C.V RAMAN GLOBAL UNIVERSITY BHUBNESWAR, ODISHA-752054

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

PATTERN RECOGNITION

A Case Study Report Submitted on STAAR

[CSE22243]

UNDER THE SUPERVISION OF :

Dr. Mamata Wagh

Text Extraction from Image using pattern recognition

❖ **SUBMITTED BY :**

NAME	**SAJAN KUMAR**
REGD. NO.	**2101020509**
ROLL.NO	**CSE21195**
BRANCH	**CSE**

❖ **TITLE OF JOURNEL PAPER:-** Intelligent character recognition using fully convolutional neural networks

- ✓ **NAMEOF AUTHOR:** Raymond Ptuchaa, Felipe Petroski Sucha , Suhas Pillai, Frank Brockler , Vatsala Singh , Paul Hutkowski
- ✓ **NAME OF JOURNEL:** Pattern Recognition
- ✓ **YEAR OF PUBLICATION:** 2019

C. V. RAMAN GLOBAL UNIVERSITY

ODISHA BHUBANESWAR

❖ DECLARATION

We hereby Declare that we have presented in a manner required for its acceptance for the partial ful-fillment for the Bachelor Degree of Technology for which it has been submitted. This approval does not necessarily endorse or accept every statement made, opinion expressed or conclusions drawn as recorded in this major project, it only signifies the acceptance of the major project for the purpose it has been submitted.

This is to certify that the work in this STAAR report entitled **"Text Extraction from Image using pattern recognition"** Submitted by **Sajan Kumar** for partial fulfilment of the requirements for the award of **Bachelor of Technology in Computer Science & Engineering** to **C. V. Raman Global University, Odisha.**

Dr. Mamata wagh
Assistant Professor
Department Of Computer Science & Engineering

ABSTRACT

This paper presents a fully convolutional network architecture for handwriting recognition, which outputs arbitrary length symbol streams from handwritten text. It normalizes input blocks to a canonical representation, avoiding costly recurrent symbol alignment correction. The method also introduces a probabilistic character error rate to correct errant word blocks when a lexicon is known. This multi-state convolutional method is the first to demonstrate state-of-the-art results on both lexicon-based and arbitrary symbol-based handwriting recognition benchmarks.

Unconstrained text recognition is a crucial computer vision task with various sub-tasks. Deep neural networks offer promise in convergent and automated feature extractors, allowing high performance with minimal domain knowledge. A data-efficient, end-to-end neural network model is proposed for generic, unconstrained text recognition. The architecture is a fully convolutional network without recurrent connections, trained with the CTC loss function, aiming for simplicity and efficiency without compromising recognition accuracy.

Despite the shift towards a paperless world, handwritten communications remain common. Traditional pen-on-paper or stylus-on-table input is often more convenient, efficient, or cost-effective. Many documents, like invoices, taxes, and questionnaires, have fill-in-the-blank fields that require digitization. Online handwriting recognition records stroke sequences.

INTRODUCTION

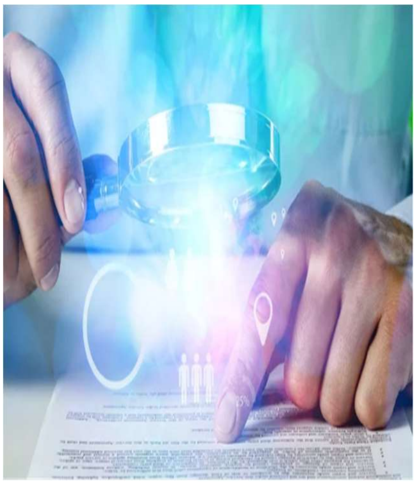
Handwritten documents, such as manuscripts, ancient texts, and books, still exist despite the dominance of electronic documents. While handwriting offers advantages, it is more difficult to access than electronic documents. However, if converted to an electronic transcript and properly indexed, information stored in handwritten documents can be easily retrieved. Text-line extraction (TLE) is a crucial part of document image processing, identifying lines of text for subsequent processing. TLE can include skew correction, text-graphics separation, word recognition, word spotting, character segmentation, script recognition, and recognition and indexing of handwritten texts.

