Université Sidi Mohamed Ben Abdellah — Année universitaire : 2024/2025

Faculté des sciences Dhar El Mahraz — **Master MLAIM**

Nom et prénom :	

Examen final – Analyse vidéo

Exercice 1

1. Expliquez comment SIFT utilise la différence de Gaussiennes (DoG) pour détecter les points-clés.

……

……

……

……

……

……

……

……

……

……

……

2. Décrivez le processus de construction des descripteurs SIFT en utilisant des histogrammes d'orientations.

……

……

……

……

……

……

……

……

……

……

……

3. Quelles bibliothèques ou frameworks permettent d'utiliser SIFT pour la reconnaissance d'objets ?

……

……

……

4. Expliquez comment SIFT peut être combiné avec un algorithme de clustering (comme K-means) pour réaliser une reconnaissance d'objets.

……

……

……

……

……

……

……

……

Exercice 2

1. Définir les caractéristiques spatiales (visuelles) et temporelles :

- Les caractéristiques spatiales……

……

……

- Les caractéristiques temporelles……

……

……

2. Expliquez pourquoi faut-il combiner les caractéristiques temporelles et spatiales de dans une architecture de reconnaissance d'actions humaines à partir de vidéos.

……

……

……

……

……

……

……

3. Expliquez comment le flot optique est utilisé pour détecter des mouvements dans une séquence vidéo.

……

……

……

……

4. Donnez quelques bibliothèques en python permettent de calculer le flot optique avec les approches variationnelles et avec le Deep Learning, et quelles fonctions spécifiques pouvez-vous utiliser ?

……

……

……

……

5. Comment optimiser le calcul du flot optique pour qu'il soit rapide et précis ?

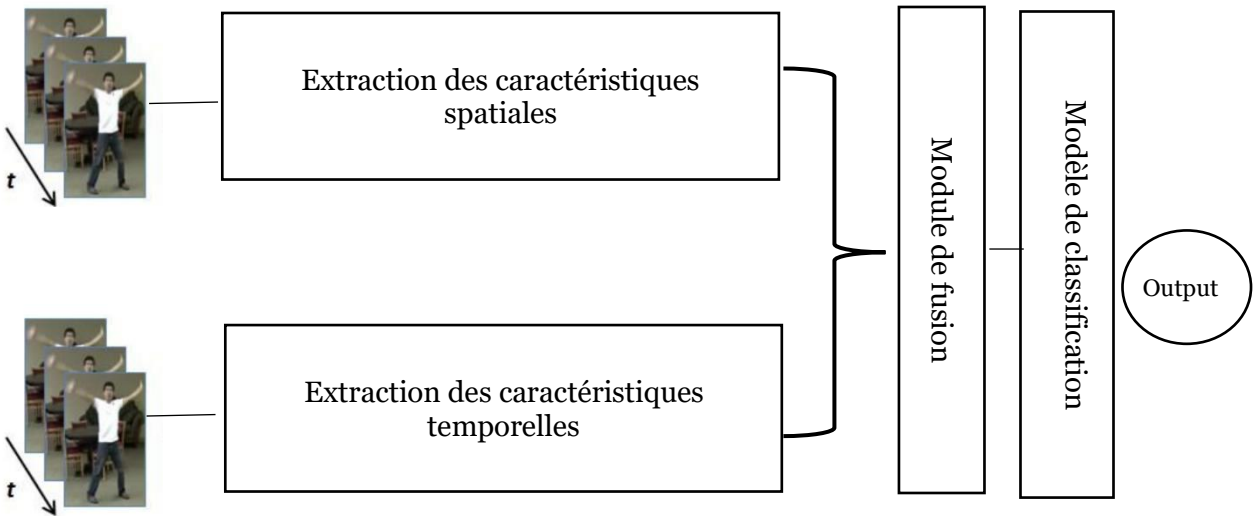
……

……

……

……

On considère une architecture pour la reconnaissance des actions d'un Homme dans une vidéo.



Entrée : Une séquence vidéo représentée par des frames successives $[F_1, F_2, \dots, F_n]$

Sortie : Pour chaque paire de frames consécutives (F_t, F_{t+1}), on calcule v_t, la représentation condensée des mouvements dans le frame t .

6. Expliquer, par une description textuelle ou par un schéma, comment construire un vecteur de caractéristiques temporelles condensées $f_{temporal}$ utilisant le flot optique et LSTM.

7. Proposer et décrire brièvement un modèle pré-entrainé pour calculer la représentation visuelle $f_{spatial}$ de chaque frame

……

……

……

……

……

……

……

……

8. On souhaite fusionner les deux représentations $f_{spatial}$ et $f_{temporal}$ pour construire un descripteur robuste décrivant le mouvement de la personne dans une vidéo, proposer et décrire un module de fusion capable d'intégrer efficacement ces deux types d'information.

……

……

……

……

……

……

……

Exercice 3

1. Expliquez brièvement le fonctionnement de YOLO. Pourquoi est-il considéré comme un modèle de détection en temps réel ?

……

……

……

……

……

……

……

……

……

……

……

2. Quel Framework ou bibliothèque est le plus couramment utilisé pour implémenter YOLO ?

……

……

……

……

3. Vous utilisez YOLO pour détecter des véhicules sur une autoroute, mais le modèle ne détecte pas bien les véhicules à grande distance. Comment pourriez-vous améliorer ses performances ?

……

……

……

……

……

……

4. On veut étendre le système de l'exercice 2 pour reconnaitre les actions de plusieurs personnes dans la vidéo. Décrire, par texte ou schéma, le système global en préservant les mêmes descripteurs $f_{spatial}$ et $f_{temporal}$ utilisés dans l'exercice précédent.