Int. J. Exp. Res. Rev., Vol. 35: 34-42 (2023)

Original Article

Peer Reviewed

(a) Open Access



International Journal of Experimental Research and Review (IJERR) © Copyright by International Academic Publishing House (IAPH) ISSN: 2455-4855 (Online) www.iaph.in



A Hybrid Approach for Complex Layout Detection of Newspapers in Gurumukhi Script Using Deep Learning

Atul Kumar^{1*} and Gurpreet Singh Lehal²

Check for updates

¹Department of Computer Science, R.G.M. Govt. College Joginder Nagar, Mandi, Himachal Pradesh, India; ²Department of Computer Science, Punjabi University, Patiala, Punjab, India

E-mail/Orcid Id:

MK, 🕲 atulkmr02@gmail.com, 💿 https://orcid.org/0000-0002-7665-1892; GSL, 🕲 gslehal@gmail.com, 💿 https://orcid.org/0000-0001-6152-8050

Article History:

Received: 30th Aug., 2023 Accepted: 06th Nov., 2023 Published: 30th Nov., 2023

Keywords:

Deep Learning, CNN, Gurumukhi, Layout, Newspapers

How to cite this Article: Atul Kumar and Gurpreet Singh Lehal (2023). A Hybrid Approach for Complex Layout Detection of Newspapers in Gurumukhi Script Using Deep Learning. International Journal of Experimental Research and Review, 35(Spl.), 34-42. DOI : https://doi.org/10.52756/ ijerr.2023.v35spl.004

good layout analysis results in better recognition results. The complexity of newspaper layout structures poses a formidable challenge in digitization. The intricate arrangement of text, images, and various sections within a newspaper demands sophisticated algorithms and techniques for accurate layout detection. The paper introduces a diverse set of methodologies from existing literature, highlighting the evolution of techniques for newspaper layout analysis. In this paper, we present a novel method to detect the complex layout of newspapers in the Gurumukhi script by using a hybrid approach. The method developed consists of two parts. In the first part, we proposed an algorithm to remove pictures and graphics from Punjabi newspaper images that involve various image preprocessing tasks based on binarization, finding contours, and erosion on the image to remove the graphics from the image. This method removes pictures from complex non-Manhattan layouts. We have tested this algorithm on 100 newspaper images, giving an accuracy of 96.22%. In the second part, a dataset of 500 newspapers was created with images labeled with five classes on which the model was trained. Finally, we have trained the deep-leaning model based on a convolutional network to detect the columns of text in newspapers. We have used four different architectures of CNN and compared their performance based on different metrics such as precision, recall, and F1 score. We have tested this method on a number of newspapers in the Gurumukhi script. We have achieved an accuracy of 95.53% with this approach.

Abstract: Layout analysis is the crucial stage in the recognition system of newspapers. A

Introduction

digitization has gained Newspaper significant importance in today's context, addressing the need to make newspapers more accessible and user-friendly. With newspapers being a vital information source, the challenge lies in efficiently accessing specific newspapers and the desired information within a vast collection of old newspapers. Therefore, the essential requirement for newspaper digitization arises, enabling users to retrieve information from their own workspace. The first step in the digitization process is to conduct a layout analysis of newspapers. As the newspaper is quite complex, detecting the correct layout is challenging. We have done

the layout analysis of a newspaper written in the Punjabi language. The primary factor driving this initiative is the predominant use of other Gurumukhi scripts for many reasons. Gurumukhi serves as the primary script for Punjab, where it is not only the first language but also ranks as the world's most spoken language. Hence, due to the widespread readership of newspapers in the Gurumukhi script, the conversion of old newspapers into digital formats becomes imperative. This transformation facilitates the accessibility of historical newspaper content on computers and other digital platforms.

Various methodologies have been proposed in the literature to detect the newspaper's layout. Chowdhury et

^{*}Corresponding Author: atulkmr02@gmail.com



Int. J. Exp. Res. Rev., Vol. 35: 34-42 (2023)



Figure 1. Workflow of the proposed system

al. (2007) proposed a technique to separate graphics by labelling components and combining components with similar features. Patil et al. (2022) created an approach based on pixel-based classification to find segment images from the text in documents. Alshameri et al. (2012) combined two algorithms-the run length smoothing algorithm and SVM algorithms with AND and OR to get a layout analysis of the image. Moreover, a dynamic horizontal projection is applied to the result based on the dynamic threshold. The benchmark dataset DocBank, which has 500K document pages and finegrained annotated images for document layout analysis, was proposed by Li et al. (2020). Vasilopoulos and Kavallieratou (2017) used morphological methods and contour tracing to detect the complex layout of documents. Wu et al. (2023) proposed a dynamic fusion network based on features for layout analysis, giving an F1 score of 89.5% on the DSSE dataset and 95.1% on the CS-150. Kosaraju et al. (2019) proposed a multiclass classifier to segment documents into various components using a convolutional neural network. Bin Makhashen and Mahmoud (2019) surveyed various layout analysis techniques, features, advantages, and disadvantages. Biswas et al. (2021) used an object recognition mask RCNN to detect and predict regions of interest. Using straightforward image processing techniques like morphological dilation, run-length smoothing algorithm,

Soma Shilpa rule-based algorithm, and and (2022) offered a novel method to identify and segment blocks discovered within the newspaper regardless of layout. Zulfiqar et al. (2019) proposed a method to detect layout using the LSTM network, giving an accuracy of around 96%. Zhu et al. (2022) experimented with various algorithms to detect the layout of documents and created a dataset of 3000 newspaper pages in US states with various complex layout structures. Oliveira et al. (2018) proposed an open-source CNN architecture for layout segmentation with post-processing to improve accuracy. Xu et al. (2021) proposed a neural network built on the Mask **R-CNN** architecture to produce object classification, bounding box identification and the creation of page object masks. Alzubaidi et al. (2021) provided a detailed overview of various deep learning approaches, including CNN architectures and their uses and characteristics. Liebl and Burghardt (2021) evaluated various architectures for layout analysis of newspapers.

In this paper, we mainly proposed an algorithm to remove the graphics and pictures from newspapers in the Gurumukhi script. After that, we trained the newspaper images on a convolutional neural network by creating the dataset and segmenting the different columns of newspapers. The next parts of this paper mainly cover Material and Methods in Section 2, Results and Discussion in Section 3, and Conclusion in Section 4.

shitalakhya Launch capsize Cargo driver, others remanded as death toll reaches 8

A harayangan (our has granted the poloe three days to question eight people, including the drive ef clargo vessel "MV Raphs)" tab. The bodies of agits people have to faint and the second second second haraway and the second second second acceleration of the second second second haraway and the second s

of eight passengers of the launch





(c)

shitalakhyalaunch capsize Cargo driver, others remanded as death toll reaches 8

Shamina R.A., Narayangan	
A Response journ has parel to pick their also is a parel of the pick the subset of the subset of the subset of the subset of the subset of the sub	
of eight pavengets of the laugh	Answers resident the trady of a station of the bareh that same in the Sold database store in terry organization



Figure 2. (a). Original Image (b) Result of the proposed algorithm after separating text from graphics (c) Original Image in Non-Manhattan layout (d) Result of the proposed algorithm after separating text from graphics

Materials and Methods

The workflow of the system has two parts. In the first part, the newspaper image is passed through an algorithm based on contours to remove the graphics from the newspaper images. In the second part, we collected around 500 images of Punjabi newspapers, performed data preprocessing and annotation into 5 classes, trained on Faster RCNN with different architectures, and performed inferences. The newspapers without graphics are then tested on the model to detect different columns of text, headline regions, and title regions. The workflow of the system is shown in Figure 1.

Proposed algorithm to Separate text from Images

The algorithm uses the concept of contour detection and performs various morphological operations to remove the graphics/photographs from the newspaper image. Firstly, newspaper images are binarized using various binarization algorithms (Sauvola and Pietikäinen, 2000). After the contours are detected, each contour's average height and width are obtained, and based on a predefined threshold, a newspaper image without graphics is generated.

Algorithm 1

1. Read the input image and convert the loaded image to grayscale.

- 2. Apply Sauvola thresholding to the grayscale image to create a binary image.
- 3. Erode the binary image using a kernel
- 4. Find contours in the eroded binary image
- 5. Initialize an empty list to store all contours.
- 6. Loop until all white pixels have been traced:
 - a. Find the first black (0) pixel as the starting point of a contour.
 - b. Initialize the current pixel as the starting point.
 - c. Initialize an empty list to store the points of the current contour.
 - d. Loop until returning to the starting point of the contour:
 - (i) Store the current pixel as a contour point.
 - (ii) Traverse in the current direction (right, down, left, up) to find the next black pixel.
 - (iii) Update the current pixel to the next pixel.
 - (iv) If the current pixel is the starting point, stop the loop.
 - (v) If the current pixel is the starting point, stop the loop.
 - e. Add the list of contour points (current_contour) to the list of all contours (all_contours).

