

If switch in XY position, the galvanometer current is zero. This means that the thermocouple is exactly equal in magnitude and opposite in sign to the standard cell. It has been shown that the 'thermocouple' and the standard cell have equal and opposite effects on the balance of e.m.f. will cancel each other. On connecting the galvanometer circuit.

At the points X' and Y', the standard cell is connected across the points c and b. If these voltages are not exactly equal, the galvanometer will indicate a deflection. By adjusting the resistor  $R_{cal}$ , the balance can be restored. This resistor  $R_{cal}$  is called the "calibration resistor" or "balance resistor". The position of the resistor  $R_{cal}$  is determined by the temperature of the standard cell. The position of the resistor  $R_{cal}$  is determined by the known temperature, since the temperature of the standard cell is proportional to the temperature of the "battery" in the circuit". The position of the resistor  $R_{cal}$  is determined by the known temperature, since the temperature of the standard cell is proportional to the temperature of the "battery" in the circuit". The position of the resistor  $R_{cal}$  is determined by the known temperature, since the temperature of the standard cell is proportional to the temperature of the "battery" in the circuit". The position of the resistor  $R_{cal}$  is determined by the known temperature, since the temperature of the standard cell is proportional to the temperature of the "battery" in the circuit". The position of the resistor  $R_{cal}$  is determined by the known temperature, since the temperature of the standard cell is proportional to the temperature of the "battery" in the circuit". The position of the resistor  $R_{cal}$  is determined by the known temperature, since the temperature of the standard cell is proportional to the temperature of the "battery" in the circuit".

which the temperature co-efficient of the standard cell is

# Correction of Annotation Affected Documents

---

Under Supervision of  
Dr. Sanjoy Pratihar

# Index

01

## Introduction

About The Problem Statement

02

## Past Work

Contributions made by researchers

## Proposed Methodology

Our Approach and Contribution

03

## Conclusion

Results and Conclusion from our approach

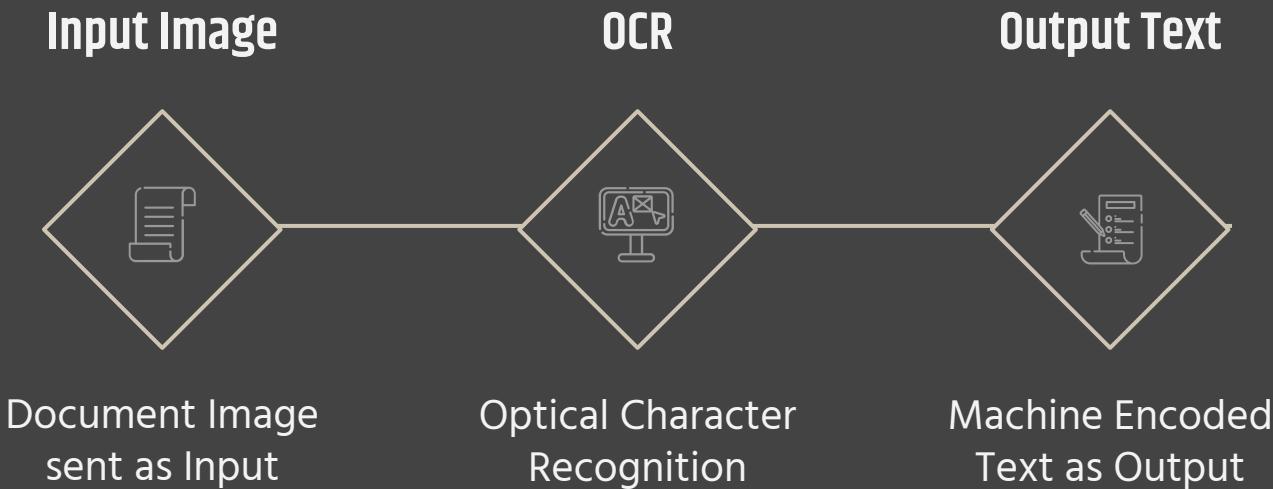
04

# Introduction

About The Problem  
Statement

01

# Working of an OCR



# The Problem

---

The text obtained by OCR systems often suffers from low accuracy owing to annotations

