

# INF6804 Vision par ordinateur H2025 – Travail

## Pratique #1 Description et comparaison de régions d'intérêt

---

### Objectifs:

- Introduire le traitement des données vidéo et image à l'aide de bibliothèques logicielles pour la vision par ordinateur.
- Permettre à l'étudiant de se familiariser avec la description d'images basée sur le pré-apprentissage contrastif de texte et d'image (CLIP).
- Permettre à l'étudiant de se familiariser avec la description d'images basée sur les histogrammes des orientations du gradient (HOG).
- Apprendre à comparer des vecteurs de caractéristiques.
- Expliquer les avantages et les inconvénients des deux approches de description d'images.

### Remise:

- *Tout votre code source* (nous devons être en mesure d'exécuter vos tests).
- Un rapport (*.pdf format* de 8 à 15 pages avec une taille de police de 10).
- Remise avant le **vendredi 7 février 2025, 17h00**, sur Moodle – *aucun retard ne sera accepté*.
- *Vous devez aussi soumettre votre rapport sur TurnItIn!*
  - Enregistrez-vous sur [www.turnitin.com](http://www.turnitin.com) en utilisant les informations disponibles sur Moodle!

### Références:

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

### Autres directives:

- Les travaux pratiques doivent être réalisés en équipes de deux. Remettez une seule version de votre travail.
-

## Présentation

Dans ce TP, vous devrez caractériser deux méthodes de description de régions d'intérêt dans des images et déterminer quelle méthode est la meilleure et dans quelles circonstances. Une description de votre travail, de vos expériences, et les réponses aux questions de cet énoncé doivent être incluses dans votre rapport.

Vous devrez comparer deux méthodes de description : la première est basée sur **les histogrammes des orientations du gradient (HOG)**, la seconde est basée sur **le pré-apprentissage contrastif de texte et d'image (CLIP)**. Vous pouvez utiliser vos notes de cours pour comprendre leurs principes de base. Pour plus de détails, explorez Internet! Pour comparer les deux méthodes, vous devrez résoudre une tâche de **recherche d'image par contenu (CBIR)**. Vous devrez caractériser chaque image à l'aide des descripteurs et trouver les images correspondant à chaque image requête en mesurant la similarité entre les descripteurs. Nous vous demandons de comparer les approches en termes d'efficacité (temps de calcul) et de performance. Vous êtes libres de choisir les métriques qui, selon vous, conviennent le mieux à chaque méthode à implémenter.

En utilisant les données fournies dans **“data.zip”**, vous devez inclure les éléments suivants dans votre rapport (noté sur 20 points):

1. Présentation des deux approches à comparer (3 points) :

Dans vos propres mots, donnez une description générale et les principes de vos deux approches.

2. Hypothèses de performance dans des cas spécifiques (2 points) :

Identifiez, sur la base de votre compréhension théorique des deux approches, quelle approche devrait être la meilleure dans au moins DEUX cas d'utilisation. Par exemple, quelle est la meilleure méthode si la taille des contenus des régions comparées est relativement uniforme? Pourquoi?

3. Hypothèses de performance concernant les boîtes englobantes (bounding boxes) (1 pts):

En sachant que vous avez accès à la boîte englobant l'objet d'intérêt dans chaque image, attendez-vous une amélioration de la performance pour les deux méthodes? Pourquoi?

4. Description des expériences, des données et des critères d'évaluation (3 points) :

Décrivez en détail les expériences réalisées pour tester les hypothèses des points 2 et 3. Quelles sont les difficultés des images requêtes dans la base de données? Quelles mesures de similarité entre descripteurs avez-vous choisies? Quelle métrique d'évaluation avez-vous utilisée?

Comparez les performances de différentes tailles de modèles CLIP et des tailles de patches variables. Montrez les résultats sous forme de graphiques

et de tableaux. Quel est l'effet d'une augmentation de la taille des patches ou des modèles?

5. Description des implémentations utilisées (2 points) :

Décrivez l'implémentation des approches étudiées. Si vous n'avez pas écrit tout le code vous-même, d'où provient-il? A-t-il nécessité des modifications? Quels sont les principaux paramètres de vos méthodes? Comment les avez-vous configurés?

6. Résultats des expériences (3 points) :

Fournissez les résultats d'évaluation de vos expériences en lien avec les hypothèses des points 2 et 3. Utilisez un format approprié — tableaux, figures, ...

7. Discussion sur les résultats et les hypothèses précédentes (3 points) :

Discutez des résultats du sixième point en lien avec les hypothèses des points 2 et 3. Quelles hypothèses sont supportées par ces résultats? Lesquelles ne le sont pas? Quels tests n'ont pas permis de conclusions? Comment pourriez-vous améliorer ces tests?

8. Lisibilité et complétude (3 points) :

Outre le contenu, le format doit être soigné et complet.

Lors des séances de laboratoire, n'hésitez pas à poser des questions — nous pouvons vous aider avec tout problème technique sur Windows/Linux, ou si vous codez en C/C++, Python ou Matlab.

Vous serez **pénalisés de 50% de la note totale si vous ne remettez pas votre code**. De plus, si votre rapport n'est pas soumis à **TurnItIn**, il ne sera pas corrigé. L'ordre de présentation des sujets ci-dessus n'a pas d'importance, tant qu'ils sont tous présents.

## Ressources

Librairies de vision par ordinateur:

- OpenCV ([https://docs.opencv.org/4.0.0/d9/df8/tutorial\\_root.html](https://docs.opencv.org/4.0.0/d9/df8/tutorial_root.html))
- scikit-image ([https://scikit-image.org/docs/stable/auto\\_examples/index.html](https://scikit-image.org/docs/stable/auto_examples/index.html))

Frameworks d'apprentissage profond:

- PyTorch (<https://pytorch.org/tutorials/>)
- Hugging Face (<https://huggingface.co/docs>)

Python:

- Guide (<https://wiki.python.org/moin/BeginnersGuide/Programmers>)

- NumPy (<https://docs.scipy.org/doc/numpy/user/quickstart.html>)

Matlab:

- Guide ([http://www.mathworks.com/help/pdf\\_doc/matlab/getstart.pdf](http://www.mathworks.com/help/pdf_doc/matlab/getstart.pdf))
- Aide-mémoire (<http://web.mit.edu/18.06/www/Spring09/matlab-cheatsheet.pdf>)

Tutoriel CLIP :

- Github ([https://github.com/gabilodeau/INF6804/blob/master/clip\\_tutorial.ipynb](https://github.com/gabilodeau/INF6804/blob/master/clip_tutorial.ipynb))