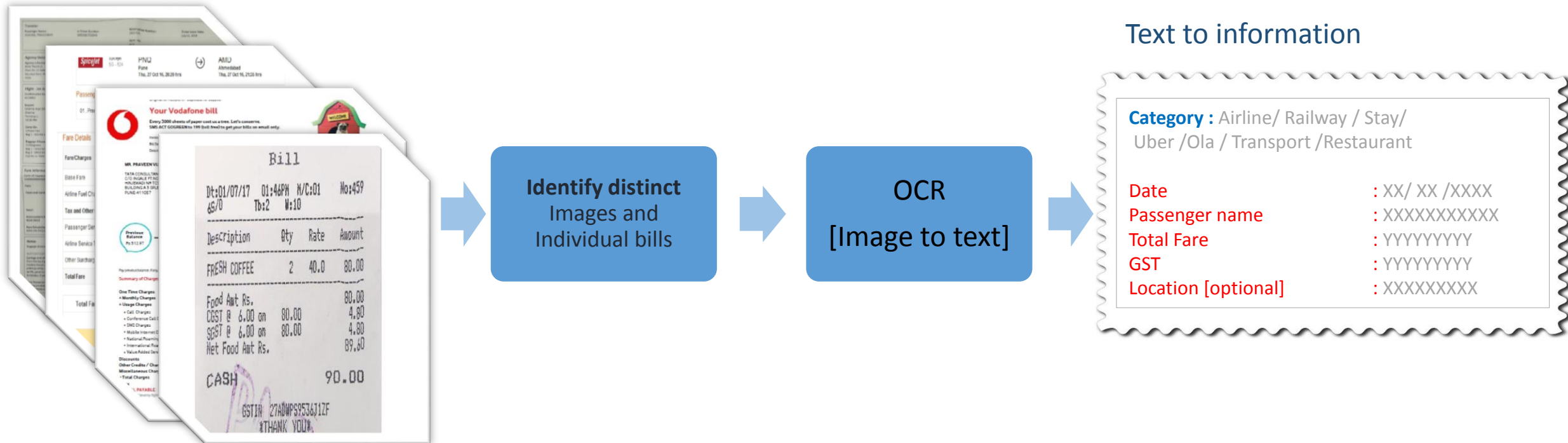


Automated Invoice processing

This module helps to **extract relevant content** from invoices of various categories **automatically** ,**efficiently** and **accurately**.

So that the automation reduces human efforts and human mistakes while typing. Hence reducing the **turnaround time** of invoice processing and giving a better **user experience**.

Sample invoices : Images, Image as PDF, PDFs



Automated Invoice processing

Variety of Invoices types

Size

Shape

Orientation

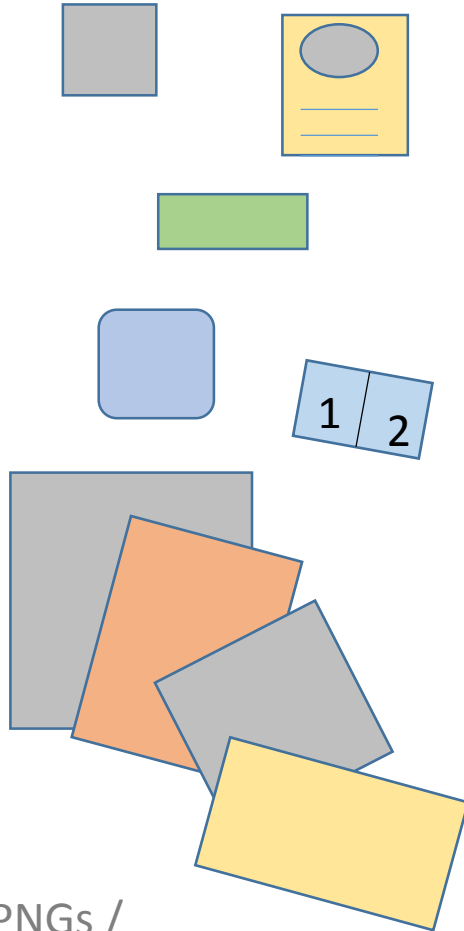
Color

Font

Font size

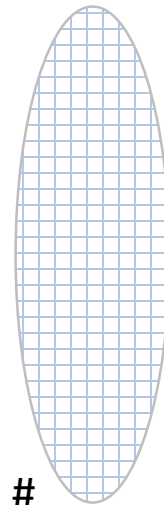
Rotation

resolution



PDFs / JPEGs / PNGs /
Images in PDFs

Quality Classifier



#

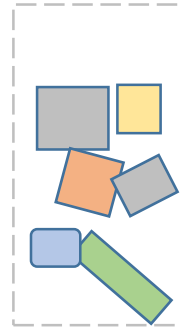
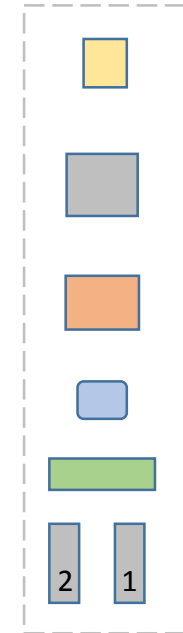


