

Information Extraction from Images using Optical Character Recognition and Named Entity Recognition

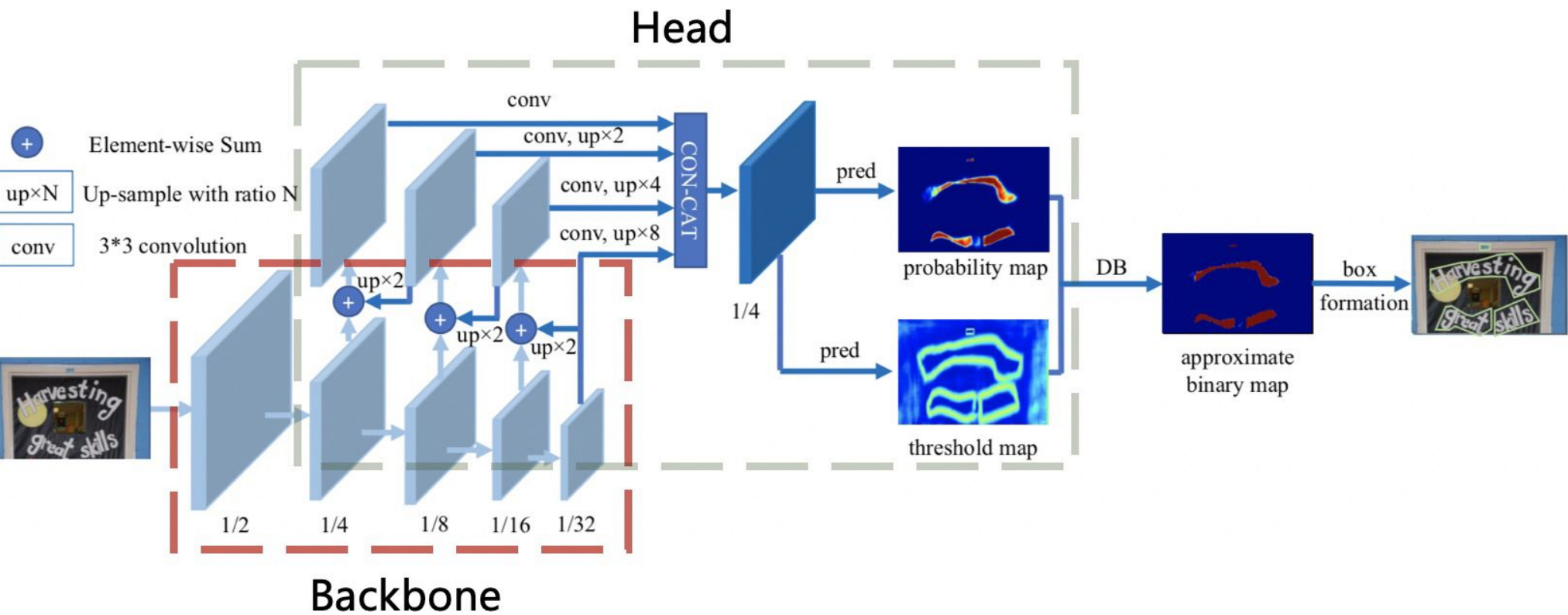
Anish Dixit, Neel Shah, Prerit Chaudhary

OPTICAL CHARACTER RECOGNITION

- OCR finds the coordinates of the character and also has to take in consideration spatial details, like baselining, spacing, variations in fonts, etc.
- Currently, many Deep learning-based text-extractors are being developed and have tackled even more variability in the input like noisy backgrounds and hand-written characters.
- In our use cases of digitized receipts and invoices, accurate identification of word clusters along with positional mapping.

Approach 1: PaddleOCR

- PP-OCR is an practical ultra lightweight OCR System.
- Text Detection -> Detected Boxes Rectification -> Text Recognition
- Each process uses complex Convolutional Layered Networks.



