

## **CS574: COMPUTER VISION USING MACHINE LEARNING**

### **Group 22: Handwritten Document Image Analysis**

Monika Khandelwal	184101026
Shavi Gupta	184101033
Shubham Jain	184101035
Tushar Geetey	184101039

## Proposed Method

### Model architecture

End to end detection and recognition of handwritten document consists of following steps:

#### Preprocessing Step

1. Binarization
2. Line Segmentation
3. Word Segmentation
4. Input Resizing and gray scale conversion

#### Recognition

1. Feature Extraction through CNN.
2. Feature Mapping through RNN.
3. Decoding text using CTC.

### Preprocessing

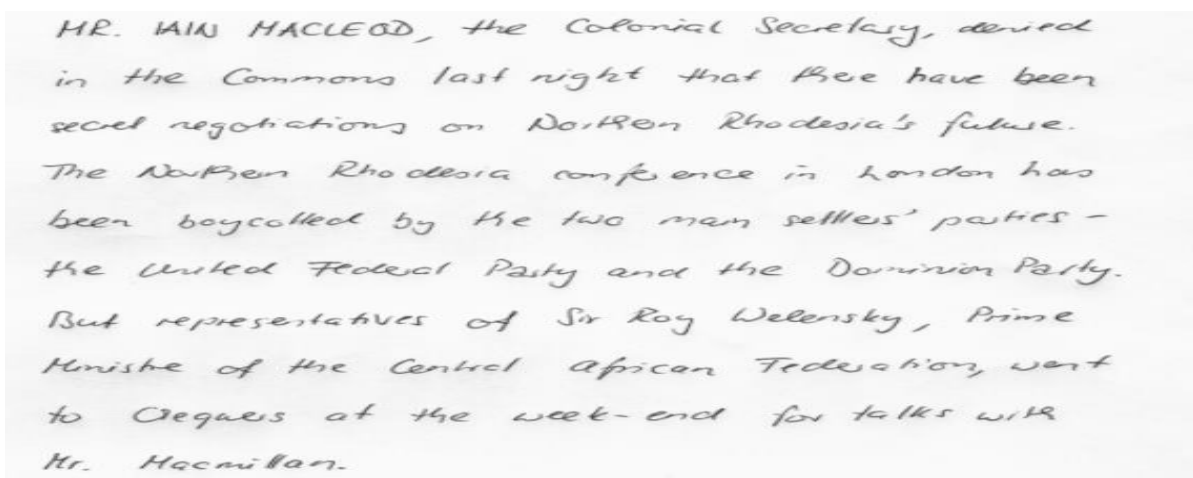
#### **Binarization**

The process of separating image foreground from page background during document image analysis is called Image Binarization. It removes stains or some faded ink marks from the background which helps in not only analyzing document but improves document readability as well.

For the Binarization purpose we have used Sauvola's adaptive document image binarization.

#### **Line Segmentation**

Given a handwritten document, to detect the words present in the handwritten document from the model we have created, the first step is to identify the region of words and before the words can be recognized, firstly line need to be determined of which the particular word is part of. This process of segmenting the document into various line is referred as Line segmentation.



**Fig. Input Document**

MR. IAIN MACLEOD, the Colonial Secretary, denied

in the Commons last night that there have been

secret negotiations on Northern Rhodesia's future.

The Northern Rhodesia conference in London has

been boycotted by the two main settlers' parties -

the United Federal Party and the Dominion Party.

But representatives of Sir Roy Welensky, Prime

Minister of the Central African Federation, went

to Cleveleys at the week-end for talks with

Mr. Macmillan.

**Fig. Output of Line Segmentation**

#### **Used Method**

Input: A Handwritten document. (We have taken forms and documents from IAM dataset)

Output: Document with separating lines and each line is stored cropped and stored separately. So that this can be fed to Word Segmentation directly. As shown in above Figure.

For this we have used the A\* path finding Algorithm explained above.

