

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



Université A. Mira de Béjaïa
Faculté des Sciences Exactes
Département d'Informatique

MÉMOIRE DE MASTER

En
Informatique

Option
Intelligence Artificielle

Thème

Proposition d'une nouvelle approche
YOLOv8+SAM pour la segmentation de pommes

Présenté par : M. IDIRI Anis
M. GHEBRIOU Oussama

Évaluer le 03 juillet 2023 par le jury composé de :

Présidente	Mme D. BOUKEREDERA	U. A. Mira Béjaïa.
Encadrant	M. F. BOUCHEBBAH	U. A. Mira Béjaïa.
Examinatrice	Mme S. AIT KACI AZZOU	U. A. Mira Béjaïa.

Béjaïa, juillet 2023.

✧ Remerciements ✧

Avant tout nous remercions dieu le tout puissant, qui nous a donné la force et la patience pour l'accomplissement de ce travail qui est pour nous le point de départ d'une merveilleuse aventure, celle de la recherche, source de remise en cause permanente et de perfectionnement perpétuel.

Nous vifs remerciements vont en premier lieu à notre encadrant Mr Bouchebbah Fatah pour avoir accepté de nous guider tout au long de ce travail, pour sa disponibilité et son implication pour l'aboutissement de ce travail.

Aux membres du jury Mme Djamila BOUKEREDERA et Mme Samira AIT KACI AZZOU pour l'intérêt qu'elles ont porté à notre travail en acceptant de l'examiner et de l'enrichir par leurs propositions.

À tous les enseignants qui ont cultivé en nous l'amour de l'Informatique. Qu'ils trouvent dans ce modeste travail un petit geste de reconnaissance qui ne sera jamais à la hauteur de la grandeur de leurs âmes.

Enfin, nous remercions, de tout cœur, tous ceux qui ont contribué de près ou de loin à la réalisation de ce mémoire.

✧ *Dédicaces* ✧

À tous ceux qui m'ont soutenu tout au long de mon parcours universitaire, qui ont cru en moi et m'ont encouragé à donner le meilleur de moi-même, je dédie ce mémoire. Votre présence et votre soutien ont été essentiels dans la réalisation de ce projet.

À mes chers parents, qui ont été mes premiers enseignants et mes plus grands soutiens. Votre amour inconditionnel, vos encouragements constants et votre confiance en moi m'ont permis d'accomplir ce parcours académique. Ce mémoire est le fruit de vos sacrifices et de votre dévouement. Je vous suis infiniment reconnaissant pour tout ce que vous avez fait et continuez de faire pour moi.

À mon frère Aghiles et mes soeurs Lydia et Yasmine, qui ont toujours été à mes côtés, me prodiguant conseils, réconfort et encouragement. Votre soutien indéfectible et vos encouragements sincères ont été une source d'inspiration constante. Ce mémoire est également le vôtre, car vous avez partagé chaque étape de mon parcours et m'avez poussé à donner le meilleur de moi-même. Je vous suis reconnaissant d'être ma famille et de m'avoir toujours soutenu.

À Adel, et à mes chers amis Sarah, Massinissa, Omar, Nabil, Oussama, Said, Rayane, Mano, Sadek, Syphax,

Mohemmed et tout les autres, qui ont illuminé mes journées d'étude par votre présence joyeuse et votre soutien indéfectible. Vos encouragements, votre compréhension et votre amitié précieuse ont été une bouée d'oxygène pendant les périodes de stress et de doute. Ce mémoire est également un témoignage de notre amitié et des liens forts que nous avons tissés. Merci d'avoir été là, de m'avoir soutenu et d'avoir rendu ce parcours mémorable grâce à votre présence. À tous ceux qui ont joué un rôle dans mon parcours académique et personnel, que ce soit de près ou de loin, je vous suis infiniment reconnaissant. Vos encouragements, votre soutien et votre amitié ont été des facteurs clés dans la réalisation de ce mémoire. Merci du fond du cœur pour tout ce que vous avez apporté dans ma vie.

M. IDIRI Anis

✧ *Dédicaces* ✧

Rien n'est aussi précieux que le lien qui unit une famille aimante et des amis fidèles, ceux qui ont partagé les joies et les peines de la vie. À travers les épreuves et les triomphes, vous êtes restés présents et soutenus les uns les autres, créant des souvenirs inoubliables.

Aujourd'hui, je souhaite dédier cette pensée spéciale à mes chers parents, à mon frère bien-aimé et à mes compagnons de vie Merouane et Nawal, ainsi qu'à mes amis Mourad, Anis, Juba et Bouzid. Vous êtes tous des piliers importants de ma vie, des êtres chers qui ont fait une différence significative.

À mes parents, je veux exprimer ma gratitude infinie pour votre amour inconditionnel, votre soutien indéfectible et les sacrifices que vous avez consentis pour mon bonheur et ma réussite. Vous avez été mes premiers modèles, me guidant avec sagesse et me montrant l'importance de la famille et de la persévérance. Votre amour et votre dévouement sont les fondations sur lesquelles je construis ma vie.

À mon frère bien-aimé Islam, tu es mon complice, mon confident et mon meilleur ami. Nos liens fraternels sont inestimables, et je suis reconnaissant de t'avoir à mes côtés. Nous avons partagé des rires, des histoires et des moments de complicité qui resteront gravés dans ma

mémoire pour toujours. Ta présence donne un sens profond à ma vie.

À Merouane Nawal, ma compagne de vie, tu es la personne avec qui je partage mes rêves, mes peines et mes victoires. Ton amour, ton soutien inconditionnel et ta présence bienveillante illuminent chaque jour de ma vie. Ensemble, nous avons construit une relation solide basée sur la confiance, la complicité et l'amour. Tu es ma source de bonheur et mon pilier, et je suis béni de t'avoir à mes côtés.

À mes amis Mourad, Anis, Juba et Bouzid, vous êtes les rayons de soleil dans ma vie. Votre présence, votre rire contagieux et votre soutien inébranlable ont enrichi mon parcours. À travers les bons moments et les moments difficiles, nous avons construit une amitié solide basée sur la confiance, le respect et le partage.

Que nos liens familiaux et amicaux continuent de se renforcer avec le temps. Puissions-nous continuer à partager des moments de joie, à surmonter les défis ensemble et à soutenir mutuellement nos aspirations et nos rêves.

À mes parents, mon frère, Merouane Nawal, Mourad, Anis, Juba et Bouzid, je vous dédie cette pensée avec une profonde gratitude et un amour sincère. Votre présence dans ma vie est un cadeau précieux, et je suis béni d'avoir chacun d'entre vous à mes côtés.

Que nos chemins se croisent encore et encore, que nous

*continuions à grandir, à rire et à créer des souvenirs
inoubliables. Je vous souhaite à tous une vie remplie
d'amour, de bonheur et de succès.*

*À mes parents, mon frère et mes chers amis, je vous
remercie du fond du cœur pour tout ce que vous
apportez à ma vie.*

M. GHEBRIOU Oussama

Table des matières

Table des matières	i
Table des figures	iv
Liste des tableaux	vi
Liste des abreviation	vii
Introduction générale	2
1 Généralités et description du contexte de travail	4
1.1 Introduction	4
1.2 Généralités sur la segmentation d'images	4
1.2.1 Techniques de segmentation d'images	4
1.3 Généralités sur l'apprentissage automatique	8
1.3.1 Types d'apprentissage automatique	8
1.3.2 Applications de l'apprentissage automatique	9
1.3.3 Modèles d'apprentissage automatique	9
1.3.4 Deep Learning	9
1.3.5 Réseaux de neurones convolutifs CNN	10
1.3.6 Transfert learning	12
1.3.7 Processus d'apprentissage automatique	17
1.3.8 Défis et limites de l'apprentissage automatique	17
1.3.9 Éthique de l'apprentissage automatique	17
1.4 Généralités sur le fruit pommes	18
1.4.1 Histoire de la pomme	19
1.4.2 Description de la pomme	19
1.4.3 Valeurs nutritives des pommes	19
1.5 Conclusion	19

2	Etat de l’art sur les méthodes de ´ segmentation de pommes dans des images en couleurs	21
2.1	Introduction	21
2.2	Segmentation d’images	21
2.2.1	Préliminaires et définition	21
2.2.2	Intérêt de la segmentation des pommes	22
2.2.3	Challenge de la segmentation des pommes	23
2.3	Quelques méthodes de segmentation de pommes	23
2.3.1	A novel green apple segmentation algorithm based on ensemble U-Net under complex orchard environment	23
2.3.2	A method of apple image segmentation based on color-texture fusion feature and machine learning	24
2.3.3	An accurate detection and segmentation model of obscured green fruits	25
2.3.4	Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot	26
2.3.5	An apple image segmentation method based on a color index obtained by a genetic algorithm	26
2.3.6	Deep learning approach for aApple edge detection to remotely monitor apple growth in orchards	27
2.3.7	Fast segmentation of colour apple image under all-weather natural conditions for vision recognition of picking robots	28
2.4	Synthèse des travaux connexes	30
2.4.1	Tableau comparatif et discussions	30
2.5	Conclusion	32
3	Nouvelle approche de segmentation de pommes : YOLOv8+SAM	33
3.1	Introduction	33
3.2	You Only Look Once (YOLO)	34
3.2.1	Présentation de YOLO	34
3.2.2	Architecture des modèles YOLO	35
3.2.3	YOLOv8	35
3.3	Segment Anything Model (SAM)	36
3.3.1	Architecture du modèle SAM	36
3.4	L’approche proposée : Combinaison de YOLOv8 et SAM (YOLOv8+SAM)	38
3.4.1	Architecture du modèle de segmentation proposé	39
3.5	Tests et évaluations	40
3.5.1	Environnement de travail	40
3.5.2	Dataset	41
3.5.3	Métriques d’évaluation	42

	iii
3.5.4 Résultats de segmentation	44
3.5.5 Evaluation du modèle YOLOv8+SAM	44
3.5.6 Discussion des résultats de la matrice de confusion	45
3.5.7 Discussion des résultats	46
3.6 Conclusion	47
Conclusion générale et perspectives	48
Bibliographie	50
	1