~~It was born and evolved from an experimental network (ARPANet) created by the US military back in the 1960s. As more and more computers have connected to this network, it has grown into the world's largest computer network.~~

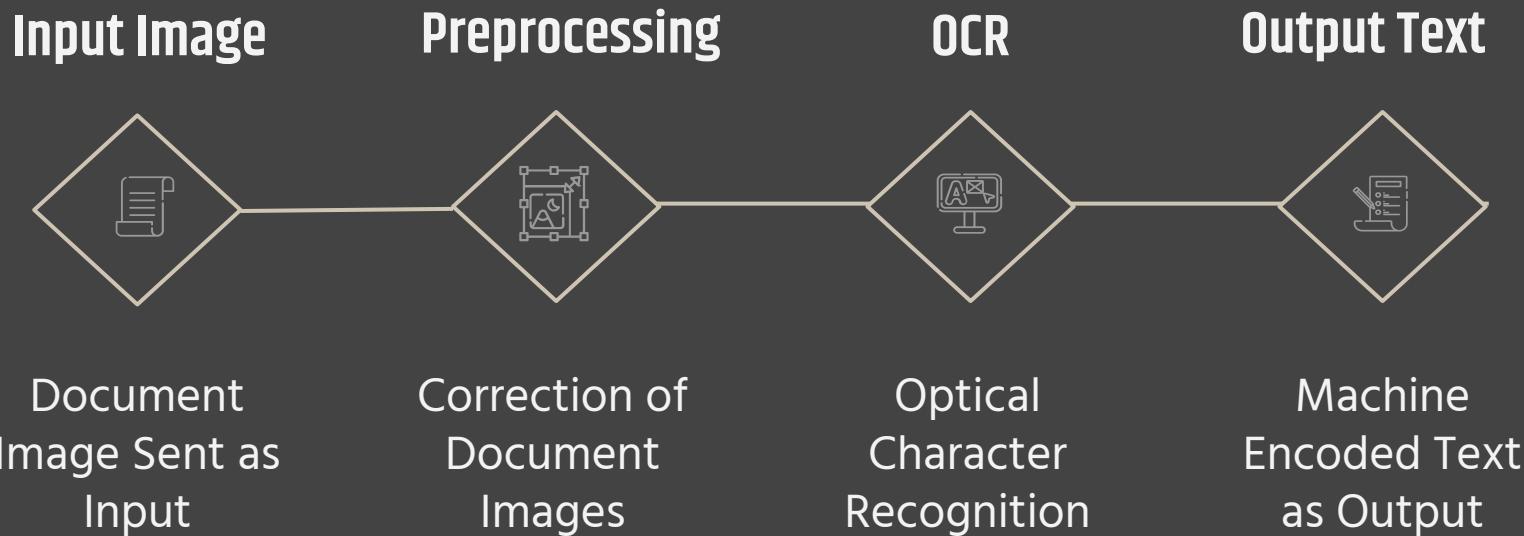
~~Today, the Internet consists of millions of smaller heterogeneous computers connected via a wired or wireless mode of communication. Most of the computers running various operating systems are able to connect to the Internet, one can receive and transmit data over the Internet. The Internet is the largest computer network in the world.~~