Intelligent Character Recognition (ICR) is a method used to decipher digitized handwritten text, which is more challenging than Optical Character Recognition (OCR), which focuses on machine-generated imagery. ICR systems extract lines of text, segment them into word blocks, and then feed them into lexicon-based recognition systems or sequence recognition systems. Lexicon-based systems are accurate but limit output to words in a training set. Sequence recognition is challenging as it converts a string of symbols and blank predictions into a word without skipping or duplicating symbols. Post processing can improve results significantly when extra context is known, such as surrounding symbols, fixed lexicon, or phone number fields. Overall, ICR is a valuable tool for recognizing handwritten text.

RELATED WORK

This paper presents a character-based classification method that uses a common CNN architecture for word identification and accurate symbol prediction. It is the first Fully Convolutional Network (FCN) method to reliably predict arbitrary symbols and words from a label. The paper also introduces a probabilistic character error rate, creates a realistic block-based dataset from the NIST single character dataset, and presents the first Fully Convolutional method to demonstrate state-of-the-art results on lexicon-based and arbitrary symbol-based handwriting recognition benchmarks.

Offline handwriting recognition traditionally involves extracting input features from image data and using classifiers like Artificial Neural Networks (ANN) or Gaussian Mixture Models (GMM) to estimate posterior probabilities. These probabilities are then inputted into a Hidden Markov Model (HMM) for transcriptions. However, HMMs struggle to model long-term dependencies in input data. Random Neural Networks (RNNs) like LSTM units can address this issue, showing remarkable abilities in sequence learning tasks like speech recognition and video summarization.



Offline recognition in 2D images can be achieved using multidimensional RNNs or region proposal networks. Wigington et al. use a line-follower algorithm to trace handwriting across a page, fed into a CNN-LSTM network. The CTC algorithm allows for inputs without prior segmentation, avoiding the need for properly segmented labeled data. This approach addresses text detection and segmentation issues.

Pham et al. proposed Multidimensional RNN with dropout to improve offline handwriting recognition performance. Other researchers use convolutional neural networks and data augmentation for text recognition, character classification, and character segmentation. Deep CNNs have been used for languages other than English, and auto encoders are used for Telugu and Chinese recognition.

PROPOSED MODEL

This work focuses on extracted handwritten symbol blocks. An input is defined as a tightly cropped grayscale image of an arbitrary 1D sequence of symbols. We use the word symbol to emphasize that the model is not limited to Latin based characters.

▶ **FCN along with CNN, RNN**

- The algorithm consists of four consecutive stages which will be described next:

❑ **Symbol alignment**

- predicts the number of symbols in a block and resamples the block to a canonical width × height.
- convolutional layer =C(d, h × w, padh × padw)
  Where d=filters, h x w =spatial size, padding=padh x padw
- P(s)=s × s, where P(s)= pooling layer of stride, s=stride

❑ **Symbol prediction**

- It reduces a 32 × 16N input image to a (2N + 1) × 111 prediction
  Where N=desired output

(i) Even filter intuition

- even tap filter = pad of $Fw/2 - 1$,where Fw is the width of the filter.

(ii) Filter receptive field

- It finds which part of the image affects one pixel on the map.

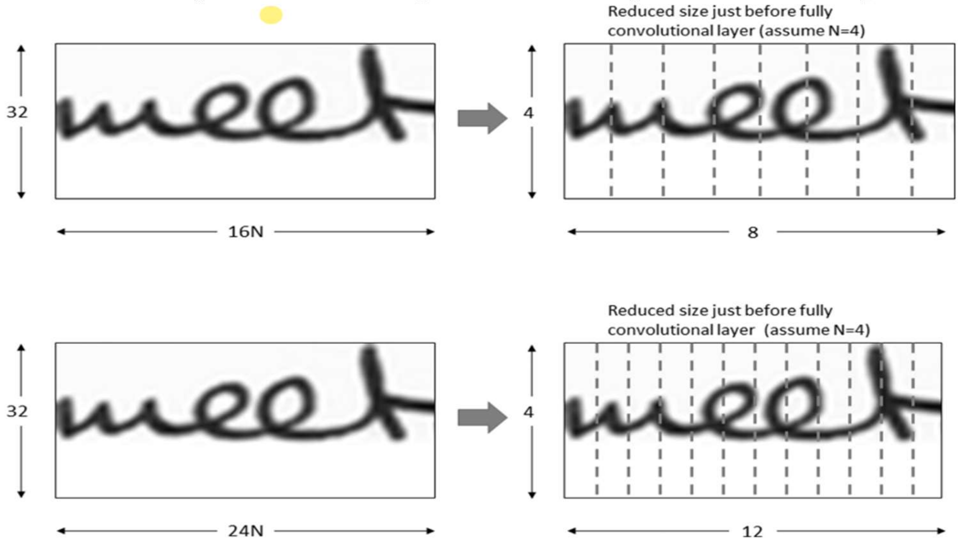


Fig. 3. Input sample representations (left) and their corresponding sizes after last pooling step (right) for input image representations of 32 × 16N (top) and 32 × 24N (bot) for a 4-symbol word block.

