# DavarOCR: A Toolbox for OCR and Multi-Modal Document Understanding

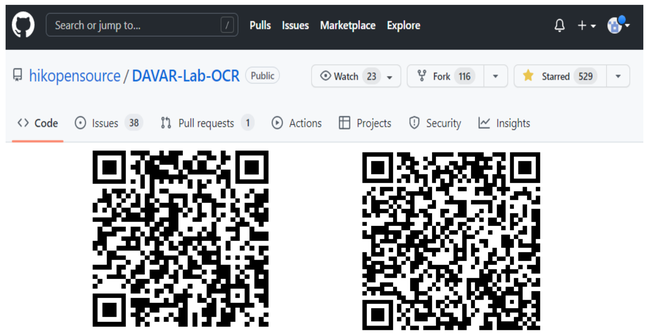
Liang Qiao, Hui Jiang, Ying Chen, Can Li, Pengfei Li, Zaisheng Li, Baorui Zou, Dashan Guo, Yingda Xu, Yunlu Xu, Zhanzhan Cheng, Yi Niu

Hikvision Research Institute  Contact Email: qiaoliang6@hikvision.com, chengzhanzhan@hikvision.com

## Introduction

DavarOCR is an open-source toolbox for OCR (Optical Character Recognition) and Document Understanding tasks.

- PyTorch + CUDA
- Apache-2.0 License
- 10 different task forms
- 20+ advanced algorithms
- Detailed Instructions and models
- Compatible with mmdetection & mmcv




**Paper**  **Github**

## Current Support Methods

| Basic OCR Tasks | | | | Document Understanding Tasks | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Text Detection | Text Recognition | End-to-End Text Spotting | Video Text | Layout Analysis | Information Extraction | Named Entity Recognition | Reading Order Detection | Table Recognition | Table Understanding |
| • EAST<br>• Mask-RCNN<br>• Text Perceptron Det | • Attention<br>• CRNN<br>• ACE<br>• SPIN<br>• RF-Learning | • Mask-RCNN<br>• Text Perceptron<br>• MANGO<br>• DLD | • YORO | • VSR | • Chargrid<br>• TRIE | • Bert NER<br>• BiLSTM+CRF | • GCN-PN | • LGPMA | • CTUNet (to be released) |

➢ More algorithm and domain support is on the way: Text Synthesis, Pre-trained Visual & Language Models, Doc VQA, etc.

➢ Welcome to public pull request (bug fixes, new features, new algorithms, etc.)

## Highlight Features

### ➢ Across Tasks Module Sharing

- Modules in mmdetection: BACKBONE, NECK, ROI_EXTRACTOR, HEAD, etc.
- New Modules in davarocr: TRANSFORMATION, EMBEDDING, CONNECT, etc.

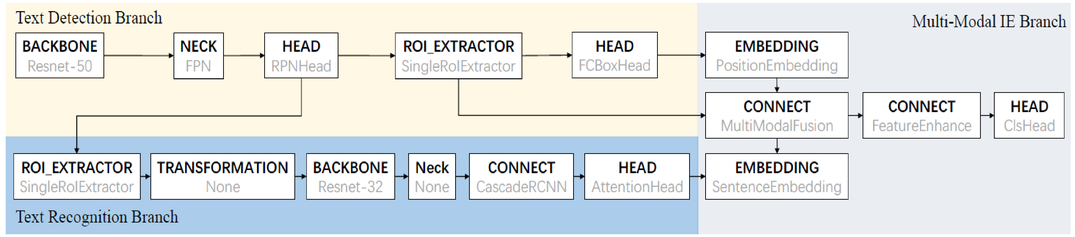



**Figure 1: The modular design for the end-to-end KIE model TRIE.**

### ➢ Uniform Data Label Format

- Unified basic formats for task sharing.
- Support for image/video/plain text.

```
{
  "Images/train/img1.jpg": {          # Relative path of images
    "height": int,                    # Image height
    "width": int,                     # Image width
    "content_ann": {                  # First-level annotations
      "bboxes": List[List[int|float]],          # Bounding box for each obj
      "cbboxes": List[List[List[int|float]]],   # Character-wised bounding boxes
      "cares": List[int],             # Whether the bboxes is ignored
      "labels": List[List[int]],      # Category for each box
      "texts": List[str]              # Transcriptions of text
    },
    "content_ann2":{                  # Second-level annotations
      ...                             # Same format with 1st-level
    }
  },
  ....
}
```

**Figure 2: Illustration of the basic data label**