Image Preprocessing layer



OCR Engine



Entity miner



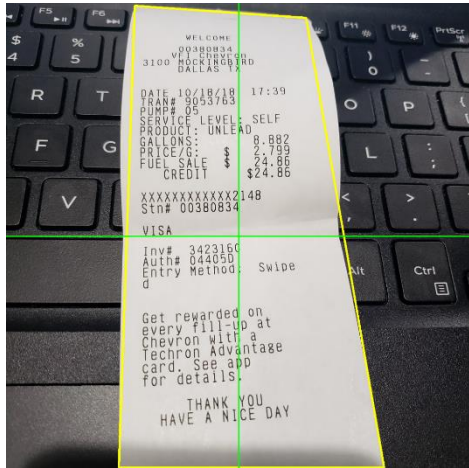
++

Quality classifier : To identify if an invoice image is good of good quality or not.
It has a machine learning component which will learn and improve as we give more and more samples.

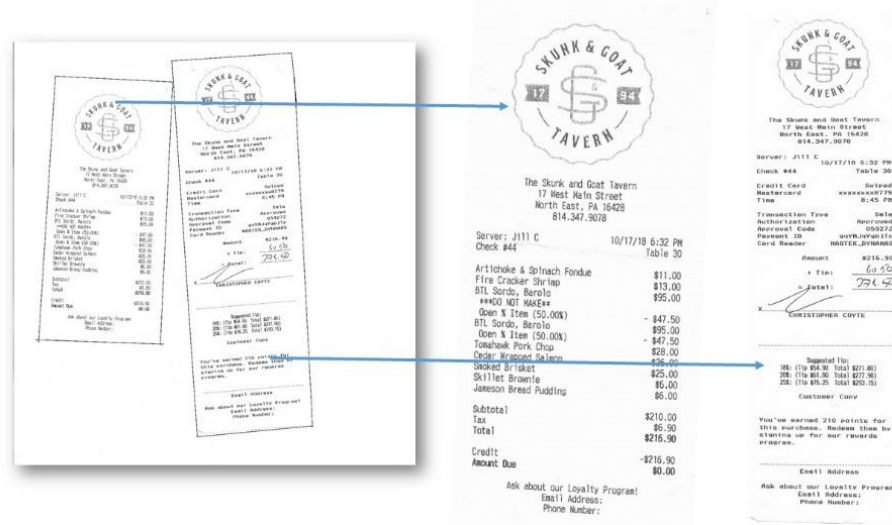
++ Entity miner : Could extract information from invoices.
Also this a configurable module, we could train by machine learning methods on what to extract.