❑ CER and vocabulary matching

step-1: Normalized character error rate

$$\text{CER} = \frac{R + D + I}{R + D + I + C}$$

Where R = number of characters replaced, D =number of characters deleted, I = number of characters inserted, and C = number of correct Characters.

step-2: CER Computation

$$C_{i,j} = \min(C_{i-1,j} + 1, C_{i,j-1} + 1, \text{Diag})$$

$$\text{where: Diag} = \begin{cases} C_{i-1,j-1}, \text{ if } P_i = L_j & \text{where } p_i = i\text{th character of prediction and } L_j = j\text{th character of label} \\ C_{i-1,j-1} + 1, \text{ otherwise} \end{cases}$$

step-3: CER-based vocabulary matching

The probabilistic CER algorithm improves the top character prediction by using character probabilities, as described in Eq. (4), where $P(p_i = L_j)$ represents the probability of the character being equal to Lj.

$$C_{i,j} = \min C_{i-1,j} + 1 - P(p_i = L_j),$$
$$C_{i,j-1} + 1 - P(p_i = \text{blank}),$$
$$C_{i-1,j-1} + 1 - P(p_i = L_j)$$

The system predicts new words and reports the Word Error Rate (WER), indicating the average word-level accuracy of a system.

$W(p) = \arg\min_{L \in V} \text{CER}(p, L) + (1/1 + C(L))$, where C(L)= frequency of occurrence of a given word

❑ Lexicon based prediction

- Lexicon CNN can predict words from a given lexicon optionally.
- It helps find common words like "the", "her", and others.

PERFORMANCE EVALUATION AND DISCUSSION

The study employed moment optimization with a learning rate of 0.01 and 64-bit batch size, fine-tuning the Block Length CNN and Symbol Prediction FCN in a combined dataset containing 24 samples from the IAM, RIMES, and NIST Special Database 19 datasets.

DATASETS:

The study examines results on the IAM, RIMES, and NIST offline handwritten datasets. The IAM dataset contains 115,320 English words by 500 authors, including training, validation, and test splits. The RIMES dataset contains 60,000 French words by over 1000 authors, with each newer release superseding prior ones. The ICDAR 2011 release is used for all tests.

IAM RESULTS

Table 2
Comparison of results on IAM dataset to previous methods.

Model	WER	CER
Dreuw et al. [50]	18.8	10.1
Boquera et al. [51]	15.5	6.90
Kozielski et al. [44]	13.30	5.10
Bluche et al. [8]	11.90	4.90
Doetsch et al. [10]	12.20	4.70
Bluche and Messina [37]	10.5	3.2
Our work	8.22	4.70
Voigtlaender et al. [19]	9.3	3.5
Poznanski and Wolf [52]	6.45	3.44
Dutta et al. [17]	**4.80**	**2.52**

- **Prediction:**

Input	Label	Prediction
5ye-es	5Ye-es	5Ye-es
President's	President's	Preseciten's
Liverpool	Liverpool	livepool
up	up	eys
only	only	outle
'	'	,
the	the	the
Ifor	Ifor	Ifor
that	that	that

The system was tested on the IAM English handwritten dataset, achieving a CER of 4.70% (8.22% WER). The model is competitive against current dataset leaders, with most errors on lower case letters. Numbers and upper case letters have close to 99% accuracy.

RIMES RESULTS

Table 4
Comparison of results on RIMES dataset to previous methods.

Database	RIMES	
Model	WER	CER
Kozielski et al. [44]	13.70	4.60
Doetsch et al. [10]	12.90	4.30
Bluche et al. [8]	11.80	3.70
Voigtlaender et al. [19]	9.6	2.8
Bluche and Messina [37]	7.9	1.9
Stunner et al. [53]	7.84	2.53
Our work	5.68	2.46
Poznanski and Wolf [52]	3.90	1.90
Dutta et al. [17]	**1.86**	**0.65**

- **Prediction:**

Predictions obtained with the symbol sequence prediction model on the RIMES dataset.

