A Case Study on

Tabular Text Data Extraction using EasyOCR

Jeyakumar N K

15/05/2024

Introduction

- EasyOCR is a python module for extracting text from image.
- It can read both natural scene text and dense text in document.
- It currently supports 80+ languages.

EasyOCR Text Detection

```
import easyocr as oc

reader = oc.Reader(['en’])    # To load the model into memory

result = reader.readtext(‘input_image.jpg')
```

Problem Statement

- Detect the tabular region present in a given input image and extract text that present inside
- Convert back the extracted text data to tabular form

Challenges

As far as I know, easyocr currently does not support table recognition. The best table recognition should be PaddleOCR's PP-Structure model. This is what I use now, and the effect is very good.

Stages Involved

Stage 1

IDENTIFYING TABULAR REGION

Sharpening Edges, Thresholding, Canny Edge Detection, Closing Operation on Canny, Contour Detection, Identifying Largest Contour, RoI, Cropping Input Image to RoI

Stage 2

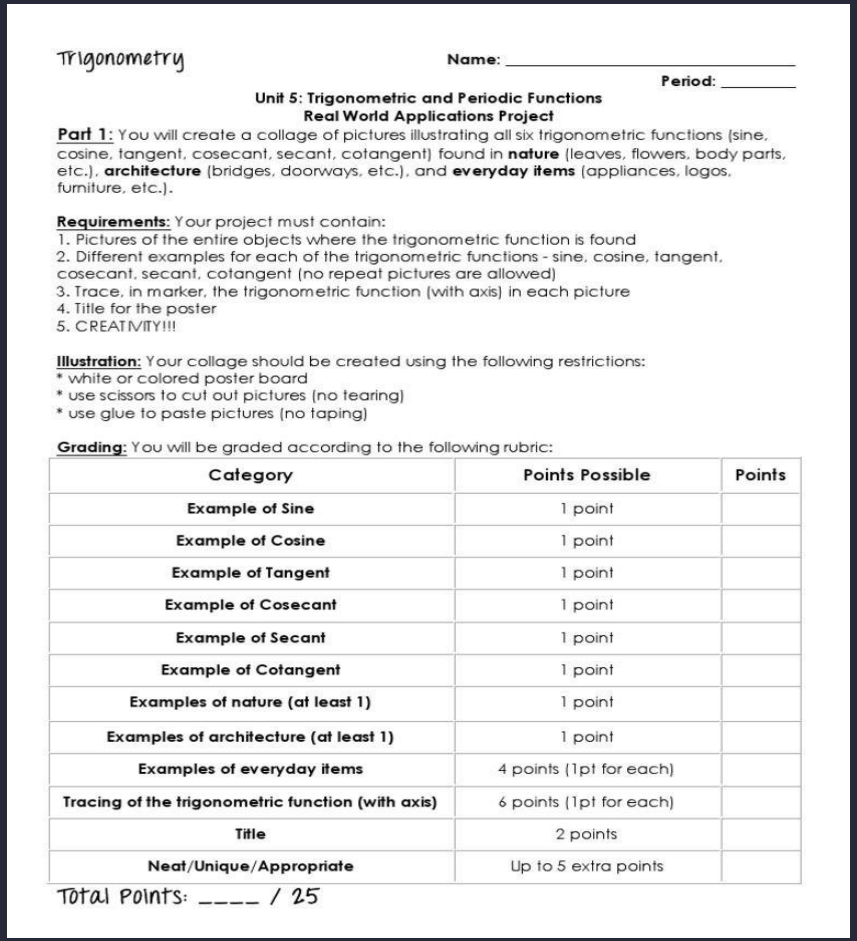
EasyOCR Text Detection

Presenting RoI to EasyOCR Module for Text Detection in Tabular Region

Stage 3

TABULAR REPRESENTATION OF DETECTED TEXT

Detecting number of columns, Appending detected row information to corresponding columns, Converting resultant list to a Pandas Data Frame



Trigonometry

Name: _____

Period: _____

Unit 5: Trigonometric and Periodic Functions
Real World Applications Project

Part 1: You will create a collage of pictures illustrating all six trigonometric functions (sine, cosine, tangent, cosecant, secant, cotangent) found in **nature** (leaves, flowers, body parts, etc.), **architecture** (bridges, doorways, etc.), and **everyday items** (appliances, logos, furniture, etc.).

Requirements: Your project must contain:
1. Pictures of the entire objects where the trigonometric function is found
2. Different examples for each of the trigonometric functions - sine, cosine, tangent, cosecant, secant, cotangent (no repeat pictures are allowed)
3. Trace, in marker, the trigonometric function (with axis) in each picture
4. Title for the poster
5. CREATIVITY!!!