Table des figures

1.1	Exemple d'utilisation du seuillage	5
1.2	Segmentation par régions	6
1.3	Segmentation par contours	6
1.4	Segmentation sémantique	7
1.5	Segmentation par instance	8
1.6	Schéma générale de l'apprentissage automatique	9
1.7	Dinstinction du deep learning	10
1.8	Architecture d'un CNN	12
1.9	Schèma général du transfert learning	13
1.10	Réseau de neurones artificiels	14
1.11	Réseau de neurones convolutifs	15
1.12	Réseau de neurones récurrents	16
1.13	Arbre de décision	16
1.14	Ensemble learning	17
1.15	Exeples de variétés de pommes	18
2.1	Schéma général de la méthode proposée par Li et al. [34].	24
2.2	Étapes générales de l'algorithme de segmentation du fruit de pomme de Zhang et al. [17].	25
2.3	Architecture matérielle du système de surveillance à distance de la croissance des pommes [59].	27
2.4	Architecture du réseau de neurones profond de détection de bords de pommes [59].	28
2.5	Exemples de segmentation d'une image prise en un jour ensoleillé [27].	29
2.6	Exemples de segmentation d'une image prise en un jour nuageux [27].	29
2.7	Exemples de segmentation d'une image prise en nuit [27].	29
3.1	Comparaison entre les versions YOLO [22]	34
3.2	Architecture YOLO [54]	36
3.3	Exemples de détection par YOLOv8 par défaut	37
3.4	Segmentation par default par le modèle SAM	38

3.5	Exemples de segmentation de SAM par défaut	38
3.6	Chaine de traitement globale de YOLOv8+Sam	39
3.7	Quelques images dans acfr-multifruit.	41
3.8	Processus d'annotation et de génération du dataset avec Roboflow	43
3.9	Résultats de la segmentation en appliquant YOLOv8+SAM	44
3.10	Matrice de confusion	45

Liste des tableaux

2.1	Comparaison des méthodes de segmentation d'images étudiées.	30
3.1	Liste de projets ayant combinés SAM avec d'autres modèles[63]	40
3.2	Résultats d'évaluation	46

Liste des abreviation

RCNN :	Réseaux de Neurones Convolutifs (Convolutional Neural Networks)
UNet :	U-Net - C'est une architecture de réseau neuronal convolutif utilisée pour la segmentation d'images biomédicales.
YOLO :	You Only Look Once - C'est une méthode de détection d'objets en temps réel.
SAM :	Segment Anything Model -Un modèle développé par Meta AI Research pour effectuer la segmentation d'objets.
Dataset :	ensemble de données
Epochs :	époques - Il s'agit d'une mesure de la durée de l'entraînement d'un modèle d'apprentissage automatique.
YOLOv8+SAM :	Combinaison du modèle YOLO et de l'approche SAM - Il s'agit de l'approche développée dans le mémoire.
AI :	Intelligence Artificielle (Artificial Intelligence) .
ML :	Machine Learning (Apprentissage automatique).
RNA :	Réseaux de neurones artificiels (Artificial Neural Networks).
RNC :	Réseaux de neurones convolutifs (Convolutional Neural Networks) .
RNR :	Réseaux de neurones récurrents (Recurrent Neural Networks) .
RNN :	Recurrent Neural Networkss (Réseaux de neurones récurrents).
CNN :	Convolutional Neural Network (Réseaux de neurones convolutifs).
ROC :	Courbe ROC (Receiver Operating Characteristic).- Une courbe graphique qui illustre la performance d'un modèle.
FAQ :	Frequently Asked Questions(Questions fréquemment posées)
CIDD :	Commission internationale du droit de la famille.
N/A :	Non applicable ou non disponible.
ImageNet :	Une vaste base de données d'images.
PyTorch :	PyTorch est une bibliothèque populaire d'apprentissage automatique .
SGD :	Stochastic Gradient Descent (descente de gradient stochastique) - un algorithme d'optimisation largement utilisé pour l'entraînement des réseaux de neurones.

GLCM :	Grey Level Co-occurrence Matrix (matrice de co-occurrence des niveaux de gris).
FCOS :	Fully Convolutional One-Stage - un modèle de détection d'objets.
FPN :	Feature Pyramid Network (réseau pyramidal de caractéristiques) - un modèle de détection d'objets.
RFPN :	Residual Feature Pyramid Network - une extension du FPN.
Mask R-CNN :	Mask Region Convolutional Neural Network (réseau neuronal convolutif régional avec masque) -une méthode de détection d'objets .
FCN :	Fully Convolutional Network (réseau neuronal entièrement convolutif) -une architecture de réseau utilisée pour la segmentation sémantique.
FC :	Fully Connected (Entièrement Connecté)
IoU :	Intersection over Union (Intersection sur Union)
FCF :	Fusion of Convolutional Features (Fusion des Caractéristiques Convolutionnelles)
RGB :	Red, Green, Blue (Rouge, Vert, Bleu)
Ncut :	Normalized Cut (Coupe Normalisée)
FCOS :	(Fully Convolutional One-Stage Object Detection) - un modèle de détection d'objets.
RFPN :	(Region Proposal Feature Pyramid Network) - un module utilisé pour la détection d'objets
ResNet :	(Residual Network) - une architecture utilisé dans l'apprentissage profond.
DenseNet :	une architecture de réseau de neurones convolutifs dense.
k-means :	un algorithme de clustering.
IPS :	Images per second
CVPR :	Computer Vision and Pattern Recognition - Une conférence dans le domaine de la vision par ordinateur .
GPT-3 :	Generative Pre-trained Transformer 3 - Un modèle de langage pré-entraîné développé par OpenAI.
YAML :	YAML Ain't Markup Language - Un format de données utilisé pour représenter des structures de données .
SA :	Semantic Segment Anything -une méthode de segmentation sémantique d'objets en utilisant le modèle SAM.
MOTRv2 :	Multi-Object Tracking and Recognition version 2 -un modèle de suivi d'objets dans des vidéos.
IEA :	Image Editing Anything - une méthode d'édition d'images .
3D-Box :	Détection de boîtes 3D .
RAM :	Random Access Memory.
VP :	Vrais positifs
VN :	vrais négatifs
FP :	Faux positifs
FN :	Faux négatifs

Introduction générale

L'agriculture occupe une place primordiale dans la satisfaction des besoins alimentaires mondiaux. Afin de favoriser le développement de cette industrie cruciale, il est impératif de mettre en place des méthodes efficaces pour optimiser les processus de production et de surveillance des cultures. Dans ce contexte, la segmentation d'images se révèle être un outil puissant permettant de détecter et de délimiter avec précision les objets d'intérêt au sein d'une image. Ce mémoire de fin d'études se focalise spécifiquement sur l'application de la segmentation d'images dans le domaine de l'agriculture, en particulier dans la segmentation des images de pommes.

L'industrie de la pomme joue un rôle économique crucial au sein du secteur agricole. Néanmoins, les processus de classification et de surveillance des pommes peuvent s'avérer complexes et exigent une segmentation précise des images. L'objectif de cette étude consiste donc à proposer une méthode de segmentation d'images efficace afin d'optimiser ces processus et de contribuer au développement de l'industrie de la pomme.

Alors comment développer une méthode de segmentation d'images performante et adaptée aux spécificités des images de pommes afin d'optimiser les processus de classification et de surveillance dans l'industrie de la pomme ?

Cette étude vise à atteindre plusieurs objectifs. Tout d'abord, elle cherche à présenter les bases théoriques nécessaires à la compréhension de la segmentation d'images, de la détection d'objets et de la vision par ordinateur, ainsi qu'à souligner l'importance économique des pommes dans le secteur agricole. Ensuite, elle se donne pour mission de réaliser un état de l'art approfondi des méthodes de segmentation d'images appliquées spécifiquement aux images de pommes, en évaluant les avantages et les limites de chaque approche. De plus, cette étude ambitionne de proposer une méthode novatrice de segmentation d'images, basée sur la détection des pommes et la segmentation, en détaillant sa mise en œuvre et en exposant les résultats obtenus lors des expérimentations. Enfin, elle contribuera à l'amélioration des processus de classification et de surveillance des pommes dans l'industrie agricole en fournissant une méthode performante et adaptée.

