Extract Text from Receipts using Deep Vision

Benjamin Maier, Hauke Lüdemann

University of Heidelberg

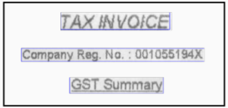
Text Extraction Pipeline

Our Focus

Extract Receipt from Background

Detect Bounding Boxes

Input for Text Recognition



Text Recognition

TAX INVOICE

Company Reg. No. : 001055194X

GST Summary

HON HWA HARDWARE TRADING
COMPANY REG. NO. : 001055194X
NO 37, JALAN MANIS 7, TAMAN SEGAR
56100 CHERAS, KUALA LUMPUR.
+603-9130 2672
GST REG : 001125220352
TAX INVOICE
CB# : 87870 21/09/2017 10:20:37 AM
M# : C2 - 0
CASHIER CASH1 -
QTY DESCRIPTION PRICE TOTAL(RM)
0.9 3/4" ALUMINIUM ROD 6.00 5.40 SR
5 PVC WALLPLUG 1.00 5.00 SR
(50PCS)
5.9 TYPE: 2 TOTAL 10.40
DISCOUNT: 0.00
ROUNDING ADJ 0.00
TOTAL INCLUSIVE GST: 10.40
CASH 10.40
GST SUMMARY
CODE % NET AMT GST TOTAL(RM)
SR SR 9.81 0.59 10.40
TOTAL 9.81 0.59 10.40
THANK YOU ! & PLEASE COME AGAIN !!
GOODS SOLD ARE NOT RETURNABLE FOR
REFUND OR EXCHANGE !!

Important Information Extraction

Name of Store
Purchased Items
Total Price

Text Detection

- First step is to localize text → bounding boxes
- Receipts with annotated bounding boxes as training data
- Detected bounding boxes as input for text recognition

Figure 1: Bounding boxes used for training.

EAST-Model: Architecture

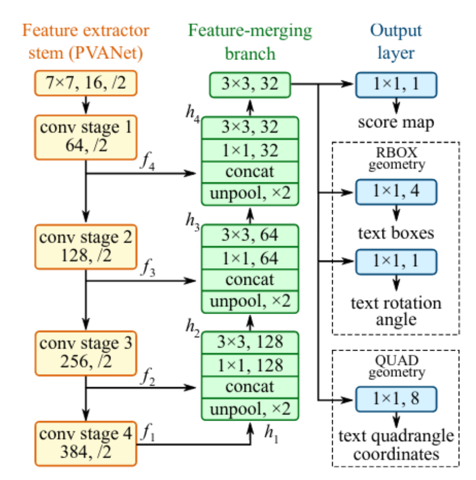


Figure 2: Architecture of the EAST-Network [6].

- EAST: Efficient and Accurate Scene Text Detector by Zhou *et al.* (2017)[6]
- Fully Convolutional Architecture
- Pre-trained classification net (VGG16) for feature extraction
- U-shaped design to merge features from different levels and keep computational cost small
- Loss function $L = L_{score} + L_{geo}$

EAST-Model: Loss function

Score Loss

- Score map S with value of 1 for every pixel inside bounding box
- Dice score for loss

$$L_{score} = 1 - \frac{2|S_{pred} \cap S_{true}|}{|S_{pred}| + |S_{true}|}$$

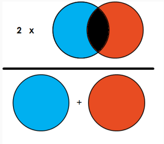


Figure 3: Illustration dice loss [4].

Geometry Loss

- Geometry map G storing offset to four corners for every pixel inside a box:
  $$G_i = \{x_1, x_2, \cdots, x_8\} \in G$$
- Loss

$$l_i = \sum_{i=1}^{8} \frac{\text{smoothed}_{\text{L1}}(x_i - \hat{x}_i)}{\text{shortest edge}}$$

$$L_{geo} = \frac{1}{N} \sum_{i=1}^{N} l_i \qquad \text{with N pixels}$$

$$\text{smoothed}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases}$$

Data set Text Detection

Data provided by:

ICDAR 2019 Robust Reading Challenge on Scanned Receipts OCR and Information Extraction

- Data set: 986 receipts in english
- Annotation: Bounding boxes and text
- 626 training images and 360 test images

Results Text Detection

- Best IoU score of 81.86%

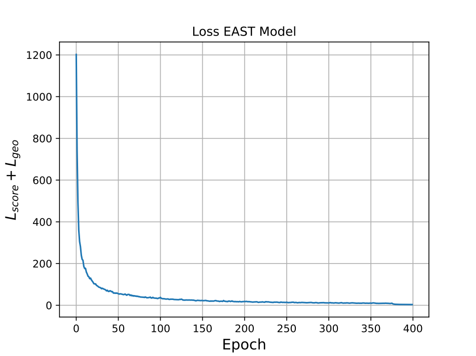


Figure 4: Loss after different number of epochs.