Illustration: Your collage should be created using the following restrictions:
* white or colored poster board
* use scissors to cut out pictures (no tearing)
* use glue to paste pictures (no taping)

Grading: You will be graded according to the following rubric:

Category	Points Possible	Points
Example of Sine	1 point	
Example of Cosine	1 point	
Example of Tangent	1 point	
Example of Cosecant	1 point	
Example of Secant	1 point	
Example of Cotangent	1 point	
Examples of nature (at least 1)	1 point	
Examples of architecture (at least 1)	1 point	
Examples of everyday items	4 points (1pt for each)	
Tracing of the trigonometric function (with axis)	6 points (1pt for each)	
Title	2 points	
Neat/Unique/Appropriate	Up to 5 extra points	

Total Points: ____ / 25

	A	B	C	D
1		Category	ints Possib	Points
2	0	Example c	point	
3	1	Example c	point	
4	2	Example c	point	
5	3	Example c	point	
6	4	Example c	point	
7	5	Example c	point	
8	6	Examples	point	
9	7	Examples	point	
10	8	Examples	points (1pt for each)	
11	9	Title	points (1pt for each)	
12	10	Neat/Uniq	2 points	
13	11	25	Up to 5 extra points	
14				

Input Image

Outcome

I. Steps in identifying Tabular Region

- Reading an Image
- Converting it to Gray Scale
- Sharpening Edges
- Thresholding
- Canny Edge Detection
- Closing Operation on Canny
- Contour Detection
- Identifying Largest Contour as Tabular Region - RoI
- Cropping Input Image to RoI

Sharpened Edges

Trigonometry

Name: _____

Period: _____

Unit 5: Trigonometric and Periodic Functions
Real World Applications Project

Part 1: You will create a collage of pictures illustrating all six trigonometric functions (sine, cosine, tangent, cosecant, secant, cotangent) found in **nature** (leaves, flowers, body parts, etc.), **architecture** (bridges, doorways, etc.), and **everyday items** (appliances, logos, furniture, etc.).

Requirements: Your project must contain:
1. Pictures of the entire objects where the trigonometric function is found
2. Different examples for each of the trigonometric functions - sine, cosine, tangent, cosecant, secant, cotangent (no repeat pictures are allowed)
3. Trace, in marker, the trigonometric function (with axis) in each picture
4. Title for the poster
5. CREATIVITY!!!

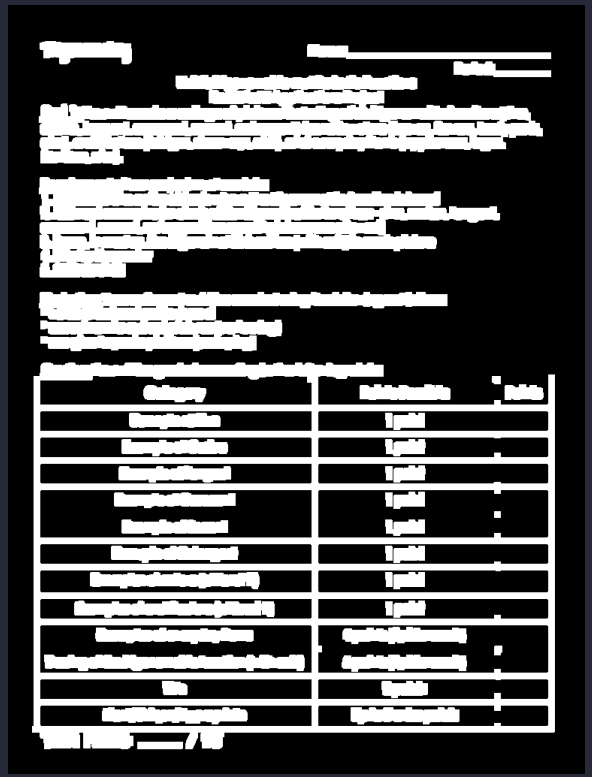
Illustration: Your collage should be created using the following restrictions:
* white or colored poster board
* use scissors to cut out pictures (no tearing)
* use glue to paste pictures (no taping)

Grading: You will be graded according to the following rubric:

Category	Points Possible	Points
Example of Sine	1 point	
Example of Cosine	1 point	
Example of Tangent	1 point	
Example of Cosecant	1 point	
Example of Secant	1 point	
Example of Cotangent	1 point	
Examples of nature (at least 1)	1 point	
Examples of architecture (at least 1)	1 point	
Examples of everyday items	4 points (1pt for each)	
Tracing of the trigonometric function (with axis)	6 points (1pt for each)	
Title	2 points	
Neat/Unique/Appropriate	Up to 5 extra points	

Total Points: ____ / 25

Dilate (or) Closing?



Trigonometry

Name _____

Period _____

Unit 6: Trigonometric and Periodic Functions

Real-World Applications Project

Goal 1: You will create a collage of pictures illustrating all six trigonometric functions (sine, cosine, tangent, cosecant, secant, cotangent) found in nature (leaves, flowers, body parts, shells, snowflakes, halogen, dominoes, etc.), and everyday items (appliances, logos, furniture, etc.).

Requirements: Your project must contain:

1. Pictures of the entire objects where the trigonometric function is found
2. Different examples for each of the trigonometric functions - sine, cosine, tangent, cosecant, secant, cotangent (no repeat pictures are allowed)
3. Basic, in words, the trigonometric function (with axis) in each picture
4. Title for the poster
5. CREATIVITY!

