

INF6804 Vision par ordinateur

H2025 – Travail Pratique 2

Segmentation d'objets vidéo

Objectifs:

- Se familiariser et expérimenter avec les algorithmes de segmentation des régions d'intérêt dans les vidéos à l'aide de modèles de segmentation à classes fixes (limités à un nombre prédéfini de classes).
- Se familiariser et expérimenter avec les algorithmes de segmentation des régions d'intérêt dans les vidéos à l'aide de modèles de segmentation zero-shot (capables de généraliser à des classes non vues sans ré-entraînement).
- Expliquer les avantages et les inconvénients des deux approches.

Remise:

- *Tout votre code source* (nous devrions pouvoir exécuter vos tests)
- Un rapport (*.pdf format* de 8 à 12 pages avec une taille de police de 10)
- Remise avant le 14 Mars 2024, 17h00, sur Moodle –*aucun retard accepté*
- *Vous devez aussi soumettre votre rapport sur TurnItIn!*
 - Enregistrez-vous sur www.turnitin.com en utilisant les informations sur Moodle!

Références:

- Voir les notes de cours sur Moodle (Chapitre 2)

Autres directives:

- Les TPs doivent obligatoirement être faits en équipe de deux, ne remettez qu'une seule version du rapport/code!
-

Introduction

Dans ce TP, vous devez caractériser deux méthodes de segmentation de régions d'intérêt dans des vidéos 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 se retrouver dans un rapport.

Travail à réaliser

Vous devez comparer deux méthodes de segmentation : l'une basée sur des modèles de segmentation à classes fixes (limités à un ensemble prédéfini de classes), tels que Mask R-CNN ou YOLO, et l'autre basée sur des modèles de segmentation zero-shot (capables de segmenter des objets de classes non vues sans ré-entraînement). Pour la segmentation zero-shot, veuillez vous référer au notebook GitHub partagé ci-dessous.

Ces méthodes ont été brièvement présentées en cours — vous pouvez utiliser vos notes de cours comme référence pour en comprendre les principes de base. Pour plus de détails, n'hésitez pas à consulter des ressources en ligne ! La tâche ici consiste à extraire uniquement les principales régions d'intérêt de la vidéo d'entrée (disponible dans le dossier "input"). Ces régions doivent être similaires à celles définies par la vérité terrain (disponible dans le dossier "groundtruth"). Elles peuvent être composées d'un ou de plusieurs objets, situés au premier plan et en mouvement.

Dans votre rapport, en utilisant les quatre vidéos de la catégorie "Baseline" de la base de données CDNET 2012, vous devez inclure les éléments suivants (noté sur 20 pts):

1. Présentation des deux méthodes (4 pts):

Dans vos propres mots, donnez la description générale et les principes des deux méthodes de segmentation : les modèles de segmentation à classes fixes et les modèles de segmentation zero-shot.

2. Hypothèses de performance dans cas spécifiques (3 pts):

Identifier, selon votre compréhension théorique des deux méthodes, laquelle des deux devrait être la meilleure dans au moins TROIS cas d'utilisation spécifiques. Par exemple, quelle est la meilleure approche à utiliser si l'objet principal est un objet rare ? Pourquoi ?

3. Description des expériences, séquences de la base de données et critères d'évaluation (2 pts):

Décrire en détail les expériences réalisées pour vérifier vos hypothèses du point précédent. Quelles vidéos ou parties de vidéos avez-vous utilisées en lien avec les cas spécifiques considérés ? Quelles sont les difficultés de ces séquences ? Quels sont vos critères d'évaluation (quantitatifs et qualitatifs) ?

4. Description des deux implémentations utilisées (2 pts):

Décrire l'implémentation des deux approches étudiées. Si vous n'avez pas tout écrit le code vous-même, d'où provient-il? A-t-il demandé des modifications? Sinon, de quel articles (ou site web) vous êtes-vous inspiré pour l'écrire? Dans tous les cas, quels sont les paramètres principaux utilisés? Comment ont-ils été choisis?

5. Présentation des résultats de tests (3 pts):

Donner les résultats d'évaluation de performance tirés de vos expériences en lien avec les hypothèses du deuxième point. Utilisez un format approprié, tableaux, figures, ...

6. Discussion des résultats et retour sur hypothèses (3 pts):

Discuter des résultats du cinquième point en fonction des hypothèses du deuxième point. Quelles hypothèses sont supportées par les résultats? Lesquelles ne le sont pas? Quels tests sont restés sans conclusions? Quels tests pourriez-vous améliorer, et comment?

7. Lisibilité et complétude (3 pts):

À part le contenu, le format doit être soigné et complet.

Lors des séances de laboratoire, n'hésitez pas à poser des questions à votre chargé — il peut vous aider avec tout problème technique si vous travaillez sur Windows/Linux, ou bien avec votre code si vous travaillez avec du C/C++ ou avec Python/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 remis sur TurnItIn, celui-ci ne sera pas corrigé. L'ordre de la présentation n'est pas important, tant que tous les éléments ci-haut sont présents. N'oubliez pas de citer vos références!

Ressources

Ensemble de données:

- CDNET 2012 dataset (<http://jacarini.dinf.usherbrooke.ca/dataset2012>). N'utilisez que les vidéos de la catégorie "Baseline".

Librairies de vision par ordinateur:

- OpenCV (https://docs.opencv.org/4.0.0/d9/df8/tutorial_root.html)
- scikit-image (https://scikit-image.org/docs/stable/auto_examples/index.html)

Librairies d'apprentissage profond:

- PyTorch (<https://pytorch.org/tutorials/>)
- Tensorflow (<https://www.tensorflow.org/tutorials>)

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)
- Aide-mémoire (<http://web.mit.edu/18.06/www/Spring09/matlab-cheatsheet.pdf>)

Modèles de segmentation zéro-shot:

- Notebook GitHub (https://github.com/gabilodeau/INF6804/blob/master/Zero_Shot_Segmentation.ipynb)