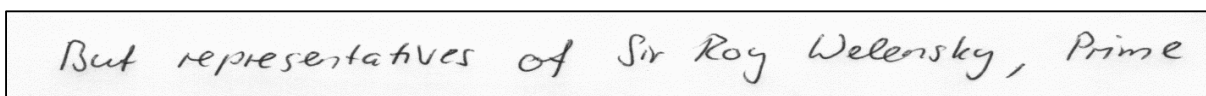
## Word Segmentation

After doing line segmentation on the handwritten document, we need to find word boundaries in each line, so that we can split the lines into meaningful words. Word segmentation is a process of finding meaningful words from a sentence or document.

Doing word segmentation on handwritten text document is complex because:-

- 1) Each character is written in distinct way.
- 2) There is not uniform spacing in handwritten text.
- 3) Scale problem (characters are of different size).

Input and Output

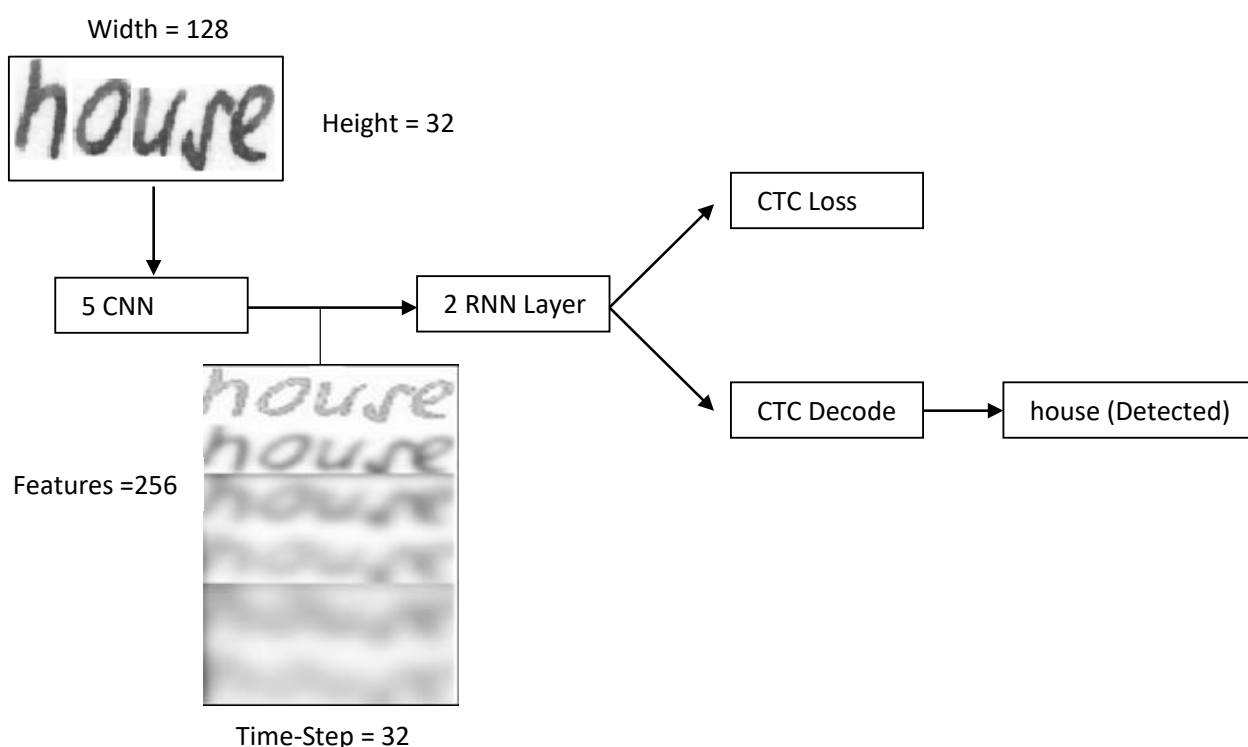


For this we have used method based on [scale space technique](#) described above.

## Input Resizing and Gray Scale conversion

The image obtained after word Segmentation may not necessarily be in desired size. That is why the image is first resized to either width of 128 or a height of 32 and then padding is done. After this Image is normalized for gray scale value.