8. Process each contour in the list of contours:

^{7.} Calculate the average height and width of all contours

18 ਤੋਂ 45 ਸਾਲ ਤੋਂ ਵੱ	ੱਧ ਉਮਰ ਵਾਲਿਆਂ ਨੂੰ	ਲੱਗੇਗਾ ਮੁਫ਼ਤ ਟੀਕਾ	ਗਰਮੀ ਤੋਂ ਰਾਹਤ	ਯੂਟੀ ਦੇ ਸਕੂਲਾਂ	ਵਿੱਚ 10 ਮਈ
ਪ੍ਰਾਈਵੇਟ ਹਸਪਤਾਲਾਂ ਨੂੰ ਬਚੀ ਹੋਈ ਵੈਕਸੀਨ ਵਾਪਸ ਕਰਨ ਦਾ ਹੁਕਮ	ਨਿਯਮਾਂ ਦੀ ਉਲੰਘਣਾ ਕਰ ਚੰਡੀਗੜ੍ਹ ਪ੍ਰਸ਼ੀਸ ਨੇ ਸ਼ਹਿਰ ਵਿੱਚ ਲਗਾਏ ਰਾਤ	ਨਿ 'ਤੇ 9 ਗ੍ਰਿਫ਼ਤਾਰ ਗ਼ਿਊਦੀ ਉਲੰਘਣ ਕਰਨ ਸ਼ਬੰਧੀ 9 ਜੀਰਆਂ	Te la	ਤੋਂ ਗਰਮੀ ਦੀਅ	ਾਂ ਛੁੱਟੀਆਂ
ਟ੍ਰਿਸ਼ਿਊਨ ਨਿਊਚ ਸਭਾਵਿਸ ਚੰਡੀਕੜ, 1 ਸਈ ਯੂਟੀ ਦੇ ਪ੍ਰਬਾਸਕ ਵੀਪੀ ਸਿੰਘ ਬਦਨੌਜ ਨੇ ਦੱਸਿਆ ਕਿ 18 ਸਾਲ ਤੋਂ ਵੱਧ ਉਮਰ ਦੇ ਨੌਜਵਾਨਾਂ ਦੇ ਰਹੌਨਾ ਟੀਰਾਰਰਨ ਦੀ ਮੁਹਿੰਮ ਨੂੰ ਚੰਡੀਗੜ੍ਹ ਪ੍ਰਸ਼ਾਸਨ ਨੇ ਰਗੋ ਤੰਡੀ ਦੇ ਦਿੱਡੀ ਹੈ। ਹੁਣ 18 ਸਾਲ ਚੋ ਵੱਧ ਉਮਰ ਵਾਲਿਆਂ ਨੂੰ ਮੁਫ਼ਤ ਟੀਕਾਰਰਨ ਹੀਤਾ ਜਾਵੇਗਾ। ਉਨ੍ਹਾਂ ਦੱਸਿਆ ਕਿ ਆਉਣ ਵਾਲੇ ਦਿਨਾਂ ਵਿੱਚ ਇਸ ਮੁਹਿੰਮ ਦੀ ਸ਼ੁਰੂਆਤ ਕਰ ਦਿੱਡੀ ਜਾਵੇਗੀ। ਪ੍ਰਬਾਸਕ ਦੇ ਸਲਾਹਕਾਰ ਮਨੇਜ ਪਰੀਦਾ ਨੇ ਦੱਸਿਆ ਕਿ ਕਰਨਾ ਦੌਰਾਨ ਵਿਦੇਸ਼' ਤੋਂ ਆਉਣ ਵਾਲਿਆਂ ਨੇ ਕਿਸੇ ਵੀ ਤਰ੍ਹਾਂ ਦੀ ਮਦਦ ਲਈ ਪੀਸੀਐੱਸ ਅਧਿਕਾਰੀ ਰਾਬੇਸ਼ ਪੇਪਰੀ ਨੂੰ ਨੋਫ਼ਲ ਅਰਸਰ ਨਿਯੁਕਤ ਕੀਤਾ ਹੈ ਜਿਨ੍ਹਾ ਵੱਲੋਂ ਵਿਦੇਸ਼ੀ ਯਾਰਟੀਆਂ ਨੂੰ ਦਿੱਤੀ	ਚੰਡਬ ਪ੍ਰਯੋਸਟੇ ਸ਼ੱਠੇ ਦੱਸ ਗਾਰੇ ਹੱਲ ਗਾਰੇ ਸੱਤੇ ਨੂੰ ਬ੍ਰਿਟਸ ਕੀਤ ਹੈ। ਮੰਪਨਾ 12 ਅਤੇ ਮਹੀ ਬਾਦਾ ਮਿੱਡੀ ਜੰਗਰ ਅਤੇ ਬਾਦਾ ਸੰਕਟਨ ਤਾਂ ਦੇ ਸ਼ਾਨੇਸ਼ਟਾ 16 ਨੂੰ ਕੁੜ ਨੇਅਕਤੀ ਤਾਤ ਦੇ ਮੁੱਖ ਕੰ ਸਨ ਜਿਲ੍ਹ ਕਿਨਾਫ਼ ਕੇ। ਦਰਜ ਕਰਕੇ ਸਿੰਘ ਕਾ ਵਿੱਚਾਂ ਹੈ। ਜਾਣ ਵਾਸ਼ੀ ਸੁਵਾਹਾ ਵਾਂ ਪ੍ਰਬਧ ਗੀਤਾ ਜਾਵੇਗਾ। ਸਿਹਟ ਵਿਚਾਗ ਦੇ ਡਾਇਰੈਂਕਟਰ ਡਾ. ਅਮਨਦੀਪ ਕੌਰ ਕੰਗ ਵੱਲੋਂ ਪ੍ਰਾਈਵੇਟ ਕਸਪਤਾਲ ਪ੍ਰਬੰਧਕਾਂ ਨੂੰ ਬਾਹੀ ਬਚੀ ਹੋਈ ਕਟੋਨਾ ਵੈਕਸੀਨ ਵਾਪਿਸ ਕੀਤੇ ਜਾਣ ਸ਼ਬੰਧੀ ਨਿਰਦੇਵ ਜਾਤੀ ਕੀਤੀ ਹੋਏ ਹਨ। ਡਾ. ਰੰਗ ਨੇ ਦੱਸਿਆ ਨਿ ਪ੍ਰਾਈਵੇਟ ਹਸਪਤਾਲਾਂ ਨੂੰ ਹੁਨ ਸਿਹਰ ਵਿਭਾਗ ਵੱਲੋਂ ਵੈਕਸੀਨ ਸਪਲਾਈ ਨਹੀ ਗੀਤੀ ਜਾਵੇਗੀ ਪ੍ਰੋਟ ਜੇਕਰ ਪ੍ਰਾਈਵੇਟ ਹਸਪਤਾਲ ਵੈਕਸੀਨ ਸ਼ਬਾਉਣੀ ਚਾਹੁਣਗੇ ਤਾਂ ਉਨ੍ਹਾਂ ਨੂੰ ਆਪਣੇ ਪੱਧਰ ਤੇ ਵੈਕਸੀਨ ਖ਼ੁਦ ਪਰੀਦਣੀ ਹੋਵੇਗੀ। ਸ਼ਹਿਰ ਵਿੱਚ ਲਗਾਤਾਰ ਵਧ ਰਹੇ	ਤਰਸ਼ਦੂ ਦੀ ਦੁਰਮਣ ਕਰਨ ਤਿਹਾਸ ਤਾ ਜਰਦਾ ਆਜਾ ਦੀ ਪ੍ਰਸ਼ੀਮ ਨੇ 51, ਬਾਦਾ ਆਈ ਦੀ ਪਰਮ, ਪ੍ਰਤੀਸ ਨੇ 11 ਜਾਣੇ ਨੂੰ ਮ੍ਰਿਤਾਬ ਸੀਤਾ ਹੈ। ਪ੍ਰਤੀਸ ਗੋਲ੍ਹੇ ਦੋਹਨ ਤੋਹਾ ਮਾਜ ਵੀ ਹੋ ਬਚੁਹੀ ਨੇ। ਤੋਂ ਗਿਰੁਤਬ ਜੀਤਾ ਪਰ ਬਾਅਦ ਵਿਚ ਜਹਾਨਤ 'ਤੇ ਕਰਨਾਵਾਇਰਸ ਦਾ ਕਸਾ ਨੂੰ ਦੁਦ ਦਿਆਂ ਦੂਦੀ ਪ੍ਰਸ਼ਾਸਨ ਨੇ 54 ਨਵੇਂ ਮਾਈ ਕ੍ਰੋ ਕਟੋਨਹੀਟ ਦੋਨ ਬਣਾਏ ਹਨ ਜਿਨ੍ਹਾਂ ਵਿੱਚ ਸੈਕਟਰ 0, 10, 11, 14, 15, 15, 18, 19, 20, 21, 25, 24, 30, 51, 52, 30, 55, 30 ਵੈਸਟ, 42, 45, 65, ਰਾਮ ਦਰਬਾਰ, ਹੱਲੋਮਾਜਦਾ, ਬਹਿਲਾਨ, ਦੂਤਾ ਲਾਹੇਰਾ, ਹਨਸ ਦੇ ਦੋਣਵੇਂ ਘਰ ਬਾਮਿਲ ਹਨ। ਇਹ ਆਦੇਸ਼ ਡਿਪਟੀ ਕਸਿਸ਼ਨ ਮਨਦੀਪ ਸਿੰਘ ਭਰਾੜ ਨੇ ਜਾਰੀ ਕੀ ਤੇ ਹਨ। ਉਨ੍ਹਾਂ ਦੱਸਿਆ ਕਿ ਜਿਹੜੇ ਇਲਾਰਿਆਂ ਵਿੱਚ ਕਰੋਨਾ ਮਬੀਜ਼ਾਂ ਦੀ ਗਿਣਰੀ ਵੱਧ ਹੋ ਉਸ ਨੂੰ ਮਾਈ ਕ੍ਰੋ ਕੋਟੋਨਮੈਂਟ ਦੋਨ ਬਣਾ ਦਿੱਤਾ ਜਾਂਦਾ ਹੈ।	<image/> <image/>	ਇਕ ਮਹੀਨਾ ਨਹੀਂ ਨੱਗਣਗੀਆਂ ਆਨਰਾਈਨ ਕਰਾਸ਼ਾਂ, ਕੋਵਿਡ-19 ਡਿਊਟੀ ਵਾਲੇ ਅਧਿਆਪਕਾਂ ਨੂੰ ਕਰਨੀ ਹੋਵੇਗੀ ਡਿਊਟੀ ਤੁਪਵਿੰਘਰ ਪਾਲ ਸੋਵੀ ਦੱਸੀਬਾ, 1 ਮਸ਼ੀ ਪ੍ਰਤੀਪ ਦੇ ਸਰਕਾਰੀ ਤੋਂ ਏਡਿਡ ਸਕੂਲਾਂ ਵਿੱਚ ਗਰਮੀਆਂ ਦੀਆਂ ਛੁੱਟੀਆਂ 10 ਮਈ ਹੋ ਕਰ ਦਿੱਤੀਆਂ ਗਈਆਂ 10 ਮਈ ਹੋ ਹੀ ਹੋ ਜਾਣਗੀਆਂ ਕਿਊਰਿ 11 ਸ਼ੁਰੀ ਦੇ ਸ਼ਹੀਨੇ ਦੀਆਂ ਛਿਟੇਗ ਦੀ ਛੁੱਟੀ ਹੀ ਕੁਟੀ ਦੇ ਸ਼ੁਰੀ ਦੇ ਕੁਟੀ ਹੈ ਕਾਣਗੀਆਂ ਕਿਊਰੀ ਹੈ ਸਦ ਗੀਆਂ ਕਿਊਰੀ ਹੈ ਸਾਇਨਾ ਕਰਨ ਦਾ ਐਲਾਨ ਕੀਤਾ ਹੈ। ਇਨ੍ਹਾਂ ਛੁੱਟੀਆਂ ਦਿਨਾਂ ਚਾਂਗ ਸਕੂਲ ਮੁਪੀਆਂ ਨੂੰ ਆਮ ਦਿਨਾਂ ਵਾਂਗ ਸਕੂਲ ਆਉਣ ਲਈ ਕਿਹਾ ਗਿਆ ਹੈ। ਇਹ ਛੁੱਟੀਆਂ ਸ਼ਹਿਰ ਦੇ 115 ਸਕੂਲਾਂ ਦੇ ਆਧਿਆਪਕਾਂ ਨੇ ਨਾਨ ਟੀਇੰਜ ਸਟਾਫ ਕੋਈ ਕਾਰ ਦੇ ਨਾਨ ਟੀਇੰਜ ਸਟਾਫ ਕੋਰ ਨਾਨ ਟੀਇੰਜ ਸਟਾਫ ਕਰਨ ਦਾ ਕੇ ਕਰਨ ਦਾ ਕੇ ਸ਼ਾਰ ਕਰਨ ਦਾ ਕੇ ਸ਼ਾਰ ਕਰਨ ਦਾ ਕਰਨ ਜਾਂ ਦੇ ਕਰਨੀਆਂ ਸ਼ਹਿਰਾ ਦੇ 15 ਸਕੂਲਾਂ ਦੇ ਆਧਿਆਪਕਾਂ ਨੇ ਨਾਨ ਟੀਇੰਜ ਸਟਾਫ ਕਰੋ ਕਰਨ ਦੀ ਇੰਜ ਸਟਾਫ ਕਰਨ ਦਾ ਕਰਨ ਦਾ ਕਰਨ ਕਰਨ ਦਾ ਕਰਨ ਦਾ ਕਰਨ ਦਾ ਕਰਨ ਜਾਂ ਦੇ ਕਰਨੀਆਂ ਦਾ ਕਰਨ ਕਰਨ ਦਾ ਕਰਨ ਜਾਂ ਜੋੜ ਕਰਨੀਆਂ ਜਾਂ ਦੀ ਕਰਨ ਦੀ ਛਿੰਦ ਸਟਾਫ ਕਰਨ ਦੀ ਛੁੱਟੀਆਂ ਹੈ ਕਰਨ ਦੀ ਛਿੰਦ ਸ਼ਾਰ ਸ਼ੁਰੀਆਂ ਜਿਹਿਰ ਦੇ 15 ਸਕੂਲਾਂ ਦੇ ਆਧਿਆਪਕਾਂ ਨੇ ਨਾਨ ਟੀਇੰਜ ਸਟਾਫ ਕਰਨ ਦਾ ਕਰਨ ਜਾਂ ਜਾਂ ਜਿੰਚ ਸਟਾਫ ਕਰਨ ਦਾ ਕਰਨ ਦਾ ਕਰਨ ਦਾ ਕਰਨ ਜਾਂ ਕਰਨ ਦਾ ਕਰਨ	ਕਰੋਨਾਵਾਇਰਸ ਦੇ ਕੇਸ ਲਗਾਤਾਰ ਵਧ ਰਹੇ ਹਨ ਜਿਸ ਹਰ ਕੇ ਕੁਣ ਅਹਿਆਪਹਾ ਗਰਮੀ ਦੀਆਂ ਛੱਟੀਆਂ ਤੋਂ ਬਾਅਦ 8 ਜੂਨ ਨੂੰ ਸਕੂਲ ਦੁਆਇਨ ਕਰਨਗੇ। ਕਿਲ੍ਹਾ ਸਿੱਖਿਆ ਅਧਿਕਾਪੀ ਨੀਨਾ ਕਾਲੀਆਂ ਨੇ ਦੱਸਿਆ ਕਿ ਜਿਹੜੇ ਅਧਿਆਪਕ ਸਕੂਲਾਂ ਵਿੱਚ ਦਾਖ਼ਲਿਆਂ ਦਾ ਕੰਮ ਦੇਖ ਰਹੇ ਹਨ ਉਨ੍ਹਾਂ ਨੂੰ ਸਕੂਲ ਮੁਖੀ ਲੋਡ ਅਨੁਸਾਰ ਸਕੂਲ ਵਿੱਚ ਦਾਖ਼ਲਿਆਂ ਦਾ ਕੰਮ ਦੇਖ ਰਹੇ ਹਨ ਉਨ੍ਹਾਂ ਨੂੰ ਸਕੂਲ ਮੁਖੀ ਲੋਡ ਅਨੁਸਾਰ ਸਕੂਲ ਸੱਦ ਸਕਦੇ ਹਨ। ਉਨ੍ਹਾਂ ਅਧਿਆਪਕਾਂ ਨੂੰ ਸਟੇਸ਼ਨ ਨਾ ਛੱਡਣ ਦੀ ਹਦਾਇਰ ਕੀਤੀ। ਇਸ ਦੇ ਨਾਲ ਹੀ ਜਿਹੜੇ ਅਧਿਆਪਕਾਂ ਦੀ ਭਿਊਟੀ ਪ੍ਰਸ਼ਾਸਨ ਨੇ ਕਰੋਨਾ ਮਹਾਸਾਰੀ ਲਈ ਹਦਾਇਰ ਕੀਤੀ। ਇਸ ਦੇ ਨਾ ਕਰਨਗੇ। ਇਕ ਹੋਰ ਅਧਿਕਾਰਾਂ ਨੇ ਦੱਸਿਆ ਕਿ ਹਾਲ ਛੱਟੀਆਂ 3 ਜੁਨ ਰਕ ਹੀ ਕੀਤੀਆਂ ਗਈਆਂ ਹਨ ਤੇ ਕਰੋਨਾ ਦੇ ਹਾਲਾਨ ਅਨੁਸਾਰ ਇਸ ਛੱਟੀਆਂ ਅੱਗੇ
				ਨੂੰ ਹੋਣਗਾਆ। ਇਸ ਵੱਲੋਂ	ਵਧਾਈਆਂ ਜਾ ਸ਼ਕਦਾਆਂ ਹਨ