Evaluation: Your collage should be created using the following specifications:

- Vertical colored poster board
- Use scissors to cut out pictures (no tearing)
- Use glue to paste pictures (no taping)

Grading: You will be graded according to the following rubric:

Category	Points Possible	Points
Example of Sine	1 point	
Example of Cosine	1 point	
Example of Tangent	1 point	
Example of Cosecant Example of Secant	1 point 1 point	
Example of Cotangent	1 point	
Examples of asymptotes (at least 1)	1 point	
Examples of undefinedness (at least 1)	1 point	
Example of everyday items Meaning of the trigonometric function (with axis)	4 points (1pt for each) 6 points (1pt for each)	
Title	1 point	
Neatness/Originality/Appropriate	Up to 5 extra points	

Total Points: _____ / 15

Largest Contour

Trigonometry

Name: _____

Period: _____

Unit 5: Trigonometric and Periodic Functions
Real World Applications Project

Part 1: You will create a collage of pictures illustrating all six trigonometric functions (sine, cosine, tangent, cosecant, secant, cotangent) found in **nature** (leaves, flowers, body parts, etc.), **architecture** (bridges, doorways, etc.), and **everyday items** (appliances, logos, furniture, etc.).

Requirements: Your project must contain:
1. Pictures of the entire objects where the trigonometric function is found
2. Different examples for each of the trigonometric functions - sine, cosine, tangent, cosecant, secant, cotangent (no repeat pictures are allowed)
3. Trace, in marker, the trigonometric function (with axis) in each picture
4. Title for the poster
5. CREATIVITY!!!

Illustration: Your collage should be created using the following restrictions:
* white or colored poster board
* use scissors to cut out pictures (no tearing)
* use glue to paste pictures (no taping)

Grading: You will be graded according to the following rubric:

Category	Points Possible	Points
Example of Sine	1 point	
Example of Cosine	1 point	
Example of Tangent	1 point	
Example of Cosecant	1 point	
Example of Secant	1 point	
Example of Cotangent	1 point	
Examples of nature (at least 1)	1 point	
Examples of architecture (at least 1)	1 point	
Examples of everyday items	4 points (1pt for each)	
Tracing of the trigonometric function (with axis)	6 points (1pt for each)	
Title	2 points	
Neat/Unique/Appropriate	Up to 5 extra points	
Total Points: ____ / 25		

Region of Interest

Category	Points Possible	Points
Example of Sine	1 point	
Example of Cosine	1 point	
Example of Tangent	1 point	
Example of Cosecant	1 point	
Example of Secant	1 point	
Example of Cotangent	1 point	
Examples of nature (at least 1)	1 point	
Examples of architecture (at least 1)	1 point	
Examples of everyday items	4 points (1pt for each)	
Tracing of the trigonometric function (with axis)	6 points (1pt for each)	
Title	2 points	
Neat/Unique/Appropriate	Up to 5 extra points	

Total Points: ____ / 25

II. EasyOCR Text Detection

- This involves passing RoI to EasyOCR

```
import easyocr as oc
reader = oc.Reader(['en’])
result = reader.readtext(roi)
```

Output of '`result`' variable:

```
Detected Test Results: [([[145, 8], [228, 8], [228, 35], [145, 35]], 'Category', 0.9775009221672206), ([[433, 11], [555,
11], [555, 31], [433, 31]], 'Points Possible', 0.9985647332397556), ([[625, 11], [677, 11], [677, 31], [625, 31]], 'Point
s', 0.9948524298369781), ([[129, 51], [247, 51], [247, 69], [129, 69]], 'Example of Sine', 0.8702974277998347), ([[476, 4
6], [523, 46], [523, 70], [476, 70]], 'point', 0.9665031258880472), ([[115, 81], [258, 81], [258, 106], [115, 106]], 'Exa
mple of Cosine', 0.9817100291745914), ([[481, 87], [521, 87], [521, 105], [481, 105]], 'point', 0.9293741400607151), ([[1
14, 117], [262, 117], [262, 141], [114, 141]], 'Example of Tangent', 0.7430791363523259), ([[476, 116], [523, 116], [523,
 140], [476, 140]], 'point', 0.9696398408886909), ([[109, 155], [269, 155], [269, 175], [109, 175]], 'Example of Cosecant
', 0.962936743247506), ([[481, 157], [521, 157], [521, 173], [481, 173]], 'point', 0.9638947131924763), ([[117, 191], [25
9, 191], [259, 211], [117, 211]], 'Example of Secant', 0.8150827687667433), ([[479, 191], [521, 191], [521, 211], [479, 2
11]], 'point', 0.9634583272227314), ([[105, 227], [273, 227], [273, 247], [105, 247]], 'Example of Cotangent', 0.81678161
88413394), ([[480, 226], [522, 226], [522, 246], [480, 246]], 'point', 0.8965821886855843), ([[77, 261], [299, 261], [299
, 281], [77, 281]], 'Examples of nature (at least 1)', 0.7540243659569185), ([[479, 261], [521, 261], [521, 281], [479, 2
81]], 'point', 0.953802130288594), ([[55, 299], [321, 299], [321, 319], [55, 319]], 'Examples of architecture (at least 1
)', 0.8848066390058598), ([[476, 296], [523, 296], [523, 320], [476, 320]], 'point', 0.6853968662412122), ([[84, 334], [2
93, 334], [293, 355], [84, 355]], 'Examples of everyday items', 0.831003563001235), ([[422, 332], [576, 332], [576, 356],
 [422, 356]], 'points (Ipt for each)', 0.5366906584483264), ([[11, 367], [360, 367], [360, 392], [11, 392]], 'Tracing of
the trigonometric function (with axis)', 0.849662599043716), ([[422, 368], [576, 368], [576, 392], [422, 392]], 'points (
Ipt for each)', 0.7108143711468203), ([[171, 407], [203, 407], [203, 423], [171, 423]], 'Title', 0.9990850638569154), ([[
465, 406], [523, 406], [523, 426], [465, 426]], '2 points', 0.731434124724312), ([[89, 437], [284, 437], [284, 463], [89,
 463]], 'Neat/Unique / Appropriate', 0.7508336074564068), ([[420, 438], [568, 438], [568, 462], [420, 462]], 'Up to 5 ext
ra points', 0.8689709825623763), ([[8, 470], [134, 470], [134, 493], [8, 493]], 'Total Polnts:', 0.883665811232402), ([[2
20, 470], [252, 470], [252, 493], [220, 493]], '25', 0.9999665355662857)]
```