# Past Work

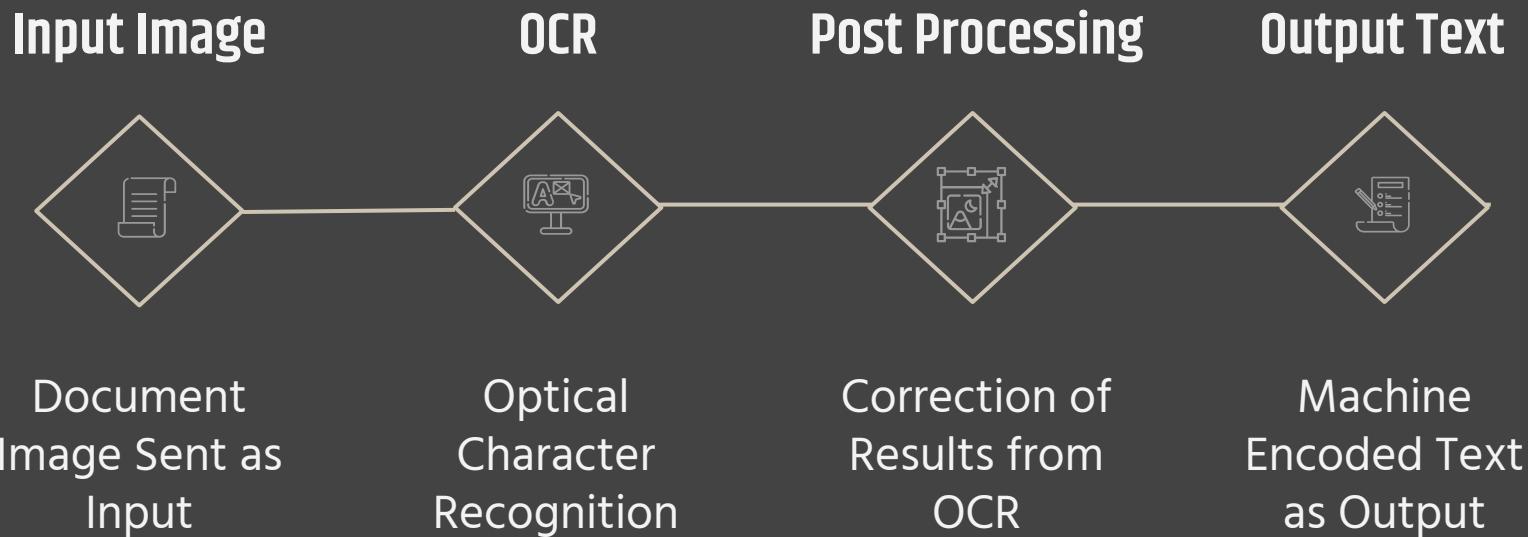
Contributions made by  
researchers

02

# Preprocessing of Annotation Affected Documents



# Post-processing of Annotation Affected Documents



# Past Work - Preprocessing

## Rovina and Seema (2016)

Removal of circular annotation, strikethrough lines and straight lines

## Pratihar et al. (2012)

Removal of hand drawn underlines in scanned document

## Adak and Bidyut (2014)

Graph based model for strikethrough removal

## Bai et al. (2004)

Three module approach for underline removal

## Oba et al. (2009)

Table ruled lines removal in business documents

## Pinto et al. (2004)

Handwritten underline removal in books

# Past Work - Postprocessing

**Evans D. et al.  
(1996)**

---

Error correction based  
on confusion mapping

**Atwell E. and  
Ellitm S. (1987)**

POS tagger and rules of  
English for capturing  
errors

**Mays and  
Damerau (1991)**

Word trigrams to  
correct real and  
non-word errors

**Golding R. et  
al. (1996)**

---

Better POS tagger by  
trigrams and Bayesian  
classifier

**Michael L. Wick  
et al. (2007)**

---

Use of a topic model to  
correct the OCR  
outputtext

**K. Kise et al.  
(1996)**

---

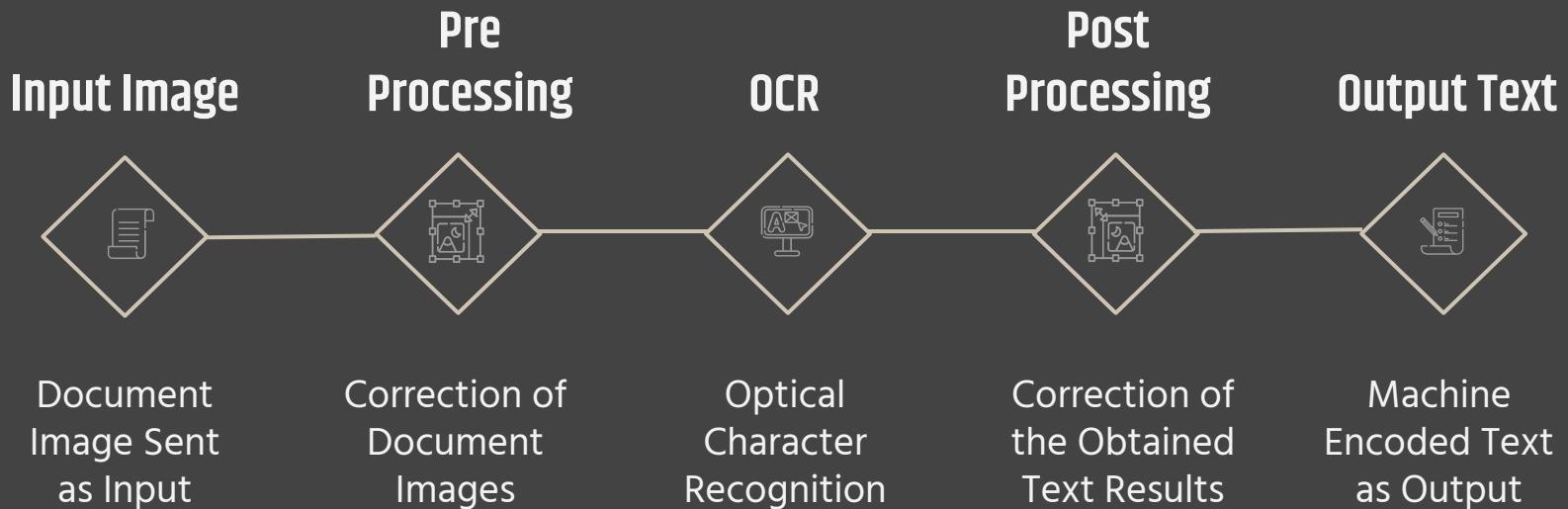
Approach based on  
syntactic and semantic  
correction

# **Proposed Methodology**

Our Approach and  
Contribution

**03**

# End-To-End Correction of Words in Annotation Affected Documents



# Proposed Methodology - Preprocessing

## Localizing Annotations

Connected Component  
Analysis

