

Détecteur de signatures sur des documents commerciaux d'Orange France

Contexte :

La direction France d'Orange souhaite optimiser son workflow de souscription de ses offres sur le segment SoHo (Small Office, Home Office, clients professionnels de professions libérales). Jusqu'à présent, l'activation de ces dernières est conditionnée par une vérification manuelle de la signature des documents. Afin d'accélérer le processus et en réduire le coût, l'opérateur souhaite automatiser la détection des signatures.

Pour répondre à ce besoin nous avons développé un POC de détecteur d'objets basé sur un modèle de computer vision. Sa finalité est de déterminer si les documents scannés appartiennent à classe « signature absente » ou « signature présente ».

Temporalité :

De juin à octobre 2020

Équipe :

2 personnes

Technologies mises en œuvre :

- Langage : Python
- Modèle de computer vision : YOLOv4
- Framework d'entraînement du modèle : Darknet
- Framework de l'API : FAST API
- Framework du démonstrateur : Flask
- Base de données : MongoDB et SQLite

Dataset :

- [Tobacco 800](#) : 1290 documents scannés.
- [Nanonets](#) : 174 documents issus des moteurs de recherche Bing et Google
- [GSA Lease Documents](#) : 772 documents scannés

Bibliographie :

- Papier de recherche sur arXiv : [YOLOv4: Optimal Speed and Accuracy of Object Detection, Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao](#)
- Benchmark papers with code: [Real-Time Object Detection on COCO](#)