La méthodologie adoptée pour cette étude comprendra les étapes suivantes. Tout d'abord, une collecte exhaustive des bases théoriques sera effectuée, englobant la segmentation d'images, la détection d'objets, la vision par ordinateur, ainsi que les particularités spécifiques aux pommes dans l'industrie agricole. Ensuite, un état de l'art sera réalisé afin d'examiner les méthodes existantes de segmentation d'images appliquées aux images de pommes, en éva-

luant soigneusement leurs avantages et leurs limites respectives. Par la suite, une méthode novatrice de segmentation d'images sera conceptualisée et mise en œuvre en se basant sur la détection des pommes et la segmentation, en utilisant l'approche "YOLOv8+SAM". Pour évaluer la performance de cette méthode, des expérimentations seront menées sur un ensemble de données d'images de pommes, en comparant les résultats obtenus avec d'autres méthodes existantes. Enfin, une analyse approfondie et une interprétation des résultats seront effectuées pour mettre en lumière les avantages et les limites de la méthode proposée, et des perspectives d'amélioration futures seront envisagées.

Ce mémoire est structuré en trois chapitres principaux :

Chapitre 1 : Fondements théoriques

Introduction à la segmentation d'images, à la détection d'objets et à la vision par ordinateur, et aussi de généralités sur les pommes et leur importance économique dans l'industrie agricole.

Chapitre 2 : État de l'art

Revue approfondie des méthodes de segmentation d'images appliquées aux images de pommes. Suivie d'une analyse comparative des avantages et des limites de chaque méthode.

Chapitre 3 : Méthode proposée

Présentation détaillée de la méthode novatrice "YOLOv8+SAM" basée sur la détection des pommes et la segmentation. Puis une description des étapes de mise en œuvre et des résultats expérimentaux, et finalement des perspectives d'amélioration et de développement ultérieur. Cette étude vise à apporter une contribution significative à la recherche existante dans le domaine de la segmentation d'images appliquée à l'agriculture, en particulier dans le domaine de la segmentation des images de pommes. Les résultats anticipés de cette recherche sont une méthode novatrice et performante de segmentation d'images, adaptée aux spécificités des images de pommes. Ces résultats contribueront à l'amélioration des processus de classification et de surveillance des pommes, favorisant ainsi le développement durable de l'industrie de la pomme.

Les limitations potentielles de cette recherche peuvent inclure des contraintes liées aux ressources disponibles, telles que la taille de l'ensemble de données utilisé pour les expérimentations. De plus, les résultats obtenus peuvent être influencés par des facteurs externes tels que les conditions d'éclairage ou la qualité des images.

Des perspectives futures peuvent inclure l'exploration de variantes de la méthode proposée, l'extension de l'étude à d'autres types de fruits ou cultures, ainsi que l'intégration de techniques d'apprentissage automatique avancées pour améliorer les performances de la segmentation d'images.

Ce mémoire met en évidence l'importance de la segmentation d'images dans le domaine de l'agriculture, en particulier dans la segmentation des images de pommes. En proposant une méthode novatrice et performante, cette recherche vise à optimiser les processus de classification et de surveillance des pommes, contribuant ainsi au développement durable de l'industrie de la pomme. Les chapitres suivants approfondiront les bases théoriques, l'état de l'art, et la méthode proposée, en fournissant des informations détaillées et des résultats significatifs.

Généralités et description du contexte de travail

1.1 Introduction

Avant d'entamer notre étude, il est primordial de rappeler les définitions de base concernant le domaine de notre recherche. Dans ce chapitre, nous aborderons les notions fondamentales liées à la segmentation d'images, à l'apprentissage automatique et à la vision par ordinateur. Nous présenterons également les généralités sur les pommes, qui constituent l'objet de notre étude.

1.2 Généralités sur la segmentation d'images

La segmentation d'images est une tâche fondamentale dans le domaine du traitement d'images et de la vision par ordinateur. Elle consiste à diviser une image en plusieurs régions ou objets, afin de faciliter leur analyse et leur compréhension. La segmentation est une étape préliminaire essentielle pour de nombreuses applications, telles que la détection d'objets, la reconnaissance de formes, la surveillance vidéo, la réalité augmentée, la médecine, etc. Cette section fournira une vue d'ensemble des concepts et des méthodes couramment utilisées pour la segmentation d'images.

1.2.1 Techniques de segmentation d'images

La segmentation d'images peut être réalisée à l'aide de diverses techniques, chacune ayant ses propres avantages et limitations. Parmi les méthodes les plus couramment utilisées, on trouve les méthodes de seuillage, la segmentation par régions, la segmentation sémantique et la segmentation par instance.

1.2.1.1 Méthodes de seuillage

Les méthodes de seuillage sont parmi les techniques les plus simples et les plus intuitives pour segmenter une image. Elles reposent sur le choix d'un seuil pour diviser les pixels en deux groupes distincts : ceux dont l'intensité est supérieure (ou inférieure) à ce seuil. Le seuillage peut être effectué de manière globale, où un seul seuil est appliqué sur

toute l'image, ou de manière adaptative, où différents seuils sont utilisés localement pour tenir compte des variations d'éclairage ou de contraste. Le seuillage peut également être basé sur l'histogramme de l'image pour déterminer automatiquement les seuils optimaux [15].

Parmi les méthodes de seuillage les plus utilisées, on trouve le seuillage simple, le seuillage à hystérésis, le seuillage par région de croissance, le seuillage basé sur l'histogramme, etc. Chaque méthode a ses propres critères de sélection du seuil et convient à des scénarios spécifiques. Par exemple, le seuillage à hystérésis est couramment utilisé pour détecter les contours d'objets dans une image en utilisant des seuils haut et bas pour délimiter les pixels d'intérêt.

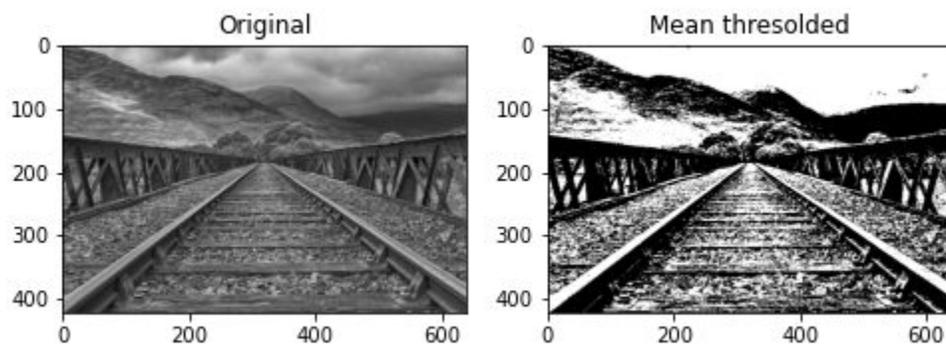


FIGURE 1.1 – Exemple d'utilisation du seuillage

La figure 1.1 illustre un exemple d'utilisation de la technique du seuillage.

1.2.1.2 Segmentation par régions

La segmentation par régions vise à regrouper les pixels en régions homogènes selon des critères prédéfinis. Ces critères peuvent être basés sur des caractéristiques telles que l'intensité, la couleur, la texture ou la cohérence spatiale. L'idée est de créer des régions contiguës qui partagent des propriétés similaires à l'intérieur d'elles-mêmes tout en étant différentes des autres régions.

Il existe différentes approches pour la segmentation par régions, notamment la croissance de régions par seuillage, les méthodes de segmentation basées sur les graphes, la classification ascendante hiérarchique (CAH), etc. La croissance de régions par seuillage consiste à sélectionner un pixel initial (graine) et à agrandir la région en ajoutant des pixels voisins qui satisfont à un critère de similarité prédéfini. Les méthodes basées sur les graphes utilisent des représentations graphiques pour modéliser les relations entre les pixels et effectuer une partition en utilisant des algorithmes de regroupement. [30]

La figure 1.2 illustre un exemple de segmentation par régions.

1.2.1.3 Segmentation par contours

La segmentation par contours vise à détecter les frontières ou les contours entre les objets présents dans l'image. Les contours fournissent des informations cruciales sur la forme, la structure et la disposition des objets. Les techniques de segmentation par contours utilisent des opérateurs de détection de contours pour identifier les discontinuités d'intensité, de

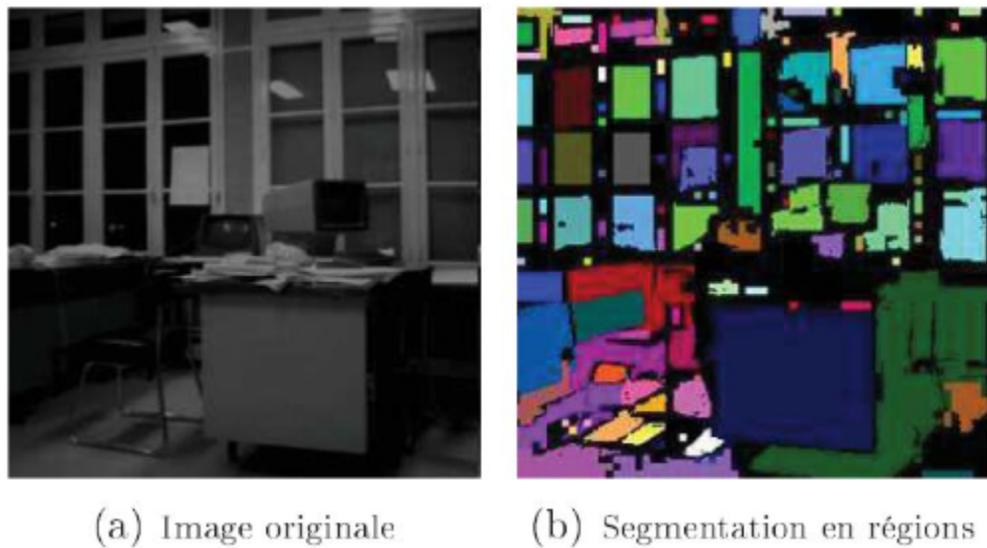


FIGURE 1.2 – Segmentation par régions

texture ou de gradient dans l'image [30].