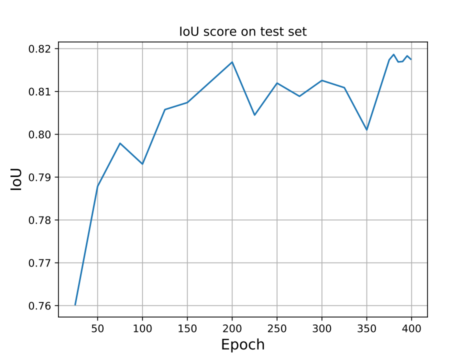


Figure 5: Intersection over Union.

Results Text-Detection - Example

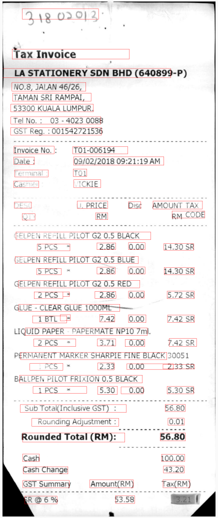


Figure 6: Text detection on test data.

Figure 7: Text detection on test image.

Text-Recognition

- Use ground truth bounding boxes for training
- Crop bounding boxes and scale as preprocessing
- Different approaches have been tested:
    - CRNN
    - DenseNet + RNN + CTC
    - DenseNet + CTC
    - Adding artificial data for training
    - Different CTC input lengths (32 and 64)

Figure 8: Example of text recognition.

Connectionist Temporal Classification (CTC) - Encoding

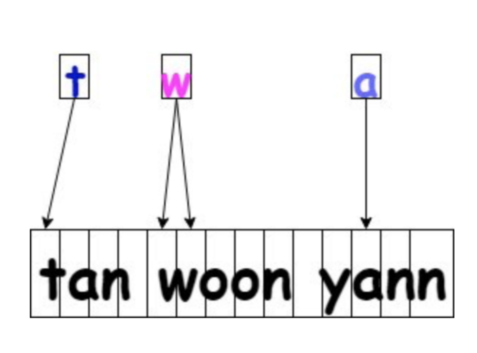


Figure 9: Text encoding mechanism for CTC Loss [2].

- Characters can span over multiple lines
- Introduce blank "-" character to encode duplicate characters (removed in decoding)

Examples:

"aa" $\rightarrow$ "a-a"

"a" $=$ "-a" $=$ "a- -" $=$ "aaa"

$\rightarrow$ different alignments represent same text

CTC Loss - Calculation

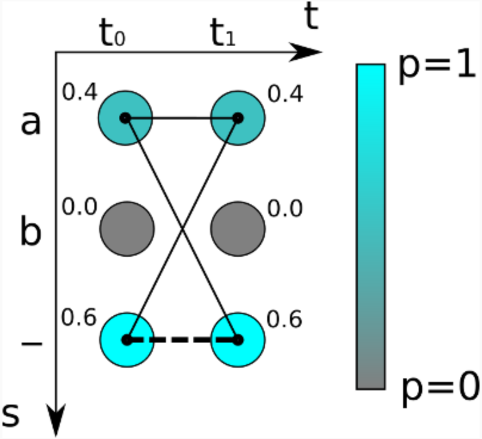


Figure 10: Example CTC Loss calculation [3].

- Output Net: Score for every character for every slot/time step
- Score: Sum output for all possible alignments of ground truth text

Example for ground truth : "a"

Possible alignments: "-a", "a-", "a"

Score: $0.4 \cdot 0.6 + 0.6 \cdot 0.4 + 0.4 \cdot 0.4 = 0.64$

$\rightarrow$ Maximize score

Data set Text Recognition

- Cropped bounding boxes for training
- 33626 images as training set
- Additionally create artificial images with data generator [1].

Results Text Recognition - Comparison of different Models

- Word Accuracy: Number of correctly predicted words/sentences
- Levenshtein Score: Similarity depending on number of character differences

	Dense Net 32	Dense Net 64	CRNN 32	CRNN 64	Tesseract
CTC Loss	0.0025	0.0059	0.1332	0.0147	–
Word Accuracy	57.6%	64.2%	54.8%	66.3%	53.6%
Levenshtein Score	85.6%	93.3%	88.4%	93.4%	90.0%

Table 1: Quantitative results of the different methods.