NAMED ENTITY RECOGNITION

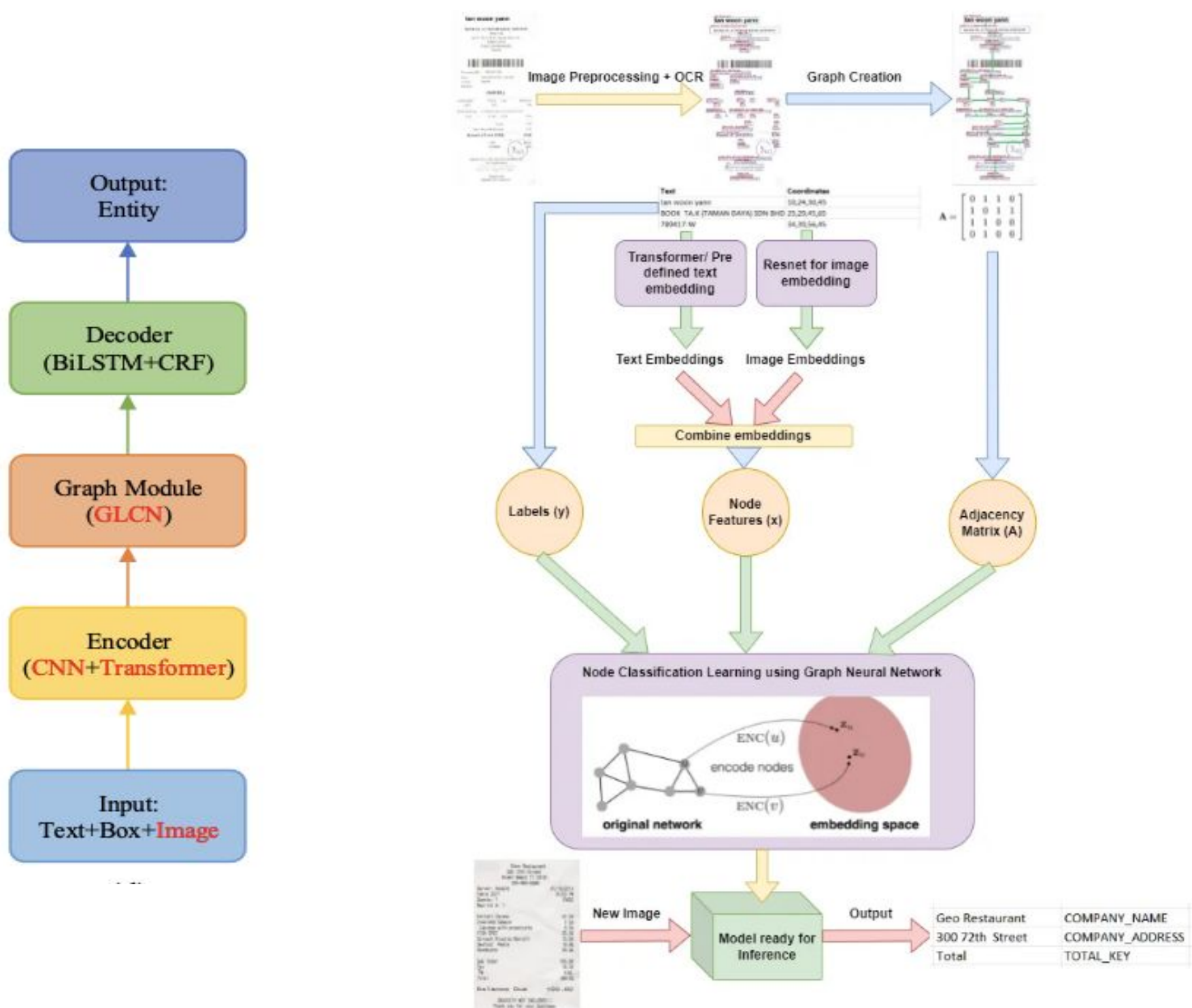
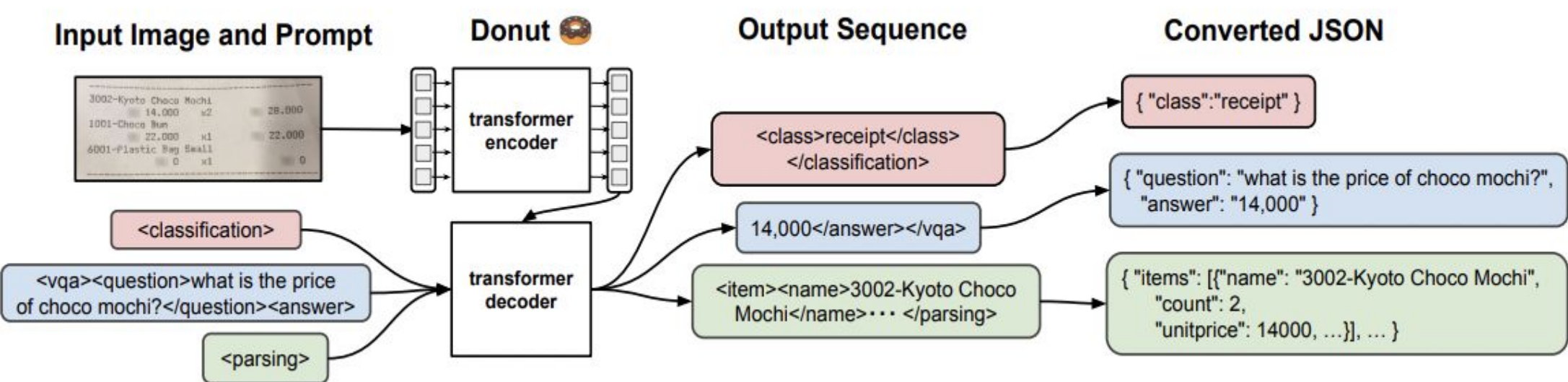
- The goal of NER is to label names of people, places, organizations, and other entities of interest in text strings..
- There are three major approaches to NER: lexicon-based, rule-based, and machine learning based.
- Various Deep Learning models of varying complexities are getting better and better at classifying words into certain buckets .