Figure 3. Labelling of Punjabi Newspaper Image using LabelImg Tool

- a. Get the contour's bounding rectangle (x, y, width, height).
- b. If the height of the bounding rectangle is greater than the threshold times the average height:
 - i. Set the corresponding region in the original image to white.
- 10. Create a mask by drawing filled contours on a white canvas
- 11. Save the modified image and the mask to the specified paths.
- 12. Display the modified image without graphics.

In Figure 2, a is an original figure, and b is an image obtained by applying the proposed algorithm with text separated from graphics. Figures 2c and 2d also show that the proposed algorithm correctly removes pictures from the Manhattan layout as well as pictures from non-Manhatten layouts. We have tested the proposed algorithm on several images of complex layouts and found that it correctly separates text from pictures. Non-Manhattan layouts are those that form any arbitrary shape in newspapers.

Dataset Collection

In the second part of this paper, we first created the dataset. The dataset for newspaper layout analysis was

collected from various internet resources like websites and image resources. These newspapers are labeled into 5 classes Image_Region, Text_Region, Advertisement _Region, Heading_Region, and Title_Region with the help of the labelling tool Labeling as shown in Figure 3. Total number of Punjabi newspaper images collected is 500 and labelled. Table 1 represents a number of different regions annotated with the help of an annotation labeling tool. Total number of annotating objects is 20873.

Table 1. Distribution of Different labels in the dataset

Labels	Number of Annotations
Text_Region	8342
Image_Region	4563
Advertisement_Region	3421
Heading_Region	4332
Title_Region	215
Total	20873

Dataset Pre-Processing and splitting

As the images of newspapers are quite large, they are converted into a standard, smaller, and uniform size. The dataset is divided into training, testing, and validation. 70% of the images are in the training dataset, 20% in the testing dataset, and 10% in the validation dataset.



Figure 4. Faster RCNN Architecture

Deep Learning architectures

We have used an object detection model called Faster RCNN with different backbones: ResNET+FPN 50, 101, and Retinanet. Detectron2 which is a Facebook AI Computer Vision research framework that has various pre-trained models used for transfer learning. We have trained different architectures with the same configuration, and after that, testing is done on the test dataset. The images without graphics generated by the proposed algorithm are used for testing to detect the column of text.

Deep Learning Architecture Faster RCNN

Faster RCNN is made of CNN, the region proposal network, and prediction as shown in Figure 4. The RPN layer gives two outputs. One is objectiveness classification, and the other is a bounding box regressor, which acts as input to the ROI layer. The output of the ROI layer is given to the fully connected classification layer, giving two outputs: a softmax classifier and a bounding box regression, giving the bounding box for each area of the newspaper image to detect the layout of the newspaper image.

