

Mesure de la Qualité d'Annotation Automatique sans Ground Truth

1 Contexte

Dans le domaine de la télésurveillance, l'analyse vidéo joue un rôle essentiel pour assurer la sécurité et la surveillance en temps réel. **Ipsotek**, développé par **Eviden**, est une plateforme avancée de video analytics qui intègre des technologies d'intelligence artificielle pour la détection et le suivi d'objets dans des flux vidéo en direct. Elle permet d'automatiser l'analyse des scènes et d'améliorer la réactivité des systèmes de surveillance.

Pour optimiser la précision des modèles de détection utilisés dans ces systèmes, la constitution d'un jeu de données annoté de haute qualité est une étape critique. Cependant, l'annotation manuelle des images est un processus long et coûteux. Une approche prometteuse consiste à utiliser l'**auto-labeling**, qui exploite des modèles de fondation puissants pour générer automatiquement des annotations.

L'auto-labeling est une technique innovante qui utilise des modèles de fondation, tels que **Grounding DINO**, pour annoter automatiquement des ensembles de données d'entraînement. Ces annotations servent ensuite à entraîner des modèles plus petits, comme **YOLO**, qui seront déployés en ligne pour des tâches de détection d'objets. Cependant, la qualité des annotations générées automatiquement est cruciale pour garantir la performance et la fiabilité des modèles déployés. Une annotation de mauvaise qualité peut entraîner des erreurs de détection, compromettant ainsi l'efficacité des systèmes basés sur ces modèles.

2 Objectif

L'objectif principal de cette étude est de développer des méthodes pour mesurer la qualité des annotations générées par des modèles de fondation utilisés pour l'auto-labeling.

2.1 Modèles, datasets et évaluation

Nous proposons de concentrer cette étude sur le modèle **Grounding DINO** (version mmdet) pour l'auto-labeling. Le dataset **LVIS**, réduit aux classes "*cellular_telephone*", "*gun*", "*pocketknife*", "*suitcase*" et "*glove*", servira de base de données à annoter.

L'évaluation de la métrique de qualité d'annotation pourra être réalisée selon plusieurs approches complémentaires afin de vérifier sa pertinence :

- **Corrélation avec l'IoU des prédictions de Grounding DINO** : Mesure de la corrélation entre la métrique recherchée et l'IoU entre les prédictions de Grounding DINO et le ground truth. Cette approche permettra d'évaluer dans quelle mesure la métrique reflète la qualité des bounding boxes.
- **Corrélation avec le taux de faux positifs** : Comparaison de la métrique recherchée avec le nombre de faux positifs dans une image, normalisé par le nombre total d'objets annotés. Cela nous permettra d'analyser si la métrique est pertinente pour la précision des classes d'objets.
- **Corrélation avec la performance globale du modèle (mAP)** : Afin d'évaluer si la métrique est un bon indicateur de qualité à l'échelle du dataset, la corrélation entre la métrique et la mAP (Mean Average Precision) pourra être reporté. Cette analyse devra être réalisée sur des sous-ensembles d'images afin d'obtenir une granularité plus fine.

Ces méthodes ne sont pas exclusives, et toute autre métrique pertinente pourra être intégrée pour compléter l'évaluation.

3 Pistes de recherche

Pour construire une métrique mesurant la qualité d'annotation, des approches basées sur la **quantification d'incertitudes** telles que [1, 2] peuvent être envisagées. Des approches basées sur la prédiction conforme [3] pourraient être aussi une piste intéressante.

Références

- [1] Deepshikha, Kumari, et al. "Monte carlo dropout for modelling uncertainty in object detection." arXiv preprint arXiv :2108.03614 (2021).

- [2] Munir, Muhammad Akhtar, et al. "Cal-DETR : calibrated detection transformer." Advances in neural information processing systems 36 (2024).
- [3] Timans, Alexander, et al. "Adaptive bounding box uncertainties via two-step conformal prediction." European Conference on Computer Vision. Cham : Springer Nature Switzerland, 2024.