## Recognition



**Fig.** A High level Architecture of Word

## Feature Extraction through CNN

1. After preprocessing the input image is  $128 \times 32$ .
2. It is fed into 5 layer of CNN which gives the feature map of size  $32 \times 256$ .

Three operation are performed on each layer:

- i. Convolution
- ii. Relu is applied
- iii. Max pooling.

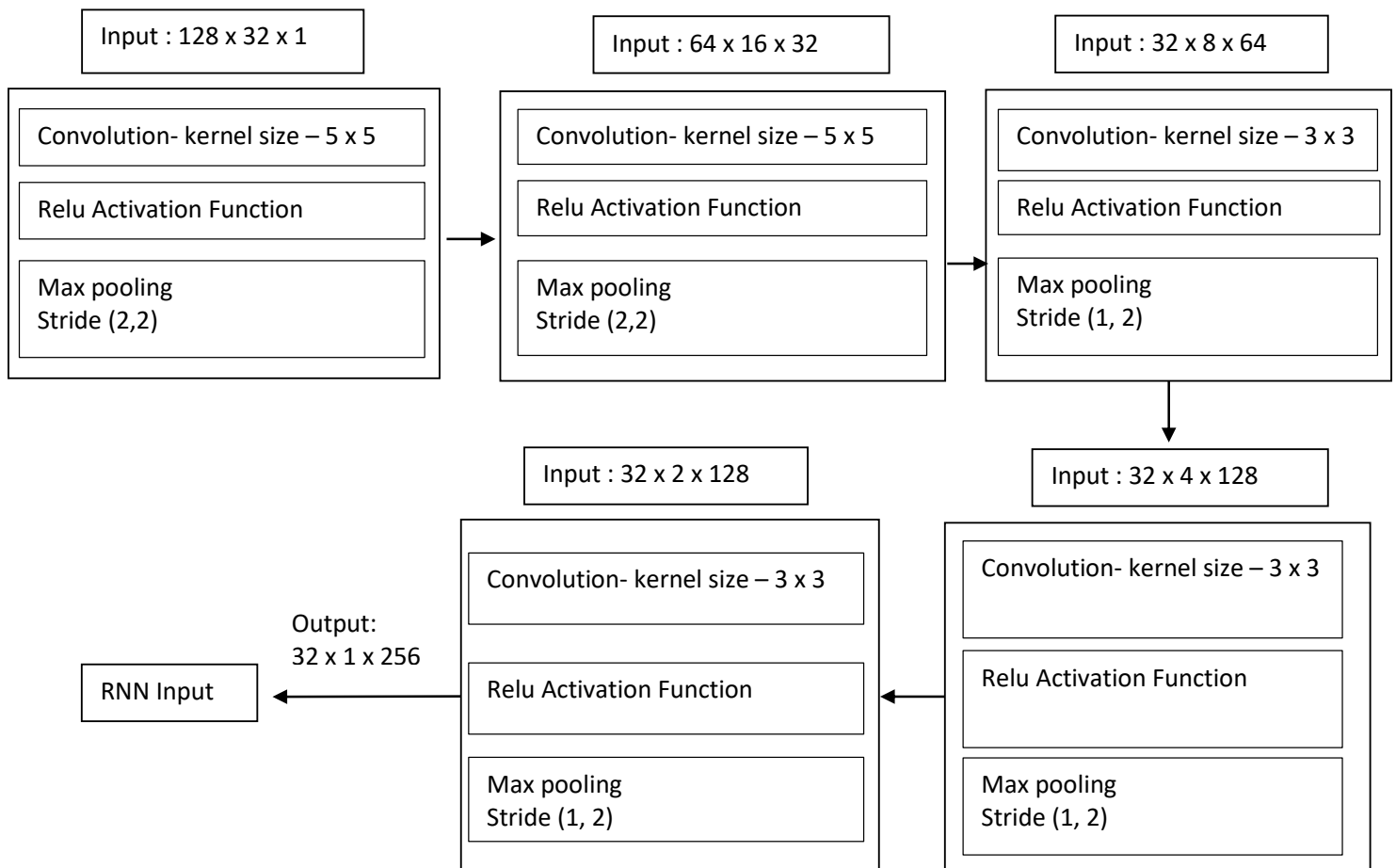
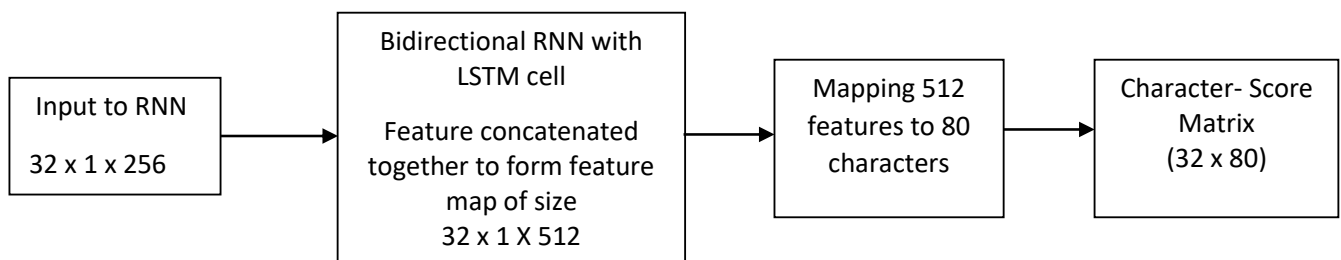


