Détection et segmentation automatisées de bactériocytes à partir d'images de microscopie optique

Nathan HUTIN[1] Chantal REVOL-MULLER[1] Karen GAGET[2] Séverine BALMAND[2] Federica CALEVRO[2] Mélanie RIBEIRO-LOPES[2] Thomas GRENIER[1]

[1]INSA-Lyon, Universite Claude Bernard Lyon 1, CNRS, Inserm, CREATIS UMR 5220, U1294, F-69621, Villeurbanne, France

[2]INSA-Lyon, INRAE, BF2I, UMR0203, F-69621, Villeurbanne, France

Résumé – Ce travail présente un outil automatisé basé sur l'apprentissage profond pour la détection et la segmentation des bactériocytes, cellules spécialisées dans l'hébergement des bactéries symbiotiques chez les insectes. L'objectif est d'automatiser la mesure de surface des cellules nettes intactes, une tâche actuellement manuelle. La segmentation automatisée présente plusieurs défis tels que la superposition des cellules, les variations de netteté et la distinction entre cellules intactes et éclatées. Un protocole d'acquisition standardisé de 11 images espacées de 8 µm en microscopie champ clair (Thunder Imager 3D, Leica) a été mis en place. Trois architectures (Mask R-CNN, YOLOv8 et Mask Frozen-DETR) ont été entraînées en validation croisée, et un post-traitement basé sur le score de Dice élimine les prédictions redondantes. Les résultats montrent des performances encourageantes en termes de mAP et de rappel, ouvrant la voie à une automatisation prometteuse après optimisation complémentaire de la précision.

Abstract – This work presents an automated deep-learning-based tool for detecting and segmenting bacteriocytes, insect-specific cells involved in bacterial symbiosis. The objective is to automate surface area measurement of intact sharp cells, currently performed manually. Automated segmentation faces several challenges such as cell overlap, focus variability, and the differentiation between intact, burst, sharp, and blurry cells. A standardized acquisition protocol capturing 11 images per cell cluster at 8 µm intervals using bright-field microscopy (Thunder Imager 3D, Leica) was established. Three architectures (Mask R-CNN, YOLOv8, and Mask Frozen-DETR) were trained using cross-validation, followed by a Dice-score-based post-processing step to remove redundant predictions. The results demonstrate encouraging performance in terms of mAP and recall, paving the way toward a promising automated solution after further optimization of precision.

1 Introduction

L'objectif de cette étude est de concevoir et développer une méthode automatisée de détection et de segmentation des bactériocytes, cellules présentes chez certains insectes, dont de nombreux insectes ravageurs de cultures. Ces cellules jouent un rôle crucial dans les échanges métaboliques entre les insectes hôtes et leurs bactéries symbiotiques, essentielles à leur survie. En particulier, les biologistes étudient l'évolution du nombre et de la taille de ces cellules chez les pucerons du pois au fil de leur développement [8]. Actuellement, la mesure précise de la taille des bactériocytes est réalisée manuellement à partir d'images acquises en microscopie optique en champ clair. Cette approche, chronophage et sujette à la subjectivité, représente un frein à la reproductibilité des recherches. L'automatisation de cette tâche permettrait un gain de temps considérable et une standardisation des analyses. Bien que des méthodes de segmentation automatique aient émergé, comme StarDist pour les noyaux cellulaires [9] ou des modèles généralistes comme SAM [1], elles ne sont pas directement applicables à des cellules non nucléées et fortement superposées comme les bactériocytes.

