TRIE: End-to-End Text Reading and Information Extraction for Document Understanding

Peng Zhang¹
Yunlu Xu¹
Zhanzhan Cheng²¹
Shiliang Pu¹
Jing Lu¹
Liang Qiao¹
Yi Niu¹
Fei Wu²

1. Hikvision Research Institute, Hangzhou, China
2. Zhejiang University, Hangzhou, China
Background

◆ VRD (Visually Rich Document)
  - Text
    - plain text
  - Visual
    - layout
    - tabular
    - font size

◆ VRD Understanding

Receipt.

Entities to extract.

Name: OJC
MARKETING SDN BHD
Date: 15/01/2019
Total: 193.00

VRD Understanding Algorithm

(a) Taxi Invoices  (b) Receipt  (c) Resume

Receipt.
Background

◆ Problem of current framework

Limitation 1: Limited visual features in IE.
Keep:
\[ x_0, y_0, \ldots, x_3, y_3, '15/01/2019' \]
\[ x_0, y_0, \ldots, x_3, y_3, '193.00' \]
Lost:
Font, Color, Layout etc.

Limitation 2: Ignoring relations between OCR & IE.

◆ Motivation

Advantage 1: Multimodal fusion in IE.
Keep:
\[ x_0, y_0, \ldots, x_3, y_3, '15/01/2019' \]
\[ x_0, y_0, \ldots, x_3, y_3, '193.00' \]
Font, Color, Layout etc.

Advantage 2: Bridging OCR & IE,
Forward: OCR boost IE
Backward: IE boost OCR
Method

Overall Architecture.

Forward Process

Backward Process

Shared Convolutions

Text Detection

ROI Align

Text Recognition

Encoder

Decoder

Position features

Visual features

Textual features

Multimodal context block

Visual Context

Textual Context

Hidden Layers

Context Feature

Information Extraction

Name: Martin
E-mail: abc123***@gmail.com
Edu-period: 2018.09-2019.06

Phone: 132***6116
E-mail: abc123***@gmail.com

2018.09--2019.06
Computer Science and Technology
Fu*** University
GPA: 3.88 / 4.0

2014.09--2018.06
Computer Science and Technology
Bachelor of Science
Fu*** University
GPA: 3.83 / 4.0

Textual Context Feature

Visual Context Feature

BLSTM
## Experiment

### Datasets

| Dataset         | Training | Testing | Entities | Layout   | Text Type |
|-----------------|----------|---------|----------|----------|-----------|
| Taxi Invoices   | 4000     | 1000    | 9        | Fixed    | Struct    |
| SROIE           | 626      | 347     | 4        | Variable | Struct    |
| Resumes         | 1978     | 497     | 6        | Variable | Semi-struct |

### Dataset Statics

(a) Taxi Invoices  
(b) Receipt  
(c) Resume

### Performance Summary

| Entities     | Chargrid(TR) | NER(TR) | GCN(TR) | Our Model |
|--------------|-------------|---------|---------|-----------|
| Code         | 89.4        | 94.5    | 97.0    | 98.2      |
| Number       | 85.3        | 92.4    | 93.7    | 95.4      |
| Date         | 89.8        | 82.5    | 93.0    | 94.9      |
| Pick-up time | 82.9        | 60.0    | **86.3**| 84.6      |
| Drop-off time| 87.4        | 81.1    | 91.0    | 93.6      |
| Price        | 93.0        | 94.5    | 93.6    | 94.9      |
| Distance     | 92.7        | 93.6    | 91.4    | 94.1      |
| Waiting      | 89.2        | 85.4    | 91.0    | 92.4      |
| Amount       | 80.2        | 86.3    | 88.7    | 90.9      |
| Avg          | **87.77**   | **85.59**| **91.74**| **93.26** |

**Taxi Invoices Dataset**

| Setting                  | Model                | F1-Score |
|--------------------------|----------------------|----------|
| Setting 1:               |                      |          |
| Prediction of bboxes     | Chargrid(TR)         | 78.24    |
| and transcript of texts  | NER(TR)              | 69.09    |
|                          | GCN(TR)              | 76.51    |
|                          | Our model            | **82.06**|
| Setting 2:               |                      |          |
| Groundtruth of bboxes    | Character-Word LSTM  | 90.85    |
| and transcript of texts  | [24]                 |          |
|                          | LayoutLM[54]         | 95.24    |
|                          | PICK[58]             | 96.12    |
|                          | Our model            | **96.18**|

**ICDAR2019 SROIE Dataset**
## Experiment

### Performance Summary

| Entities     | Chargrid(TR) | NER(TR) | GCN(TR) | Our Model |
|--------------|--------------|---------|---------|-----------|
| Name         | 43.4         | 42.7    | 42.9    | 45.7      |
| Phone        | 87.0         | 86.6    | 83.3    | 88.0      |
| E-mail       | 70.9         | 69.6    | 68.0    | 74.9      |
| Edu-period   | 77.1         | 68.7    | 62.2    | 81.4      |
| University   | 74.7         | 86.0    | 82.3    | 87.4      |
| Major        | 72.1         | 80.4    | 78.7    | 80.8      |
| Avg          | 70.87        | 72.33   | 69.57   | 76.3      |
| Speed(ips)   | 1.13         | 1.69    | 1.62    | 1.76      |

### Discussion

- Effects of multimodal features on IE.
- Effects of E2E framework on text reading.
- Effects of layers and heads in textual context block.
见远行更远
See Far, Go Further