

Métriques permettant de mesurer l'impact d'échantillons sur la performance des modèles de détection d'objets

1 Contexte

Dans le domaine de la télésurveillance, l'analyse vidéo joue un rôle crucial pour assurer la sécurité et la surveillance en temps réel. **Ipsotek**, la plateforme de vidéo analytics d'Eviden, se distingue par sa capacité à intégrer des modèles avancés de détection d'objets pour identifier et suivre des objets dans des flux vidéo en direct. Un modèle de détection d'objets, tel que **YOLO** (You Only Look Once), est un réseau de neurones convolutif qui permet de détecter et de classifier plusieurs objets dans une image en une seule passe. YOLO est particulièrement apprécié pour sa rapidité et son efficacité, ce qui le rend idéal pour des applications en temps réel comme la surveillance vidéo.

Cependant, la performance de ces modèles dépend fortement de la qualité des données d'entraînement. La diversité des images, la précision des annotations, et le domaine des données influencent directement la précision et la fiabilité du modèle. Une image mal annotée ou de mauvaise qualité peut entraîner des erreurs de détection, compromettant ainsi l'efficacité du système de surveillance.

2 Objectif

L'objectif de ce projet est de développer un ensemble de métriques de qualité qui évalue l'impact potentiel des images et de leurs annotations sur la performance du modèle YOLO. Ces métriques devraient être capables de prédire si une image et son annotation amélioreront ou dégraderont la performance du modèle.

2.1 Modèles, datasets et évaluation

Nous proposons de concentrer l'étude sur le modèle YOLO v11, pré-entraîné sur le dataset COCO. L'objectif sera de construire un jeu de données d'entraînement ($N \leq 1000$) sélectionné à l'aide des métriques développées, afin d'améliorer la mAP@50 du modèle YOLO après un fine-tuning sur ce nouveau jeu de données. Cette amélioration sera évaluée sur deux domaines différents : le jeu de test de COCO et celui de Cityscapes. Nous proposons de limiter les classes aux catégories communes aux deux jeux de données, à savoir : "*Person*", "*car*", "*truck*", "*bus*" et "*traffic light*". Les images du jeu d'entraînement à créer peuvent être basées sur le jeu de validation de COCO ou sur les jeux d'entraînement et de validation de Cityscapes, ou provenir de toute autre source, mais ne doivent pas provenir des jeux de test de COCO ou Cityscapes. Une sélection aléatoire d'échantillons provenant des même sources et comportant le même nombre d'échantillons servira de modèle baseline.

3 Pistes de recherche

Plusieurs stratégies peuvent être envisagées pour construire des métriques permettant d'évaluer l'impact des échantillons sur les performances du modèle. Cela peut inclure par exemple les méthodes inspirées de l'**active learning** [1], basées sur des critères d'*incertitude* [2] et de *diversité*, ou des approches issues de l'**IA explicable (XAI)**. Les méthodes d'**influence des données**, telles que [3, 4], semblent particulièrement prometteuses.

Références

- [1] Elezi, Ismail, et al. "Not all labels are equal : Rationalizing the labeling costs for training object detection." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.
- [2] Timans, Alexander, et al. "Adaptive bounding box uncertainties via two-step conformal prediction." European Conference on Computer Vision. Cham : Springer Nature Switzerland, 2024.
- [3] Pruthi, G., Liu, F., Kale, S., & Sundararajan, M. (2020). *Estimating Training Data Influence by Tracing Gradient Descent*. NeurIPS.
- [4] Hu, Yuzheng, et al. "Most influential subset selection : Challenges, promises, and beyond." arXiv preprint arXiv :2409.18153 (2024).