Ainsi, la segmentation automatisée des bactériocytes soulève plusieurs défis scientifiques importants. Premièrement, ces cellules sont fréquemment agglomérées en amas, générant des superpositions rendant complexe leur séparation et identification individuelles. De plus, l'acquisition par microscopie optique entraîne inévitablement des variations de netteté significatives dues à la faible profondeur de champ, conduisant à des images où de nombreuses cellules apparaissent floues. Il est donc nécessaire de sélectionner précisément les cellules nettes, ce qui nécessite une méthode fiable et automatique de discrimination. Deuxièmement, l'absence initiale de protocole standardisé pour l'acquisition des images limitait la reproductibilité et la comparabilité des résultats entre expériences. Enfin, la préparation mécanique des lames provoque régulièrement l'éclatement d'un nombre significatif de cellules, complexifiant encore la tâche en raison de la difficulté à discriminer automatiquement ces cellules éclatées des cellules intactes, pourtant morphologiquement très similaires.

Pour surmonter ces difficultés, plusieurs approches complémentaires sont envisagées. Tout d'abord, un protocole d'acquisition standardisé accompagné d'un protocole rigoureux de segmentation manuelle ont été définis et mis en œuvre (détaillés dans la section 2). Ensuite, des architectures modernes de réseaux de neurones, adaptées à la détection et la segmentation d'instances, ont été explorées (section 3). Les solutions proposées seront évaluées et discutées dans la section 4.

2 Jeu de données

2.1 Acquisition

Un protocole d'acquisition standardisé a été établi à l'aide d'un microscope en champ clair (Thunder Imager 3D Live Cell®, Leica), permettant l'acquisition d'images couleur RGB

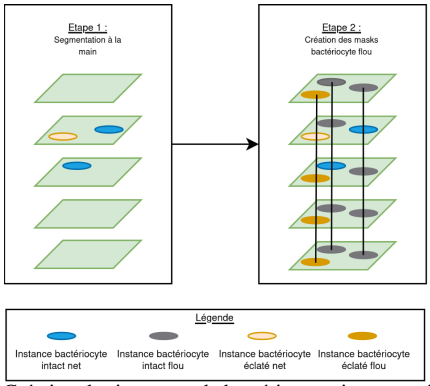

FIGURE 1 : Création des instances de bactériocytes intacts et éclatés flous.

avec une résolution de 4000×3000 pixels. Pour chaque amas de bactériocytes, une pile de 11 images a été capturée, avec un espacement régulier de 8 µm entre les différents plans focaux sur une même lame. La segmentation manuelle des bactériocytes nets, qu'ils soient intacts ou éclatés, a été réalisée sur le logiciel 3D Slicer V5.4.0. Chaque cellule a été segmentée une seule fois sur le plan où elle apparaissait le plus nettement parmi les 11 disponibles (fig. 1). Ce protocole a permis de constituer un jeu de données composé de 195 amas cellulaires, représentant au total 392 instances de bactériocytes intacts nets et 159 instances de bactériocytes éclatés nets. Les segmentations des cellules floues ont été générées artificiellement en projetant les segmentations nettes des intactes et éclatées sur les autres plans focaux de la pile d'images, en copiant à l'identique la morphologie des segmentations nettes.

La figure 2 présente visuellement la distinction entre les différentes catégories de bactériocytes : intact net, intact flou, éclaté net et éclaté flou, illustrant les variations de netteté dues à la focalisation et les modifications morphologiques causées par l'éclatement des cellules.

2.2 Prétraitement

Les images acquises ont été prétraitées afin d'optimiser leur utilisation dans l'entraînement des modèles. Tout d'abord, la résolution des images a été réduite à 400×300 pixels, ce qui a permis à la fois d'accroître la variabilité inter-pixels utile à l'apprentissage des modèles et de réduire significativement le coût computationnel lié au traitement des données. Ensuite, des segmentations artificielles des cellules floues ont été générées à partir de la pile complète des images, comme illustré sur la figure 1. Ce procédé a permis d'obtenir un jeu de données plus représentatif de la diversité des états de focalisation des cellules rencontrés en pratique.

