

STAGE MASTER/INGENIEUR

Thème : Apprentissage semi et faiblement supervisé pour de l'analyse vidéo

1. CONTEXTE

Au sein de SNCF Voyageurs, le Centre d'Ingénierie du Matériel réalise des missions d'expertise dans diverses spécialités lors de l'acquisition et la rénovation de matériels roulants. Intégré à l'équipe ETF1, le pôle « Intelligence Artificielle » expertise et conçoit des systèmes de vision artificielle ainsi que des solutions de surveillance audio et vidéo.

La problématique de disponibilité de jeu de données avec une labélisation forte pour entraîner de manière supervisé les modèles de deep learning est commune à tous les projets du pôle. L'acquisition de ce type de données est couteuse en temps et introduit des biais d'annotation. Pour contrer ces problématiques les modèles sont entraînés avec des données synthétiques ou scénarisées. Malheureusement, même si ce genre de données est intéressant pour valider la pertinence d'un modèle, ces données ne reprennent pas toute la diversité d'un environnement commercial.

Avant de pouvoir déployer les modèles en production il est donc nécessaire de les entraîner sur des données provenant de services commerciaux. La difficulté est que ces données sont soit faiblement labélisés ou partiellement labélisés. L'objectif de ce stage est donc d'évaluer et de mettre en place des approches faiblement et semi supervisé afin d'entraîner des modèles précédemment validé avec des entraînements supervisés.

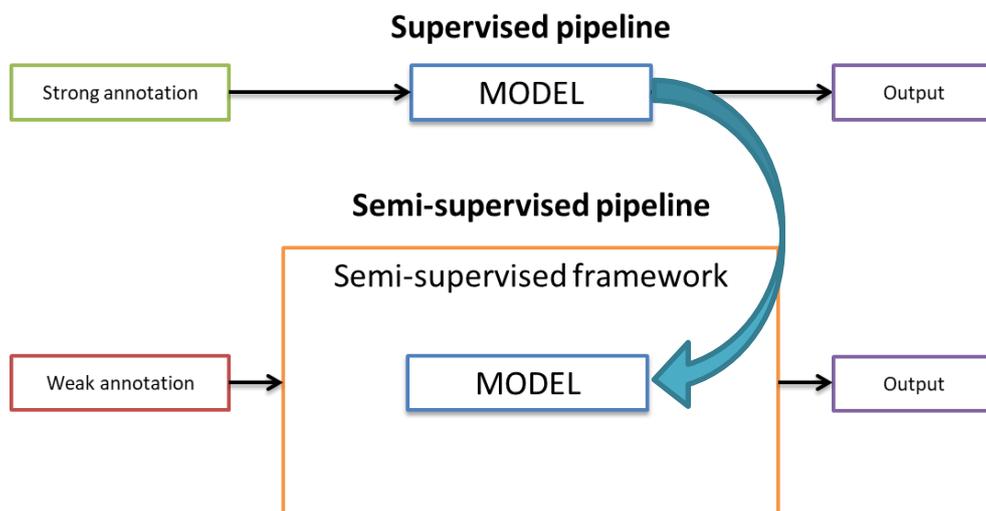
2. OBJECTIFS

Appliqué à un cas d'usage vidéo le stage aura pour but d'étudier deux approches afin d'évaluer les atouts, les faiblesses et la pertinence au regard du cas d'usage

1. Faiblement supervisé
2. Semi supervisé

Il sera nécessaire de se baser sur les travaux déjà menés dans le pôle et les techniques récentes dans la communauté scientifique.

La finalité du stage sera le développement d'un framework logiciel permettant d'entraîner facilement, sur des jeux de données faiblement ou partiellement annotés, des modèles validés via un entraînement supervisé.



3. DEROULEMENT

Le stage sera découpé en trois parties :

- La première partie « littérature » permettra de découvrir les différentes approches et d'établir l'état de l'art dans le domaine
- La seconde partie « implémentation » sera l'application sur des cas d'usage SNCF des méthodes précédemment repérées
- Enfin la troisième et dernière partie sera dédiée au développement d'un framework afin de pouvoir entraîner tous les modèles

La publication en conférence sera possible et encouragée.

Concernant le matériel une unité de calcul équipée d'un GPU dernier cri (V100, RTX 8000, RTX 6000, Titan RTX, 2080Ti) sera mise à disposition du stagiaire pour évaluer ses implémentations.

Enfin des déplacements ponctuels sont prévus dans le cadre de certains essais ou réunions.

4. PROFIL RECHERCHE

Le stage s'adresse aux étudiants en Master 2 informatique ou traitement de signal ou en dernière année d'école d'ingénieur (BAC+5).

Les compétences techniques ciblées pour ce stage sont :

- Système d'exploitation : Linux
- Langages : Python
- Bibliothèques : OpenCV, TensorFlow ou PyTorch

Les compétences fonctionnelles suivantes sont demandées :

- Intelligence artificielle : Deep Learning (DNN, CNN, RNN)
- Computer Vision

Le stagiaire devra faire preuve d'autonomie et d'initiative ainsi que d'un bon sens relationnel. Curiosité technique et un esprit de synthèse seront également nécessaires.

5. PERIODE ET DUREE

Date de début souhaitée du stage : Février 2021

Durée du stage : 6 mois

6. CONTACTS

MARTEAU Tony : tony.marteau@sncf.fr

AFANOU Sitou : sitou.afanou@sncf.fr

Mots-clés : deep learning, weakly-supervised learning, semi-supervised learning, active learning, multiple instance learning

