Object Detection and Localization for Document Profiling Systems

—

Yapay Zeka Dersi Sunumu

Bekir Bakar

Sample Document

Sample Document

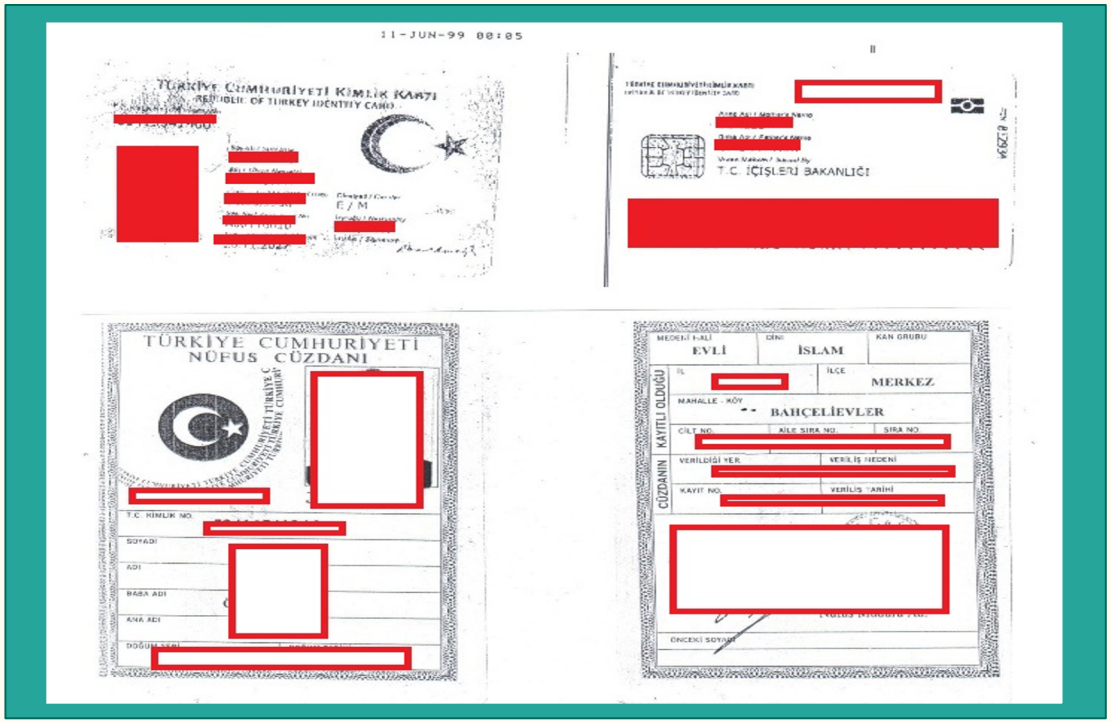
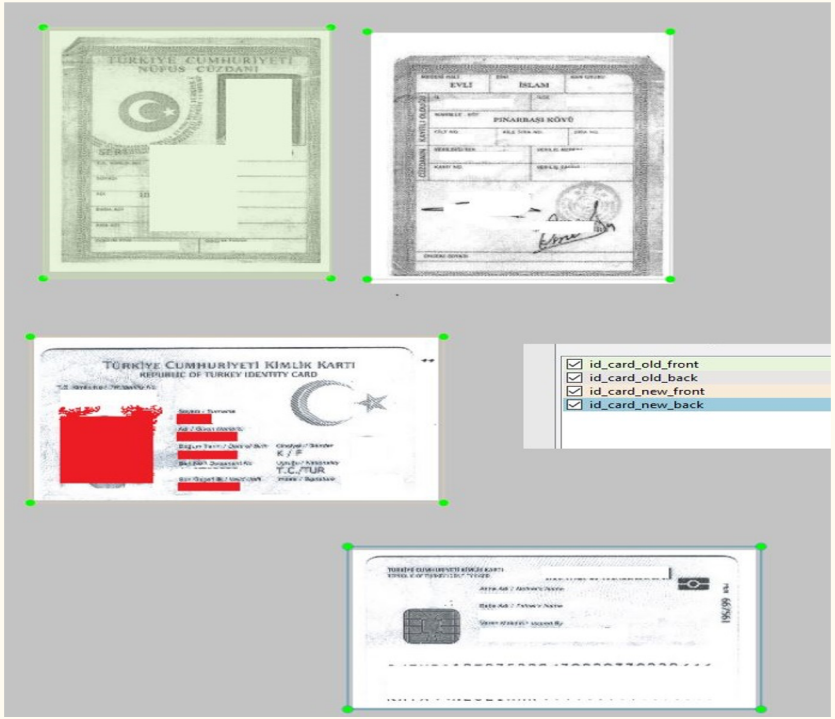


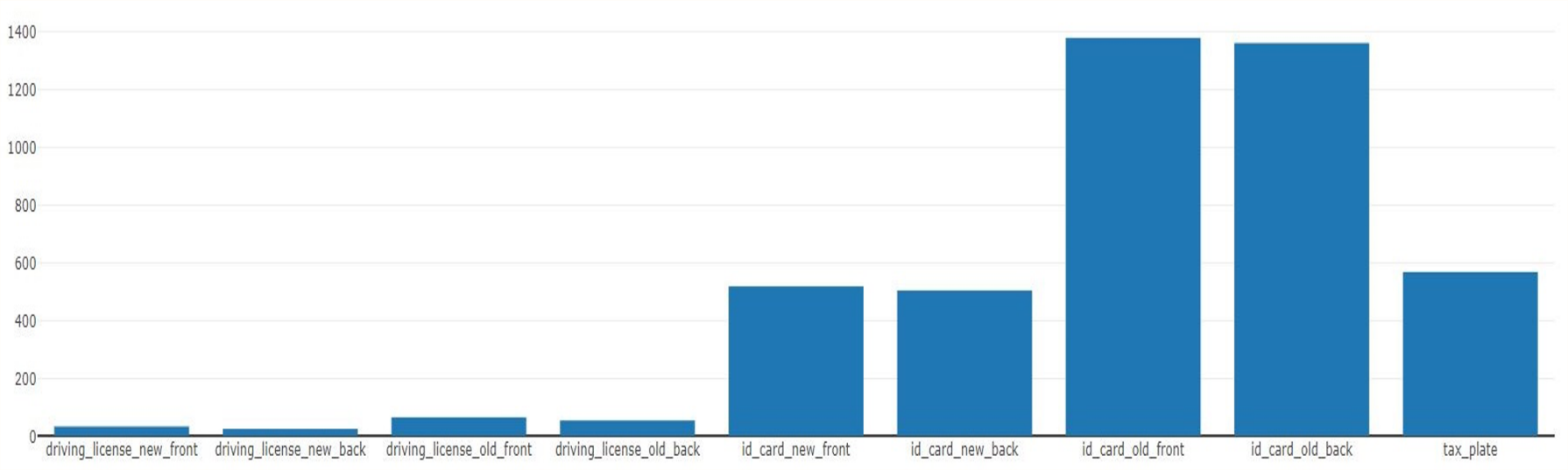
Image Annotation Sample Output




```
<object>
  <name>id_card_old_front</name>
  <bndbox>
    <xmin>38</xmin>
    <ymin>60</ymin>
    <xmax>260</xmax>
    <ymax>353</ymax>
  </bndbox>
</object>

<object>
  <name>id_card_new_front</name>
  <bndbox>
    <xmin>28</xmin>
    <ymin>421</ymin>
    <xmax>347</xmax>
    <ymax>615</ymax>
  </bndbox>
</object>
```

Image Dataset



Deciding clases
Sparse dataset
Data augmentation

Front and back of documents
Which images stands together?

Dataset

1. Pre-processing
    a. Noise Removal
    b. Skew Correction
    c. Perspective Correction
2. Fixed Size Images
        i. Upscaling
        ii. Downscaling
3. Annotation

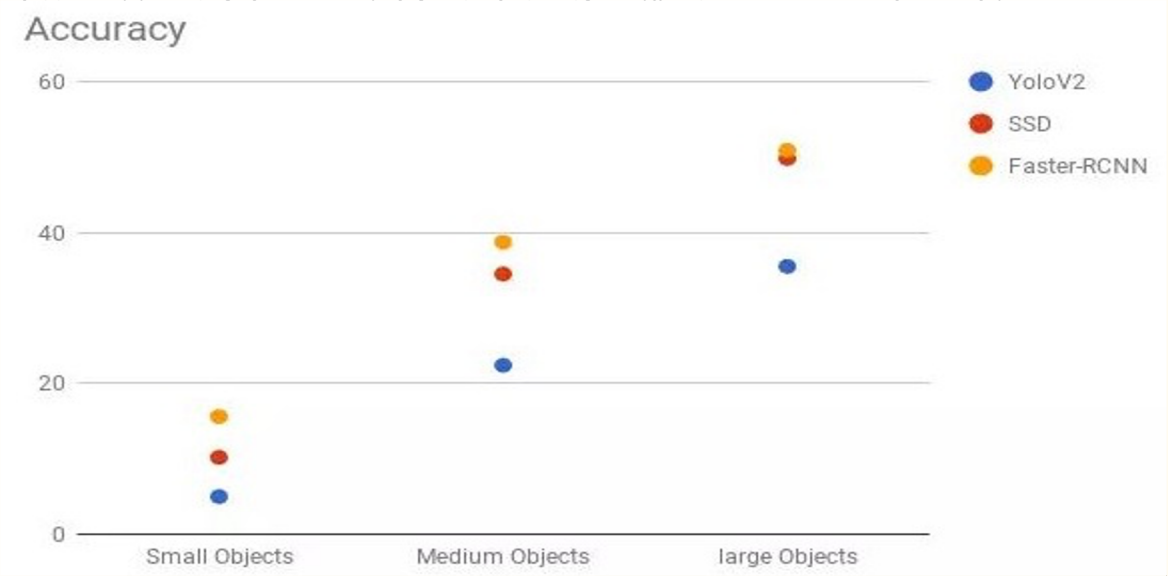
Object Detection

- HOG Features
- SPP-net
- CNN based Detectors
- SSD (regression based detector)
- YOLO (regression based detector)

Compare Detectors

https://cv-tricks.com/object-detection/faster-r-cnn-yolo-ssd/

Compare Detectors



https://cv-tricks.com/object-detection/faster-r-cnn-yolo-ssd/

Single Shot Detector

- Runs single network
- Accuracy - speed trade-off

1. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, October). Ssd: Single shot multibox detector. In European conference on computer vision (pp. 21-37). Springer, Cham.

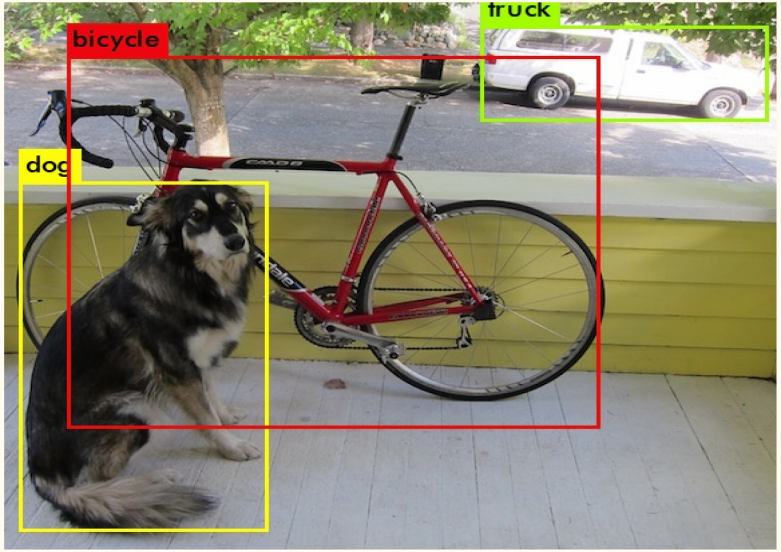
Regional Convolutional Neural Network (CNN)

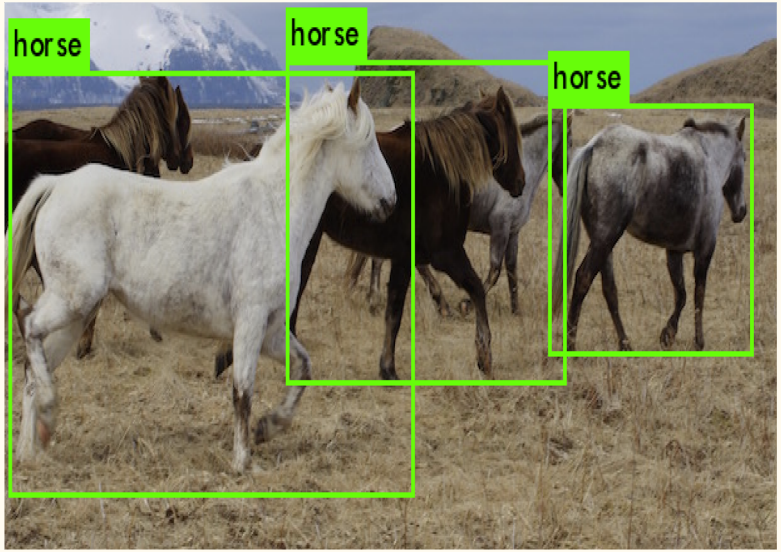
- Mask R-CNN [1]
- Faster R-CNN [2]
- Cascade R-CNN [3]

- According to literature: High accuray slow speed
- Since regression based models succeed acceptable results, CNN based detector loss popularty

1. He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).
2. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. arXiv preprint arXiv:1506.01497.
3. Cai, Z., & Vasconcelos, N. (2018). Cascade r-cnn: Delving into high quality object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 6154-6162).

You Only Look Once (YOLO)





1. https://pjreddie.com/darknet/yolo/
2. https://youtu.be/MPU2HistivI
3. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

You Only Look Once (YOLO) - Overview

- Prior detection systems repurpose classifiers or localizers to perform detection.
- Apply the model to an image at multiple locations and scales.
- High scoring regions of the image are considered detections.
- A single neural network to the full image. This network divides the image into regions and predicts bounding boxes and probabilities for each region.
- Bounding boxes are weighted by the predicted probabilities.

1. https://pjreddie.com/darknet/yolo/
2. https://youtu.be/MPU2HistivI
3. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

You Only Look Once (YOLO) - Versions

- Version 1: Proposal Initial Version
- Version 2 : YOLO9000 / YOLOv2
  - Inclusion of batch Normalization layers after each Conv Layer
  - It has 30 layers in comparison to YOLO v1 26 layers.
  - Anchor Boxes were introduced.
- Version 3: YOLOv3
  - 06 layers neural network
  - Detection on 3 scales for detecting objects of small to very large size
  - 9 anchor boxes taken; 3 per scale. Hence more bounding boxes are predicted than YOLO9000 & YOLOv1
  - MultiClass problem turned in MultiLabel problem
  - Certain changes in the Error function.
  - Quite good with small objects
- Ultranaltyics Implementations: https://github.com/ultralytics/yolov3

1. https://pjreddie.com/darknet/yolo/
2. https://youtu.be/MPU2HistivI
3. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

You Only Look Once (YOLO) - Implementation

- Annotation over Dataset
- Train
- Evaluate
- Report Results

1. https://pjreddie.com/darknet/yolo/
2. https://youtu.be/MPU2HistivI
3. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

Tools and Technologies

- Python, C/C++
- Darknet
- PyTorch
- Tensorflow
- Keras
- Vott Image Annotation - LabeImage

Conclusion - Throwbacks - Future Works

- No OCR data from bounding boxes
- Lack of data (hand crafted methods)
- Rule based validator in case of wrong output
- System speed
- System accuray
- Named Entity Recognition
- Rule Based Models

Final

Info

- Bekir Bakar
- 193331541008
- Elektrik Elektronik Mühendisliği
- Bursa Teknik Üniversitesi

Contact

- b.bakar@outlook.com