Retinanet

RetinaNet typically uses a deep convolutional neural network (CNN) as its backbone. Popular choices include ResNet, ResNeXt, and other similar architectures. RetinaNet incorporates a feature pyramid network (FPN) on top of the backbone to handle objects of different sizes shown in Figure 5. RetinaNet employs anchor boxes at multiple scales and aspect ratios to propose potential object locations in the feature pyramid. These anchor boxes act as priors for object localization. RetinaNet employs separate classification and regression heads for each anchor. The classification head predicts the probability of each anchor containing an object of a specific class. The regression head predicts adjustments to



Figure 5. Architecture of Retinanet based on Resnet and Feature Pyramid Network to predict the Layout DOI: https://doi.org/10.52756/ijerr.2023.v35spl.004

the anchor box coordinates to fit the object's bounding box. The key innovation of RetinaNet is the focal loss function, designed to address the issue of class imbalance.

Results and Discussions

The experiments are conducted to remove the images of the Gurumukhi script from the newspapers based on the proposed algorithm. Around 100 newspaper images are tested using the proposed method. Table 2 shows the results and accuracy. We have trained the model based on Faster RCNN based on the configuration shown in Table 3.

Table 2. Performance of the proposed algorithm(Algorithm 1) to Separate text from Images

No of Newspapers	Total pictures in Newspaper images	Pictures removed	Accuracy
100	545	519	96.22

Table 3. Training System configuration ofConvolution Neural network

Epoch : 300	Batch size: 4
Max Size Iteration: 300	Backbone: X101,R50
Optimizer: Default SGD	Number of classes:6
with Detectron2	
Data Loader Worker: 4	Learning Decay: 0.001

Evaluation Criteria

1. The IOU, or Jaccard Index, is a metric that is the ratio of the area of intersection (the overlapping region) to the area of union (the combined region) of the predicted and ground truth regions used to test the performance.

IoU = (Area of Intersection) / (Area of Union) (1)

2. Precision: Precision is a metric that measures the accuracy of positive predictions made by a model.

In the context of object detection, precision represents the fraction of the predicted bounding boxes that are correct. It is calculated as the ratio of true positives (correctly predicted objects) to the sum of true positives and false positives (incorrectly predicted objects).

Precision = (True Positives) / (True Positives + False Positives) (2)

3. Recall: Recall, also known as sensitivity or true positive rate, measures the ability of a model to capture all positive instances. In object detection, it represents the fraction of the actual objects that the model correctly

detected. Recall is calculated as the ratio of true positives to the sum of true positives and false negatives (missed objects). Recall = (True Positives) / (True Positives + False Negatives) (3)

4. The F1-score is another metric that combines precision and recall into a single value, which can be useful when you want to strike a balance between the two. It is calculated as the harmonic mean of precision and recall:

F1-score = 2 * (Precision * Recall) / (Precision + Recall) (4)

The different architectures are compared in Table 4 on the test dataset of newspaper images. The model's performance with X101 in this paper is better on mAP and F1_score, and the model's performance is more stable than other architectures. We have tested the model's performance on different architectures on 50 newspaper images. The accuracy of a model for layout detection using different architectures is shown in Table 5 and Figure 6.

 Table 4. Comparison results of Faster RCNN with different architectures

Architecture	Precision	Recall	F1
			Score
X101-FPN	0.87	0.81	0.838
R50-FPN	0.79	0.77	0.78
Retinanet R50	0.76	0.75	0.75
Retinanet101	0.81	0.83	0.82

Table 5. Performance of model with different

 architectures tested on 50 newspaper images

Model	Number of Text Columns	Detected Text Columns	Accuracy
X101-FPN	560	535	95.53
R50-FPN	560	521	93.03
Retinanet R50	560	502	89.64
Retinanet101	560	525	93.75





Figure 6. Validation Accuracy of different architectures for Layout Detection

ਪੰਚਾਇਤ ਚੋਣਾਂ

89% वैवाफेटो बेको पावटी वेटा डिसाड ਉੱਕੇ मंची मजीका संघीका पर ত হাৰতা নমতা থাৰাই মঁ বাজ লৈও টুঁতে খুটাম বাঁথা খিলাইত উত ইবাম নাজ্য খুমিন বাঁথা থেচে মাতা ঘটা খাখাত তিথিতলকা বা মুখ দেই বাঁৱা মা। ৰাম্মান জাতে মান বাঁকা মিৰা খুল খঁৱা মানা অভিজেলমেন ব विक संस्थित आगर भारती से तरह बाजारकों काल कार्सीयका जिस 40.40 मोट ਵਿਸ਼ ਅੰਨ ਤੋਂ ਕੱਧ ਕਿਤੇ ਵੀ ਨਹੀਂ ਜਿੱਤ ਸਕੀ। 1988 ਆਜਾਂਦ ਉਮੀਦਵਾਰਾ ਨੇ ਬਾਜੀ ਮਾਰੀ ਹੈ। ਲਮਾਜਵਾਦੀ ਪਾਰਟੀ ਨੇ ਭਾਜਪਾ ਨੂੰ ਪੂਰੀ ਟੇਂਡਰ ਮਿੱਡੀ ਹੈ। ਜੇਡਰ ਵਿਰੋਧੀ ਪਿਤਾ ਅਮੇ ਭਾਜਪਾ ਦਾ ਮੁਤਾਬਨਾ ਕੀਜ਼ਾ ਜਾਵੇ ਜਾਂ ਭਾਜਪਾ ਨੇ ਮੁਸ਼ਕਿਨ ਨਾਨ ਲ ਕੀਸਵੀ ਸੀਟਾਂ ਜੋ ਜਪਾਦਾ ਰੁਹਾਬਲਾ ਹੈ। ਜਿੱਤ ਧਾਪਤ ਨੀਟੀ ਹੈ।

ਹੀ ਵਿੱਚ ਪਾਪਤ ਹੀ ਸੀ। ਹੋ। ਇਹ ਨਹੀਂ ਜੋ ਇਸ ਹਫ ਕੇ ਵੀ ਸਹਿੰਤ ਧੁਸ਼ੂਰਨ ਹਨ ਕਿਉਂਕਿ ਸੁੜੇ ਦੀ ਲਗਭੱਗ 77.75 ਫ਼ੀਸਦੀਂ ਆਪ ਦੀ ਪਿੱਡਾਂ ਵਿਚ ਬਹਿੰਦਾਂ ਹੈ। 2017 ਵਿਚ ਭਾਜਪਾ ਨੇ ਵਿਚਾਨ ਜਗਾ ਦੀ ਜਾਂ ਕਰ ਗੇਠਾਂ ਵਿਚਾਲ 20 ਜਿ ਜਿੱਜ ਪ੍ਰਾਪਕ ਕਰ ਕੇ ਨਾ ਡੀਸਰੀ ਡੇ ਵੱਧੋਂ ਵੋਟਾਂ ਦੱਸਿਲ ਕੀ ਡੀਆਂ ਸਨ 2019 ਵੀਤੇ ਜੀ ਪੰਚਾਬ ਨੇ ਸ਼ੁਰੂ ਦੇ ਇਹਨ ਤਰੀਜਿਆਂ ਨੂੰ ਕਿਸਾਜ਼ ਅੰਦਰਨ ਜਾਰ ਜਿਸ ਕੇ ਦੱਸਿਆ ਜਾ ਜਿਸ਼ਾ ਦੀ ਪੰਚਾਬ ਕੇ ਸ਼ੁਰੂ ਦੇ ਇਹਨਾਜ਼ ਅੰਦਰਨ ਦੇ ਤਰਿਆਰਾ, ਉੱਡਾ ਪ੍ਰਦੇਸ਼ ਅਤੇ ਉਸਸੱਤ ਸਿਹਾ 'ਚ ਪੰਜਾਬ ਕੇ ਸ਼ੁਰੂ ਚੱਬੇ ਤਿਸਾਨ ਅਰੇਲਾ ਦੇ ਰਚਿਆਰਾ, ਉੱਡਾ ਪ੍ਰਦੇਸ਼ ਅਤੇ ਉਸਾਪਣੇ ਵਿਚ ਵੇਰ ਸਦ ਕਾਰ ਕਰਦਾ ਦੀ ਕਿਆਗ ਨੇ ਗਿਆ ਹੈ, ਇੱਕ ਦੀ ਪੰਜਾ ਕਿਆ ਇਹ ਪ੍ਰੋਡਿਆ ਦਾ ਇਹ 'ਚ ਇੱਕ ਦੀ ਪੰਤਾਇਹ ਦੇਤਾ ਦੇ ਨਹੀਂ ਜਿਆਂ ਦਾ ਸ਼ੁਰੂ ਦੇ ਦੀਆਂ ਸ਼ੁਰੂ ਵਿਚ ਦੇਣ ਵਾਰਮਾਂ ਅਤੇ ਵਿਚ ਨਾ ਭਾਰੇ ਦੇ ਦੇ ਉੱਤੇ ਬੱਧੇ-ਸਿਹਾ ਪੰਜਾ 'ਚ 'ਚ ਪਿੰਡਾ ਹੈ ਜਾਣਕਾ ਪੰਤਾ ਮਹੀਰਿਆਂ ਦੀ ਰਿਸੇਸ਼ ਚੀਆਂ ਬਹੁਤਾਂ ਉੱਕੇ ਬੱਧੇ-ਸਿਹਾਰ ਦੀ ਸ਼ਾਂਦ ਨਹੀਂ ਹੁਣਾ ਦੀ ਭਾਜਪਾ ਦੀ ਵਾਰ ਹੈ ਪਿੱਛੇ ਜੰਤੂਸ਼ਨਾ ਕਿਸਾਨ ਮੰਦਰੇ ਦੇ ਅੰਜਾਨ ਕੀਤਾ ਹੈ ਕਿ ਸਿਰਤ ਹੋਵਾਂ ਜਾਣਕਾ ਪੰਤ ਮਹੀਰਿਆਂ ਦੀ ਰਿਸੇਸ਼ ਚੀਆਂ ਬਹੁਤਾਂ ਉੱਕੇ ਬੱਧੇ-ਸਿਹਾਰ ਦੀ ਸ਼ਾਂਦ ਨਹੀਂ ਜਾਣਕਾ 'ਚ ਕਾਂ ਤਿਹਾ ਵਿਚ ਬਹੁਤਾ ਵਿਚ ਹਾ ਜਾ ਜੀ ਕਾਰਸ਼ਾ ਸਿਲਕਾ ਸ਼ਹਿਮ ਇਹ ਕਰਨਜ਼ ਸ਼ਰਦਾ ਕਿਸਾ ਨੇ ਅੰਜ ਕਾਂ ਬੱਧੇ ਸ਼ਿਸ਼ਨਾ ਅੰਦੂ ਕਜਾਂ ਉਂਕਨ ਨੇ ਸ਼ਾਮਦਾ ਬੁੱਧੇ ਜੀਆਂ ਰਿਸ਼ਨਾ ਸ਼ਰਦਾ ਕਿ ਸ਼ਰ ਨਹੀਂ ਜਿਆਂ ਉੱਕੇ ਗਿਆਨ ਲਈਨਨ ਦਾ ਸ਼ਰਦਿਆਂ ਵੀਪਾ ਹਿੰਗ 202 ਦੀਆਂ ਵਿਧਾਨ ਸ਼ਰਾ ਦੇ ਬੋਰਾ ਨਹੀਂਜਿਆਂ ਉੱਕੇ ਗਿਆਨ ਲਏਨਨ ਦਾ ਸ਼ਰਦਿਆਂ ਦੇ।

