OCR - Date

A common problem with digital transformation is to automatically digitize the documents, this is time consuming as someone manually data-entries the information, but as AI emerged, there are a lot of ways to do so without human involvement. Optical character recognition (OCR) is the process of reading the numbers, characters and words from images.
We'd like you to tackle the problem of extracting dates from images.

What we are expecting:

1. Preprocess the data (if needed).
2. Build a suitable network to solve this problem.
3. Choose a suitable loss function and performance metric.
4. Clean and well-commented code
5. Convert model to onnx and TRT
6. An endpoint using Django to deploy your work. (You don't have to make a html or front-end code, just a basic endpoint would be fine) [Optional]
7. PDF report describing your work in detail.

You can find the dataset for this task here

Notes:

- It's not an obligatory thing to come up with the best accuracy, the sole target of this task is to capture your coding capabilities within the time spent in doing the task and your thought process when approaching this problem.
- Please don't use pre-implemented packages such as Tesseract, we'd like you to showcase your skills in building the pipeline.
- The dataset consists of images and their text for a supervised learning approach.
- You are free to use whatever preprocessing technique, network architecture and loss function you find suitable for the task.

Bonus script:

- Use C++/TF to preprocess your images