

TP7

Objectifs :

L'objectif de ce travail pratique est d'analyser et de comparer différents types de **fusion de caractéristiques** pour le suivi d'objets dans des vidéos à partir de flux vidéo RGB. Vous allez expérimenter avec des techniques de **fusion des caractéristiques spatiales** (via CNN) et **temporelles** (via LSTM) et évaluer leur performance pour le suivi d'objets en fonction de critères comme la précision, la robustesse et la vitesse.

Contexte :

Le suivi d'objets dans des vidéos est un défi majeur dans de nombreuses applications de vision par ordinateur, telles que la surveillance, les véhicules autonomes, et les systèmes de réalité augmentée. Les vidéos RGB, qui contiennent des informations sur l'apparence visuelle des objets et leur mouvement, sont couramment utilisées pour ce type d'analyse. Cependant, les vidéos peuvent être affectées par des changements de lumière, des occlusions, ou des objets qui apparaissent et disparaissent rapidement. La fusion de caractéristiques extraites de différentes manières permet de renforcer la capacité de suivi.

Dans ce TP, vous allez explorer plusieurs types de fusion de caractéristiques, en utilisant uniquement des vidéos RGB pour effectuer l'analyse vidéo et le suivi d'objets.

Partie 1 : Préparation des données

1. Chargement des vidéos :

- Téléchargez un jeu de données vidéo contenant des objets en mouvement.
Par exemple, vous pouvez utiliser des datasets comme **UAV123**, **OTB**.

2. Prétraitement des vidéos :

- Redimensionnez les vidéos à une taille uniforme (par exemple, 224x224 pixels).
- Normalisez les pixels des images en soustrayant la moyenne et en divisant par l'écart-type.
- Appliquez un filtrage pour réduire le bruit (par exemple, un filtre Gaussien).

Partie 2 : Détection d'objets dans les vidéos

1. Utilisation de la détection d'objets :

- Utilisez un modèle de détection d'objets comme **YOLOv5** pour détecter les objets dans chaque image de la vidéo.
- Chaque objet détecté sera identifié par des **boîtes de délimitation** (bounding boxes) et une **étiquette de classe** (par exemple, "personne", "voiture").

Partie 3 : Extraction des caractéristiques visuelles avec CNN

1. Utilisation d'un modèle CNN pour l'extraction des caractéristiques :

- Implémentez un réseau CNN pré-entraîné comme **ResNet-50**, **VGG16**, ou **MobileNet** pour extraire les caractéristiques visuelles des objets détectés dans chaque image.
- Découpez les objets détectés dans chaque image et passez-les à travers le modèle CNN pour obtenir des **descripteurs de caractéristiques** (feature vectors).

2. Fusion des caractéristiques :

- Testez différentes méthodes de fusion des caractéristiques :
 - **Fusion intermédiaire (mid fusion)** : Fusionnez les caractéristiques extraites des objets détectés avant de les envoyer dans le modèle LSTM.
 - **Fusion tardive (late fusion)** : Traitez les caractéristiques extraites pour chaque objet indépendamment avec un modèle LSTM, puis combinez les sorties des LSTM pour chaque objet à la fin.