ਗੁਰੁਦੇਵ ਰਾਬਿੰਦਰ ਨਾਥ ਟੈਗੋਰ ਨੂੰ ਯਾਦ ਕਰਦਿਆਂ 5 96<mark>%</mark>वर्गाउर से भाषीयल जिसलाह से प्रेन 'व

5 96% ন গোরাম ভিন রেই দায়াওক কর্মা কর্মাজা কর্মা নার্মাজ ਦੇ ਸ਼ਾਹਕਾ ਦਾ ਸਿਹਮਾ ਕਿਸ ਕਾਂਸ ਸਮਸੰਦਿਕਾ ਯਾਗਵਾਨ ਸ਼ਾਸ ਕੀਤੀਆਂ ਜਾਂਸੀਆਂ ਹਨ ਕਿਸ਼ੇਰਿ ਹੋਨ ਜੋ ਰਿਕਾਰਾਂ ਇਹ ਸਿੱਖ ਨੇ ਨਰਮਾਂ ਕਿਸ਼ੇਰਿ ਹਨ ਜੋ ਰਿਕਾਰਾਂ ਇਹ ਸਿੱਖ ਨੇ ਨਰਮਾਂ ਸ਼ਾਸਰੀ ਜਦੇਸ਼ੀ ਦੇ ਦੀ ਹੈ ਜਿਹੜੀ ਜੀਤਾਂ ਸੱਤੇ ਦੇ ਵਿੱਚ ਮੰਗਣਾਨ ਦੀ ਬਣਤੀ ਵਿੱਚੋਂ ਸੁੜੀ ਜਦੇ ਦੇ ਛੋਟੇ ਵਿਚੋਂ ਮੰਗਣਾਨ ਦੀ ਬਣਤੀ ਵਿਚੋਂ ਸੁੜੀ ਜਦੇ ਦੇ ਛੋਟੇ ਵਿਚੋਂ ਪੈਰਾਨ ਦੀ ਬਣਤੀ ਵਿਚੋਂ ਸੁੜੀ ਜਦੇ ਦੇ ਛੋਟੇ ਵਿਚੋਂ ਬਹਿਮਾਂ ਕਰਨਾਂ, ਜਾਤੀ ਤੋਂ ਸਮਤੀ ਨੇ ਕਰਦਾਨ ਕਰਦੀ ਹੈ। ਮੰਗਣਾਨ ਦੀ ਬਣਤੀ ਵਿਚੋਂ ਸੁੜੀ ਜਦੇ ਦੇ ਛੋਟੇ ਵਿਚੋਂ ਬਹਿਮਾਂ ਕਰਨਾਂ, ਜਾਤੀ ਤੇ ਸਮਤੀ ਨੇ ਕਰਦਾਨ ਕਰੀ ਦੀ ਸ਼ਾਂ ਸ਼ਹਿਰਾਂ ਸ਼ਰਦਾਂ ਸਿਹਨਾ ਸ਼ਰੀ ਜਾਂ ਜਾਨੀਂ ਉਹਨਾਂ ਦਿ ਵੱਚੋਂ ਵਿਚ ਸ਼ਰੀ ਦੇ ਕਰਦਿਆਂ ਦੇ ਸ਼ਰਦਾਨਾ ਬੀਰਵਾਰਨ ਸ਼ਾਹਿਆਂ ਸ਼ਰੀ ਦੇ ਸ਼ਰਦਾਨਾ ਸ਼ਰੀ ਦਾ ਸਿੱਖਾਂ ਸ਼ਰੀ ਸ਼ਰਦਾਨ ਸ਼ਰੀ ਦੇ ਦੇ ਕਰਦੇ ਹੀ ਜੋ ਹਨਾਂ ਬਰਿਤ ਦੇ ਵਿਚੋਂ ਹੀ ਸੱਤਾ ਸ਼ਹਿਰਾਂ ਸ਼ਰਦਾਂ ਸਿਰਤੀ ਦੇ ਸ਼ਰਦਾਨਾ ਬੀਰਵਾਰਨ ਕਰਦੀ ਦੇ ਕਰਦੀ ਹੈ। ਇੱਕ ਕੋਈ ਉਹ ਦੇ ਵਿਚੋਂ ਹੀ ਸ਼ਰੂਵੇਡ ਸ਼ਹਿਰਾਂ ਸ਼ਹਿਰਾਂ ਸ਼ਰਦਾਂ ਹੋ ਸ਼ਰਦਾ ਬਣਾ ਜਾਇ ਕਾਰੀ ਸ਼ਰੂਵੇਡ ਸ਼ਹਿਰਾਂ ਕਿਸੇਆਂ ਸ਼ਰਿਕਾਂ ਸ਼ਰਦਾ ਹੋ ਗਾਰਾਂ ਛੋਟੀ ਵਿਰਾਸਨਾ ਸ਼ਿਹਨਾਂ ਸ਼ਰੂ ਹ P ਕੇਜ ਸਾਡੇ ਸਨਮੁੱਖ ਹੈ। ਉਨ੍ਹਾਂ ਦੀ ਸਮੁੱਖੀ ਨਿਖਜ ਵਿਚ ਦੁਨੀਆਂ ਭਰ ਦੇ ਵਿਦਵਾਨਾਂ

5.96% বিধেন্দেশ পাৰাৰ প্ৰথ বিগতে বিগত বা নিৰ্মান প্ৰথি সাম্পনি বিভিন্ন পৰিমা ঘৰলা বি মানিৰ মাৰ্গ কি না বুৰুলি কাৰ্বিকা পাৰ বিগতে বিগ বি মানিৰ মানিৰা হৈ বি বি বি বি বিগল পা মতন বিগ পিছা মাৰ্গ মহাৰ কৰা হৈ হা বাৰ্গ নিন্দিলৰ মানৱা হৈ বিত কৰা ১৯ সহ ১০০ এ, কয় নামান্ধ মানৱা (ব এন কৰা ঠাক ট নামী গোমানিক ত তীৰ ঠা ਉন্দু তী মূৰগৰিৱা টা গাঁৱা ই, ৰাজী ৰাজন সম্ভাৱনা হয়, না ভিষ্ণা হা, ৰাজী নামানে নিৰা খোৱা হোৱা হোৱা হয়নী ৰাজনো মানানে নিৰা খোৱা হোৱা হোৱা হিয়া হিয় মুন্যান্তৰ দী নাৰ না হাৰবা হোৱা নাছিল, বন্ধানীয়া

ਅਸ ਪੱਲਮ ਦੇ ਇਸਕਾਨ ਇਸਟੁੱਸ ਕਾ ਸਿੱਕੇ ਸ਼ਹਿੰਦਾ ਵਰਨਰੇ ਕਾ ਉਨ ਨੂੰ ਮਹਿਲਾ ਸੀਮਦਾ ਕੀਜ਼ਾ ਕਿ ਅਸੀਂ ਸਾਡੇ ਹਰਮਾਂ ਨਹੇ ਸੀਇਸਾਬਾਟਾਂ ਪੀਆ ਵਿਕਾਸ਼ਾ ਅਸੀਂ ਪਟੇਸ਼ਵਾਂ ਨੂੰ ਇਹ ਕਰਤੇ ਨਾਜਨੀ ਨੂੰ ਵਿਚ ਨਾਸ਼ਨਿਕ ਸੇਅਦ ਦੀ ਵਿਚਸਟ ਕਾਰੇਗੇ ਵਿਚੋਂ ਕਰਤੇ ਕਾਂ ਕਿਹੜ ਵਾਰਤੀ ਪੈਂਟਰਮੈਸ਼ਨਲ ਪ੍ਰਤੇਵਰਸਿਟੀ ਦਾ ਸਾਣੇ ਨਜ਼ਾ ਵਿਚ, ਸ਼ੱਖਾਰ ਇਹ ਨਾਨ ਵਾਨੇ 'ਚ ਸਾਰੇ ਕਿਸ ਕਿਸ ਹੈ ਸੁਰੁਵੇਚ ਤਾਇਣਾ ਨਾਲ ਟੇਵੀਰ ਨੇ ਸ਼ਾਂਤੀ ਨਿਲੇਤਲ ਵਿਚ ਵਾਰਲ ਦੇ ਜਭ ਬਣਾਂ 'ਚੋਂ ਨਿਰੇਸ਼ਾਨ, ਸਾਇਲ ਅਤੇ ਉਨ੍ਹਾਂ ਦੀ ਨਿਰਾਲਗੇ ਨਾਲ ਸ਼ਹਿਤ ਵਿਚਸ਼ਾਂ ਦਾ ਅਰਿਮੀਜ਼ ਕਰਮ ਸ਼ਹਿਜਾਤ ਹੋ ਦੀ ਸ਼ਹਿਮੀਆ ਨੇ ਗਿਲ ਦੇ ਸ਼ਵਿੰਦ ਵਾਲਤੀ गांव 100 फोला दो दुस्लान का दोगा के साथ का जिनक दोनी दुस्लीन है पुरावे के विशेष के लावी 2 दुसे ही विद्यालय करी रही विगा। दो बाराठा भवी कार पर भारता बोटा वहुएक का विवेद उसकी विद से किस्तार का सुरोधन बढ़ सिक्ताल र सोता दो किस्तान की बाराज में किसी किस कही नहीं कि प्रकार के बाराज से किसी का कही नहीं का प्रकार

