C. V. Raman Global University
ODISHA BHUBANESWAR INDIA

Pattern Recognition

Text Extraction from Image using pattern recognition

Under the Guidance of:
Dr. Mamata Wagh

INTELLIGENT CHARACTER RECOGNITION USING FULLY CONVOLUTIONAL NEURAL NETWORKS

STUDENT NAME: **SAJAN KUMAR**

REGISTRATION NO: **2101020509**

Outlines

- INTRODUCTION
  - (i) Motivation
  - (ii) Model proposes
  - (iii) Application
- LITERATURE REVIEW
- TECHNICAL DETAILS
  - (i) Work analysis
  - (ii) FCN along with CNN, RNN
  - (iii) Results
- SUMMARY
- REFERENCES

Introduction

- A method to find accurate character and symbol.
- Finding and isolating handwritten symbols from a close-up grayscale picture.
- Sequence recognition over handwritten text recognition.

Motivation:

- Latest text extraction technology from images improves efficiency, accessibility, and innovation in real-life applications across industries.
- For ex: clinical decision support, Road Safety etc

Model proposed:

- character based classification.
- It predicts both arbitrary symbols as well as words from a lexicon.

Application:-

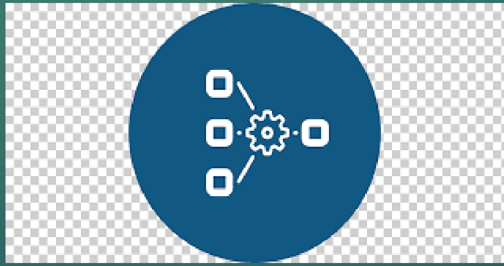
- Vehicle recognition and tracking
- Medical imaging and analysis
- Captcha solving

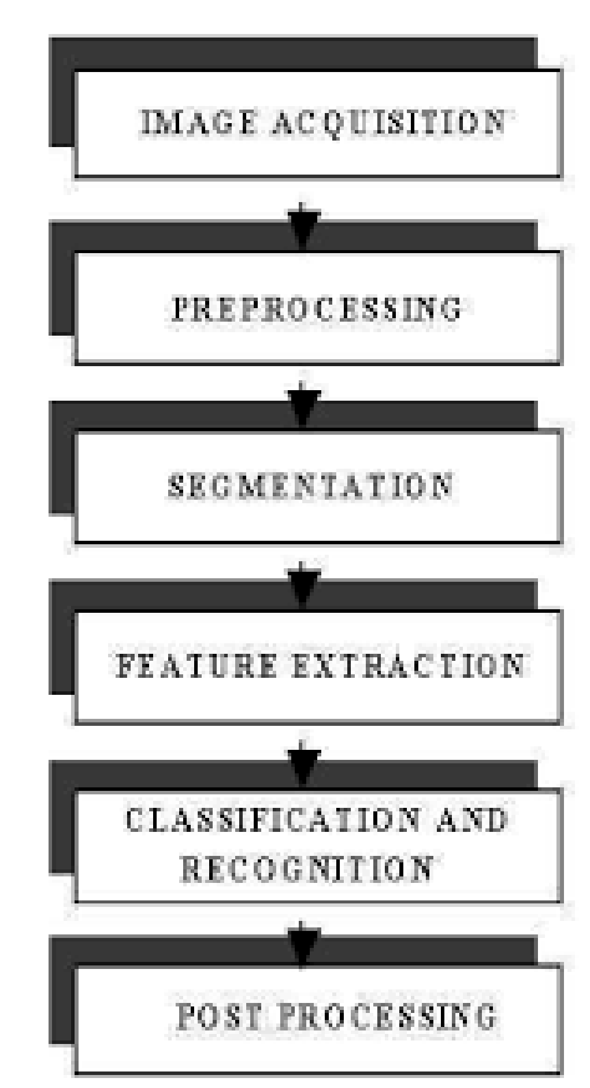
Literature Review

TITLE	AUTHOR	YEAR	ADAVANTAGE	LIMITATIONS
Distance Transform based Text-line Extraction from Unconstrained Hand-written Document Images	Suman Kumar Bera, Soumyadeep Kundu, Neeraj Kumar, Ram Sarkar	2021	Placing paragraphs in a one-page document has always yielded impressive results in most cases.	In multi-page documents, if there's a seperator between paragraphs, the method might mistake it for a new line.
Accurate, Data-Efficient, Unconstrained Text Recognition with Convolutional Neural Networks	Mohamed Yousef, Khaled F. Hussain, Usama S. Mohammed	2020	It worked very well on both short and long lines of text and also can handle different handwriting styles, sizes, and orientations with robustness.	Existing line recognition methods can't handle paragraphs or multiple lines without line segmentation algorithms.

