

SUBJECT 7:

OCR-Free Document Understanding



Department
R&D



Duration
At least 6 months

The most traditional answer to the problem of Visual Document Understanding (VDU) relies on analyzing/parsing the OCR output of images. Even though these OCR-based approaches have demonstrated quite good performance. Nevertheless, they present some drawbacks as OCR is computationally expensive, inflexible across languages/ document types..., not to mention the fact that an inaccurate OCR result will propagate errors in the subsequent processes.

Thus, researchers are on the verge of designing and delivering new solutions that perform VDU tasks (such as extracting information from documents: like taxable amounts...). As an example, Clova AI Research have published the Donut model which is an end-to-end VDU solution and has made it available for use with the HuggingFace library.

The goal of this PFE project is to deepen the research on the OCR-free VDU approaches and to experiment different strategies, in order to create a model that will improve the extraction of invoice fields such as date, reference, items, VAT... (or some of them) to be able to anticipate the evolution of information extraction (IE) and to be prepared for those changes that will lead to more accurate and faster results.

Technologies & Environment of work :

C#

Python

OCR

ML

VDU