ति के बाद के दिल्ला के प्राणने सुनुयंख तनियन रुप रुरात से लिस विजन्म घर कि में दिया से रुप विजन

राज्य का तो फा या जिसके समय के राजे के सम

ਅਤੇ ਪੱਛਮ ਦੇ ਵਿਦਵਾਨ 'ਸਕਦੇਟ ਹ ਕੇ ਇੱਕ ਸੋਇਅ

5.96% ह ta tag क्रम द केंग्रे प्रसिद्ध रिय मा गतासामा सा घर मंत्रा मित्रापत सित्रापत होधना फिला जिली ने असे सत हम दिस नामा पत बेल किता ते, जिस सर सीनी सी अपस्यत 8 धन ਜੱਬ ਜਿਹਾ ਹੈ, ਉਸ ਦਲ ਸਾਂਦਾ ਜਾਂ ਆਬਾਲ ਲਾਭ ਗੁਰੂਦਾ ਕਾਰਿਕਰ ਨਾਲ ਨਾਲ ਜਾਂ ਗੱਲ ਉਨ ਜੱਢਾ ਕਰਦਾ ਨਿੱਡਾ 11 ਸਾਂਦਾ ਨੇ ਜਾਂ ਉਨ੍ਹਾਂ ਦੀਆਂ ਗੱਲਾਵਾਂ ਦਾ ਹੈ ਨਿਹਾਲ ਨੇ ਜਾਂ ਦੇ ਦਿੱਤਰ ਨਾਲ ਦਾ ਕਰਦਾ ਕਰਦਿਆਂ ਪੰਜਾਬੀ ਸੁਰੀਬਟੀਮਟੀ ਦੇ ਦਿਸ਼ੇਸ਼ਰ ਖਾਸ਼ ਜਿਲਦਾ ਨੇ ਨੇ ਦੱਸ ਦਿਵਾ ਸੇਵੇਦਨਤਾਂ, ਦਿੱਤਨ ਕਰਮਨਾ ਅਤੇ ਇਹ ਅਜਿਹੀ ਵਿਜੇ ਹੈ ਜਿਸ ਵਿਚ ਕਈ ਨਿਸਮ ਦਾ ਮੰਗੀਚ ਹੈ, ਰਣਾ ਵਾਂਗ ਨਾਲ ਕਰਨ ਦੀ ਸਕਾਰੀ ਨੇ ਪੁੱਤ ਦਾ ਦੂਰਜ ਨਾਨ ਗੱਲਾਰ ਪੱਤਾ ਲੱਗਦਾ ਨੇ ਦਿੱਤ ਨਿਸ਼ੇਸ਼ ਦਾ ਮੰਗੀਚ ਹੈ, ਰਣਾ ਵਾਂਗ ਨਾਲ ਕਰਨ ਦੀ ਸਕਾਰੀ ਨੇ ਪੁੱਤਾ ਦਾ ਸੁਰਜ ਨਾ ਗੱਲਾਰ ਪੱਤਾ ਲੱਗਦਾ ਨੇ ਦਿੱਤ ਨਿਸ਼ ਕਰਦੇ ਜੀਤਰ ਨੇ ਪਾਣੀ ਦਾ ਨਰੀਂ ਪ੍ਰੈਦੀਆਂ ਬਾਹਾ ਅਜਿਹੀ ਅਜੀਮ ਕਰਸਿਆ ਹੈ ਸਾਣ ਕਰ ਹਿੱਡ ਪੰਜਾਬੀ ਕੁਰੀਬਰਜਿਟੀ, ਪਰਿਆਲਾ ਸ਼ਰਮੀ ਜੀਬਰ ਰਹੇ ਨੇ ਮਜੀ ਉਨ੍ਹਾਂ ਦਾ ਜਨਜ਼ ਜਿਨਤਾ ਉਸ ਜਾਂ ਮਿਲਾ ਕੇ ਜਾ ਦਾ ਨੀਂ ਸਿਹਿਕਾਂ ਅਨਿਹਾਂ ਨੇ ਸ਼ਰਮੀਆਂ ਨੇ ਦਾ ਸਿਹਿਲ ਕਰੂਸਰ ਮਨ ਉਹ ਨੇ ਸਰਕੀ ਪੱਧਰ ਦਾ ਸਮਨਾਨ ਕਰਬਾ ਰਹੀ ਨੇ ਅਸੀਂ ਉਨ੍ਹਾ ਦਾ ਨਸ਼, ਇਨਾ ਉਸ ਸੰਮੇ ਮਨਾ ਰਹੇ ਹਾ ਜੱਦੇ ਇਹਿਰਤਾ ਅਜਿਹੀਆਂ ਸਮਸੀਅਰਾਂ ਨੂੰ ਮੁਲਾਬਾਰ ਕੱਖਣ ਅਤੇ ਉਨ੍ਹਾਂ ਵਿਚ ਕਿਵਿੰਪ ਨੇ ਨੋਲ ਲਗਾ ਗਸਾਸ਼ਨ ਨਸੀਂ ਕਾਂਸੇ ਨਸ਼ ਨਿਗਨ ਦੀ ਪੀਲ੍ਹਾਂ ਬਿਜ ਸਾਮਲ ਜਨਾ ਦੀ ਮੰਗ ਤਰਦਾ ਹੈ। **มันสล**ะ คยธรา-15438

ਖ਼ਾਮੋਸ਼ੀ ਦੇ ਦੌਰ ਦਾ ਏਕਲਵਯ ਹਰਿਭਜਨ ਸਿੰਘ ਰੈਣ

81% a uar seuto teta was et your ਕਨ ਵਾਲਾ ਪੰਜੋਨਕ ਹੀਰਵਪਨ ਪਿੱਛ ਰੋਣੂ ਗੋਵੀਰ ਦਰਾ ਵਾਲਾ ਅਖਨਪਛਾਦ ਕਵੀ ਸੀ। ਉਹ ਦੁਰਨਾਹ वि सप्तव्याच, माजर, मनस्वयान, अठपुन, विकासी, सविजन, तुबु ठाफर्ख, वाधीन, ब्रेन्ट्रे सप्त अन्द्रे स्वीस ताल तिथ से क्यीके स्ट चेल्ट सी। सिल्प्साट से प्रेय ਸਨਮ ਇਨ 'ਤੇ ਵਿਸ਼ੇਸ਼

বেৰি ক'বা গাইজ ছিবজা ভূৱন কাজা উল্লু জাল

aper.punjabitribunechline.com

5 97% तिफल्ला के प्रेतिकल्व फेल के किल्ला के प्राप्त के किल्ला के प्राप्त के किल्ला के प्राप्त के किल्ला के किला ਰਗਰੂਪਾਂ ਤੋਂ ਬਾਹਰ ਸ਼ਹਿਣਾ ਪਨਜ਼ੂਰ ਸਾ, ਨਾ ਸ਼ਾਸਣ ਜ਼ਾਰਾ ਪੰਨ ਹੁਣਾ ਘੰਤਾ ਦਾ ਹੀ ਪਹਾਜ਼ਾਜ਼ਰ ਦੇ ਪਾਹਜ মাজ ১৯ বনশ থার জা বি মনীনাগর র' এবেন মন্তর্জার রাল চিনিসাজ দে মুন কনের মন্ত্রায় বিরুদ্ধ বিরাগজ্ঞের নি বিরাগ বাদেশে বিরা বিরাগ পায়ু বিষাগজ্ঞের নি বিরাগ বাদেশে বিরা বিরাগ পায়ু বিষা ও মে বিরাগী ঘটনার ও বিরাগ বিরাগ মন্ত্রায় বাদ বিরাগি বিরাগ বিরাগে বিরাগে নি বির্বাগন বিরাগে বিরাগে বিরাগে বিরাগে বিরাগে সের নিয়া মান বিরাগে বিরাগে বিরাগে বিরাগে বিরাগ বিরাগে বিরাগে বিরাগে বিরাগে বিরাগে বিরাগ বিরাগে বিরাগের বিরাগে বিরাগের বিরাগে বিরাগে বিরাগে বিরাগে বিরাগে বিরাগেরে মান্টার বিরাগে বিরাগে বিরাগে বিরাগে ব্যক্ত বিরাগে বিরাগে বাদেশের মার্বেরের মন্ত্রারা মন্ত্রায় বিরাগেরে না বিরাগের বিরাগে বিরাগে মন্ত্রায় মন্ত্রায়।