Output of ‘`result`’ variable:

([[145, 8], [228, 8], [228, 35], [145, 35]], 'Category', 0.9775009221672206)

([[145, 8], [228, 8], [228, 35], [145, 35]], 'Category', 0.9775009221672206)

([[145, 8], [228, 8], [228, 35], [145, 35]], 'Category', 0.9775009221672206)

([[145, 8], [228, 8], [228, 35], [145, 35]], 'Category', 0.9775009221672206)

II. EasyOCR Text Detection

Category	Points Possible	Points
Example of Sine	1 point	
Example of Cosine	1 point	
Example of Tangent	1 point	
Example of Cosecant	1 point	
Example of Secant	1 point	
Example of Cotangent	1 point	
Examples of nature (at least 1)	1 point	
Examples of architecture (at least 1)	1 point	
Examples of everyday items	4 points (1pt for each)	
Tracing of the trigonometric function (with axis)	6 points (1pt for each)	
Title	2 points	
Neat/Unique/Appropriate	Up to 5 extra points	
Total Points: ____ / 25		

After Preprocessing

Category	Points Possible	Points
Example of Sine	1 point	
Example of Cosine	1 point	
Example of Tangent	1 point	
Example of Cosecant	1 point	
Example of Secant	1 point	
Example of Cotangent	1 point	
Examples of nature (at least 1)	1 point	
Examples of architecture (at least 1)	1 point	
Examples of everyday items	4 points (1pt for each)	
Tracing of the trigonometric function (with axis)	6 points (1pt for each)	
Title	2 points	
Neat/Unique/Appropriate	Up to 5 extra points	
Total Points: ____ / 25		

But..

```
Detected Test Results: ([[[141, 3], [232, 3], [232, 40], [141, 40]], 'Category ,', 0.9228015098251716), ([[430, 4], [556,
 4], [556, 36], [430, 36]], 'Points Possible ', 0.5910789239018812), ([[125, 42], [248, 42], [248, 71], [125, 71]], "'Exa
mple of Sine", 0.5827103059987683), ([[465, 42], [523, 42], [523, 71], [465, 71]], 'Tpoint', 0.5132014665955953), ([[113,
 75], [261, 75], [261, 112], [113, 112]], 'Example of Cosine', 0.7083658303662448), ([[464, 78], [522, 78], [522, 108], [
464, 108]], 'ipoli', 0.41084367936867405), ([[112, 116], [262, 116], [262, 140], [112, 140]], 'Example ot Tanigent', 0.64
4011376378862), ([[466, 116], [522, 116], [522, 140], [466, 140]], 'Ipoinf', 0.29725171867162), ([[106, 150], [270, 150],
 [270, 178], [106, 178]], "Example ot Cosecant'", 0.8935830221264257), ([[114, 188], [260, 188], [260, 212], [114, 212]],
 'Example ct Secant', 0.581145718038942), ([[466, 188], [522, 188], [522, 214], [466, 214]], 'Lpoint', 0.9299289758012137
), ([[103, 218], [274, 218], [274, 248], [103, 248]], 'Example oi Cotangenf', 0.7973716506747129), ([[466, 220], [522, 22
0], [522, 248], [466, 248]], 'Ipoini', 0.40269861700833043), ([[74, 254], [306, 254], [306, 286], [74, 286]], 'Exampies o
i_nafure (af jogsf f);', 0.10818344332200303), ([[478, 254], [524, 254], [524, 286], [478, 286]], 'paini -', 0.3686596907
134466), ([[52, 292], [322, 292], [322, 324], [52, 324]], 'Examples ot archiecture (atleasi W', 0.19552054484353737), ([[
466, 294], [524, 294], [524, 326], [466, 326]], 'Jpoin _', 0.440099256753988), ([[82, 332], [294, 332], [294, 356], [82,
356]], "Examples of everyday items'", 0.5135205667166407), ([[402, 332], [576, 332], [576, 358], [402, 358]], '#4 points
(Tpt for eachl', 0.42892416517349746), ([[13, 363], [360, 363], [360, 393], [13, 393]], 'Tracing oi lhe igonomelric funct
ion (wih &xis)', 0.2598808575501405), ([[410, 364], [576, 364], [576, 394], [410, 394]], 'U poinls  ipi for eachy', 0.272
1005127198858), ([[171, 405], [207, 405], [207, 425], [171, 425]], "Title'", 0.39498669619739274), ([[456, 404], [530, 40
4], [530, 428], [456, 428]], '12 points}', 0.64431310363352), ([[90, 436], [286, 436], [286, 462], [90, 462]], 'NeatJUniq
ue/ Appropiate .', 0.4951969285837001), ([[418, 436], [568, 436], [568, 462], [418, 462]], '2P fo 5 extra points', 0.2299
2901222990567), ([[10, 470], [134, 470], [134, 493], [10, 493]], 'Total Polnts:', 0.9263859212815321), ([[202, 470], [256
, 470], [256, 493], [202, 493]], '1i051', 0.03198374489602229), ([[622.5430712064641, 9.377528171635387], [677.5817300484
72, 4.472037318251267], [678.4569287935359, 32.62247182836461], [623.418269951528, 36.52796268174873]], 'Points', 0.91803
04640198443), ([[463.6027607015021, 151.34524533285406], [520.4883107813158, 145.32419886156282], [523.3972392984979, 175
.65475466714594], [466.5116892186842, 181.67580113843718]], 'Ipoini-', 0.2905128447667485)]
```