Technical Details

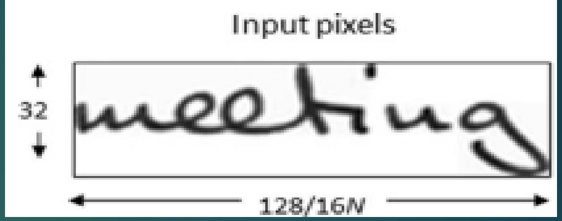
- Datasets are:
  - IAM
  - RIMES
  - NIST
- Work analysis:
  - Traditional feature extraction along with HMM
    - ANN or GMM
    - Long term dependency
  - Advanced feature extraction
- LSTM:
  - Naïve approach along with RNN
  - Multidimensional RNNs
- CNN-LSTM network
  - used CTC techniques





Technical Details

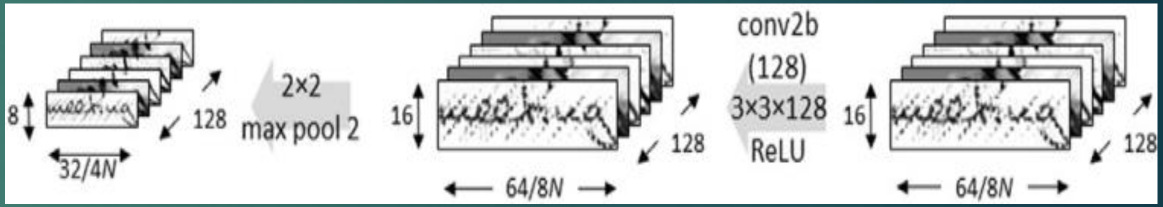
- FCN along with CNN, RNN
- The algorithm consists of four consecutive stages which will be described next:



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)

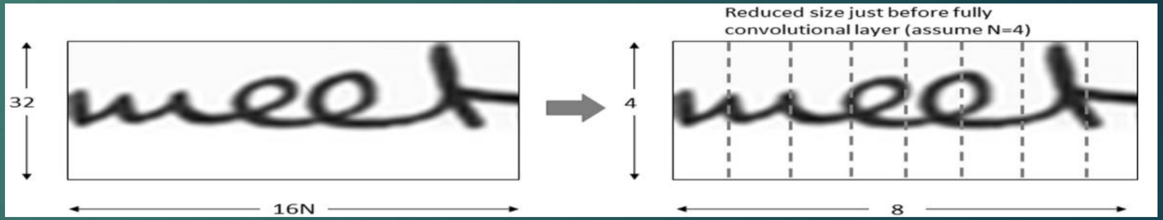
   Where d=filters, h x w =spatial size, padding=padh x padw
- P(s)=s × s, where P(s)= pooling layer of stride, s=stride



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.

Technical Details

CER and vocabulary matching

step-1: Normalized character error rate

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

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

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

where:

$$\mathrm{Diag} = \begin{cases} C_{i-1,j-1}, & \text{if } P_i = L_j \\ C_{i-1,j-1} + 1, & \text{otherwise} \end{cases}$$

where p_i = ith character of prediction and L_j = jth character of label

step-3: CER-based vocabulary matching

$W(p) = \arg\min_{L \in V} \mathrm{CER}(p, L) + (1/1 + C(L))$, 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.

Technical Details

- IAM Results:
- Prediction:

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

- Comparison

Comparison of results on IAM dataset to previous methods.

Model	WER	CER
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**

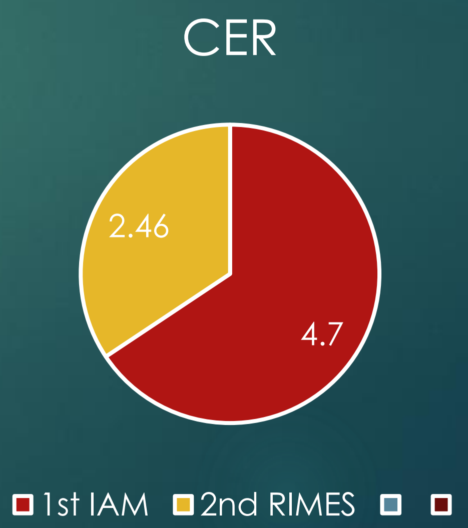
- RIMES Results:
- Prediction:

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

- Comparison

Comparison of results on RIMES dataset to previous methods.

Model	WER	CER
Voigtlaender et al. [19]	9.6	2.8
Our work	5.68	2.46
Poznanski and Wolf [52]	3.90	1.90
Dutta et al. [17]	**1.86**	**0.65**



- Comparison of results on RIMES and IAM:
  - (i) error-rate
  - (ii) accuracy
  - (iii) performance
  - (iv) probabilistic CER

Fine-tuned	Lex. CNN	Prob. CER	IAM	RIMES
X	X	X	**4.70(8.22)**	**2.46(5.68)**
X		X	5.05(8.62)	2.55(5.98)
X	X		6.50(18.30)	4.15(15.91)
X			7.09(17.77)	4.74(19.91)
			8.86(21.80)	5.03(20.05)

Summary

- The proposed model introduces a novel character prediction method, diverging from conventional recurrent neural networks.
- This approach suggests an alternative to using RNNs followed by CTC for character cleanup.
- It proposed a novel approach for character correction.
- It helps for recognize speech and symbols using punctuation.
- Without using RNNs , advanced results have been achieved on IAM and RIMES datasets.

- 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.

References

[1] Ptucha, R., Petroski Such, F., Pillai, S., Brockler, F., Singh, V., & Hutkowski, P. (2019). Intelligent character recognition using fully convolutional neural networks. Pattern Recognition, 88, 604–613. doi:10.1016/j.patcog.2018.12.017

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

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

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

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