## Inpainting

Regeneration of Affected  
Text

**01**



**02**



**03**



**04**



**05**

## Input

Annotated  
Document Image

## Annotation Masks

Undirected Path  
Opening Operation

## OCR

Cleaned image for  
OCR Input

# Preprocessing Input Image

It means, a 1 cm. change in the height of the reservoir will cause 100 cm. change in the height of the small tube. Sensitivity can be increased by slanting the arm of the small tube, because the liquid will have a greater movement over the scale for the same vertical travel and this will facilitate the reading. By using a float in the small tube, attached with a hinge and pointer, the pressure change can be recorded over suitable scale.

2. Pressure Measurement Using Elastic Deformation : Pressure sensing element using *Elastic Deformation* can be of the following types :

- (a) Metallic Diaphragm.
- (b) Bellows.
- (c) Bourdon tube.

(a) Metallic Diaphragm ~~Diaphragms are elastic elements and very commonly used as pressure sensing elements. The diaphragm is essentially a thin circular plate stretched and fastened at its periphery. Diaphragms are made from elastic metal alloys, such as bronze, phosphor bronze, beryllium, copper and stainless steel and may be flat or corrugated. The time constant of the diaphragm should be of the order of .001 cm. Mechanical output of the transducers can be converted into electrical signal. The maximum displacement of the diaphragm should be of the order of .001 cm.~~

Diaphragm pressure gauges are used in trucks, tractors and in other heavy duty applications.