2.3 Augmentation de données

Pour augmenter le nombre d'exemples disponibles pour l'entraînement, une stratégie d'augmentation de données a été mise en œuvre, inspirée de la méthode RandAugment [3]. Cette approche a toutefois nécessité quelques adaptations spécifiques aux contraintes de ce projet. Ainsi, certaines transformations proposées initialement dans RandAugment, telles que la modification de netteté (*sharpness*), ont été exclues afin d'éviter de rendre artificiellement nettes toutes les cellules. De même, les transformations *shearX* et *shearY* ont été éliminées car elles déformaient excessivement la morphologie des cellules. Toutes les transformations affines retenues ont été appliquées conjointement aux images et aux segmentations afin de préserver la correspondance spatiale entre celles-ci et leurs vérités terrain. Enfin, pour des raisons pratiques et techniques (certaines architectures utilisées ne supportant pas l'augmentation à la volée), les données augmentées ont été générées avant l'entraînement, produisant ainsi 24 variantes transformées de chaque image originale.

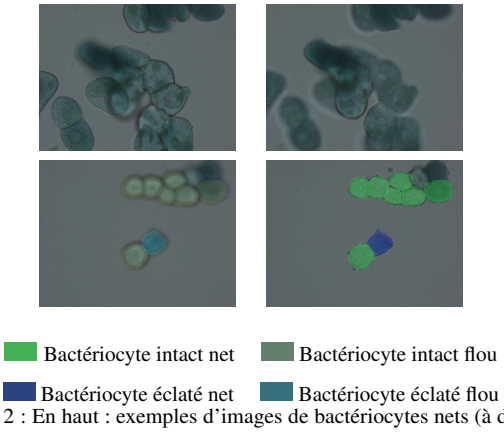

FIGURE 2 : En haut : exemples d'images de bactériocytes nets (à droite) et flous (à gauche). En bas : exemples d'annotations manuelles.

3 Méthode

3.1 Choix des architectures

Afin de réaliser efficacement la tâche de détection et de segmentation d'instances des bactériocytes, trois architectures représentatives des principales approches actuelles en apprentissage profond ont été sélectionnées, entraînées et comparées.

La première architecture sélectionnée est YOLOv8 [7], réputée pour sa vitesse d'inférence élevée et ses performances reconnues en détection d'objets en temps réel. YOLOv8 appartient à la famille des détecteurs d'objets *one-stage*, c'est-à-dire des modèles capables de prédire directement les coordonnées et les classes des objets à partir d'une seule passe sur l'image d'entrée. Pour pouvoir mener à bien une évaluation robuste par validation croisée à cinq plis, des modifications spécifiques ont été apportées à la classe dataset de l'implémentation originale, afin d'éviter des copier coller d'image inutle mais nécessaire pour l'achitecture du code original. L'entraînement a été effectué à l'aide de la librairie Ultralytics [5].

La deuxième architecture mise en œuvre est Mask R-CNN [4], qui constitue une référence dans le domaine de la segmentation d'instances grâce à ses performances solides et documentées sur de nombreux jeux de données standards. Mask R-CNN est une méthode dite *two-stage*, dérivée de Faster R-CNN, et caractérisée par une étape intermédiaire de génération de propositions (*Region Proposal Network*), suivie d'une étape de classification et de segmentation précise des régions détectées. L'implémentation utilisée provient de la librairie Detectron2 [11], développée par Facebook Research.

Enfin, la troisième architecture considérée est Mask Frozen-DETR [6], une variante récente de l'approche DETR (DEtection TRansformer), qui utilise une architecture basée sur les Transformers [10]. DETR constitue une avancée significative par rapport aux méthodes traditionnelles basées sur des ancrages (*anchors*) en formulant la détection et la segmentation comme une prédiction directe d'ensembles d'objets grâce aux

mécanismes d'attention. Mask Frozen-DETR est particulièrement prometteur pour sa capacité à capturer les relations complexes entre objets et gérer naturellement la variabilité spatiale des instances. Pour utiliser efficacement cette architecture pour notre méthode, nous avons modifié la loss lorsque l'entrainement se fait sur le jeu de donnée avec un catégorie.