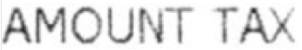
Result Text Recognition - Examples

Correct Predictions

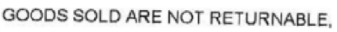
TAX INVOICE

53300 KUALA LUMPUR.

AMOUNT TAX

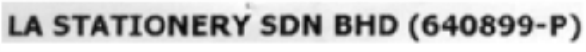


GOODS SOLD ARE NOT RETURNABLE,



Wrong Predictions

LA STATIONERY SDN BHD (G40899-P)



NOE (7 IDCATED THE ITEM SOLD HAS BEEN RELAED

PERMAIET MARKCR SHARPE FNE BLAK 001

CASHNIEOY :

0. PRICE

Results Text Recognition - Example Complete Receipt

318 02013

Tax Invoice

LA STATIONERY SDN BHD (640899-P)
NO.8, JALAN 46/26,
TAMAN SRI RAMPAI,
53300 KUALA LUMPUR.
Tel No. : 03 - 4023 0088
GST Reg. : 001542721536

Invoice No. : T01-006194
Date : 09/02/2018 09:21:19 AM
Terminal : T01
Cashier : VICKIE

DESC U. PRICE Disc AMOUNT TAX
QTY RM RM CODE

GELPEN REFILL PILOT G2 0.5 BLACK
5 PCS * 2.86 0.00 14.30 SR
GELPEN REFILL PILOT G2 0.5 BLUE
5 PCS * 2.86 0.00 14.30 SR
GELPEN REFILL PILOT G2 0.5 RED
2 PCS * 2.86 0.00 5.72 SR
GLUE - CLEAR GLUE 1000ML
1 BTL * 7.42 0.00 7.42 SR
LIQUID PAPER - PAPERMATE NP10 7ml.
2 PCS * 3.71 0.00 7.42 SR
PERMANENT MARKER SHARPIE FINE BLACK 30051
1 PCS * 2.33 0.00 2.33 SR
BALLPEN PILOT FRIXION 0.5 BLACK
1 PCS * 5.30 0.00 5.30 SR

Sub Total(Inclusive GST) : 56.80
Rounding Adjustment : 0.01
Rounded Total (RM): 56.80

Cash 100.00
Cash Change 43.20
GST Summary Amount(RM) Tax(RM)
SR @ 6 % 53.58 3.21

True text

TAX INVOICE
LA STATIONERY SDN BHD (640899-P)
NO.8, JALAN 46/26,
TAMAN SRI RAMPAI.
53300 KUALA LUMPUR.
TEL NO. : 03-4023 0088
GST REG. : 001542721536
INVOICE NO. : T01-006194
DATE: 09/02/2018 09:21:19AM
TERMINAL : T01
CASHIER : VICKIE
DESC U. PRICE DISC AMOUNT TAX
QTY RM RM CODE
GELPEN REFILL PILOT G2 0.5 BLACK
5 PCS * 2.86 0.00 14.30 SR
GELPEN REFILL PILOT G2 0.5 BLUE
5 PCS * 2.86 0.00 14.30 SR
GELPEN REFILL PILOT G2 0.5 RED
2 PCS * 2.86 0.00 5.72 SR
GLUE - CLEAR GLUE 1000ML
1 BTL * 7.42 0.00 7.42 SR
LIQUID PAPER - PAPERMATE NP10 7ML.
2 PCS * 3.71 0.00 7.42 SR
PERMANENT MARKER SHARPIE FINE BLACK 30051
1 PCS * 2.33 0.00 2.33 SR
BALLPEN PILOT FRIXION 0.5 BLACK
1 PCS * 5.30 0.00 5.30 SR
SUB TOTAL(INCLUSIVE GST) : 56.80
ROUNDING ADJUSTMENT : 0.01
ROUNDED TOTAL (RM): 56.80
CASH 100.00
CASH CHANGE 43.20
GST SUMMARY AMOUNT(RM) TAX(RM)
SR @ 6% 53.58 3.21