Whatever the operations applied on Input image results with the detection of wrong textual information

III. Tabular Representation

It involves,

- Detecting number of columns
- Appending detected row information to corresponding columns
- Converting resultant list to a Pandas Data Frame

No. of Columns?

[([[143, 7], [230, 7], [230, 35], [143, 35]], 'Category', 0.7159639437158294),

([[432, 8], [556, 8], [556, 32], [432, 32]], 'Points Possible', 0.8149897168967521),

([[625, 11], [677, 11], [677, 31], [625, 31]], 'Points', 0.9109005020581996),

([[129, 51], [247, 51], [247, 69], [129, 69]], 'Example of Sine', 0.8702974277998347),

([[480, 50], [521, 50], [521, 69], [480, 69]], 'point', 0.793118081793997),

([[115, 81], [258, 81], [258, 106], [115, 106]], 'Example of Cosine', 0.9817100291745914),

([[481, 87], [521, 87], [521, 105], [481, 105]], 'point', 0.9293741400607151),

([[114, 117], [262, 117], [262, 141], [114, 141]], 'Example of Tangent', 0.7430791363523259),

([[476, 116], [523, 116], [523, 140], [476, 140]], 'point', 0.9696398408886909),

([[108, 154], [269, 154], [269, 175], [108, 175]], 'Example of Cosecant', 0.8046274847951407),

No. of Columns?

```
def detect_columns(result, start = 0):
    cols = []
    col_ref = []
    x1 = []
    y1 = []
    for tup in range(start, len(result)):
        x, y = result[tup][0][0]
        x1.append(x)
        y1.append(y)

    for i in range(len(result)):
        beg = y1[start]-10
        end = y1[start] + 10
        if (y1[i] >= beg) and (y1[i] <= end):
            cols.append(result[i][1])
            col_ref.append(x1[i])

    return (x1, y1, cols, col_ref)
```

Appending Rest of Data to Corresponding Columns

```
for tup in range(len(cols)+counter, len(result)-counter):
    x_val = int(x1[tup])
    beg = x_val - 30
    end = x_val + 30
    if (x_val > beg) and (x_val < end):
        for i in range(len(cols)):
            if (x_val > (col_ref[i] - 100)) and (x_val < (col_ref[i] + 100)):
                table[i].append(result[tup][1])
                break
        else:
            continue
```

([[143, 7], [230, 7], [230, 35], [143, 35]], 'Category', 0.7159639437158294),

([[432, 8], [556, 8], [556, 32], [432, 32]], 'Points Possible', 0.8149897168967521),

([[625, 11], [677, 11], [677, 31], [625, 31]], 'Points', 0.9109005020581996),

([[129, 51], [247, 51], [247, 69], [129, 69]], 'Example of Sine', 0.8702974277998347),

([[480, 50], [521, 50], [521, 69], [480, 69]], 'point', 0.793118081793997),

([[115, 81], [258, 81], [258, 106], [115, 106]], 'Example of Cosine', 0.9817100291745914),