Tous les entraînements réalisés ont été validés selon une approche de validation croisée à cinq plis, permettant une évaluation robuste et une mesure pertinente des capacités de généralisation des modèles. Plusieurs expériences ont été réalisées en variant les catégories d'objets incluses dans le jeu de données, tout en conservant systématiquement la catégorie de référence, les bactériocytes nets intacts, afin de déterminer la meilleure stratégie pour leur détection et leur segmentation. Le tableau 1 résume les configurations expérimentales explorées.

Expérience	BNI	BFI	BNE	BFE
BNI	✓	×	×	×
BNI vs. BFI	✓	✓	×	×
BNI vs. BNE	✓	×	✓	×
BNI vs. BFI vs. BNE vs. BFE	✓	✓	✓	✓
BNI vs. non souhaité	✓	✓		

TABLE 1 : Résumé des différentes expériences réalisées avec les catégories cellulaires incluses (BNI : Bactériocyte Net Intact, BFI : Bactériocyte Flou Intact, BNE : Bactériocyte Net Éclaté, BFE : Bactériocyte Flou Éclaté).

3.2 Post-traitement des prédictions

Les prédictions fournies par les architectures entraînées nécessitent un traitement complémentaire pour éliminer les redondances dues à la présence d'une même cellule sur plusieurs plans focaux. Pour résoudre ce problème, un post-traitement basé sur le calcul du score de similarité de Dice a été mis en place (cf. Fig. 3). Le processus consiste à comparer les masques prédits entre les différentes images focales au sein de chaque amas cellulaire. Lorsque plusieurs masques prédits sont détectés comme représentant la même cellule, seule la prédiction présentant le score de confiance maximal donnée par le modèle est conservée, les autres étant supprimées. En outre, une étape supplémentaire permet d'écarter les prédictions incorrectes en supprimant systématiquement les instances dont les contours atteignent les bords de l'image, ces prédictions partielles n'étant pas exploitables pour les mesures ultérieures.

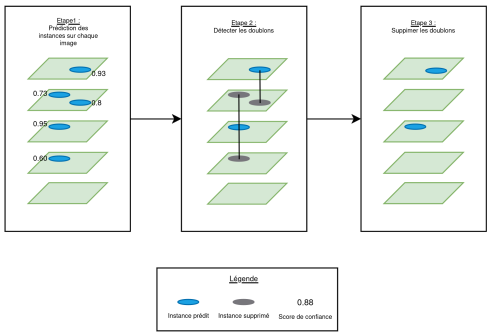

FIGURE 3 : Post-traitement éliminant les prédictions redondantes et basé sur le score de Dice entre les masques obtenus à différents plans focaux.

4 Résultats

L'évaluation des performances obtenues par les différentes architectures repose principalement sur deux types de métriques : d'une part, le score de Dice [2], utilisé pour évaluer la qualité des masques de segmentation en comparant la similarité entre les masques prédits et les vérités terrain au niveau pixel, et d'autre part, les métriques de détection telles que la précision moyenne (*mean Average Precision*, mAP), la précision et le rappel au niveau des instances.

Le mAP est calculé à partir de l'indice d'intersection sur union (IoU, *Intersection over Union*) entre la prédiction et la vérité terrain. Cet indice permet de déterminer si une prédiction donnée (une instance de bactériocyte) est un vrai positif ou un faux négatif, en fonction d'un seuil fixé (75 et 90 dans notre cas). On construit une courbe précision-rappel dont l'intégrale fournit l'*Average Precision* (AP) pour chaque catégorie, définie par la formule suivante :

$$AP = \int_0^1 p(r)\,dr \quad \text{où } p(r) \text{ est la fonction précision-rappel.}$$

Le mAP correspond à la moyenne des AP sur l'ensemble des catégories considérées :

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \quad \text{avec } N \text{ le nombre de catégories.}$$