Parmi les opérateurs de détection de contours couramment utilisés, on trouve le filtre de Canny, la transformée de Hough, le laplacien du gaussien, le gradient de Sobel, etc. Le filtre de Canny est largement utilisé pour sa capacité à détecter avec précision les contours tout en réduisant les réponses parasites. La transformée de Hough est utilisée pour détecter des formes géométriques spécifiques, telles que des lignes ou des cercles, dans une image [15].

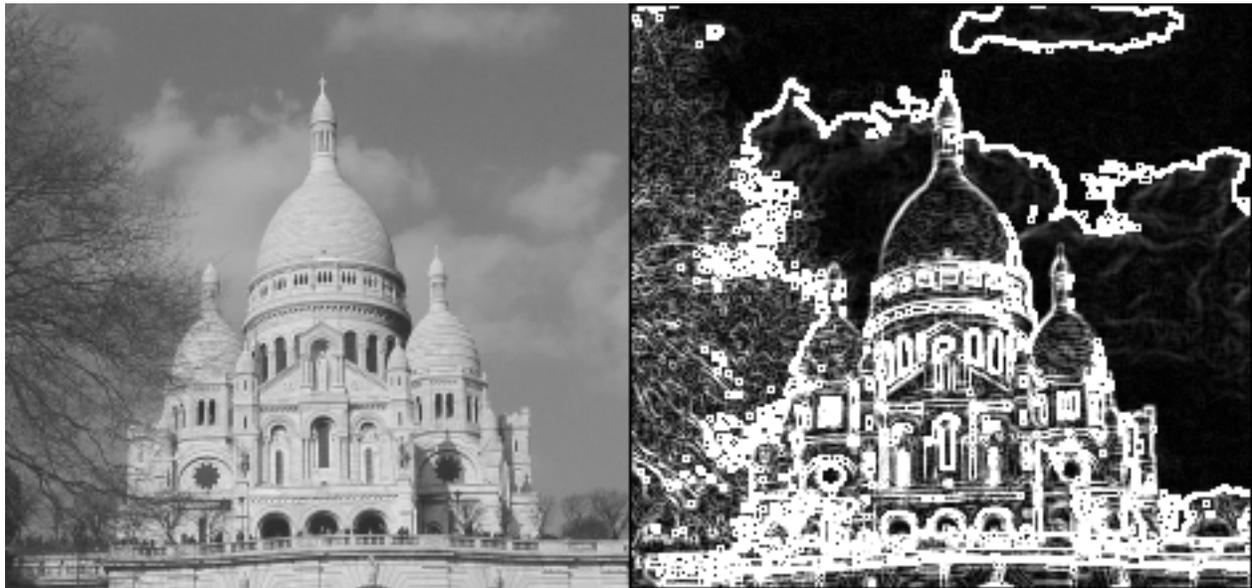


FIGURE 1.3 – Segmentation par contours

La figure 1.3 illustre un exemple de segmentation par contours.

1.2.1.4 Segmentation sémantique

La segmentation sémantique vise à attribuer des étiquettes sémantiques à chaque pixel de l'image, en regroupant les pixels en fonction de leur signification sémantique. Contrairement à la segmentation par régions, qui se concentre sur la similarité des pixels en termes de propriétés basiques telles que l'intensité ou la couleur, la segmentation sémantique vise à comprendre le contenu de l'image en identifiant et en étiquetant les différentes classes d'objets présentes [29].

Les méthodes de segmentation sémantique combinent souvent des modèles d'apprentissage automatique, tels que les réseaux de neurones convolutifs (CNN), avec des informations contextuelles et des connaissances a priori pour attribuer des étiquettes aux pixels. Ces modèles sont généralement entraînés sur de grandes bases de données annotées pour apprendre à reconnaître les objets et les structures sémantiques dans l'image [29].



FIGURE 1.4 – Segmentation sémantique

La figure 1.4 illustre un exemple de segmentation sémantique.

1.2.1.5 Segmentation par instance

La segmentation par instance consiste à segmenter les objets dans une image en attribuant des étiquettes distinctes à chaque instance d'objet plutôt qu'à des classes sémantiques générales. Cette technique permet d'identifier et de différencier les différentes occurrences des mêmes objets dans l'image.

Les méthodes de segmentation par instance combinent généralement des approches de détection d'objets avec des algorithmes de segmentation. Elles peuvent utiliser des modèles basés sur les régions de propositions (region proposal-based models) ou des architectures entièrement convolutionnelles. Ces méthodes permettent d'obtenir des masques précis pour chaque instance d'objet dans l'image [31].

La figure 1.5 illustre la segmentation par instance.

1.2.1.6 Méthodes de segmentation basées sur l'apprentissage profond

Les méthodes de segmentation basées sur l'apprentissage profond utilisent des réseaux de neurones profonds pour effectuer la segmentation d'images. Ces réseaux, tels que les réseaux de neurones convolutifs (CNN) et les réseaux de neurones entièrement convolutifs

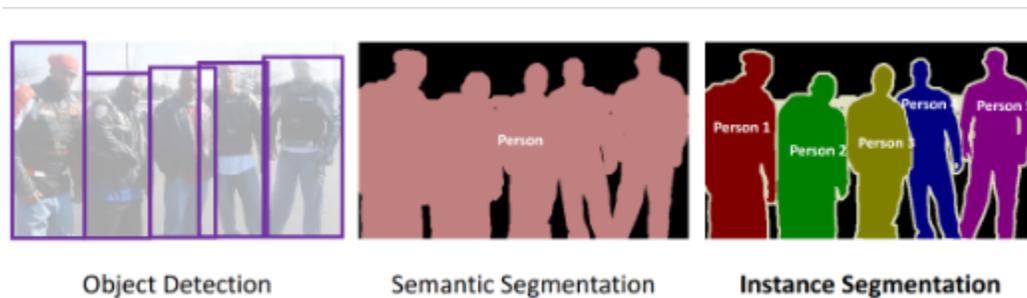


FIGURE 1.5 – Segmentation par instance

(FCN), sont capables d’apprendre des représentations de haut niveau à partir des données d’entraînement. Ils peuvent être utilisés pour effectuer à la fois la segmentation sémantique et la segmentation par instance en exploitant la capacité des réseaux à extraire des informations contextuelles et à localiser les objets dans une image [43].

La segmentation d’images est une étape cruciale dans le traitement d’images et la vision par ordinateur. Dans cette section, nous avons présenté un aperçu des techniques de segmentation d’images, notamment le seuillage, la segmentation par régions, la segmentation sémantique et la segmentation par instance. Ces méthodes constituent les fondements de la segmentation d’images et sont utilisées dans de nombreux domaines d’application. La sélection de la méthode de segmentation appropriée dépend des caractéristiques de l’image, du contexte de l’application et des objectifs spécifiques de l’analyse.

1.3 Généralités sur l’apprentissage automatique

L’apprentissage automatique (ou machine learning) est une branche de l’intelligence artificielle qui permet à un système informatique d’apprendre à partir de données, sans être explicitement programmé. Les algorithmes d’apprentissage automatique permettent à un système de trouver des modèles et des relations dans les données d’entraînement, et de les utiliser pour faire des prédictions ou des classifications sur de nouvelles données.[26]

1.3.1 Types d’apprentissage automatique

Il existe plusieurs types d’apprentissage automatique, chacun ayant ses propres avantages et limites. L’apprentissage **supervisé** consiste à apprendre à partir d’exemples étiquetés, où chaque exemple est associé à une réponse connue. L’apprentissage **non supervisé** consiste à trouver des structures et des modèles dans les données non étiquetées. Enfin, l’**apprentissage par renforcement** est une technique où le système apprend par essais et erreurs, en interagissant avec un environnement dynamique.[56]

La figure 1.6 illustre le schéma générale de l’apprentissage automatique avec tout ses types.