5 95% ਹੈ ਘਟੀ ਕੇ ਵਾਵੇ ਨਾਲ ਸਹਿਲਾ ਖ਼ਰੀਮਰ ਸ਼ੁਰਿਨ ਦਾ ਸੁਖਨਾ ਲਿਆ, ਪਰ ਗਰੀਜੀ ਨੇ ਜਾਤੀ ਵਿਕਰਧਿਆਂ ਦਾ ਗੱਸ ਜਾਨ ਇਨਸਾ, ਪਰ ਗਰੀਜੀ ਨੇ ਜਾਤੀ ਵਿਕਰਧਿਆਂ ਦਾ ਗੱਸ ਜਾਨ ਇਨਸਾਰ ਦੀ ਪਰਿਸ਼ ਦੇਵ ਹੀ ਸ਼ੋਹਾ ਨਿੱਤੀ। ਇਹ ਵਾਹਤੀ ਬਹੁਤ ਦੀ ਸ਼ਹਿਤੀ ਆਵਾਰਤਾਤੀ ਘਟਾ ਨੇ ਜੁਸ਼ ਨੇ ਜਾਤੀ ਪਿਤਰ ਦੇ ਕੀ ਹੁਟੀ ਹੈ ਵਿਲੇ ਦਾ ਸੁਆਰਾ ਲੈਨਨਾ ਨਾਲ ਗਸਵੀਨ ਇੱਥਾ ਹੈ ਜਾਨਿਕਟਾਰਜੇ ਨੇਕਰਾ ਨੇ ਨਾਲ ਗਸਵੀਨ ਇੱਥਾ ਹੈ ਜਾਨਿਕਟਾਰਜੇ ਨੇਕਰਾ ਨੇ ਨਾਲ ਗਸਵੀਨ ਇੱਥਾ ਹੈ ਜਾਨਿਕਟਾਰਜੇ ਨੇਕਰਾ ਨੇ ਨਾਲ ਗਸਵੀਨ ਦਿੱਸਾ ਹੈ ਜਾਨਿਕਟਾਰਜੇ ਨੇਕਰਾ ਜਿੰਦਾ ਹੈ ਉਂਜਾਡੇ ਤੋਂ ਦੁੜ੍ਹਾ ਮਾਡੀ ਜ਼ਿੰਦਗੀ ਦੇ ਤਾਲਾ ਮਾਰਸਾਸ ਨਾਲ ਇਨ੍ਹਿਆ ਤੇਣੂ ਗਾਂਦਰਾ ਹੀ ਸ਼ਹੁਣ 'ਬਾਰ ਮਾ

ਤੇ ਕਵਿਤਾ ਉਸ ਲਈ ਤੁਹੋਮਲ ਨਹੀਂ ਸੀ। ਟਾਵਿ ਸਿਰਜਣਾ ਉਸ ਲਈ ੍ਹਹਮਲ ਉਸ ਕਈ ਅਤੇ ਸਵੇ ਚ ਉਸ ਨੇ ਸ਼ਸ਼ਾਲਾਂ ਵਿੱਚੋਂ ਦੇ ਨਹੀਂ ਜੀ। ਕਾਵਿ ਸਿਰਜਣਾ ਉਸ ਲਈ ਹਾਲ ਸੱਚੇ ਦਾ ਆਰਾਨ-'ਚੈਂਡਨ, ਲੂਹ ਤਰੇਬੀ ਜੱਤ ਹੈ ਸਭੋ ਨਾਲ ਭੇਜ਼ ਸੀ। ਸਾਢੇ ਦੂ ਚਾਣ ਦਹਾਇਆਂ ਵਿੱਚ ਉਸ ਨੇ ਸ਼ੌਰ ਸ਼ਾਵਿ ਸ਼ੀਵਰਿ ਸ਼ਾਸਿਆ ਕਰਕੇ ਵਿਕੇਸ਼ੀ ਪ੍ਰਬਾਦ ਹਾਲੇ ਪ੍ਰਤਿਕੀਰ ਸ਼ਰੀ ਵਜੋਂ ਆਸ਼ਸ਼ ਹੈ – ਹੁੰਦੀ (2000) भगत (मिट्र" (1975), 'धनायम खंदन पुनर्स' (1988), 'भारतीय ज मेरी घट फिली' (1996), पुलनाज ह গ্রহেশার র মনী হার ছিলো" (২০০১, গ্রলারার র ৬ব্যার লাগ (১০০৫) জুচিয়া ন মনীর (১০০৫) চারে भेगे जिमे से सार्व्य (2011) भ

5.96% _{etal} Nan 8 k a 418 etal 1990 ਸਿਵਾਰਪਾਤ ਕਰ ਉਸ ਹੀ ਰਾਜ ਦੇ ਖਾਰ ਵਿਚਾ ਕੁਰ ਹੁੰਦੀ ਹੈ। ਬਾਰਟੀ ਤੋਂ ਪੱਛਮੀ ਦਾਰਗਨਿਸਾਂ, ਚਿੰਗਲ ਬਰੰਪਬਾਬਾ, ਪੰਰਾਇਸ ਸਾਇਰ ਅਤੇ ਰਾਜਇਨ, ਮੁਵਾਬਾਰਲ ਦੀਆਂ ਬਸ਼ਨੇਸ਼ਾਂ ਮਿੱਬਾ ਦੇ ਹਵਾਲੇ ਉਸ ਦੀ ਹਵਿਡਾ ਵਿਚਾ ਸ਼ਾਇ ਹੋਏ ਨਸ ਬਾਵਾ ਬਰਵੰਡ ਕਿ ਸ਼ਰਬਾਦ, ਖ਼ਾਰਦ ਨਾਲ ਨਾਲ ਦਾ ਨਾਲ ਦਾ ਨਾਲ ਮਾਰਾਡਰ ਦੀਆਂ ਅਸ਼ੇਨਾਂ ਜਿੱਥਾਂ ਦੇ ਦਰਾਲੇ ਉਸ ਦੀ ਨਵਿਡਾ ਵਿਚ ਨਾਲੇ ਹੋਏ ਹੋਏ ਨਾਲ ਬਾਕਾ ਬਰਕੇਡ ਹੋ ਪੰਜਾਬੀ ਸ਼ਵਿਤਾ ਨੂੰ ਡਾਟਰੀ ਅਵਚੇਤਨ ਰਾਸ ਸਿਸਤਾਰਾ ਹੈ। ਨਿਗਾਰੇਜ਼, ਤੇਨੂ ਹੁਰੇਰ ਵਾਜਸੀ ਸਦੀ ਸਿਸਤਾਰਾ ਹੈ। ਨਿਗਾਰੇਜ਼, ਤੇਨੂ ਹੁਰੇਰ ਵਾਜਸੀ ਸਦੀ ਸਾਲ ਪ੍ਰਤੀਬਾਰਾ ਨਾ ਗਿਆ। ਤੇ ਆਪਣੀ ਕੱਧ ਦੀ ਨਹੀਂ ਨਿਰਾ ਉਹ ਆਪਣੇ ਬਹੁਤੇ ਸਮਕਾਲੀਆਂ ਵਾਗ ਬਾਲੀਲਾ ਕਵ ਨਹੀਂ ਸੀ। ਉਸ ਨੇ ਬਣਜਾਲੀ ਪ੍ਰਗਰੀਵਾਰੀ ਮੰਗਤੀ ਜਵਿਤਾ ਦੇ ਮਮੀਰ ਨੂੰ ਹੋ ਸਦਾਲਿਆਂ ਨਹੀਂ, ਤਾਂ ਉਸ ਦੇ भित्राप्त हु उद्धा तेते. ततुन स्टित्य विषयवयप्रात प्रावजीनी अज्ञाल अवीसे अने

हित्तावीपांच गावतीले क्वांत क्वांत के ति के राजीत संभारू राज से तिस्वीधांचन जवते देखू में कविवा स्वयाई देंद के आवंतित, राजरी, प्राराम, संविधायणक पर्य केंग्रेज सावीकर 5 वेतावसी राज द्वार्थनिय पूर्वों के 1 दिल्लाक्या सावस के वाजरातील मुराम 40 दिन से बायारी पांच कर्मन संविध रिया विवाल कर पाछ का भोरती है में देखी कि फेजरी यो बादिवन स्वर दुन दिसाबीता के प्रारा के प्राराम केंग्रे की से सुधी राज का सावसावरणी किस्सार के देखा की स्वराम प्रारा प्रसित्त का द्वार साव प्रतियोग कि स्वराम प्रारा प्रसार कि प्राराम कर दिसावर के ब्रेज की सावना जे साव प्रसार किये प्रतियाग के ब्रेज की

5 96% & NO TER TREATS & SURTO ਵਿਚ ਉਹ ਵਿਸ਼ਾਬੀਨਰ, ਕਾਰਤਵਾਦ ਤੋਂ ਉਪਰੋਗੇ ਜੀਇਜਾਰਾਰ ਨੂੰ ਸਾਰੇਘਟਰੀ ਅਨਾਲ ਹੋਏ ਹਾਇ ਕਿ ਮਹੁੰਸ਼ ਕਾਰਪੋਰਲ ਦੀ ਆਵਾਦਾ ਪੁੱਜੀ ਦੁਸ਼ਖਾਦਾ ਸੰਚਾਇਨ ਕਾਰਪੋਰਲ ਦੀ ਆਵਾਦਾ ਪੁੱਜੀ ਦੁਸ਼ਖਾਦਾ ਸੰਚਾਇਨ ਸਾਂਚਰਨਵਿਧੂਣੀ ਤੀਰ ਵਿਚ ਹਰਨ ਵਿੱਠਾ ਨੇ ਨਰਨਾਨ ਦੇ ਸੰਚਰਨਵਿਧੂਣੀ ਤੀਰ ਵਿਚ ਹਰਨ ਵਿੱਠਾ ਨੇ ਨਰ ਸੰਚਰਨਵਿਧੂਣੀ ਤੀਰ ਵਿਚ ਹਰਨ ਵੱਡਾ ਨੇ ਨਰ ਵੱਕ ਨੇ, ਕੋਸਟ ਤੋਂ ਪੁਲਬਾਂ ਨੂੰ ਜਿੱਤਨ ਲਈ ਪੂਜੀ ਤੇ ਬਾਹਬ ਦਾ ਸ਼ਲੋਕਾ ਦੀ ਜੀ ਤੇਜ ਵਿਚਾਉਂਦਾ ਹੈ ਆਪਰਮੱਟਾਂ (ਉਪਰਾ ਡੇ ਪਰੋਡ) ਸ਼ਰਿਕਾ ਸਿੱਚ ਹ ਸੁਲਸਾ ਹੈ। 'ਮੈਂ ਤੇ ਗਲੋਕਾ ਦਿਸਾਰੇ ਤਿੰਬਾ ਲਿਖਾਰ ਹੈ contractives his and the mount of आजालन ठोजोआ फ्रिस कुभ नराभव के वस्त्रो अने जुननजी प्रत्याहर दिधन नजन रूपी प्रेराधनीका नी किंग अन्द्र ताल बॉफर्स क्रियोस्ट ਸਟਾਪਾਰਕਾ ਦਾ ਜਿਸਾ ਮੇਟ ਗੱਲ ਵਧਦਾ ਮੁਹਾਸ਼ਮ ਵਾਨੇ ਸ਼ਲੀਬ ਬਾਜਦ ਦੇ ਸਿੰਘਨ ਸਾਡੇ ਸਿੰਘ ਰਾਖੀ ਲੋੜ ਦਾ ਰਾਇ-ਪਾਰਤ ਕਦੇ ਇਰਲਕਦ ਵਾਂਗ ਲੱਗਣਾ ਹੋਣ ਤੋਂ ਇਤਸਾਦੀ ਕੇ, ਵਦੇ ਜੀਤਰੇਤਾ ਦੇ ਕਵੀਬ ਖਾਗ ਤਰਮ ਕੁਆਰ ਬਦਦਾ ਦੇ ਅਤੇ ਬਦੇ ਆਪਣੇ ਜਿੰਦੇ ਦੇ ਦਰਸੇ ਕਿਹਾ ਨੂੰ ਆਪਣੇ ਇਹ ਕੀਤੇ ਦਾ ਜਿੰਦੇ ਕੋਰੇ ਨੂੰ ਆਪਣੇ ਇਹ ਕੀਤੇ ਦਾ ਜਿੰ-ਕੋਰੋਨਾਡੀਨ ਕਾਰਤਸ਼ਨੀਆਂ ਇਹਾਂ ਦੇ ਕਾਰ ਸੋਹੇ ਵਰੋਂ/ਜਾਡੋਕਾ/ਜਵਾਨਾਮਵੀਂ? ਵਿਹਾਂ ਤੋਂ ਕਾਂਤ ਜੱਤੇ