Les résultats expérimentaux obtenus sont regroupés dans le tableau 2. Pour Mask R-CNN, les meilleures performances de détection après l'étape de post-traitement (en termes de mAP@75 et mAP@90) ont été observées pour les expériences incluant les catégories BNI, BNI vs. BNE, ainsi que BNI vs. non souhaité. En particulier, l'entraînement visant explicitement à distinguer les bactériocytes nets des éclatés (BNI vs. BNE) a permis une amélioration notable de la détection des bactériocytes nets intacts après post-traitement. Toutefois, ces expériences affichent une précision relativement modeste, proche de 0.5, lié à un taux élevé de faux positifs. Ce faible taux de précision est néanmoins compensé par un rappel très élevé (autour de 0.96), indiquant que la quasi-totalité des cellules cibles sont détectées. Les expériences cherchant à distinguer les cellules nettes des floues ont, quant à elles, donné des résultats moins convaincants. Globalement, la qualité des segmentations produites par Mask R-CNN reste élevée, avec des scores de Dice supérieurs à 0.95 dans les meilleurs cas.

Les résultats obtenus avec YOLOv8 montrent une tendance différente : ce modèle semble avoir réussi à mieux généraliser dans certaines configurations, notamment pour les expériences BNI seules ou BNI vs. BNE. L'expérience utilisant uniquement les cellules nettes (BNI seulement) a permis d'atteindre une valeur de rappel acceptable (0.894), malgré une précision modérée. La qualité des segmentations mesurée par le score de Dice est globalement satisfaisante, variant autour de 0.90. Ces résultats suggèrent que YOLOv8 pourrait avoir une meilleure aptitude à gérer les subtilités visuelles entre les catégories nettes et floues dans certains contextes spécifiques.

Enfin, les résultats préliminaires obtenus avec l'architecture DETR, bien que non évalués en validation croisée complète, se révèlent prometteurs. En particulier, l'expérience distinguant les cellules intactes des éclatées (BNI vs. BNE) a affiché des résultats encourageants, avec des scores de précision (0.747) et de rappel (0.810) élevés, malgré un mAP légèrement inférieur à celui des deux autres architectures testées. Ces résultats initiaux soulignent le potentiel de l'approche DETR pour une application pratique future, notamment grâce à sa capacité à saisir efficacement les différences subtiles entre catégories.

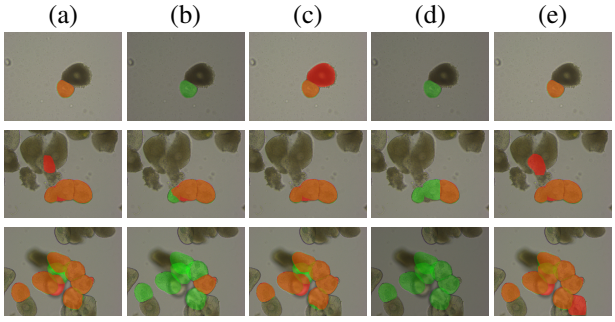

FIGURE 4 : Prédictions des bactériocytes nets intacts pour chaque configuration d'entraînement Mask R-CNN : (a) BNI, (b) BNI vs BFI, (c) BNI vs BNE, (d) BNI vs BFI vs BNE vs BFE, (e) BNI vs non souhaité.

Une difficulté commune rencontrée par tous les modèles est la distinction entre cellules nettes et floues. Cette difficulté peut s'expliquer par la nature subjective de cette distinction ainsi que par la forte similarité visuelle intrinsèque à ces deux états. La visualisation qualitative des prédictions (Fig. 4) confirme ces tendances, montrant clairement une bonne correspondance entre les prédictions et les vérités terrain pour les meilleurs entraînements, ainsi qu'une aptitude satisfaisante à gérer les cas de chevauchements de cellules tout en évitant les faux positifs sur les cellules éclatées.