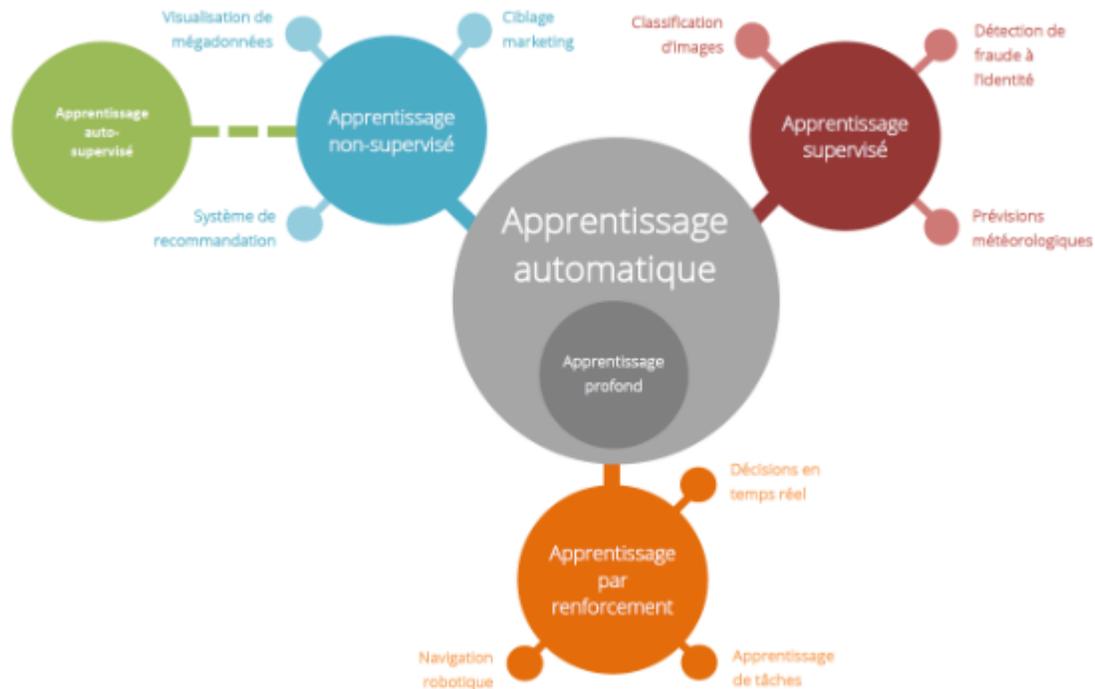


FIGURE 1.6 – Schéma générale de l'apprentissage automatique

1.3.2 Applications de l'apprentissage automatique

L'apprentissage automatique est utilisé dans de nombreux domaines, tels que la reconnaissance d'images, le traitement du langage naturel, la recommandation de produits, la détection de fraudes financières, et la prédiction des tendances du marché. Les algorithmes d'apprentissage automatique sont également utilisés dans des applications plus larges telles que la reconnaissance de malwares et la prédiction des catastrophes naturelles[41].

1.3.3 Modèles d'apprentissage automatique

Les modèles d'apprentissage automatique sont des représentations mathématiques des relations et des modèles dans les données. Les modèles d'apprentissage supervisé incluent les modèles linéaires, les arbres de décision, les réseaux de neurones, et les modèles à support de vecteurs. Les modèles d'apprentissage non supervisé incluent les algorithmes de clustering et les algorithmes de réduction de dimensionnalité.

1.3.4 Deep Learning

Le Deep Learning, également connu sous le nom d'apprentissage profond, est une branche de l'intelligence artificielle qui se concentre sur l'apprentissage automatique de modèles de données en utilisant des réseaux de neurones artificiels profonds. Ces réseaux de neurones, inspirés du fonctionnement du cerveau humain, sont capables d'apprendre et d'extraire des représentations hiérarchiques complexes à partir des données.

L'un des principaux avantages du Deep Learning est sa capacité à automatiser l'apprentissage à partir de données brutes, éliminant ainsi le besoin de spécifier explicitement les

caractéristiques à extraire. Les modèles de Deep Learning sont capables d'apprendre à reconnaître des motifs et des structures complexes directement à partir des données, ce qui les rend particulièrement adaptés à des tâches telles que la vision par ordinateur, le traitement du langage naturel, la reconnaissance de la parole et bien d'autres.

Les réseaux de neurones profonds sont généralement composés de multiples couches, permettant ainsi la représentation de niveaux d'abstraction croissants. Les couches d'entrée du réseau reçoivent les données brutes, telles que des images pixelisées, tandis que les couches cachées successives apprennent à extraire des caractéristiques de plus en plus complexes à partir de ces données. Enfin, la couche de sortie produit les prédictions ou les résultats souhaités.

L'apprentissage du réseau se fait par rétropropagation, un algorithme qui ajuste les poids des connexions entre les neurones afin de minimiser l'erreur entre les prédictions du réseau et les étiquettes réelles des données d'entraînement. Ce processus itératif permet au réseau de s'améliorer progressivement en ajustant ses paramètres pour obtenir des prédictions plus précises.

Le Deep Learning a révolutionné de nombreux domaines, notamment la reconnaissance d'objets, la détection d'objets, la traduction automatique, la génération de texte et bien d'autres. Grâce à sa capacité à traiter de grandes quantités de données et à extraire des modèles complexes, le Deep Learning a ouvert de nouvelles perspectives dans la résolution de problèmes difficiles et a contribué à des avancées significatives dans de nombreux domaines de recherche et d'application.

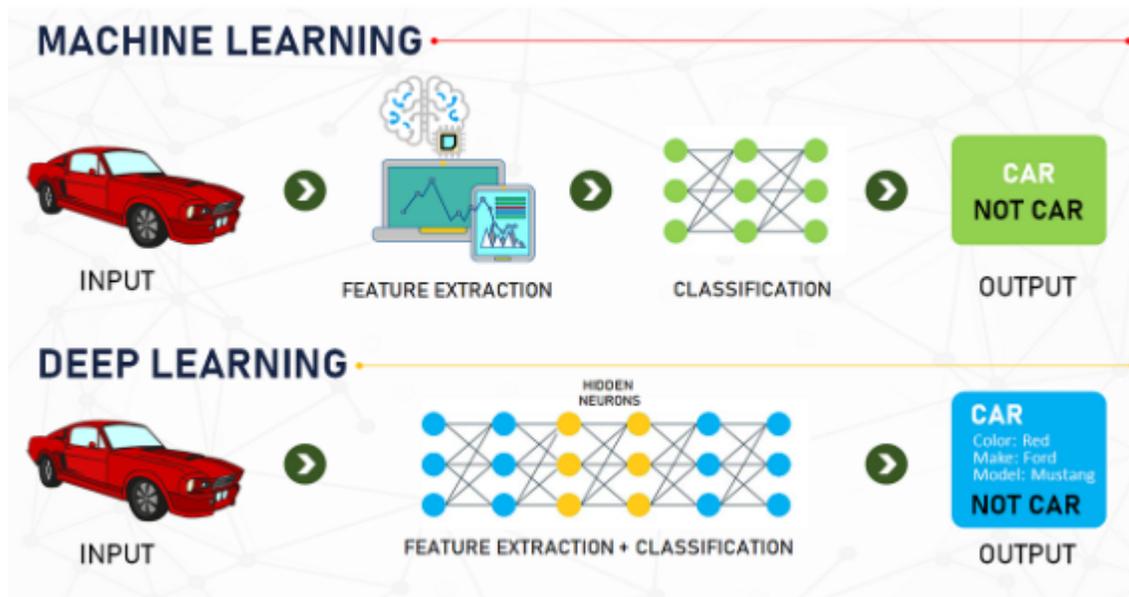


FIGURE 1.7 – Distinction du deep learning

La figure 1.7 illustre une brève comparaison du deep learning et du machine learning pour mieux appréhender son apport.

1.3.5 Réseaux de neurones convolutifs CNN

Les CNN, ou Convolutional Neural Networks (réseaux de neurones convolutifs), sont une architecture spécifique de réseaux de neurones profonds largement utilisée dans le domaine de la vision par ordinateur. Ils ont révolutionné la façon dont nous traitons les données visuelles

et ont obtenu des performances remarquables dans des tâches telles que la classification d'images, la détection d'objets, la segmentation d'images, etc.

L'une des principales caractéristiques des CNN est leur capacité à capturer et à exploiter les motifs spatiaux présents dans les données visuelles. Contrairement aux réseaux de neurones traditionnels, qui traitent les données de manière séquentielle, les CNN utilisent des couches de convolution pour effectuer des opérations de filtrage sur les images d'entrée. Ces filtres apprennent à extraire des caractéristiques pertinentes, telles que des bords, des textures ou des formes, à différentes échelles et orientations.

La structure en couches des CNN permet également une représentation hiérarchique des caractéristiques. Les premières couches du réseau captent des motifs de bas niveau, tels que des bords ou des coins, tandis que les couches supérieures combinent ces motifs pour former des représentations de plus haut niveau, telles que des formes complexes ou des objets.

L'apprentissage des CNN se fait généralement par rétropropagation, où les poids des connexions sont ajustés pour minimiser l'erreur entre les prédictions du réseau et les étiquettes réelles des données d'entraînement. Les CNN sont souvent entraînés sur de grands ensembles de données annotées, tels que ImageNet, qui contient des millions d'images réparties sur des milliers de catégories.

Grâce à leur capacité à apprendre des caractéristiques discriminantes directement à partir des données, les CNN ont surpassé les méthodes traditionnelles de vision par ordinateur dans de nombreuses tâches. Leur architecture convient particulièrement bien à l'analyse d'images, car ils peuvent détecter automatiquement les motifs et les structures pertinents sans nécessiter de spécification manuelle des caractéristiques.

Les CNN ont été largement adoptés dans divers domaines, notamment la reconnaissance faciale, la conduite autonome, la médecine, la surveillance vidéo, etc. Leur capacité à extraire des informations significatives à partir d'images a ouvert de nouvelles opportunités dans l'analyse visuelle et a contribué à des avancées majeures dans de nombreux domaines d'application.