रिवरोप सकारपुध्धी ताग्वा जूं तावी ज्यं अर्थे मेंने ड्रॉडक जेभ उंडे स्वान्त/प्रवच प्रेडां कार-ग्व... WHERE COLCO DESCE

Figure 7. Layout Detection Results on Punjabi Newspaper Image. Different blocks of text and the percentage of detection and label are shown



Figure 8. Layout Detection Results on Punjabi Newspaper Image. Different blocks of text and the percentage of detection and labels are shown.

We tested the model inference on different kinds of layouts by removing images or instances. In the case of X101-FPN, the model achieved an accuracy of 95.53%, which means it correctly detected 95.53% of the actual text columns. These metrics provide insights into the performance of the object detection models in identifying text columns within the dataset. A high accuracy percentage indicates that the model is proficient at correctly recognizing text columns, while lower accuracy percentages may suggest areas for improvement in the models or the dataset. The sample images detected are shown in Figures 7 and 8.

Conclusion

In the realm of newspaper digitization, particularly in the context of the Gurumukhi script, we presented a comprehensive approach to newspaper layout analysis for digitizing newspapers in the Gurumukhi script. This approach is critical to making historical newspapers accessible and user-friendly for modern audiences. We proposed an algorithm to effectively remove graphics and pictures from Puniabi newspaper images, achieving an accuracy of 96.22% on a test set of 100 images with diverse layouts, including non-Manhattan layouts. This algorithm also works on other newspaper scripts, like Hindi and English. The second part of the study focused on training a deep learning model, Faster RCNN, with different architectures to detect columns of text in newspapers. The dataset, consisting of 500 labeled Punjabi newspaper images, was used for training and testing. The model achieved high accuracy, with the bestperforming architecture (X101-FPN) reaching an accuracy of 95.53% in detecting text columns. Evaluation metrics such as precision, recall and F1 score were employed to assess the performance of the deep learning models. The X101-FPN architecture demonstrated superior performance across these metrics, indicating its effectiveness in accurately identifying text columns within the newspapers.

The proposed methodology contributes to the field of newspaper digitization, particularly for languages like Gurumukhi, by addressing the challenges posed by complex layouts. The visual examples highlight the robustness of the proposed system for handling the complex layouts of Punjabi newspapers. The algorithm effectively removes graphics, and the deep learning model accurately detects different text regions within the newspapers. The qualitative assessment emphasizes the system's ability to handle diverse layouts, including nonmanhattan, and showcases its potential for digitizing newspapers in the Gurumukhi script. The combination of algorithmic techniques and deep learning models presented in this paper provides a robust solution for accurate layout detection, facilitating the digitization of historical newspapers for enhanced accessibility.

Future work could involve further refining the algorithm and model, exploring additional architectures, and expanding the dataset for improved generalization. Overall, this research lays the foundation for advancements in digitising newspapers in diverse scripts and layouts.

Conflict of Interest

The authors declare no conflict of interest.

References

- Alshameri, A., Abdou, S. M., & Mostafa, K. (2012). A combined algorithm for layout analysis of Arabic document images and text lines extraction. *International Journal of Computer Applications*, 49(23), 30–37. https://doi.org/10.5120/7945-1282
- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A. Q., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8(53). https://doi.org/10.1186/s40537-021-00444-8
- BinMakhashen, G. M., & Mahmoud, S. A. (2019). Document Layout analysis. ACM Computing Surveys, 52(6), 1–36. https://doi.org/10.1145/3355610
- Biswas, S., Riba, P., Lladós, J., & Pal, U. (2021). Beyond document object detection: instance-level segmentation of complex layouts. *International Journal on Document Analysis and Recognition* (*IJDAR*), 24(3), 269–281.

https://doi.org/10.1007/s10032-021-00380-6

Chowdhury, S. P., Mandal, S., Das, A. K., & Chanda, B. (2007). Segmentation of Text and Graphics from Document Images.*International Conference on Document Analysis and Recognition ICDAR*, *Curitiba*, *Brazil*, 2, 619-623.

https://doi.org/10.1109/icdar.2007.4376989

Kosaraju, S., Masum, M., Tsaku, N. Z., Patel, P., Bayramoglu, T., Modgil, G., & Kang, M. (2019). DoT-Net: Document Layout Classification Using Texture-Based CNN. 2019 International Conference on Document Analysis and Recognition (ICDAR), Sydney, NSW, Australia, pp. 1029-1034.

https://doi.org/10.1109/icdar.2019.00168

Int. J. Exp. Res. Rev., Vol. 35: 34-42 (2023)

- Li, M., Xu, Y., Cui, L., Huang, S., Wei, F., Li, Z., & Zhou, M. (2020). DocBank: A Benchmark Dataset Document Layout Analysis. for arXiv. https://doi.org/10.18653/v1/2020.coling-main.82
- Liebl, B., & Burghardt, M. (2021). An Evaluation of DNN Architectures for Page Segmentation of Historical Newspapers. 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, pp. 5153-5160.

https://doi.org/10.1109/icpr48806.2021.9412571

- Oliveira, S. A., Séguin, B., & Kaplan, F. (2018). dhSegment: A Generic Deep-Learning Approach for Document Segmentation. Arxiv. https://doi.org/10.1109/icfhr-2018.2018.00011
- Patil, S., Vijayakumar, V., Mahadevkar, S., Athawade, R., Maheshwari, L., Kumbhare, S., Garg, Y., Dharrao, D., Kamat, P., & Kotecha, K. (2022). Enhancing Optical Character Recognition on Images with Mixed Text Using Semantic Segmentation. Journal of Sensor and Actuator *Networks*, 11(4), 63.

https://doi.org/10.3390/jsan11040063

- Sauvola, J., & Pietikäinen, M. (2000). Adaptive document image binarization. Pattern Recognition, https://doi.org/10.1016/s0031-33(2). 225-236. 3203(99)00055-2
- Soma, S., & Shilpa. (2022). A Novel Approach for Newspaper Block Segmentation using Run-Length Smoothing Algorithm. 2022 IEEE 2nd International Conference on Mobile Networks and

Wireless Communications (ICMNWC), Tumkur, Karnataka, India, pp. 1-6.

https://doi.org/10.1109/icmnwc56175.2022.10031933

Vasilopoulos, N., & Kavallieratou, E. (2017). Complex layout analysis based on contour classification and operations. morphological Engineering Applications of Artificial Intelligence, 65, 220-229.

https://doi.org/10.1016/j.engappai.2017.08.002

- Wu, X., Ma, T., Du, X., Hu, Z., Yang, J., & He, L. (2023). DRFN: A unified framework for complex document layout analysis. Information Processing and Management, *60*(3), 103339. https://doi.org/10.1016/j.ipm.2023.103339
- Xu, C., Shi, C., Bi, H., Liu, C., Yuan, Y., Guo, H., & Chen, Y. (2021). A page object detection method based on mask R-CNN. IEEE Access, 9, 143448-143457.

https://doi.org/10.1109/access.2021.3121152

- Zhu, W., Sokhandan, N., Yang, G., Martin, S., & Sathyanarayana, S. (2022). DocBed: A Multi-Stage OCR Solution for Documents with Complex Layouts. Proceedings of the AAAI Conference on Artificial Intelligence, 36(11), 12643–12649. https://doi.org/10.1609/aaai.v36i11.21539
- Zulfigar, A., Ul-Hasan, A., & Shafait, F. (2019). Logical Layout Analysis using Deep Learning. Digital Image Computing: Techniques and Applications (DICTA), Perth, WA, Australia, 1-5. https://doi.org/10.1109/dicta47822.2019.8946046

How to cite this Article:

Atul Kumar and Gurpreet Singh Lehal (2023). A Hybrid Approach for Complex Layout Detection of Newspapers in Gurumukhi Script Using Deep Learning. International Journal of Experimental Research and Review, 35(Spl.), 34-42. DOI: https://doi.org/10.52756/ ijerr.2023.v35spl.004



CO ISE This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.