([[481, 87], [521, 87], [521, 105], [481, 105]], 'point', 0.9293741400607151),

Outcome

	Category	Points Possible	Points
0	Example of Sine	point	None
1	Example of Cosine	point	None
2	Example of Tangent	point	None
3	Example of Cosecant	point	None
4	Example of Secant	point	None
5	Example of Cotangent	point	None
6	Examples of nature (at least 1)	point	None
7	Examples of architecture (at least 1)	point	None
8	Examples of everyday items	points (Ipt for ea	None
9	Title	points (Ipt for ea	None
10	Neat/Unique / Appropriate	2 points	None
11	25	Up to 5 extra poi	None

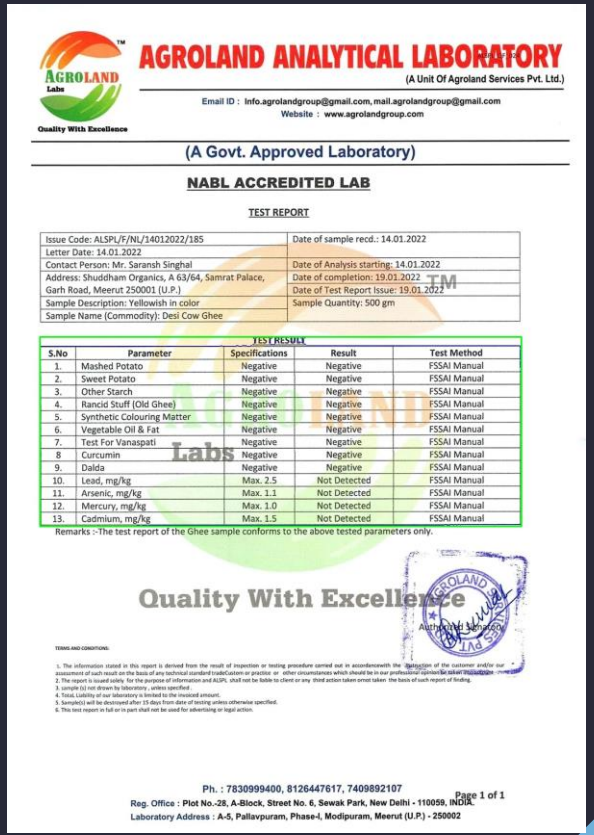
Lab Test	Result	Reference Range
Total bilirubin	20.9 mg/dL	0.2 - 1.2 mg/dL
Conjugated bilirubin	12.5 mg/dL	0.0 - 0.5 mg/dL
Alkaline phosphatase	327 Units/L	40 - 150 Units/L
Gamma glutamyltransferase	185 Units/L	9 - 64 Units/L
Alanine aminotransferase (ALT)	34 Units/L	0-55 Units/L
Aspartate aminotransferase (AST)	29 Units/L	5-34 Units/L
Prothrombin time	18.7 seconds	12.1 - 14.8 seconds
International normalized ratio (INR)	1.6	0.9-1.1
Albumin	2.5 g/dL	3.4 - 5.0 g/dL
White blood cell count with differential	8.9×10^3/uL	$3.5\text{-}10.5 \times 10^3$/uL
Neutrophils	81%	35-70%
Lymphocytes	11.30%	20-50%
Monocytes	18%	3-15%

	Lab Test	Result	Reference Range
0	Total bilirubin	20.9 mg/ dL	0.2 -1.2 mg,
1	Conjugated bilirubin	12.5 mg/ dL	0.0 - 0.5 mg
2	Alkaline phosphatase	327 Units / L	40
3	Gamma glutamyltransferase	185 Units / L	150 Units/ L
4	Alanine aminotransferase (ALT)	34 Units/L	64 Units/ L
5	Aspartate aminotransferase (AST)	29 Units/ L	0-55 Units/L
6	Prothrombin time	18.7 seconds	5-34 Units/L
7	International normalized ratio (INR)	1.6	121
8	Albumin	2.5 g/dL	14.8 seconds
9	White blood cell count with	8.9 x 10'/uL	0.9-1.1
10	differential	81 %	34 - 5.0 g/dL
11	Neutrophils	11.30%	3.5-10.5 X 103 /ul
12	Lymphocytes	18%	35-70%
13	Monocytes	None	20-50%

Challenges

- In some cases the region of interest was detected along with the caption information of the table. Since here I'm considering the text data that is present in first few tuples as the columns for a given input image.

- To overcome this, I just measured the length of columns detected. If it is less than 2 then that will be stored in the '`header`' variable.



AGROLAND ANALYTICAL LABORATORY

(A Unit Of Agroland Services Pvt. Ltd.)