1.3.5.1 Architecture d'un CNN

L'architecture d'un réseau de neurones convolutifs (CNN) est une composante clé de l'apprentissage profond, spécifiquement conçue pour traiter efficacement des données telles que des images. Voici une description générale de l'architecture d'un CNN :

1. Couches de convolution : Les couches de convolution sont les éléments fondamentaux d'un CNN. Elles consistent en des filtres (ou des noyaux) qui sont appliqués sur l'image en effectuant des opérations de convolution. Chaque filtre extrait des caractéristiques spécifiques de l'image en détectant des motifs tels que des contours, des textures ou des formes. Les couches de convolution permettent d'apprendre des représentations hiérarchiques de l'image, en passant d'éléments de bas niveau à des éléments de plus haut niveau d'abstraction.

2. Couches de pooling : Les couches de pooling sont utilisées pour réduire la dimensionnalité des caractéristiques extraites par les couches de convolution. Elles permettent de sélectionner les éléments les plus significatifs et de réduire la taille spatiale des cartes de caractéristiques. Le pooling peut être réalisé à l'aide d'opérations telles que le max pooling, où la valeur maximale d'une région est conservée, ou le average pooling, où la moyenne des valeurs est calculée.

3. Couches entièrement connectées : Les couches entièrement connectées sont situées à la fin de l'architecture du CNN. Elles prennent les caractéristiques extraites précédemment

et les transforment en vecteurs de caractéristiques, qui sont ensuite utilisés pour effectuer des classifications ou des prédictions. Ces couches sont similaires aux couches d'un réseau de neurones traditionnel, où chaque neurone est connecté à tous les neurones de la couche précédente.

4. Couches de normalisation : Les couches de normalisation sont souvent utilisées pour accélérer et stabiliser l'apprentissage. Elles aident à normaliser les activations des couches précédentes, ce qui facilite la convergence du modèle pendant l'entraînement.

5. Fonctions d'activation : Les fonctions d'activation, telles que ReLU (Rectified Linear Unit), sont appliquées après chaque couche de convolution ou couche entièrement connectée. Elles introduisent une non-linéarité dans le modèle, permettant ainsi de capturer des relations complexes et non linéaires entre les caractéristiques extraites.

L'architecture spécifique d'un CNN peut varier en fonction de la tâche à accomplir et des contraintes du problème. Certains réseaux CNN populaires incluent AlexNet, VGGNet, GoogLeNet (Inception), ResNet, et DenseNet, parmi d'autres. Ces architectures présentent différentes configurations et niveaux de complexité pour répondre à des besoins spécifiques en termes de précision, de vitesse ou de consommation de ressources.

En résumé, l'architecture d'un CNN est conçue pour extraire des caractéristiques pertinentes à partir d'images en utilisant des couches de convolution, de pooling, et des couches entièrement connectées. Cette conception permet au modèle d'apprendre des représentations adaptées à la tâche spécifique et d'obtenir des performances optimales dans des domaines tels que la classification d'images, la détection d'objets, et la segmentation d'images.

La figure 1.8 illustre un exemple de cette architecture.

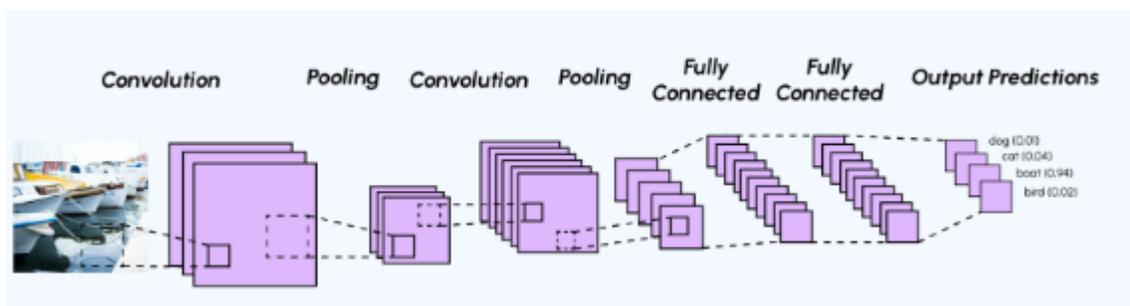


FIGURE 1.8 – Architecture d'un CNN

1.3.6 Transfert learning

Le transfert learning, transfert d'apprentissage (ou apprentissage par transfert) est une technique d'apprentissage automatique qui consiste à exploiter les connaissances acquises pour résoudre une tâche donnée et les appliquer à une tâche différente mais connexe. Plutôt que de commencer le processus d'apprentissage à partir de zéro, le transfert d'apprentissage permet aux modèles de bénéficier de modèles pré-entraînés ou des connaissances acquises lors de tâches précédentes.

Dans le transfert d'apprentissage, un modèle est initialement entraîné sur un grand ensemble de données et une tâche complexe, telle que la classification d'images ou le traitement du langage naturel. Les connaissances et les représentations apprises lors de cet entraînement

initial sont ensuite transférées à une tâche différente mais connexe. Cette approche est particulièrement utile lorsque la tâche cible dispose d'une quantité limitée de données annotées, car le modèle pré-entraîné peut fournir un avantage initial et améliorer la généralisation.

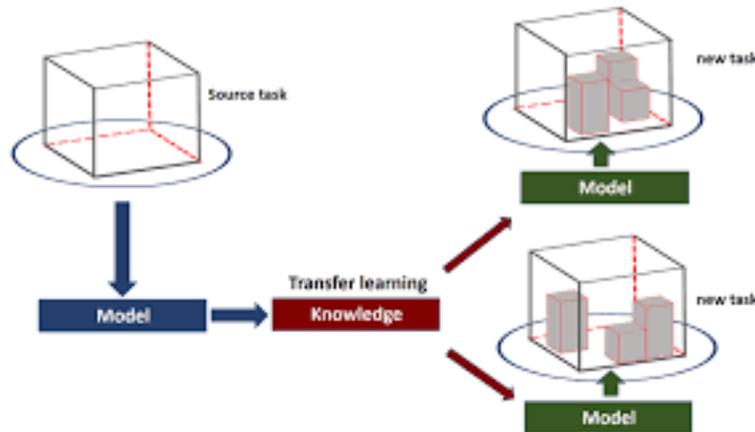


FIGURE 1.9 – Schéma général du transfert learning

La figure 1.9 illustre le passage d'un modèle pré-entraîné vers un modèle personnalisé. Les principales étapes du transfert d'apprentissage sont les suivantes :

Pré-entraînement : Un grand ensemble de données est utilisé pour entraîner un modèle sur une tâche complexe. Par exemple, un réseau neuronal convolucional (CNN) peut être entraîné sur un grand ensemble de données d'images pour la classification d'images.

Extraction de caractéristiques : Après le pré-entraînement, le modèle appris est utilisé comme extracteur de caractéristiques. Les données d'entrée pour la tâche cible sont passées à travers le modèle pré-entraîné, et les activations d'une ou plusieurs couches intermédiaires sont utilisées comme caractéristiques. Ces caractéristiques capturent les motifs généraux appris lors du pré-entraînement.

Fine-tuning (affinage) : Les caractéristiques extraites sont utilisées comme entrée pour un nouveau modèle, qui est ensuite entraîné sur la tâche cible en utilisant un ensemble de données annotées plus petit. Le nouveau modèle est généralement initialisé avec les poids pré-entraînés, et seuls les paramètres des couches finales ou de certaines couches spécifiques sont mis à jour lors de l'affinage.

En utilisant le transfert d'apprentissage, les modèles peuvent obtenir de meilleures performances et une convergence plus rapide sur la tâche cible par rapport à l'entraînement à partir de zéro, en particulier lorsque la tâche cible dispose de données limitées. Le modèle pré-entraîné agit comme une base de connaissances, capturant les motifs et les structures générales qui peuvent être utiles pour des tâches connexes.

1.3.6.1 Réseaux de neurones artificiels

Les réseaux de neurones artificiels (RNA) sont un type d'architecture de l'apprentissage automatique inspiré par le fonctionnement du cerveau humain. Les RNA sont composés de couches de neurones qui communiquent entre elles pour effectuer des tâches complexes telles que la reconnaissance d'images, la reconnaissance de la parole et la traduction automatique. Les RNA peuvent être entraînés en utilisant des algorithmes d'apprentissage tels que la

rétropropagation, qui ajustent les poids des connexions entre les neurones pour améliorer la précision des prédictions.[26]

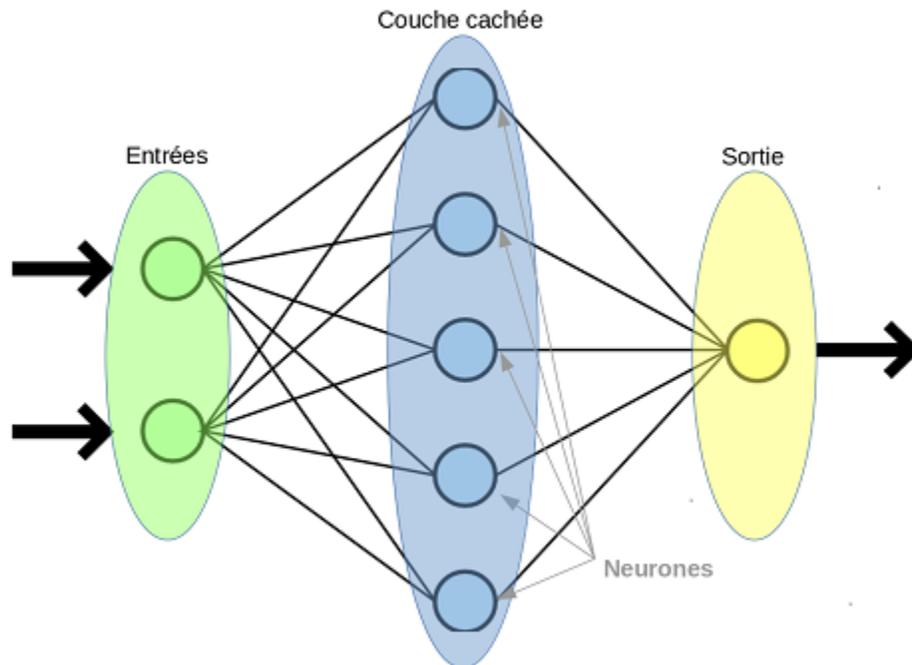


FIGURE 1.10 – Réseau de neurones artificiels

La figure 1.10 illustre une architecture simple d'un RNA.

