



HAL
open science

Annotation collaborative zero-shot par deux modèles pré-entraînés dans l'étude du flux de bulles de gaz

Trong Nghia Ngo, Christophe Vial, Adib Essid, Jean-Sébastien Guez, Thierry
Thierry Chateau

► **To cite this version:**

Trong Nghia Ngo, Christophe Vial, Adib Essid, Jean-Sébastien Guez, Thierry Thierry Chateau. Annotation collaborative zero-shot par deux modèles pré-entraînés dans l'étude du flux de bulles de gaz. ORASIS 2025, ISEN Yncréa Ouest, Jun 2025, Le Croisic, France. <hal-05131220>

HAL Id: hal-05131220

<https://hal.science/hal-05131220v1>

Submitted on 10 Jul 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Annotation collaborative zero-shot par deux modèles pré-entraînés dans l'étude du flux de bulles de gaz

T. N. Ngo¹ C. Vial¹ A. Essid¹ JS. Guez¹ T. Chateau¹

¹ Université Clermont Auvergne, Clermont Auvergne INP, CNRS, Institut Pascal, 63000 Clermont-Ferrand, France

trong_nghia.ngo@doctorant.uca.fr

Résumé

Un pipeline d'annotation automatique est proposé pour la détection des flux de bulles, exploitant des modèles pré-entraînés afin d'accélérer considérablement le processus d'annotation.

Mots Clef

Segmentation d'instances, Détection des objets de petite taille, Vision par ordinateur, Modèle vision-langage

Abstract

An automatic annotation pipeline is proposed for bubble flow detection, exploiting pre-trained models to significantly speed up the annotation process.

Keywords

Instance segmentation, tiny object detection, computer vision, vision-language model.

1 Introduction

Dans le cadre des recherches menées sur le coefficient de transfert gaz-liquide [1], la compréhension du comportement des flux de bulles dans les bioréacteurs joue un rôle crucial. Divers coefficients, tels que le coefficient de transfert de masse volumétrique, la surface interfaciale et le diamètre moyen de Sauter [2], sont en effet étroitement liés à la détection et la mesure des bulles de gaz. La première étape de ce processus consiste à créer des données de référence pour alimenter les modèles de détection. Cette étape introduit un nouveau défi, ce sont la petitesse, la surpopulation et le chevauchement des bulles, par exemple, une image donnée contenant de 600 à 1900 bulles dont les tailles varient de 1x2 pixels à 54x29 pixels. Tout cela cause de la pénibilité à l'annotation.

Une solution consiste à générer des données synthétiques annotées. Différentes études proposent des générateurs de données synthétiques de type GAN [3], notamment BubGAN [4] et BF-GAN [5]. BubGAN génère des bulles synthétiques et leurs masques correspondants. Cette méthode repose sur des contraintes prédéterminées telles que le rapport hauteur/largeur, l'angle de rotation, la distribution, la fraction vide et intègre des informations sur les masques de

bulles. BF-GAN propose un générateur d'images de flux de bulles et fournit un détecteur de bulles, en prenant en compte des contraintes physiques comme la vitesse des bulles et leur distribution de tailles, en plus des contraintes géométriques.

Bien que des bulles artificielles aient récemment montré un haut niveau de réalité, les bulles en écoulement peuvent présenter des morphologies plus variées en fonction d'autres paramètres opératoires ou propriétés physiques comme la viscosité, la tension de surface, etc. Pour résoudre ces problèmes, notre stratégie en deux étapes, nommée FSOC basée sur FastSAM[6] et OpenCLIP [7], permet de localiser et de classer des bulles directement sur des images d'expérimentation. Notre approche offre la possibilité de raccourcir le temps d'annotation.

2 Méthode et Données

2.1 Pipeline proposé

Dans cet article, nous présentons une pipeline d'annotation automatique à l'aide de FastSAM et OpenCLIP. La Figure 1 illustre ce pipeline d'annotation par FSOC avec deux étapes : localisation et classification.

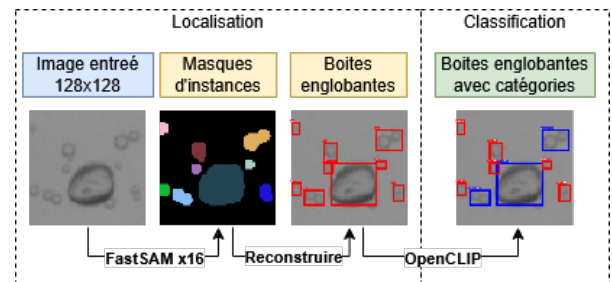


FIGURE 1 – Pipeline d'annotation FSOC proposé - les bulles sont zoomées pour être plus visibles.