Fig. 5 Layer Architecture of Convolution Neural Network Architecture

## Feature Mapping through RNN

The feature map obtained from CNN is of size  $32 \times 256$ . RNN propagates useful information(character-scores) through this sequence and map it to a matrix of size  $32 \times 80$ . We have used bidirectional RNN using LSTM cell because information can be propagated through longer distance. The IAM dataset consists of 80 different characters that is why each time step (32 in no.) is mapped to 80 entries.



**Fig.** Flow through RNN layers

## Decoding text using CTC

The output of RNN is character score matrix. This is fed to CTC along with the ground truth value of text. There are two main operation of CTC.

1. Train – Loss value calculation to train Neural Network by trying all possible alignment of character sequence.
2. Infer – Decode RNN output to get the final text using Best Path Decoding or Word beam Search.

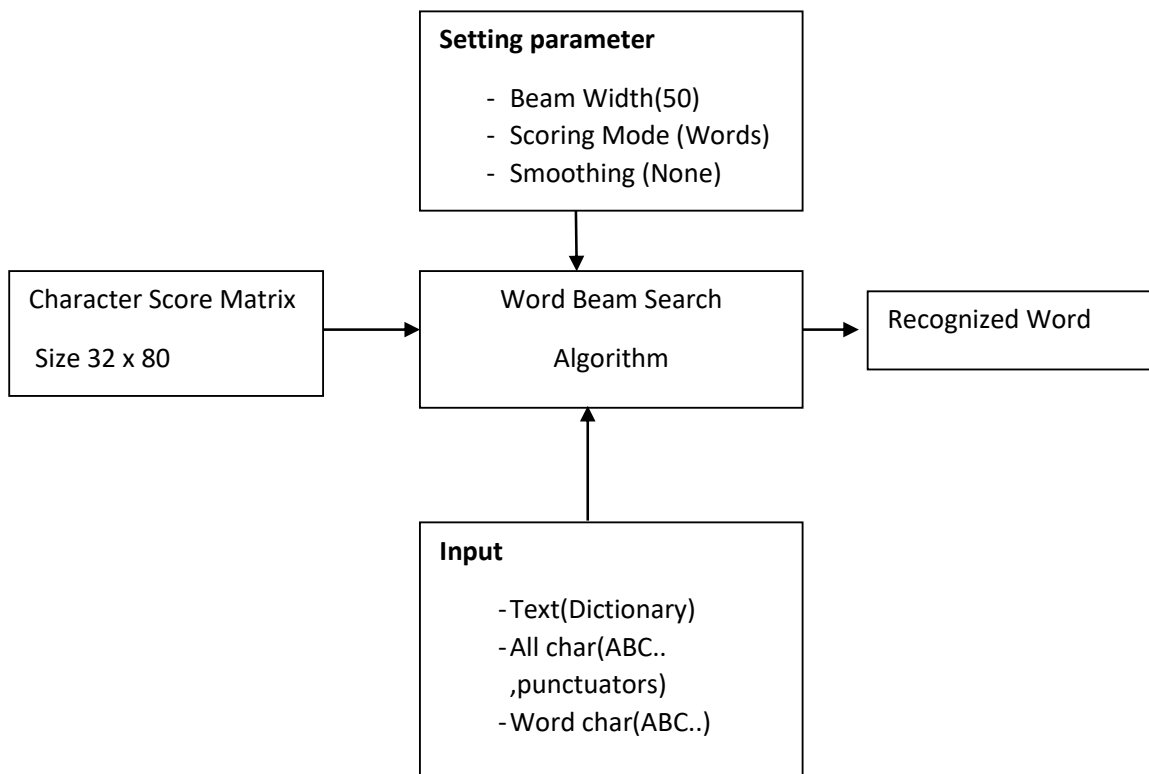
**Best Path Decoding:** It uses the output generated by the Neural Network