1.3.6.2 Réseaux de neurones convolutifs

Les réseaux de neurones convolutifs (RNC) sont une architecture de l'apprentissage automatique utilisée pour la reconnaissance d'images. Les RNC sont composés de couches de neurones qui effectuent des opérations de convolution pour extraire les caractéristiques les plus importantes des images. Les RNC ont été utilisés pour des tâches telles que la reconnaissance de visages, la reconnaissance de plaques d'immatriculation et la détection d'objets dans des vidéos.[37]

La figure 1.11 illustre une architecture simple d'un RNC.

1.3.6.3 Réseaux de neurones récurrents

Les réseaux de neurones récurrents (RNR) sont une architecture de l'apprentissage automatique utilisée pour les données séquentielles telles que la reconnaissance de la parole et la traduction automatique. Les RNR sont composés de neurones qui communiquent entre eux de manière récurrente pour prendre en compte les informations précédentes dans la séquence de données[64]. Les RNR ont été utilisés pour des tâches telles que la génération de texte, la prédiction de la prochaine étape dans une séquence et la reconnaissance de l'émotion dans la parole.

La figure 1.12 illustre une architecture simple d'un RNR.

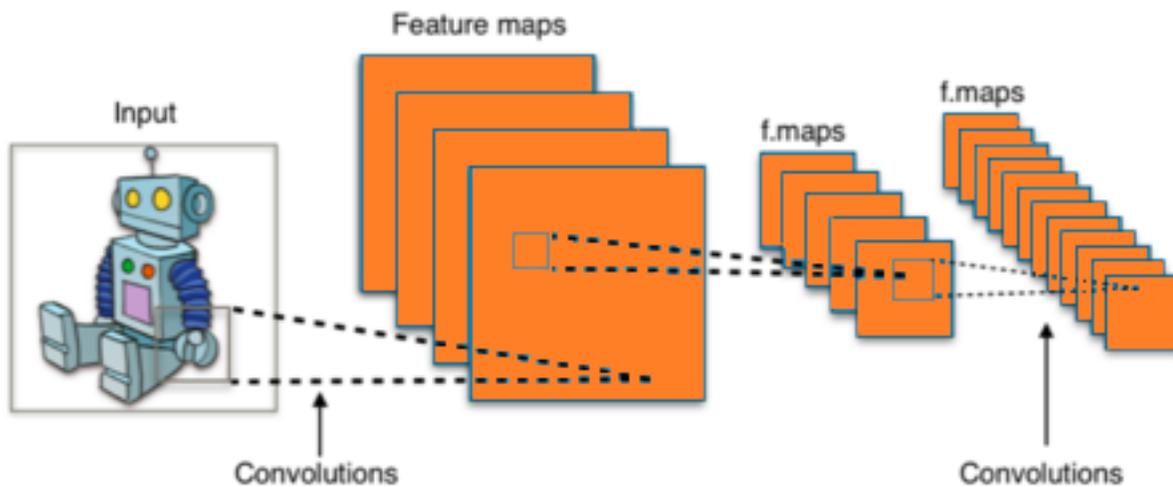


FIGURE 1.11 – Réseau de neurones convolutifs

1.3.6.4 Arbres de décision

Les arbres de décision sont une architecture de l'apprentissage automatique utilisée pour des tâches de classification et de prédiction. Les arbres de décision sont composés de nœuds et de feuilles. Les nœuds représentent des décisions à prendre en fonction des données d'entrée, tandis que les feuilles représentent les résultats finaux de la classification[57]. Les arbres de décision ont été utilisés pour des tâches telles que la détection de fraude, la prédiction de la durée de vie des machines et la prédiction des résultats des élections.

La figure 1.13 illustre un exemple d'un arbre de décision.

1.3.6.5 Ensemble learning

L'ensemble learning est une architecture de l'apprentissage automatique qui consiste à combiner plusieurs modèles d'apprentissage automatique pour améliorer la précision des prédictions[64]. Les modèles individuels sont souvent des arbres de décision, des RNA ou d'autres types de modèles. L'ensemble learning peut être utilisé pour des tâches de classification, de régression ou de clustering.

Les architectures de l'apprentissage automatique que nous avons décrites ci-dessus ne sont pas les seules qui existent, mais ce sont les plus importantes et les plus couramment utilisées dans l'industrie. Les avantages de ces architectures sont leur flexibilité, leur capacité à travailler avec de grandes quantités de données et leur aptitude à découvrir des relations complexes entre les données. Cependant, chaque architecture a ses propres avantages et inconvénients, et la sélection de l'architecture appropriée dépend de la tâche à accomplir et des données disponibles.

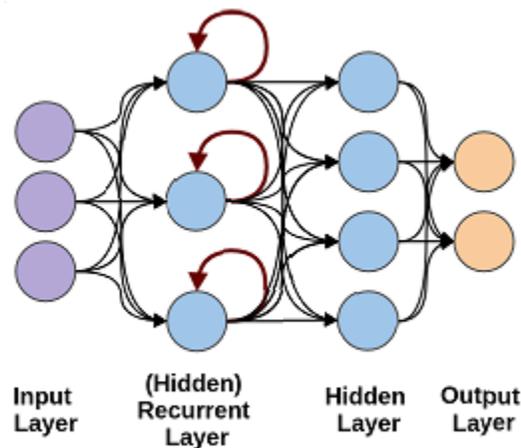


FIGURE 1.12 – Réseau de neurones récurrents

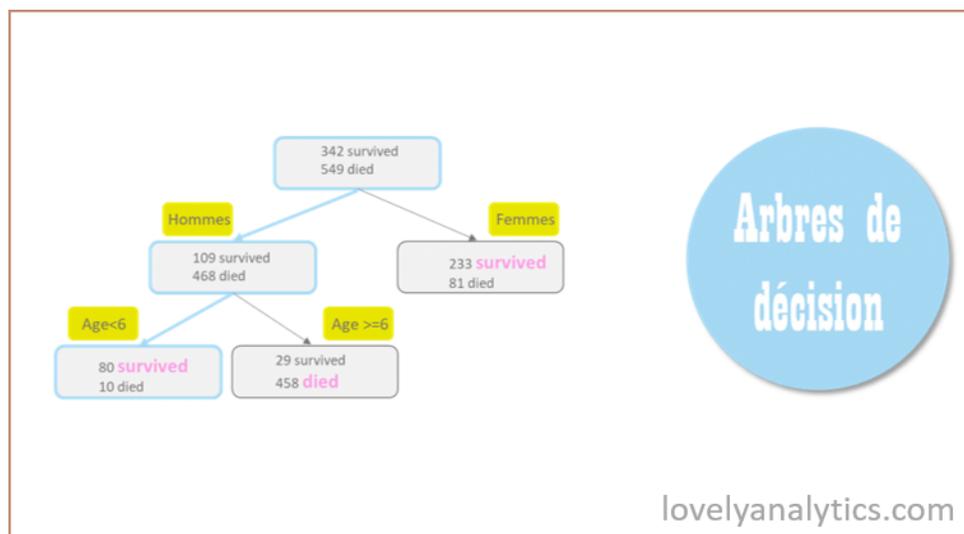


FIGURE 1.13 – Arbre de décision

En outre, il existe des architectures hybrides qui combinent plusieurs types de modèles d'apprentissage automatique pour obtenir de meilleurs résultats. Par exemple, les réseaux de neurones convolutifs peuvent être combinés avec des réseaux de neurones récurrents pour traiter des données à la fois spatiales et temporelles.

Il est important de noter que le choix de l'architecture de l'apprentissage automatique ne garantit pas toujours la précision des prédictions. Les résultats dépendent de nombreux facteurs tels que la qualité des données, la taille de l'ensemble de données, la complexité de la tâche, la capacité de calcul et la qualité de l'algorithme d'apprentissage.[64]

La figure 1.14 illustre un exemple d'architecture d'un ensemble learning.

En conclusion, l'apprentissage automatique est une discipline en constante évolution qui offre de nombreuses architectures pour résoudre une grande variété de tâches. Les architectures décrites dans cette section sont parmi les plus importantes et les plus couramment utilisées dans l'industrie. Cependant, il est important de comprendre les avantages et les inconvénients de chaque architecture et de les utiliser de manière appropriée en fonction de la tâche à accomplir et des données disponibles.

