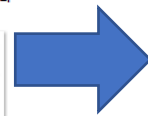
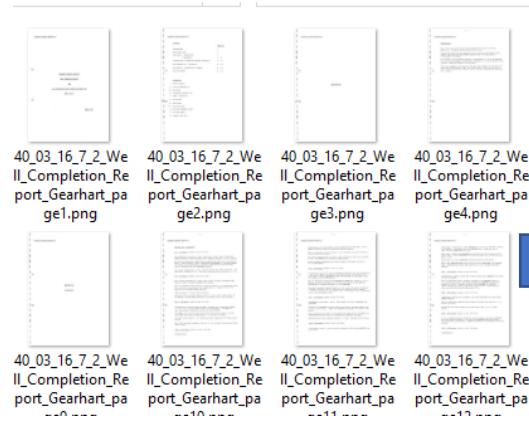
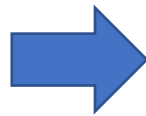


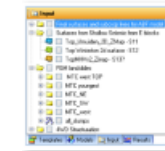
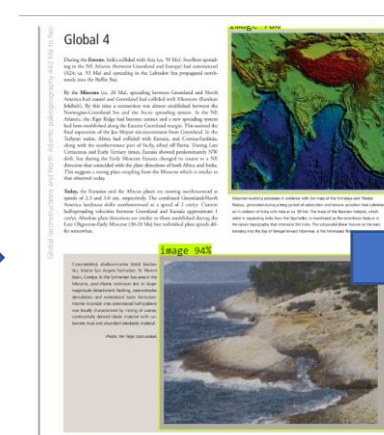
• Create a python script for document layout analysis and extraction

- 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_imagewidth_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)

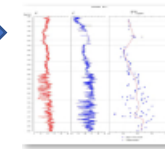
2_01_1_3_4_Completion_Report.pdf
 3_2_2_3_COMPLETION_REPORT_AND_LC
 5_01_30_6_12_Completion_Report_and_
 5_34_10_17_Completion_Report.pdf
 6_01_31_2_11_Completion_Report_and_
 7_01_30_6_13_Completion_Report_and_
 10_30_3_3_Completion_Report.pdf
 11_30_9_2_COMPLETION_REPORT.pdf
 12_01_Completion_Report_and Comple
 13_7117_9_2_COMPLETION_REPORT_AN
 14_01_31_2_12_Completion_Report_and
 15_01_Completion_Report_and Compo:
 17_01_Completion_Report.pdf
 18_01_Completion_Report.pdf
 19_01_Completion_Report_and Compo:



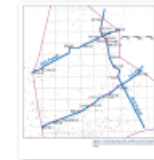
nages/8 (latest)



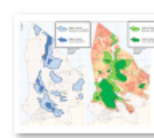
0.84_tno 2013
r10060_report_incl.
appendices_high
res.pdf_-14500305...



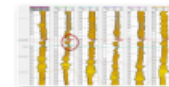
0.84_TNO report
TNO 2015
R10343_JUSTRAT_FI
NAL REPORT INCL...



0.84_TNO Report
2017 R10425 -
MMU.pdf_47059477
6_p26.png_dla_1....



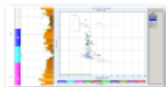
0.84_TNO report
TNO 2015
R10343_JUSTRAT_FI
NAL REPORT INCL...



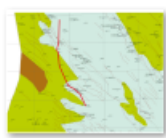
0.84_TNO Report
2017 R10425 -
MMU.pdf_47059477
6_p27.png_dla_1....



0.84_TNO report
TNO 2015
R10343_JUSTRAT_FI
NAL REPORT INCL...



0.84_TNO Report
2017 R10425 -
MMU.pdf_47059477
6_p30.png_dla_2....



0.84_TNO report
TNO 2015
R10343_JUSTRAT_FI
NAL REPORT INCL...