1. Take the most likely character at every time step.
2. Undoes the encoding by removing duplicate and blank space.

**Beam Search:** It explores a graph(remaining output) by expanding only the most promising node. Thus at each step most likely output are explored. Beam search allow arbitrary character to get detected.

**Token Passing:** It actually uses the dictionary and searches for the most probable sequence of words from dictionary. Main disadvantage is it cannot handle arbitrary sequences like number, punctuators and words are predicted from dictionary only. This avoid spelling mistakes.

**Word Beam Search:** It uses the advantage of Token passing and Beam search. Whenever the special character are detected, beam search is used otherwise Token passing technique is used. This help in removing the disadvantages of Token passing and Beam search.



**Fig.** Block Diagram showing working of Word Beam Search

## Results:

For input form image from IAM dataset following results are obtained as output from our recognition system:

The document is first segmented in lines followed by word segmentation. The result obtained by (Number of word detected correctly) using various decoder are stated below:

### Image1:

The Bow-wave Theory. This assumes that all fishing gear, when moving, sends before it a kind of scaring effect, probably through waves or vibrations in the water or along the ground. Underwater films suggest that the footrope of a trawl does this. Films have also shown plaice moving before a seine net in just the same way after being gathered inwards by the ropes.

Line Number	Number of words in line	Best Path	Beam Search	WordBeam Search
1	6	5	6	6
2	6	6	6	6
3	8	5	6	6
4	6	5	5	6
5	5	5	5	6
6	7	6	6	7
7	6	6	6	6
8	9	7	7	8
Total number of words	53	45	47	51
Accuracy		84.9%	88.67%	96.22%

Image2:

MR. IAIN MACLEOD, the Colonial Secretary, denied in the Commons last night that there have been secret negotiations on Northern Rhodesia's future. The Northern Rhodesia conference in London has been boycotted by the two main settlers' parties - the United Federal Party and the Dominion Party. But representatives of Sir Roy Welensky, Prime Minister of the Central African Federation, went to Oleguer at the week-end for talks with Mr. Macmillan.



Line Number	Number of words in line	Best Path	Beam Search	Word Beam Search
1	9	4	4	4
2	6	3	3	5
3	7	1	1	4
4	8	6	6	8
5	8	2	2	4
6	7	6	6	6
7	7	3	3	4
8	9	7	7	9
<b>Total number of words</b>	61	32	32	44
<b>Accuracy</b>		52.4%	52.4%	72.13%

Since our implementation is segmenting form into words and we use ground truth value of word instead of form, the accuracy is calculated manually for different forms from IAM Dataset.

## Conclusion

In this report we have proposed a method to convert the handwritten document into digital text. Firstly the document is segmented into lines, for this we have used A\* path finding algorithm with modified cost function. Secondly the lines are segmented into word by analyzing extent of blob using scale space. The segmented words are fed to CNN followed by RNN and CTC to recognize the words. Further we have used decoding algorithm word beam search that uses language model and dictionary to improve accuracy of the recognized word and also it allow non-arbitrary character like numbers or punctuator marks. Also we have analyzed various architectures, preprocessing steps to check for accuracy.