Expérience	mAP75	mAP90	Dice	Précision$_{75}$	Rappel$_{75}$
Mask R-CNN					
BNI	0.771 ± 0.074	0.583 ± 0.113	0.953 ± 0.044	0.501 ± 0.087	$\mathbf{0.961 \pm 0.023}$
BNI vs. BFI	0.147 ± 0.072	0.127 ± 0.057	0.810 ± 0.308	0.765 ± 0.157	0.170 ± 0.078
BNI vs. BNE	$\mathbf{0.802 \pm 0.095}$	$\mathbf{0.614 \pm 0.101}$	$\mathbf{0.954 \pm 0.045}$	0.562 ± 0.104	0.960 ± 0.030
BNI vs. BFI vs. BNE vs. BFE	0.147 ± 0.095	0.137 ± 0.100	0.941 ± 0.105	$\mathbf{0.935 \pm 0.090}$	0.149 ± 0.102
BNI vs. non souhaité	0.761 ± 0.093	0.585 ± 0.140	0.953 ± 0.046	0.511 ± 0.065	0.954 ± 0.020
YOLOv8					
BNI	$\mathbf{0.757 \pm 0.038}$	$\mathbf{0.533 \pm 0.036}$	$\mathbf{0.931 \pm 0.02}$	0.435 ± 0.058	$\mathbf{0.894 \pm 0.058}$
BNI vs. BFI	0.562 ± 0.16	0.378 ± 0.13	0.877 ± 0.051	0.652 ± 0.077	0.656 ± 0.182
BNI vs. BNE	0.668 ± 0.042	0.526 ± 0.049	0.912 ± 0.027	0.508 ± 0.092	0.790 ± 0.088
BNI vs. BFI vs. BNE vs. BFE	0.584 ± 0.196	0.366 ± 0.165	0.900 ± 0.018	0.693 ± 0.125	0.671 ± 0.240
BNI vs. non souhaité	0.602 ± 0.172	0.390 ± 0.146	0.919 ± 0.032	$\mathbf{0.695 \pm 0.112}$	0.686 ± 0.219
DETR					
BNI vs. BNE	**0.631**	**0.277**	**0.916**	0.747	**0.810**
BNI vs. BFI vs. BNE vs. BFE	0.260	0.093	0.915	**0.838**	0.273

TABLE 2 : Résultats des expériences d'apprentissage avec Mask R-CNN (conf : 0.8), YOLOv8 (conf : 0.5) et DETR (conf : 0.9). Les résultats sont présentés après post-traitement, uniquement sur la catégorie des BNI. Les indices 75 et 90 sont les seuils d'IoU utilisés pour déterminer les vrais positifs. La détection et la segmentation sont évaluées relativement à la pile d'images.

5 Conclusion

Cette étude démontre le potentiel prometteur des approches basées sur l'apprentissage profond pour automatiser efficacement la détection et la segmentation des bactériocytes sur des images de microscopie optique. Parmi les modèles étudiés, Mask R-CNN a affiché les meilleures performances globales après post-traitement, en particulier lorsqu'il était entraîné spécifiquement à différencier les bactériocytes nets des cellules éclatées. Toutefois, la précision relativement modeste obtenue lors des expériences les plus performantes indique clairement la nécessité de poursuivre les travaux afin de réduire significativement le nombre de faux positifs, condition essentielle à une utilisation opérationnelle par les biologistes.

YOLOv8 présente quant à lui l'intérêt majeur d'une vitesse d'inférence élevée, particulièrement utile dans un contexte applicatif où de grands volumes d'images doivent être analysés rapidement. Cependant, les résultats en détection restent légèrement inférieurs à ceux obtenus par Mask R-CNN. L'architecture DETR, testée de manière préliminaire, révèle un fort potentiel grâce à ses résultats prometteurs en termes de précision et de rappel, justifiant ainsi des études complémentaires approfondies pour pleinement exploiter ses capacités. La difficulté rencontrée par toutes les approches pour distinguer efficacement les cellules nettes des cellules floues souligne la complexité intrinsèque de cette tâche. Cette problématique pourrait être abordée différemment, notamment en envisageant une prédiction directe d'un score de netteté pour chaque instance, plutôt qu'une classification binaire stricte.