Afin de résoudre le problème de détection, nous avons d'abord opté pour l'application de FastSAM, un modèle de segmentation de toutes les instances. Ce modèle s'appuie sur l'architecture convolutive de YOLO-Seg v.8 [8], qui s'entraîne sur 2% des données SA-1B [9]. Les modèles sous-prédisent souvent des objets petits à cause de leur petitesse. Pour limiter l'impact de ce problème, nous réali-

sons une inférence d’image donnée en l’élargissant, avec un facteur de 16. Ensuite, FastSAM infère des masques des bulles puis reconstruit les boîtes englobantes.

Une fois les bulles localisées, une classification est effectuée à l’aide du modèle OpenClip, un modèle vision-langage pré-entraîné sur les données LAION-5B [10] permettant de classifier des boîtes fournies par FastSAM.

2.2 Préparation de données

Dans notre configuration d’expérimentation, une colonne transparente en PVC (polychlorure de vinyle) sert de milieu de réaction. De l’eau distillée est injectée à la hauteur de la colonne, tandis qu’un flux de gaz (air ou hydrogène) est introduit en sens opposé. Un diffuseur en céramique ou en métal assure la dispersion des bulles. Pour éviter la distorsion des bulles lors de l’acquisition des images, une boîte cubique est placée au tour de la colonne. Les images en niveaux de gris sont capturées par une caméra sur un fond éclairé par une lumière blanche.

Le jeu de données comprend 12 images de taille 470×1216 pixels. Ces images sont annotées manuellement, avec un temps moyen d’environ 4 heures par image. Dans ces données, deux catégories sont proposées : les bulles isolées (10300 instances) et les groupes de plusieurs bulles (1993 instances). La première catégorie consiste à des bulles complètement isolées et la deuxième concerne des ensembles dans lesquels les bulles sont très proches les unes des autres ou se chevauchent. La mesure de la taille de chaque catégorie est différente, car les bulles isolées sont plus faciles à estimer que celles dans un groupe de bulles. Ce jeu de données est découpé dans 780 patchs de taille uniforme de 128x128 pixels, avec un chevauchement de 25% de la largeur et de la hauteur du patch en utilisant [11]. Ensuite, ce jeu de données est divisé en 455/325 pour le jeu de données d’entraînement et le jeu de données de validation lors de l’apprentissage.

3 Expérimentations

3.1 Configuration d’expérimentation

Deux sessions d’apprentissage sont réalisées avec des données générées par FSOC et des données manuelles. La validation est effectuée sur donnée manuelles à chaque epoch. Ces sessions sont effectuées avec Faster RCNN [12] sur 12 epochs. Dans le cadre de notre démarche d’évaluation, nous avons appliqué des métriques de AITOD, telles que présentées dans les travaux de [13] et [14]. Ces métriques, spécifiques à la détection d’objets de petites tailles, sont déclinées en différentes tranches de taille : très très petit (de 0 à 64 pixels), très petite (de 64 à 256 pixels), petite (de 256 à 1024 pixels) et moyenne (à partir de 1024 pixels).

3.2 Résultats

Comme indiqué dans le Tableau 1, le temps d’annotation par FSOC est largement inférieur à celui de l’annotation manuelle, étant 98% plus rapide. En outre, le Tableau 2 présente un aperçu des diverses métriques. Les mAPs de Faster RCNN sur les données par FSOC restent faibles. En

général, la catégorie de la bulle isolée est mieux détectée que celle du groupe de plusieurs bulles en raison des déséquilibres entre deux catégories.

TABLE 1 – Comparaison du temps de traitement d’une image selon chaque approche

Données	Temps (min)	Temps (max)	Temps (moyen)
Manuel	180 min	300 min	240 min
FSOC	1 min 30 s	5 min	3 min

TABLE 2 – Comparaison des résultats pendant l’apprentissage sur le jeu de données de validation (avec l’annotation manuelle)

Données	Catégorie	mAP_50	mAP_vt	mAP_t	mAP_s	mAP_m
Manuel	General	0.852	0.366	0.590	0.667	0.729
	Bulle isolée	0.874	0.367	0.63	0.728	0.85
	Plusieurs bulles	0.83	0.366	0.549	0.606	0.607
FSOC	General	0.414	0.062	0.16	0.265	0.374
	Bulle isolée	0.53	0.11	0.24	0.362	0.55
	Plusieurs bulles	0.297	0.014	0.08	0.169	0.198

Comme illustré dans la figure 2 : le modèle entraîné sur les données générées par FSOC est capable de localiser la plupart des objets dans l’image donnée, sans distinction de catégories. Cependant, son application à des classes spécifiques soulève des défis. Pour les très petits objets, des faux négatifs apparaissent, tandis que pour les groupes de bulles, les prédictions donnent des faux positifs en raison de classification incorrecte par OpenCLIP. En outre, les boîtes englobantes semblent souvent plus petites que les boîtes de vérité terrain à cause de la reconstruction de boîtes à partir des masques de FastSAM.