Approach 2: CNN + Transformer + Graph Convolution

- Key Information Extraction from complex documents using Graph convolution neural networks.
- Information Extraction -> Graph -> Text Classification
- Encoder: CNN + Transformer = Generate text and Image Embeddings
- Graph learning + Graph convolution = Richer semantic representation containing the textual and visual features
- Decoder: BiLSTM + CRF = Text classification and Tagging
- GCN embeds node feature vectors to represent the input node as a point in an N-dimensional space mapping the similar neighbor nodes to classify the text into classes

Approach 1: DonutProcessor

- Donut consists of an image transformer encoder and an autoregressive text transformer decoder.
- Performs document understanding tasks such as document image classification, form understanding, and visual question answering.

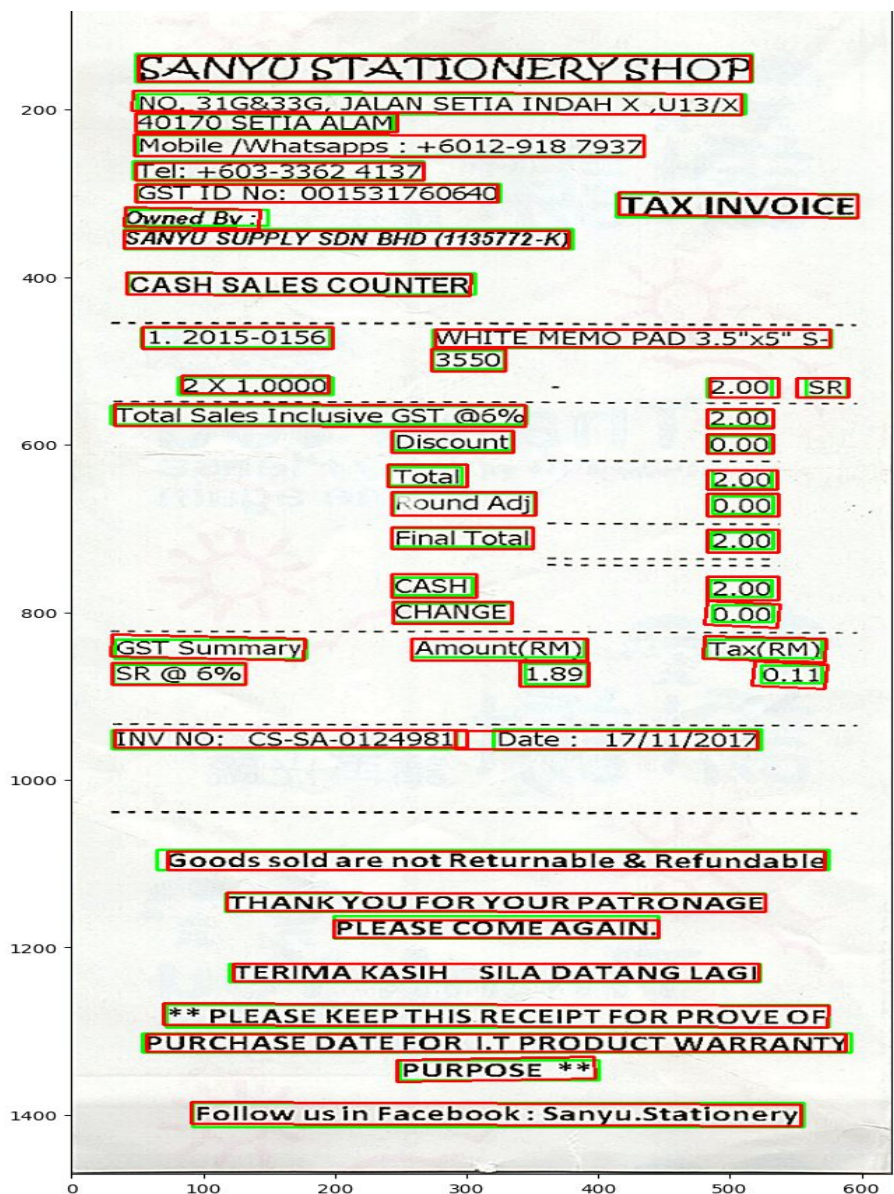


Results

- Mean Levenshtein Ratio: 0.934

```
{ "address": "NO. 31G&33G, JALAN SETIA INDAH X,U13/X 40170 SETIA ALAM",  
  "company": "SANYU STATIONERY SHOP",  
  "date": "17/11/2017",  
  "total": "2.00" }
```

Mean Average Precision: 0.827
Intersection over Union: 0.7432



Results

- Mean Levenshtein Ratio: 0.672

```
{ "company": "SANYU STATIONERY SHOP,company",  
  "address": "NO. 31G&33G\\, JALAN SETIA INDAH X \\,U13/X,address4",  
  "date": "17/11/2017" }
```

name	mEP	mER	mEF	mEA
company	0.246154	0.178771	0.20712	0.178771
address	0.355742	0.129131	0.189482	0.129131
total	0.406542	0.0722591	0.122708	0.0722591
date	0.527397	0.776471	0.628144	0.776471
overall	0.420058	0.210233	0.28022	0.210233