Input	Label	Prediction
permet	permet	puent
vous	vous	vur
XEXGRS2.	XEXGR52	XEXGGRS2
commandées,	commandées	commandores
ces	ces	Cs
effet ,	effet	effett
département	département	tiprtement
salutations	salutations	salutations

Our model achieved a 2.46% CER on the RIMES dataset, outperforming current leaders. Table 4 shows performance against current leaders, using a vocabulary of top 500 words from the training set with a minimum confidence level of 70%.

DISCUSSION

The study reveals that generically trained models perform better than fine-tuned models for specific datasets. Vocabulary matching with probabilistic CER is more important than pre-processing with a Lexicon CNN, but both methods yield best results. All models can detect 110 unique symbols and output words are not limited to a pre-defined lexicon. The size of the lexicon used by the Lexicon CNN is application-dependent, with insurance forms requiring larger dictionaries and tax forms using smaller ones. Future research should explore larger lexicons.

This study demonstrates the use of fully convolutional methods to process various lengths of offline handwriting imagery. The streams are broken into parts, measured in length, and resampled to a canonical representation compatible with a fully convolutional network. This divide-and-conquer approach is input length agnostic and does not suffer from exploding or vanishing gradients. The method requires fewer computational resources but yields similar results. Unlike methods like Poznanski and Wolf, which use simple CNNs, the method of Dutta et al. uses both convolutional and bidirectional LSTM, test time augmentation, deslanting preprocessing, and CTC post processing. However, sequence-to-sequence methods, such as Battenberg et al. and Sueiras et al., are subject to gradient propagation issues when processing long streams of data.

CONCLUSIONS

This proposed model offers a novel character prediction strategy that uses a large symbol set without lexicon constraints, contrasting with the conventional recurrent neural network approach. This approach suggests an alternative to using RNNs followed by CTC for character cleanup. it suggested a new way to fix characters with recurrent neural networks, useful for speech recognition and similar tasks and it also used punctuation and Latin characters to recognize common words and symbol blocks, achieving advanced results on the English-based IAM and French-based RIMES datasets, despite avoiding recurrent neural networks and small symbol sets.

Ex: Amazon Textract is a one of the leading technology in this field. So, in Future research should expand is broad and promising, with potential applications across diverse domains, driving innovation in artificial intelligence and computer vision technologies. The study explores the use of larger contextual filters, sentence-level features like sent2vec, hierarchical processing, fully convolutional recurrent networks, and their application to various languages and resource-constrained applications, aiming to improve knowledge at various levels.

REFERENCES

[1] Ptucha, R., Petroski Such, F., Pillai, S., Brockler, F., Singh, V., & Hutkowski, P. Intelligent character recognition using fully convolutional neural networks. Pattern Recognition, Elsevier, 2019.

[2] Suman Kumar Bera; Soumyadeep Kundu; Neeraj Kumar; Ram Sarkar, Distance transform based text-line extraction from unconstrained handwritten document images . Expert Systems with Applications, (), –. doi:10.1016/j.eswa.2021.115666, 2021.

[3] Yousef, Mohamed; Hussain, Khaled F.; Mohammed, Usama S. Accurate, Data-Efficient, Unconstrained Text Recognition with Convolutional Neural Networks. Pattern Recognition, (), 107482–. doi:10.1016/j.patcog.2020.107482, 2020.

[4] K. Dutta, M.-M. Krishnan Praveen, C. Jawahar, Improving CNN-RNN hybrid networks for handwriting recognition, in: Frontiers in Handwriting Recognition (ICFHR), 2018 18th International Conference on, IEEE, 2018.

[5] A. Poznanski, L. Wolf, Cnn-n-gram for handwriting word recognition, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[6] P. Doetsch, H. Ney, et al., Improvements in rwth's system for off-line handwriting recognition, in: 2013 12th International Conference on Document Analysis and Recognition, IEEE, 2013, pp. 935–939 Computer Science from RIT in 2013. Ray was awarded an NSF Graduate Research Fellowship in 2010 and his Ph.D. research earned the 2014 Best RIT D

[7] T. Bluche, H. Ney, C. Kermorvant, A comparison of sequence-trained deep neural networks and recurrent neural networks optical modeling for handwriting recognition, in: International Conference on Statistical Language and Speech Processing, Springer, 2014, pp. 199–210

[8] https://en.wikipedia.org/wiki/Long_short-term_memory

[9] https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks/