

• Create a python script for document layout analysis and extraction

- Convert pdf documents in a folder to png images with 95dpi (documentname _pagenumber.png)
- Run yolov5s (small) pretrained object detection model in tensorflow over the images
 - Yolov5s_name_name.pt weights model will be taken from a user specified directory
- Extract the recognized objects inside the bounding box above a user defined confidence threshold (0.1) as images in native resolution as png
- Save the images in a new directory within the document directory with the name: detectionconfidence_imagenumber_documentname_pagenumber.png
- Create a richly annotated python script that is easy to read and run for a novice
- Make variables for dpi values (pdf2image library)