Behind the scenes : Image Preprocessing layer

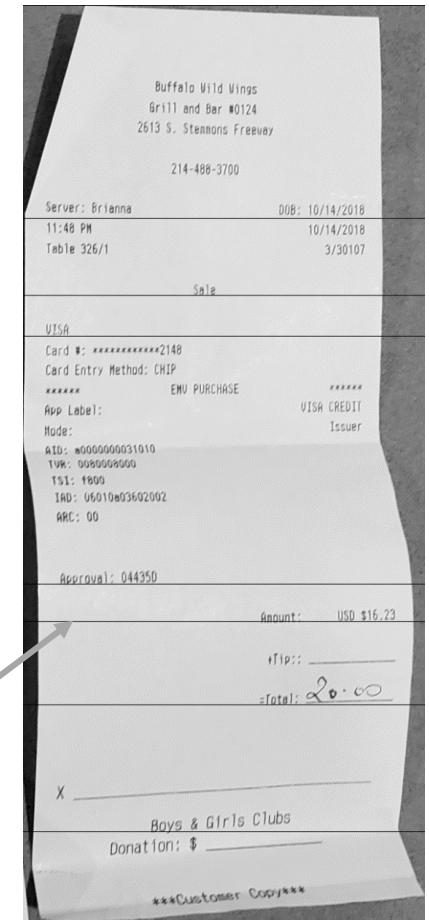
Deskewing



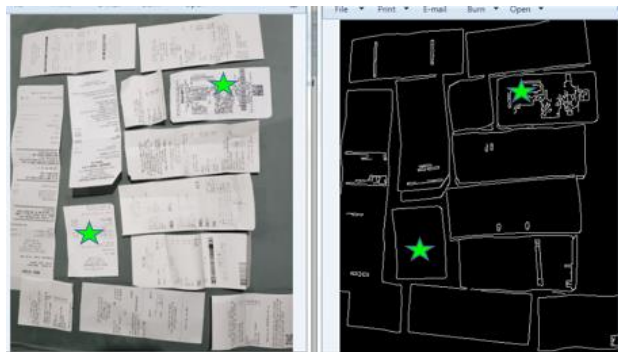
Handling multiple bills and skewness



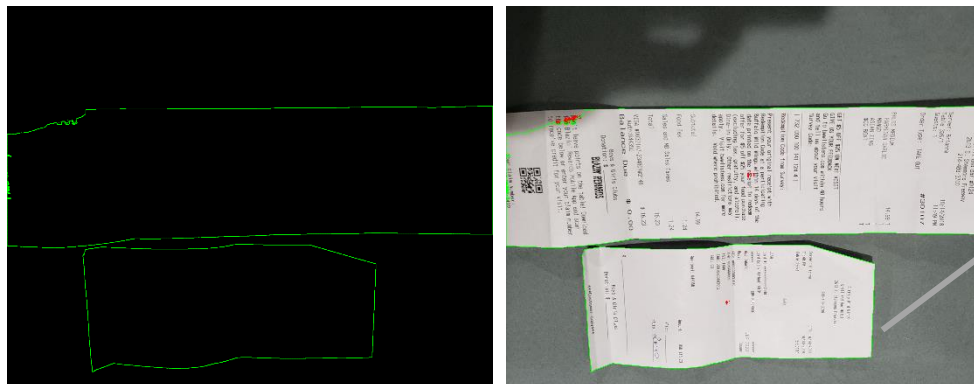
Rotation & Segmentation



Handling multiple bills



Extracting multiple bills





Successful Samples

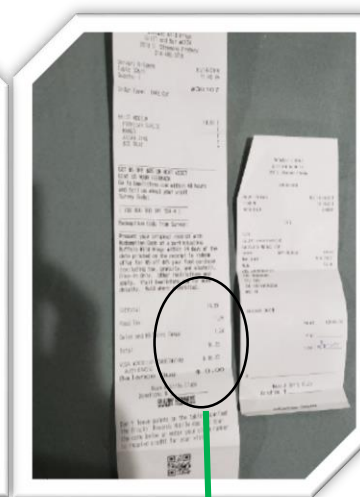
small bills



wide bills



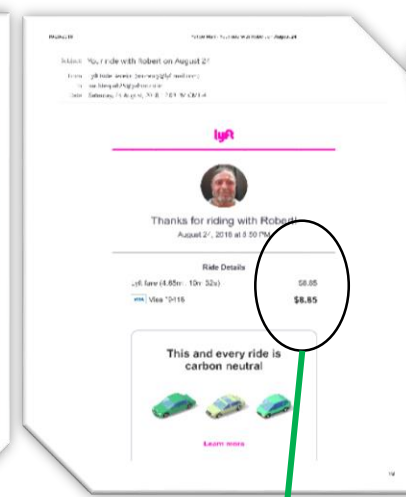
multiple bills



color



font



disoriented



Image
Preprocessing
layer

17.00
£14.17
£17.00
£17.00
£17.00

84.76
2.12
2.12
89.00

16.23
\$ 16.23
0.00

Amount
£43.60

\$8.85
\$8.85

£5.87
Thanks for choosing

Entity miner # Amount

[17.00 | 14.17]

[89.00 | 84.76 | 2.12]

[16.23 | 0.00]

[43.60]

[8.85]