Predicted Text

TAX INVOICE
LA STATIONERY SDN BHD (G40899-P)
NO.8, JALAN 46/26,
TAMAN SRI RAMPAI,
53300 KUALA LUMPUR.
TEL NO. : 03 -4023.0088
GST REG. : 001542721536
INVOICE NO. : TO01-006194
DATE : 09/02/2018 09:21:19 AM
TER MINAL : T01
CASHNIEOY : /ICKIE
DES0 0. PRICE DISC AMOUNT TAX
QT / RM RM CODE
GELPEN REHIUL PILOT G2 0.5 BACK
5.PCS * 2.86 0.00 14.30 SR
GEUPEN REFILL PILOT G2 0.5 BLUE
5 PCS * 2.86 0.00 14.30 SR
GELPEN REHILL PILOT G2 0.5 RED
2 PCS E X 2.86 0.00 5.72 SR
GLUE - CLEAR GLUE 100OMT
1 BTL * 7.42 0.00 7.42 SR
LIOUID PAPER PAPERMATE NP10 7ML.
2 PCS X 3.71 0.00 7.42 SR
PERMAIET MARKCR SHARPE FNE BLAK 001
PCS X 2.33 0.00 2.33 SR
BALLPEN PILOT FRIXION 0.5 BLACK
1 PCS X 5.30 0.00 5.30 SR
SUB TOTA(INCLUSIVE GST) : 56.80
ROUNDING ADJUSTMENT : 0.01
ROUNDED TOTAL (RM): 56.80
CASH 100.00
CASH CHANGE 43.20
GST SUMMARY AMOUNT(RM) TAX(RM)
SR D 6 % 53.58 .21

Figure 11: Example text detection on complete receipt.

Summary

- Successfully implemented Text Detection and Text Recognition
- Dense Net and CRNN perform very similar on Text Recognition
- CRNN is faster then Dense Net
- CTC input size important for good predictions

Outlook

- Data augmentation for Text Detection $\rightarrow$ detect rotated bounding boxes
- Test different architectures of Dense Net and CRNN to improve prediction of long text further
- Improve image preprocessing $\rightarrow$ improve contrast
- General hyper parameter optimization

References

Edouard Belval. *TextRecognitionDataGenerator*. https://github.com/Belval/TextRecognitionDataGenerator. [Online; visited 23.07.2020]. 2019.

Dinh Sang and Le Cuong. “Improving CRNN with EfficientNet-like feature extractor and multi-head attention for text recognition”. In: Dec. 2019, pp. 285–290. ISBN: 978-1-4503-7245-9. DOI: 10.1145/3368926.3369689.

Harald Scheidl. *An Intuitive Explanation of Connectionist Temporal Classification.* https://towardsdatascience.com/intuitively-understanding-connectionist-temporal-classification-3797e43a86c. [Online; visited 20.07.2020]. 2018.

Ekin Tiu. *Metrics to Evaluate your Semantic Segmentation Model.* https://towardsdatascience.com/metrics-to-evaluate-your-semantic-segmentation-model-6bcb99639aa2. [Online; visited 20.07.2020]. 2019.

Sik-Ho Tsang. *Review: DenseNet — Dense Convolutional Network (Image Classification).* https://towardsdatascience.com/review-densenet-image-classification-b6631a8ef803. [Online; visited 20.07.2020]. 2018.

Xinyu Zhou et al. “EAST: An efficient and accurate scene text detector”. In: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. Vol. 2017-Janua. 2017, pp. 2642–2651. ISBN: 9781538604571. DOI: 10.1109/CVPR.2017.283. arXiv: 1704.03155.

DenseNet

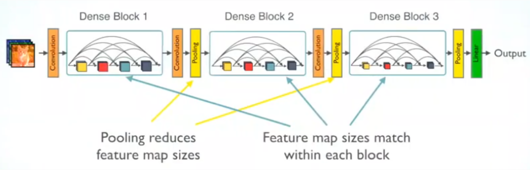


Figure 12: Basic structure of the DenseNet [5].

- Convolutional part build on dense blocks
- Like ResNet each layer receives additional data from all preceding layers, but using concatenation
- Down sample until height of each column is equal 1

