INF6804 Vision par ordinateur
H2026 – Travail Pratique #1
Reconnaissance d'actions sportives via Embeddings de caractéristiques

---

Objectifs :

- Apprendre à effectuer la reconnaissance d'actions en utilisant une approche par *embeddings* de caractéristiques, sans entraînement de bout en bout.
- Implémenter et comparer un extracteur de caractéristiques classique avec un extracteur basé sur l'apprentissage profond.
- Implémenter et analyser deux stratégies de classification simples : le centroïde le plus proche et les k-plus proches voisins (k-NN).
- Analyser les compromis entre les méthodes en termes de taux d'exactitude (accuracy), de vitesse et de robustesse.
- Formuler des hypothèses et les évaluer de manière critique basée sur les résultats expérimentaux.

Remise :

- Soumettre avant **le 6 février à 17h00** – *les retards ne seront pas acceptés*
- **Où soumettre :**
  - *Code Source :* Soumettez votre code sur **Moodle** (nous devons être capables d'exécuter vos tests).
  - *Rapport :* Soumettez votre rapport (*format .pdf*, 8 à 15 pages, taille de police 10) sur **Gradescope**.

Références :

- Voir les notes de cours sur Moodle (Chapitre 1)

Autres directives :

- Les travaux doivent être faits en équipes de deux, soumettez une seule version de votre travail !

Présentation

Dans ce travail pratique, vous allez construire et évaluer un système de reconnaissance d'actions sportives. Au lieu d'entraîner un réseau profond à partir de zéro, vous utiliserez une approche par *embeddings* de caractéristiques. Vous devez comparer deux méthodes et déterminer laquelle est la meilleure pour notre tâche, qui consiste à trouver la bonne étiquette de sport pour une vidéo. La première méthode est basée sur les **Histogrammes des Orientations du Gradient (HOG)**, la seconde est basée sur un **Réseau de Neurones Convolutif pré-entraîné (ResNet)**. Vous pouvez utiliser vos notes de cours comme référence pour comprendre leurs principes de fonctionnement de base, et vous pouvez chercher en ligne pour plus de détails.

Pour comparer les deux méthodes, vous devrez évaluer à quel point le modèle est précis pour classer la vidéo dans la bonne catégorie. Le flux de travail est le suivant : pour un ensemble de vidéos de référence, vous extrairez des caractéristiques de trames individuelles pour construire une représentation pour chaque classe de sport. Ensuite, pour une nouvelle vidéo de test, vous extrairez ses caractéristiques et la classerez en fonction de sa similarité avec les représentations des classes de référence. Afin de classer la vidéo, nous comparons deux stratégies de classification :

- **Vecteur Moyen (Centroïde) :** Représenter chaque classe par le vecteur de caractéristiques moyen de toutes ses trames de référence.
- **k-Plus Proches Voisins (k-Nearest Neighbors - k-NN) :** Stocker toutes les caractéristiques de référence et classer les trames d'une vidéo de test par un vote majoritaire de leurs voisins les plus proches.

En utilisant les classes *("Push ups", "Handstand Pushups", "Biking", "Playing Cello", "Playing Violin")* du **jeu de données UCF101**, vous devez inclure les éléments suivants dans votre rapport (noté sur 20 pts) :

Structure du Rapport et Questions

1. **Présentation des Méthodes (2 pts) :**

   Dans vos propres mots, décrivez les principes fondamentaux des deux extracteurs de caractéristiques que vous comparez :

   - Comment fonctionne **HOG** ? Quel type d'information encode-t-il principalement (ex: texture, forme, couleur) ?
   - Comment fonctionne un **ResNet** ? Quel type de caractéristiques attendez-vous de ses couches finales ?
   - Expliquez brièvement les classifieurs **Centroïde le plus proche** et **k-Plus Proches Voisins (k-NN)**. Quelle est la différence principale ?

2. **Hypothèses de Performance (2 pts) :**

   Basé sur votre compréhension théorique, formulez des hypothèses sur la performance attendue :

   - **Comparaison des Extracteurs :** Quel extracteur de caractéristiques (HOG ou ResNet) prédisez-vous qui atteindra un taux d'exactitude (accuracy) plus élevé sur les classes de sport données ? Par exemple, quel modèle sera meilleur pour détecter la classe "Biking".
   - **Comparaison des Classifieurs :** Pour un type de caractéristique donné (ex: caractéristiques ResNet), quelle stratégie de classification (Centroïde ou k-NN) vous attendez-vous à voir mieux performer ? Expliquez les compromis potentiels en termes d'exactitude, d'utilisation de la mémoire et de vitesse de prédiction.

3. **Hypothèses sur la Robustesse (2 pts) :**

   Dans cette question, vous effectuez un "test de robustesse" sur votre meilleur modèle en augmentant les vidéos de test. Avant de lancer l'expérience, émettez des hypothèses :

   - Quel extracteur de caractéristiques (HOG ou ResNet) croyez-vous sera le plus robuste aux **changements d'éclairage** (ex: un grand changement de luminosité) ? Pourquoi ?
   - Lequel croyez-vous sera le plus robuste à un **changement géométrique** comme une rotation de 90 degrés ? Pourquoi ?

4. **Description des Expériences et Implémentation (3 pts) :**

   Décrivez votre configuration expérimentale :

   - Comment avez-vous échantillonné les trames des vidéos pour créer vos ensembles de référence et de test ?
   - Décrivez l'implémentation. Quelles bibliothèques avez-vous utilisées (ex: OpenCV, Scikit-learn, Hugging Face) ? Quels étaient les paramètres clés pour HOG et ResNet (ex: taille de cellule pour HOG) ? Comment avez-vous défini la valeur de k pour k-NN ?
   - Quelles métriques d'évaluation avez-vous utilisées pour mesurer la performance (ex: taux d'exactitude global, matrice de confusion) ? Justifiez votre choix.

5. **Résultats d'Expérimentation (4 pts) :**

   Présentez vos résultats clairement en utilisant des tableaux et des figures.

   - Fournissez un tableau comparant le taux d'exactitude (accuracy) de la classification pour les quatre combinaisons (HOG+Centroïde, HOG+kNN, ResNet+Centroïde, ResNet+kNN).
   - Montrez la matrice de confusion pour la combinaison la plus performante.
   - Fournissez un tableau ou un graphique montrant les résultats de vos expériences de "test de robustesse" de la Question 3 (performance sur les données originales vs augmentées).

6. **Discussion et Analyse (4 pts) :**

   Analysez vos résultats en profondeur :

   - Discutez de vos résultats de la Question 5 en relation avec vos hypothèses des Questions 2 et 3. Vos prédictions étaient-elles correctes ? Si non, pourquoi les résultats auraient-ils pu différer de vos attentes ?
   - Quelle méthode (HOG ou ResNet) s'est avérée supérieure ? Analysez la matrice de confusion de votre meilleur modèle : quelles classes ont été facilement distinguées, et lesquelles ont été confondues ? Pourquoi cela pourrait-il être le cas ?
   - Quels sont les compromis finaux mesurés (exactitude, temps d'extraction des caractéristiques, temps de classification) entre les différentes approches ? Quelle combinaison recommanderiez-vous pour une application réelle et pourquoi ?

7. **Lisibilité et complétude (3 pts) :**

   En plus du contenu, le format doit être bien structuré et complet.

Ressources

- **Jeu de données :** UCF101 : https://www.crcv.ucf.edu/research/data-sets/ucf101/.
- **Descripteurs Classiques :** OpenCV, scikit-image.
- **Frameworks d'Apprentissage Profond :** PyTorch, Hugging Face Transformers, TensorFlow.