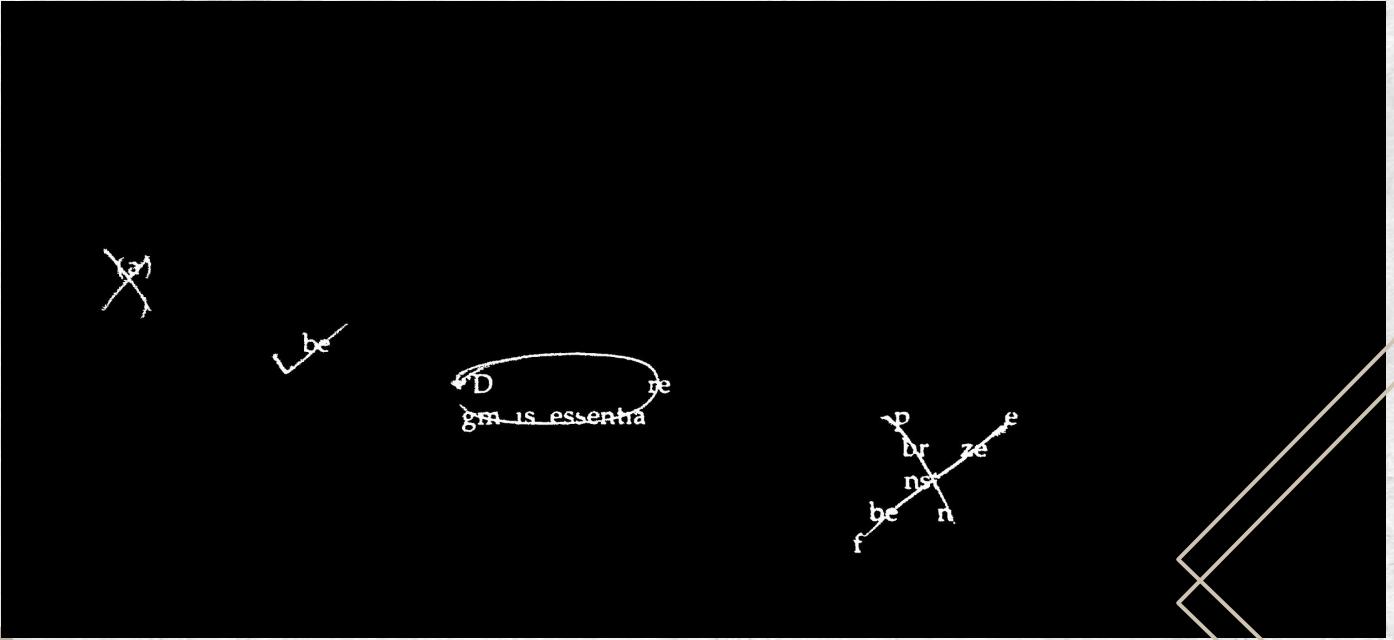
# Preprocessing - Localizing Annotations

Detection of Annotations using Connected Component Analysis  
by Statistical Mode of Area of Components

---

1. Get all the connected components from the image
2. Calculate the mode of the areas of the bounding boxes of the connected components.
3. Store the connected components which have an area greater than five times the mode for processing in further stages and remove all other components.