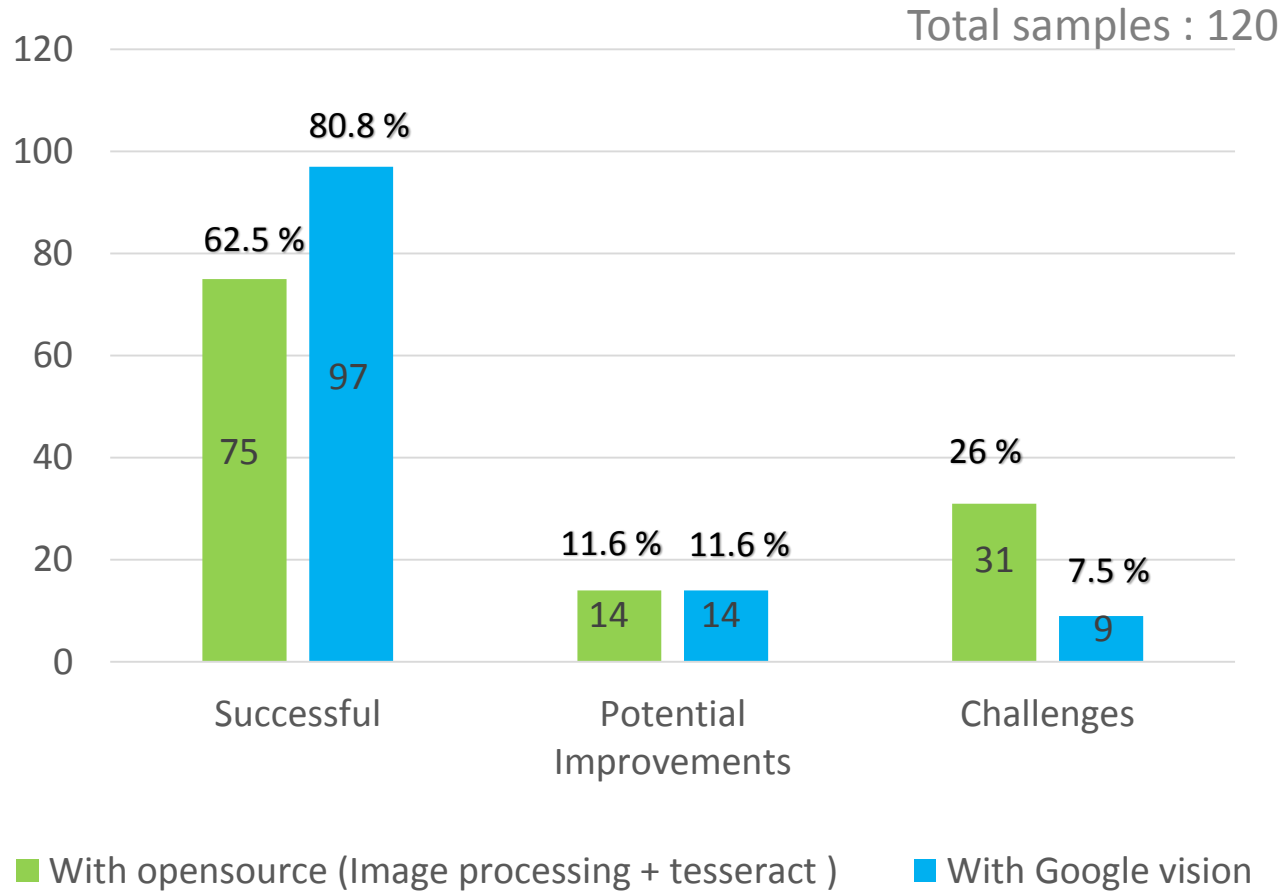
[5.87]

Date

Name

Customized entities

Success Metrics



	Success Cases	Potential Improvements	% Accuracy
Open source	75 of 120	14 of 120	62.5% → 74.1%
Google vision	97 of 120	14 of 120	80.8% → 92.5%

Highlights

- Open Source implementation could solve around 70% of all solvable invoices.
- Google vision could produce better results on much harder cases.
- Combination to prove optimum results.
- Performance to Cost ratio to be evaluated.

Potential Improvements

- By improving entity mining



- By better recognition of Rupee , dollar , pound symbols

- ₹ is recognized as 2 or 7 or 1.
- £ is recognized as 4.

- By training with invoice/bill type fonts

- 60 as 68.

Challenging Samples

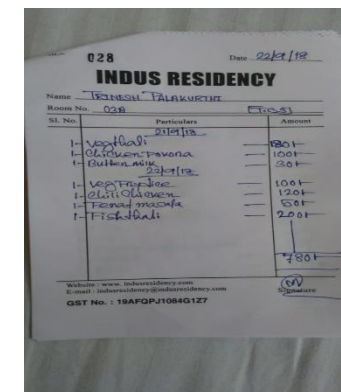
A



B



C



D



E



- Blurred/faded bills : B
- Documents overlapping in the scan : A, E
- Poor resolution : A, B
- Multiple languages : D
- Handwritten : C