Les perspectives ouvertes par ces travaux sont nombreuses et prometteuses. Parmi celles-ci, l'exploitation simultanée des informations contenues dans l'ensemble des 11 images des piles focales apparaît particulièrement intéressante. En effet, une telle approche pourrait permettre d'intégrer pleinement l'information tridimensionnelle disponible, évitant ainsi l'étape de post-traitement actuel. Dans ce contexte, l'exploitation plus poussée des mécanismes d'attention déjà présents dans les architectures à base de Transformers, telles que DETR, ainsi que leur intégration éventuelle au sein de modèles hybrides, par exemple une variante de Mask R-CNN enrichie d'attention, pourrait améliorer considérablement les performances. Cela permettrait notamment aux modèles d'exploiter pleinement les relations spatiales et temporelles existant entre les différents plans focaux. L'objectif à terme est de déployer une solution performante sous forme d'extension pour le logiciel Slicer.

Références

[1] Anwai ARCHIT, Luca FRECKMANN, Sushmita NAIR, Nabeel KHALID, Paul HILT, Vikas RAJASHEKAR, Marei FREITAG, Carolin TEUBER, Genevieve BUCKLEY, Sebastian von HAAREN *et al.* : Segment anything for microscopy. *Nature Methods*, pages 1–13, 2025.

[2] Jeroen BERTELS, Tom EELBODE, Maxim BERMAN, Dirk VANDERMEULEN, Frederik MAES, Raf BISSCHOPS et Matthew B. BLASCHKO : *Optimizing the Dice Score and Jaccard Index for Medical Image Segmentation : Theory and Practice*, page 92–100. Springer International Publishing, 2019.

[3] Ekin D. CUBUK, Barret ZOPH, Jonathon SHLENS et Quoc V. LE : Randaugment : Practical data augmentation with no separate search. *CoRR*, abs/1909.13719, 2019.

[4] Kaiming HE, Georgia GKIOXARI, Piotr DOLLÁR et Ross GIRSHICK : Mask r-cnn. *In Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[5] Glenn JOCHER, Ayush CHAURASIA et Jing QIU : Ultralytics yolo, january 2023. *URL https ://github. com/ultralytics/ultralytics*, 3, 2022.

[6] Zhanhao LIANG et Yuhui YUAN : Mask frozen-detr : High quality instance segmentation with one gpu, 2023.

[7] Dillon REIS, Jordan KUPEC, Jacqueline HONG et Ahmad DAOUDI : Real-time flying object detection with yolov8, 2023.

[8] Mélanie RIBEIRO LOPES, Karen GAGET, François RENOZ, Gabrielle DUPORT, Séverine BALMAND, Hubert CHARLES, Patrick CALLAERTS et Federica CALEVRO : Bacteriocyte plasticity in pea aphids facing amino acid stress or starvation during development. *Frontiers in Physiology*, 13:982920, 2022.

[9] Uwe SCHMIDT, Martin WEIGERT, Coleman BROADDUS et Gene MYERS : Cell detection with star-convex polygons. *In MICCAI 2018 : 21st int. conference, Spain, September 16-20, 2018, proceedings, part II 11*, pages 265–273. Springer, 2018.

[10] Ashish VASWANI, Noam SHAZEER, Niki PARMAR, Jakob USZKOREIT, Llion JONES, Aidan N GOMEZ, Łukasz KAISER et Illia POLOSUKHIN : Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[11] Yuxin WU, Alexander KIRILLOV, Francisco MASSA, Wan-Yen LO et Ross GIRSHICK : Detectron2. `https://github.com/facebookresearch/detectron2`, 2019.