# Preprocessing Localized Annotations



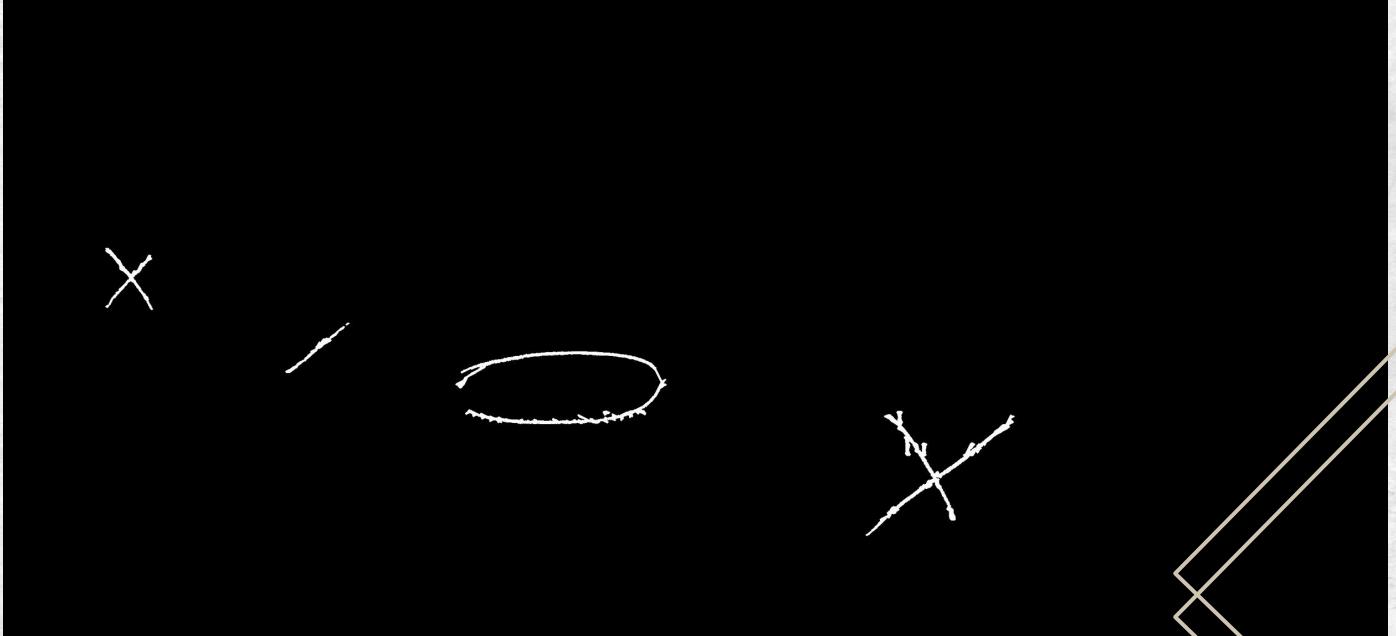
# Preprocessing - Annotation Mask

Separation of Localized Annotations from Letters using  
Undirected Path Opening Morphological Operation

---

1. Opening Operation is applied over all possible paths in all possible directions of a specified length
2. Length is equal to the maximum of height and width of the localized annotations' bounding box
3. We constrain the direction after every alternate step to restrict movement of path only in the initial main direction

# Preprocessing Annotation Masks



# Preprocessing - Inpainting

Masks of Detected Annotations are used for Inpainting to Regenerate Affected Text

---

1. Inpainting is based on current state of the art Fast Marching Method
2. Starts from the boundary of this region and goes inside the region gradually filling everything in the boundary first.
3. Pixel is replaced by normalized weighted sum of all the known pixels in the neighbourhood of radius = 0.5

# Preprocessing Inpainting

It means, a 1 cm change in the height of the reservoir will cause 100 cm change in the height of the small tube. Sensitivity can be increased by slanting the arm of the small tube, because the liquid will have a greater movement over the scale for the same vertical travel and this will facilitate the reading. By using a float in the small tube, attached with a hinge and pointer, the pressure change can be recorded over suitable scale.

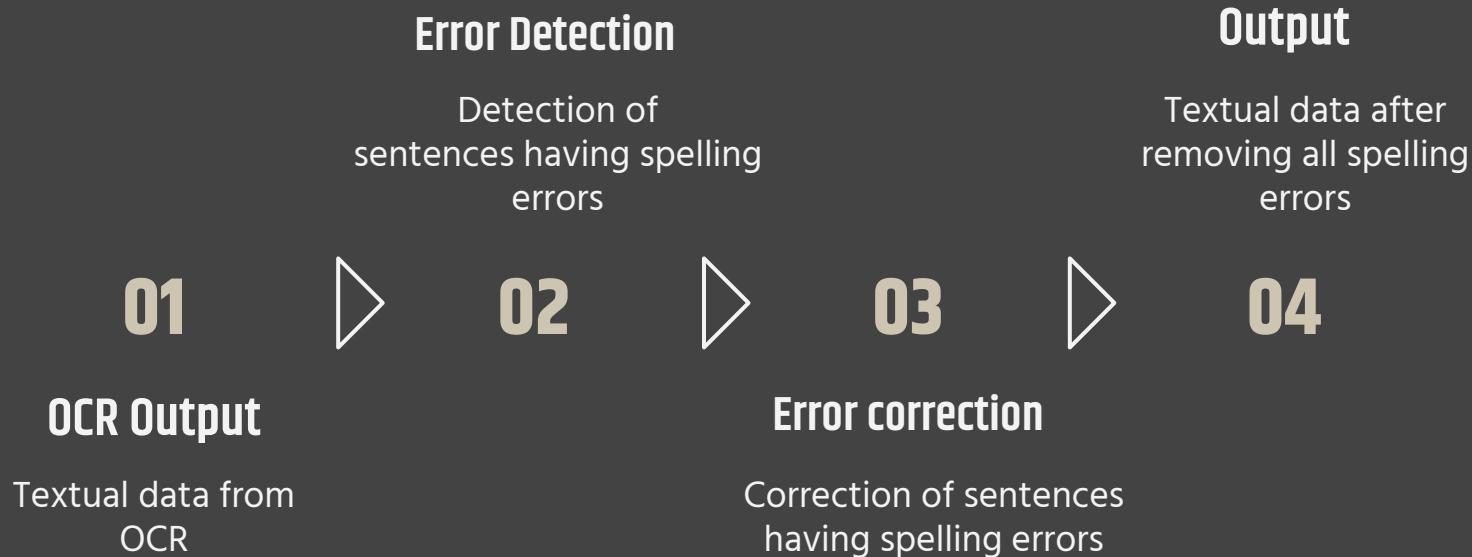
**2 Pressure Measurement Using Elastic Deformation :** Pressure sensing element using *Elastic Deformation* can be of the following types .