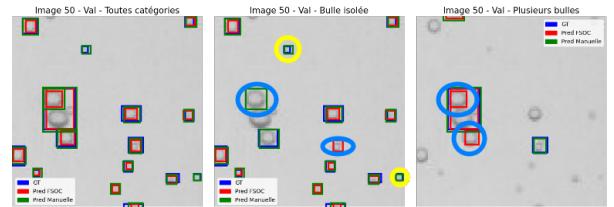


FIGURE 2 – Illustration de la prédiction sur une image de validation pour trois scénarios. Les ellipses jaunes représentent des faux négatifs et les ellipses bleues, des faux positifs.

4 Conclusion et Perspectives

Dans les premières expérimentations, nous démontrons la faisabilité d’un pipeline d’annotation FSOC et réduisons significativement le temps d’annotation. Actuellement, les performances sur les données générées par FSOC sont limitées par rapport à celles obtenues manuellement, car les sorties de FastSAM sont insuffisamment traitées. En outre, le modèle de classification OpenCLIP ne fonctionne pas de manière optimale.

Afin de rendre le pipeline FSOC utilisable, il serait opportun d’approfondir les études sur le post-traitement des boîtes englobantes générées par FastSAM. Nous envisageons également de réaliser une session d’apprentissage simultanée de FastSAM et d’OpenClip sur quelques données annotées manuellement afin d’améliorer les performances. Ces pistes font l’objet de nos recherches actuelles et seront au cœur de nos futurs développements.

Références

- [1] M. Keramati et .al. Intensification of ex-situ biome-
thanation in a bubble column bioreactor by addition
of colonized biochips. *Bioresource Technology Re-
ports*, 27 :101938, 2024.
- [2] B. Sanogo et .al. Exploring the impact of proteins
and surfactant on oxygen mass transfer in gas-liquid
bioreactors : an experimental investigation. *Chemical
Engineering Science*, page 121146, December 2024.
- [3] I. J. Goodfellow et .al. Generative adversarial net-
works. *Advances in Neural Information Processing
Systems*, 3, 06 2014.
- [4] F. Yucheng et .al. Bubgan : Bubble generative ad-
versarial networks for synthesizing realistic bubbly
flow images. *Chemical Engineering Science*, 204, 04
2019.
- [5] W. Zhou et .al. BF-GAN : Development of
an ai-driven bubbly flow image generation model
using bubbly generative adversarial networks, 2024.
arXiv :2502.06863.
- [6] X. Zhao et .al. Fast segment anything, 2023.
arXiv :2306.12156.
- [7] I. Gabriel et al. OpenCLIP, July 2021. Version
0.1. Zenodo. [https://doi.org/10.5281/
zenodo.5143773](https://doi.org/10.5281/zenodo.5143773).
- [8] G. Jocher et .al. Ultralytics yolov8, 2023. [https://
github.com/ultralytics/ultralytics](https://github.com/ultralytics/ultralytics).
- [9] A. Kirillov et .al. Segment anything, 2023.
arXiv :2304.02643.
- [10] C. Schuhmann et .al. LAION-5B : an open large-
scale dataset for training next generation image-text
models. In *Proceedings of the 36th International
Conference on Neural Information Processing Sys-
tems, NIPS '22*, New Orleans, LA, USA, 2022. Cur-
ran Associates Inc.
- [11] F. C. Akyon et .al. Slicing aided hyper inference
and fine-tuning for small object detection. In *Procee-
dings of the 2022 IEEE International Conference on
Image Processing (ICIP)*, pages 966–970, Bordeaux,
France, 2022.
- [12] S. Ren et .al. Faster R-CNN : Towards real-time ob-
ject detection with region proposal networks. *IEEE
Transactions on Pattern Analysis and Machine Intel-
ligence*, 39(6) :1137–1149, 2017.
- [13] J. Wang et a. Tiny object detection in aerial images.
In *2020 25th International Conference on Pattern
Recognition (ICPR)*, pages 3791–3798, Milan, Italy
(virtual), 2021.
- [14] J. Wang. Coco api for ai-tod dataset. [https://
github.com/jwwangchn/cocoapi-aitod](https://github.com/jwwangchn/cocoapi-aitod),
2021.