Email ID : Info.agrolandgroup@gmail.com, mail.agrolandgroup@gmail.com
Website : www.agrolandgroup.com

Quality With Excellence

(A Govt. Approved Laboratory)

NABL ACCREDITED LAB

TEST REPORT

Issue Code: ALSPL/F/NL/14012022/185	Date of sample recd.: 14.01.2022
Letter Date: 14.01.2022	
Contact Person: Mr. Saransh Singhal	Date of Analysis starting: 14.01.2022
Address: Shuddham Organics, A 63/64, Samrat Palace, Garh Road, Meerut 250001 (U.P.)	Date of completion: 19.01.2022
	Date of Test Report Issue: 19.01.2022
Sample Description: Yellowish in color	Sample Quantity: 500 gm
Sample Name (Commodity): Desi Cow Ghee	

TEST RESULT

S.No	Parameter	Specifications	Result	Test Method
1.	Mashed Potato	Negative	Negative	FSSAI Manual
2.	Sweet Potato	Negative	Negative	FSSAI Manual
3.	Other Starch	Negative	Negative	FSSAI Manual
4.	Rancid Stuff (Old Ghee)	Negative	Negative	FSSAI Manual
5.	Synthetic Colouring Matter	Negative	Negative	FSSAI Manual
6.	Vegetable Oil & Fat	Negative	Negative	FSSAI Manual
7.	Test For Vanaspati	Negative	Negative	FSSAI Manual
8	Curcumin	Negative	Negative	FSSAI Manual
9.	Dalda	Negative	Negative	FSSAI Manual
10.	Lead, mg/kg	Max. 2.5	Not Detected	FSSAI Manual
11.	Arsenic, mg/kg	Max. 1.1	Not Detected	FSSAI Manual
12.	Mercury, mg/kg	Max. 1.0	Not Detected	FSSAI Manual
13.	Cadmium, mg/kg	Max. 1.5	Not Detected	FSSAI Manual

Remarks :-The test report of the Ghee sample conforms to the above tested parameters only.

Quality With Excellence

Authorised Signatory

TERMS AND CONDITIONS:

1. The information stated in this report is derived from the result of inspection or testing procedure carried out in accordancewith the Instruction of the customer and/or our assessment of such result on the basis of any technical standard tradeCustom or practice or other circumstances which should be in our professional opinion be taken into account.
2. The report is issued solely for the purpose of information and AGPL shall not be liable to client or any third action taken error taken the basis of such report of finding.
3. sample (s) not drawn by laboratory , unless specified .
4. Total Liability of our laboratory is limited to the invoiced amount.
5. Sample(s) will be destroyed after 15 days from date of testing unless otherwise specified.
6. This test report in full or in part shall not be used for advertising or legal action.

Ph. : 7830999400, 8126447617, 7409892107

Page 1 of 1

Reg. Office : Plot No.-28, A-Block, Street No. 6, Sewak Park, New Delhi - 110059, INDIA.

Laboratory Address : A-5, Pallavpuram, Phase-I, Modipuram, Meerut (U.P.) - 250002

Challenges

- That header data will reside outside the table. This helped me to retain column information as in the input image.

```
if len(cols) < 2:
    # st.write("This block is about to execute..")
    header = cols[0]
    counter += 1
    x1, y1, cols, col_ref = detect_columns(result, counter)
```

By default, this ‘start’ is set to zero

```
def detect_columns(result, start = 0):
```

IEST RESULT

	Parameter	Specifications	Result	SNo	IEST RESULT
0	Mashed Potato	Negative	Negative	FSSAI Manual	FSSAI Manual
1	Sweet Potato	Negative	Negative	FSSAI Manual	FSSAI Manual

Limitations

- This is not applicable for table with too many empty cells, but it outperforms in case an entire column is empty as shown in the previous slide. It is **prone to None value**
- If a null value is present in between any cells in a particular column then the data that has to be placed below the null value will be **shifted up to Null value's position** and the place for null value will be preserved at the end of the table.
- **Preprocessing/ Resizing** original image comparably affects the EasyOCR's model ability to produce a correct outcome.
- As mentioned in the EasyOCR documentation, the detected text will not follow **natural human reading** so depending on the context the columns may interchange in the output by preserving its data items in the corresponding columns.
- **Watermarks/ background** information present in an image may affect the positioning of data items in the table slightly.

