahour, A., & Bruno, T. (2007). Text line segmentation of historical documents: A survey. *International Journal on Document Analysis and Recognition*, *9*, 123–138. https://doi.org/10.48550/arXiv.0704.1267

Manigandan, T. V. V. D. V. N. B., Vidhya, V., Dhanalakshmi, V., & Nirmala, B. (2017). Tamil character recognition from ancient epigraphical inscription using OCR and NLP. In *Proceedings of the International Conference on Energy, Communication, Data Analytics and Soft Computing*, 1008–1011.

Mohana, H. S., Navya, K., Rajithkumar, B. K., & Nagesh, C. (2014). Interactive segmentation for character extraction in stone inscriptions. In *Proceedings of the International Conference on Current Trends in Engineering and Technology*, 87–96.

Murthy, K. S., & Kumar, H. (2004). Nearest neighbouring clustering based approach for line and character segmentation in epigraphical scripts. In *The Proceedings of International Conference on Cognitive Systems*, 132–140.

Preethi, P., Praneeth Kumar, K., Sumukha, M., & Mamatha, H. R. (2019). Denoising epigraphical estampages using nested run length count. In: *Ubiquitous Communications and Network Computing. UBICNET* 2019. Lecture Notes of the Institute for

Computer Sciences, *Social Informatics and Telecommunications Engineering*, 198–206.

Reza, M. M., Bukhari, S. S., Jenckel, M., & Dengel, A. (2019). Table localization and segmentation using GAN and CNN. In *Proceedings of the International Conference on Document Analysis and Recognition Workshops*, 152–157.

Sapirstein, P. (2019). Segmentation, Reconstruction, and Visualization of Ancient Inscriptions in 2.5D. *Journal on Computing and Cultural Heritage*, *15,* 1–30. https://doi.org/10.1145/3286977

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *Proceedings of International Conference on Learning Representations*, 1–16.

Sowmya, A., & Kumar, H. (2015). Enhancement and segmentation of historical records. In *Proceedings of the International conference on CS and IT*, 5(13).

Sridevi, N., & Subhashini, P. (2012). Segmentation of text lines and characters in ancient tamil script documents using computational intelligence techniques. *International Journal of Computer Applications*, *52*, 7–12. https://doi.org/10.5120/8268-1826

Susmaga, R. (2004). Confusion matrix visualization. In *Intelligent Information Processing and Web Mining*. *Advances in Soft Computing*, 107–110.

Szegedy, C., Loffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-V4, inception-ResNet and the impact of residual connections on learning. In *Proceedings of Association for the Advancement of Artificial Intelligence*, *4*, 4278–4284.

Szegedy, C., Vanhoucke, V., Loffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2818–2826.

Zirari, F., Ennaji, A., Nicolas, S., & Mammass, D. (2013). A document image segmentation system using analysis of connected components. In *Proceedings of International Conference on Document Analysis and Recognition*, 753–757.

How to Cite: Preethi, P. & Mamatha, H. R. (2023). Region-Based Convolutional Neural Network for Segmenting Text in Epigraphical Images. *Artificial Intelligence and Applications* 1(2), 103–111, https://doi.org/10.47852/bonviewAIA2202293