

Multimodal Transformer-based Optical Character Recognition

Information extraction from rich unstructured documents is a complex but key process in various businesses. For years, Analytics Consulting DataLab within BNP Paribas CIB has been developing many solutions that are widely used within the group.

For instance, the AI Lab developed cutting-edge approaches based on pre-trained Language Model Transformers in order to provide relevant automatic extraction [1]. However, the team faces an industrial challenge: Some documents come in image format, requiring an **Optical Character Recognition** (OCR) step before further text processing. The extraction performances therefore **greatly depend on the output quality of the OCR** model, which is all the more important when the image is **noisy** or of **low-resolution**.

In such environment, this project aims at exploring state-of-the-art methods on **Optical Character Recognition**, applied on internal, synthetic or open-source datasets.

The students will have the opportunities to:

- Familiarize with the literature on **Convolution-based methods** like **PP-OCrv3** [2]
- Implement **Multimodal Text-Image Transformers methods** such as **TrOCR** [3]
- Perform benchmarks on TrOCR with **current internal & external OCR solutions**
- Integrate an end-to-end TrOCR-based pipeline for the team's different solutions

In parallel, various axes are possible for further investigation, for example

- Explore **Image Denoising** [4][5] or **Super Resolution methods** [6] to improve the image quality as pre-processing
- Explore **Statistical & Probabilistic Deep Language models** [7] to correct the output texts as post-processing

This project will allow the student to gain in expertise in various domain especially: advanced Deep Learning, Computer Vision and Natural Language Processing. This industrial research will have a direct impact on all the use cases of the platform in production and on other internal solutions as well.

References:

- [1] LayoutLMv3: Pre-training for Document AI with Unified Text and Image Masking, Yupan Huang, Tengchao Lv, Lei Cui, Yutong Lu, Furu Wei, 2022
- [2] PP-OCrv3: More Attempts for the Improvement of Ultra Lightweight OCR System, Chenxia Li, Weiwei Liu, Ruoyu Guo, Xiaoting Yin, Kaitao Jiang, Yongkun Du, Yuning Du, Lingfeng Zhu, Baohua Lai, Xiaoguang Hu, Dianhai Yu, Yanjun Ma, 2022
- [3] TrOCR: Transformer-based Optical Character Recognition with Pre-trained Models, Minghao Li, Tengchao Lv, Jingye Chen, Lei Cui, Yijuan Lu, Dinei Florencio, Cha Zhang, Zhoujun Li, Furu Wei, 2022
- [4] Simple Baselines for Image Restoration, Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, Jian Sun, 2022
- [5] SwinIR: Image Restoration Using Swin Transformer, Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, Radu Timofte, 2021
- [6] SwinFIR: Revisiting the SwinIR with Fast Fourier Convolution and Improved Training for Image Super-Resolution, Dafeng Zhang, Feiyu Huang, Shizhuo Liu, Xiaobing Wang, Zhezhu Jin, 2022
- [7] Survey of Post-OCR Processing Approaches, Thi Tuyet Hai Nguyen, Adam Jatowt, Mickael Coustaty, Antoine Doucet, 2021



Practical Information:

Supervisor: RATNAMOGAN Pirashanth, Data Scientist, pirashanth.ratnamogan@bnpparibas.com

Location: Grands Moulins de Pantin, 9 Rue du Débarcadère, 93500 Pantin (Possible partial work from home)

Compensation: Competitive salary

How to apply: Please submit your application at paris.cib.analytics.consulting.careers@bnpparibas.com



BNP PARIBAS

The bank for a changing world



analytics
consulting

Classification : Internal