Limitations

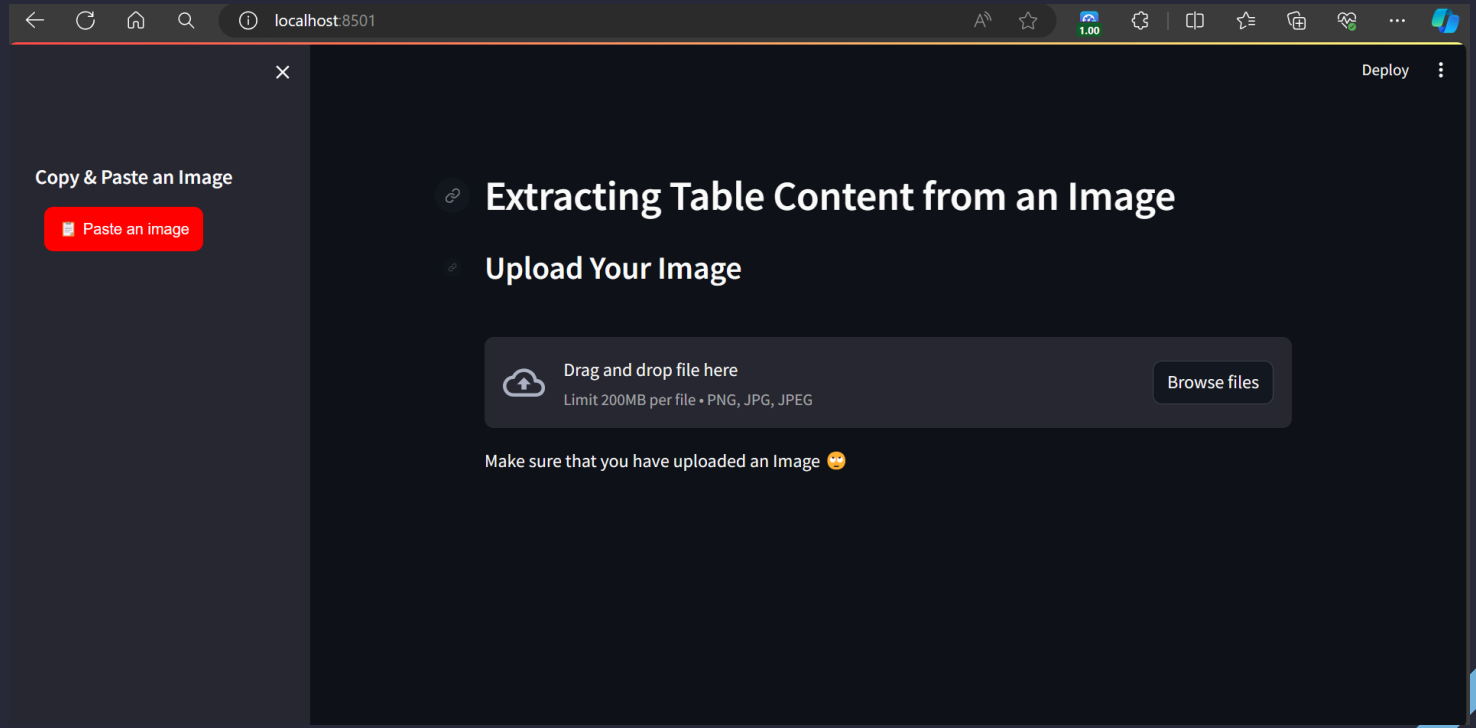
TEST RESULT

S.No	Parameter	Specifications	Result	Test Method
1.	Mashed Potato	Negative	Negative	FSSAI Manual
2.	Sweet Potato	Negative	Negative	FSSAI Manual
3.	Other Starch	Negative	Negative	FSSAI Manual
4.	Rancid Stuff (Old Ghee)	Negative	Negative	FSSAI Manual
5.	Synthetic Colouring Matter	Negative	Negative	FSSAI Manual
6.	Vegetable Oil & Fat	Negative	Negative	FSSAI Manual
7.	Test For Vanaspati	Negative	Negative	FSSAI Manual
8	Curcumin	Negative	Negative	FSSAI Manual
9.	Dalda	Negative	Negative	FSSAI Manual
10.	Lead, mg/kg	Max. 2.5	Not Detected	FSSAI Manual
11.	Arsenic, mg/kg	Max. 1.1	Not Detected	FSSAI Manual
12.	Mercury, mg/kg	Max. 1.0	Not Detected	FSSAI Manual
13.	Cadmium, mg/kg	Max. 1.5	Not Detected	FSSAI Manual

	SNo	Parameter	Specifications	Result	Test Method
0	Mashed Potato	Mashed Potato	Negative	Negative	FSSAI Manual
1	Sweet Potato	Sweet Potato	Negative	Negative	FSSAI Manual
2	Other Starch	Other Starch	Negative	Negative	FSSAI Manual
3	Rancid Stuff (Old Ghee)	Rancid Stuff (Old Ghee)	Negative	Negative	FSSAI Manual
4	Synthetic Colouring Matter	Synthetic Colouring Matter	Negative	Negative	FSSAI Manual
5	Vegetable Qil & Fat	Vegetable Qil & Fat	Negative	Negative	FSSAI Manual
6	Test For Vanaspati	Test For Vanaspati	Negative	Negative	FSSAI Manual
7	Curcumin	Curcumin	Negative	Negative	FSSAI Manual
8	Dalda	Tabb	Negative	Negative	FSSAI Manual
9	10.	Dalda	Max 2.5	Not Detecl	FSSAI Manual
10	Lead,mg/kg	Lead,mg/kg	Max 1.1	Not Detecl	FSSAI Manual
11	11.	Arsenic,mglkg	Max: 1.0	Not Detecl	FSSAI Manual
12	Arsenic,mglkg	Mercurymglkg	Max 1.5	Not Detecl	FSSAI Manual
13	12.	Cadmium, mglkg	None	None	None

WebApp

- Used Streamlit for hosting this web application locally in my device.



Conclusion

In conclusion, while EasyOCR may not be suitable for tables with excessive empty cells, it excels when entire columns are empty. Preprocessing or resizing images may impact its accuracy slightly, but overall, EasyOCR reliably preserves data integrity, even if column order may vary. Despite potential challenges like watermarks, EasyOCR remains a valuable tool for efficient and accurate table data extraction.

Thank You!!