- (a) Metallic Diaphragm
- (b) Bellows
- (c) Bourdon tube

**(a) Metallic Diaphragm** Diaphragms are elastic elements and very commonly used as pressure sensing elements. The diaphragm is essentially a thin circular plate stretched and fastened at its periphery. Diaphragms are made from elastic metal alloys, such as iron, phosphor bronze, beryllium, copper and stainless steel and may be flat or corrugated. The time constant of the diaphragm should be of the order of .001 cm. Mechanical output of the transducers can be converted into electrical signal. The maximum displacement of the diaphragm should be of the order of .001 cm.

Diaphragm pressure gauges are used in trucks, tractors and in other heavy duty applications.

# Proposed Methodology - Postprocessing



# Post-processing OCR Output

it means, a 1cm change in the height of the reservoir will cause 100 cm change in the height of the small tube Sensitivity can be increased by slanting the arm of the small tube, because the liquid will have a greater movement over the scale for the same vertical travel and this will facilitate the reading. By

using a float in the small tube, attached with a hinge and pointer, the pressure change can be recorded over suitable scale

2 Pressure Measurement Using Elastic Deformation : Pressure sensing element using Elastic Deformation can be of the following types .

- (e) Metallic Diaphragm
- (bo) Bellows
- (c) Bourdon tube|

(a) Metallic Diaphragm "Diaphragms are elastic elements and very commonly used as pressure sensing elements. The diaphragm is essentially a thin circular plate stretched and fastened at its periphery. Diaphragms are made from elastic metal alloys, such as bronze, phosphor bronze, beryllium, copper and stainless steel and may be flat or corrugated. The time constant of the diaphragm should be of the order of .001 sec. Mechanical output of the transducers can be converted into electrical signal. The maximum displacement of the diaphragm should be of the order of .001 cm.

Diaphragm pressure gauges are used in trucks, tractors and in other heavy duty applications.

# Post-processing - Error Detection

Dictionary based error detection is used to detect sentences having spelling error

---

1. The entire text from OCR is broken into list of sentences.
2. If any word of a sentence is not found in the dictionary then that sentence is marked as erroneous sentence.
3. This is achieved by a python library called pyenchant.

# Post-processing Error Detection

it means, a \_\_\_ change in the height of the reservoir will cause 100 cm change in the height of the small tube Sensitivity can be increased by slanting the arm of the small tube, because the \_\_\_ will have a greater movement over the scale for the same vertical travel and this will facilitate the reading. By using a float in the small tube, attached with a hinge and pointer, the pressure change can be recorded over suitable scale

2 Pressure Measurement Using Elastic Deformation : Pressure sensing element using Elastic Deformation can be of the following types .