Text Recognition Models - Loss

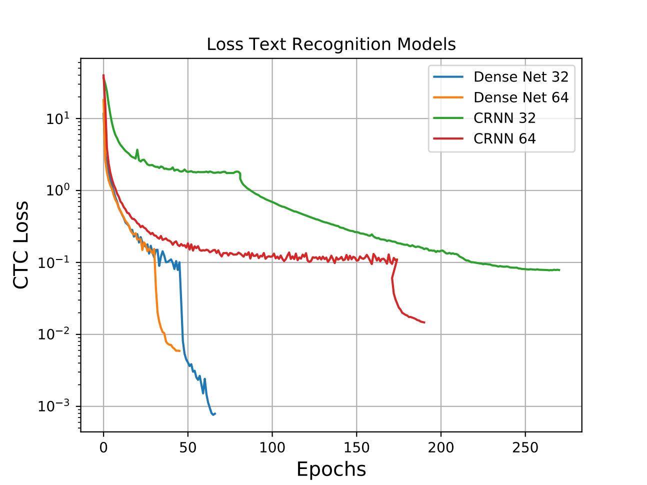
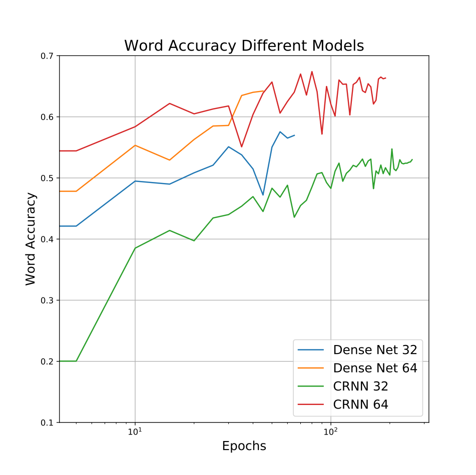
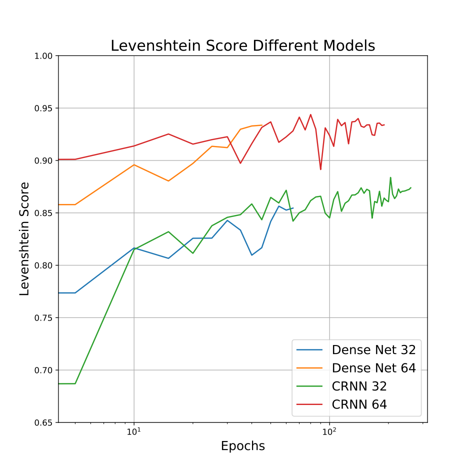


Figure 13: CTC loss for Text Recognition Models.

Results Text Recognition - Score





Results Text Recognition - Example Complete Receipt


0E
AEON CO. (M) BHD (126926-H)
3RD FLR, AEON TAMAN MALURI SC
JLN JEJAKA, TAMAN MALURI
CHERAS, 55100 KUALA LUMPUR
GST ID : 002017394688
SHOPPING HOURS
MON-SUN:1000 HRS - 2200 HRS
VALUED CUSTOMER: 1130307913

1x 000005469765 6.65SR
TOPVALU FLOOR C
1x 000001101575 5.60SR
CIF REGULAR 50

Sub-total 12.25
Total Sales Incl GST 12.25
Total After Adj Incl GST 12.25
CASH 50.00
Item Count 2 Change Amt 37.75
Invoice No: 2018031430090060035
GST Summary Amount Tax
SR @ 6% 11.55 0.70
Total 11.55 0.70
14/03/2018 12:43 3009 006 0060035
0284846 AMIRUL

True text	Predicted Text
AEON CO. (M) BHD (126926-H)	AEON CO. M) BHD (126926-H!
3RD FLR, AEON TAMAN MALURI SC	3RD FLR. AEON TAMAN MALURI SC
JLN JEJAKA, TAMAN MALURI	JLN JEJAKA. TAMAN MALURI
CHERAS, 55100 KUALA LUMPUR	CHERAS. 55100 KUALA LUMPUR
GST ID: 002017394688	GST D : 002017394688
SHOPPING HOURS	SHOPPING HOURS
MON-SUN: 1000 HRS - 2200 HRS	MON-SUN:1000 HRS - 2200 HRS
VALUED CUSTOMER: 1130307913	VAI UED CUSTOMER: 1130307913
1X 000005469765 6.65SR	1X 000005469765 6.65SR
TOPVALU FLOOR C	TOPVALU FIOOR C
1X 000001101575 5.60SR	1X 000001101575 5.60SR
CIF REGULAR 50	CIF REGULAR 50
SUB-TOTAL 12.25	SUB-TOTAL 12.25
TOTAL SALES INCL GST 12.25	TOTAL SALES INCL GST 12.25
TOTAL AFTER ADJ INCL GST 12.25	TOIAL AFTER ADI INCL GST 12.25
CASH 50.00	CASH 50.00
ITEM COUNT 2 CHANGE AMT 37.75	IFEM COUNT 2 CHAN9E AMT 37.75
INVOICE NO:2018031430090060035	INVOICE NO : 2018031430090060035
GST SUMMARY AMOUNT TAX	GST SUAMARY AMOUNT TAX
SR @ 6% 11.55 0.70	SR 6% 11.55 0.70
TOTAL 11.55 0.70	TOTAL 11.55 0.70
14/03/2018 12:43 3009 006 0060035	14/03/2018 12:43 3009 006 0060035
0284846 AMIRUL	0284846 AMIRUL

Figure 14: Example text detection on complete receipt.