{e) Metallic Diaphragm

( \_\_\_ ) Bellows

(c) \_\_\_ tube

(a) Metallic Diaphragm "Diaphragms are elastic elements and very commonly used as pressure sensing elements. The diaphragm \_\_\_ essentially a thin circular plate stretched and fastened at \_\_\_ periphery. Diaphragms are made from elastic metal alloys, such as \_\_\_ , phosphor bronze, beryllium, copper and stainless steel and may be flat or corrugated. The time constant of the diaphragm should be of the order of .001 cm. Mechanical output of the transducers can be \_\_\_ into electrical signal. The maximum displacement of the diaphragm should be of the order of .001 cm.

Diaphragm pressure gauges are used \_\_\_ trucks, tractors and in other heavy duty applications.

# Post-processing - Error Correction

Context based spelling error correction is achieved by a deep learning model

---

1. We used a Seq2Seq architecture with Bidirectional RNN in the encoding layer and Bahdanau Attention in the decoding layer.
2. Bidirectional RNN makes the model robust by considering previous words as well as succeeding words to the erroneous word.
3. Attention mechanism helps to generalize well to test data as it gives more weightage to context surrounding the current word compared to the other words in the sentence

# Post-processing Error Correction

It means, a LCM change in the height of the reservoir will cause 100 cm change in the height of the Small tube Sensitivity can be increased by slanting the arm of the small tube, because the liquid will have A greater movement over the scale for the same vertical travel and this will facilitate the reading. By

Using a float in the small tube, attached with a hinge and pointer, the pressure change can be recorded over Suitable scale

2 Pressure Measurement Using Elastic Deformation : Pressure sensing element using Elastic Deformation can be of the following types.

- (a) Metallic Diaphragm
- (B) Bellows
- (c) Bourbon tube

(a) Metallic Diaphragm "Diaphragms are elastic elements and very commonly used as pressure Sensing elements The diaphragm is essentially a thin circular plate stretched and fastened at its Periphery. Diaphragms are made from elastic metal alloys, such as bronze, phosphor bronze, beryllium, Copper and stainless steel and may be flat or corrugated. The time constant of the diaphragm should be of The order of .001 cm. Mechanical output of the transducers can be converted into electrical signal. The Maximum displacement of the diaphragm should be of the order of .001 cm.

Diaphragm pressure gauges are used in trucks, tractors and in other heavy duty applications.

# Conclusion

Results and Conclusion  
from our approach

# 04

# Highlights



## End-To-End

Combination of Image Processing and NLP produces better results



## Versatile

Works for various different types of annotations



## Novelty

Use of Path Opening operation to get inpainting masks

# Statistics

## Sources

Number of Sources for Images

## Pages

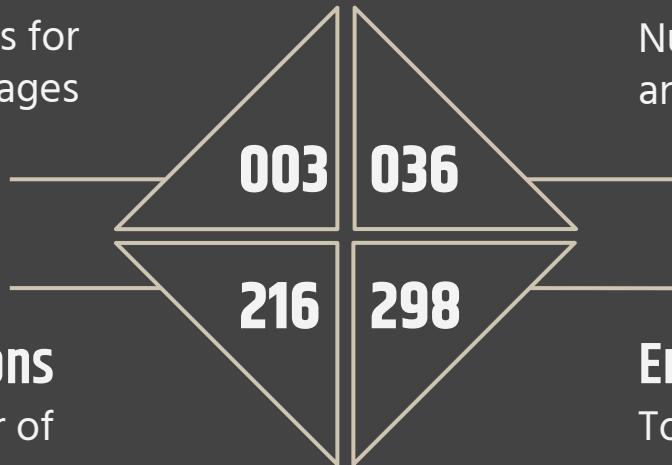
Number of Pages annotated

## Annotations

Total Number of Annotations

## Errors

Total Number of Errors



# Results

	Original	Post Processing without Preprocessing	Post Processing after Preprocessing
Errors Due To Annotations	227	56	11
Other Errors	71	19	06
Total Errors	98	75	17

Error  
Correction  
Percent

**94.32 %**

# CONCLUSIONS

We make the following observations from our experiments

---

1. With Natural Language Processing used in unison with Image Processing the error correction rate improves significantly
2. Opening operation when applied to paths is an effective method to separate letters from annotations thereby also providing a great strategy for obtaining masks for inpainting
3. Our proposed end to end method generalizes well on a plethora of documents taken from different sources with different types of noises.

# Thank You!