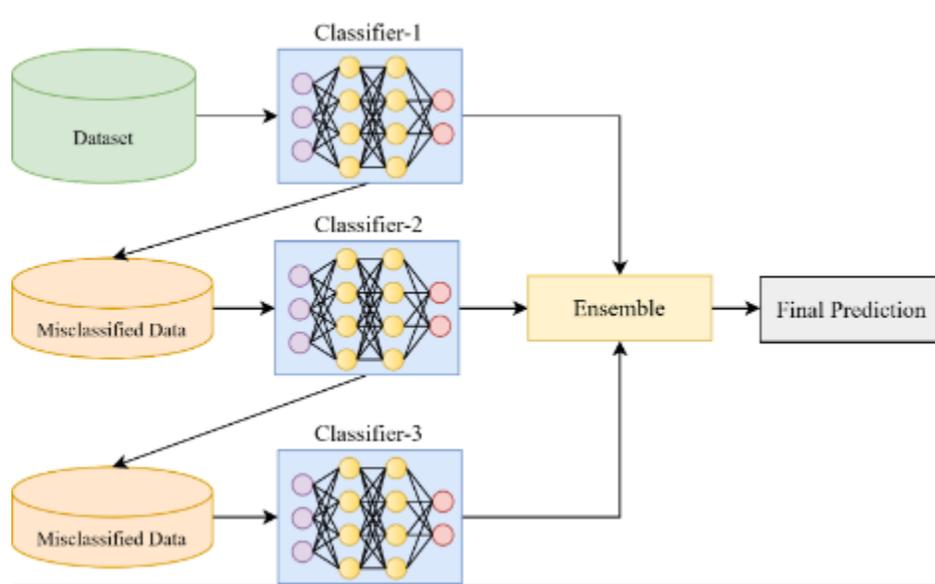


FIGURE 1.14 – Ensemble learning

1.3.7 Processus d'apprentissage automatique

Le processus d'apprentissage automatique comprend plusieurs étapes, telles que la collecte et la préparation des données, la sélection du modèle, l'entraînement du modèle, l'évaluation des performances du modèle, et l'ajustement du modèle[41]. Les performances d'un modèle sont évaluées en utilisant des mesures telles que la précision, le rappel, la courbe ROC, et la matrice de confusion.

1.3.8 Défis et limites de l'apprentissage automatique

Malgré ses nombreuses applications et avantages, l'apprentissage automatique présente également des défis et des limites. Tout d'abord, l'apprentissage automatique nécessite une quantité importante de données d'entraînement de haute qualité pour produire des résultats précis. Les données doivent également être représentatives de la population à laquelle le modèle sera appliqué, sinon le modèle peut être biaisé.[37]

De plus, les modèles d'apprentissage automatique peuvent être difficiles à interpréter, car ils peuvent être très complexes et ne pas fournir une explication claire de leurs prédictions. Cela peut être un problème dans les applications critiques telles que la médecine, où les prédictions doivent être compréhensibles et justifiables.[25]

Enfin, l'apprentissage automatique est souvent utilisé dans des applications sensibles, telles que la surveillance et la reconnaissance faciale, ce qui soulève des préoccupations concernant la vie privée et la sécurité des données. Il est important de prendre en compte ces défis et limites lors de l'utilisation de l'apprentissage automatique dans des applications réelles.[48]

1.3.9 Éthique de l'apprentissage automatique

L'éthique de l'apprentissage automatique est un domaine émergent qui étudie les implications sociales, politiques et éthiques de l'utilisation de l'apprentissage automatique. L'apprentissage automatique peut être utilisé pour des applications bénéfiques, mais il peut

également renforcer des préjugés et des discriminations, et poser des risques pour la vie privée et la sécurité.[25]

Il est important d'adopter une approche responsable et éthique lors de l'utilisation de l'apprentissage automatique, en utilisant des données représentatives et en évitant les biais et les discriminations. Il est également important de veiller à la transparence et à la responsabilité des systèmes d'apprentissage automatique, et de fournir des explications claires sur les prédictions et les décisions prises par les modèles.[64]

L'apprentissage automatique est un domaine en constante évolution avec des applications dans de nombreux domaines, y compris la médecine, l'ingénierie, la finance, et bien d'autres encore. Bien que l'apprentissage automatique ait le potentiel d'améliorer la vie de nombreuses personnes, il est important de prendre en compte les défis et les limites de cette technologie et d'adopter une approche responsable et éthique lors de son utilisation.

1.4 Généralités sur le fruit pommes

La pomme est un fruit d'une grande importance dans le monde entier en raison de sa position en tant que troisième fruit le plus produit [13]. Cela en fait un objet d'étude particulier pour les chercheurs. Selon l'Organisation des Nations unies pour l'alimentation et l'agriculture (FAO) [7], la pomme contient des composés bénéfiques pour la santé, ce qui en fait un enjeu de santé publique important. Elle est également un enjeu économique pour les producteurs, étant donné sa forte demande sur le marché et la production annuelle totale de 69 millions de tonnes dans le monde.

Les cellules constituant la pomme se distinguent par leur forme géométrique presque identique. Chaque cellule est entourée d'une paroi cellulosique, principalement composée de pectines, qui peut varier en rigidité. La composition chimique de la pomme varie en fonction de facteurs tels que la variété, les conditions climatiques, la localisation des fruits sur l'arbre et leur degré de maturité.



FIGURE 1.15 – Exemples de variétés de pommes

La figure 1.15 illustre des images de pommes prises dans des vergés, on peut distinguer trois différentes variétés, Fuji, Granny Smith et Pink Lady.

1.4.1 Histoire de la pomme

La pomme, fruit du pommier *Malus domestica*, est classée dans la famille des Rosacées et dans la sous-famille des Pomoïdées. Chargée d'histoire et de symbolisme, elle est originaire d'Asie centrale, plus précisément du Kazakhstan, où ses ancêtres poussent encore aujourd'hui. Déjà consommée en Chine il y a 3000 ans, la pomme s'est ensuite répandue chez les Arabes, les Grecs et les Romains, qui ont multiplié les variétés disponibles[8].

1.4.2 Description de la pomme

La pomme est un fruit charnu, ayant une forme presque sphérique et une pulpe uniforme. Son poids varie en fonction des variétés et des conditions de croissance. À maturité, ses couleurs peuvent varier du vert au rouge foncé, en passant par différentes nuances de vert pâle, jaune, orange ou des teintes panachées.

1.4.3 Valeurs nutritives des pommes

La pomme est un fruit dont la composition est équilibrée et variée. Elle contient principalement de l'eau (85%) et des glucides (14%) [38], ainsi que des fibres alimentaires. Les pommes sont également riches en nutriments et en composants non nutritifs, tels que les polyphénols et autres composés phytochimiques, qui possèdent une large gamme d'activités biologiques [11]. Ces composés peuvent contribuer à des effets bénéfiques sur la santé, notamment en réduisant le risque de maladies cardiovasculaires. En effet, les antioxydants présents dans les pommes aident à prévenir l'oxydation des lipides circulants dans le sang et réduisent le taux de cholestérol sanguin. Cependant, la consommation excessive de pommes ne doit pas être considérée comme un substitut à un régime alimentaire équilibré et une activité physique régulière.

Il a été démontré dans plusieurs études que la consommation régulière de pommes peut avoir des effets protecteurs contre certaines maladies telles que le cancer et d'autres maladies dégénératives associées au vieillissement, grâce à sa composition riche en vitamine C. De plus, la pomme est une source de vitamine K, qui est essentielle à la production de protéines impliquées dans la coagulation sanguine et dans la formation osseuse. La pomme est donc un fruit de composition variée et équilibrée, riche en nutriments et en composés phytochimiques, qui peuvent contribuer à des effets bénéfiques sur la santé, notamment en réduisant le risque de maladies cardiovasculaires.[6]

1.5 Conclusion

Dans ce premier chapitre, nous avons exploré le domaine de la segmentation d'images appliquée aux images de pommes dans le contexte de l'agriculture. Nous avons souligné l'importance de la segmentation d'images pour la détection et la délimitation précises des objets d'intérêt, en mettant en évidence son utilité pour optimiser les processus de production et de surveillance des cultures.

Notre étude se concentre sur la segmentation d'images de pommes, un sujet d'une grande pertinence et d'un fort intérêt pour l'industrie agricole. La problématique qui se pose est de

développer une méthode efficace de segmentation d'images spécifiquement dédiée aux images de pommes, en surmontant les limitations et les défis existants dans ce domaine.

Nous avons établi les objectifs suivants : présenter les bases théoriques de la segmentation d'images, réaliser un état de l'art des méthodes existantes, proposer une méthode novatrice basée sur la détection des pommes, et contribuer à l'amélioration des processus de classification et de surveillance des pommes. Ces objectifs nous permettront de développer une approche efficace et adaptée pour la segmentation des images de pommes.

Dans les chapitres suivants, nous présenterons en détail notre méthode novatrice de segmentation d'images pour les images de pommes, intitulée "YOLOv8+SAM". Nous détaillerons les étapes de mise en œuvre, les résultats obtenus lors de nos expérimentations, ainsi que les perspectives d'amélioration envisagées.

Dans le chapitre suivant nous présentons notre état de l'art sur différentes méthodes de ségmentations de pomme.

Etat de l'art sur les méthodes de segmentation de pommes dans des images en couleurs

2.1 Introduction

La littérature propose plusieurs méthodes de segmentation de pommes, qui dépendent du type d'image à segmenter et des objectifs visés. Ce chapitre présente un état de l'art sur la segmentation de pommes en utilisant des images en couleurs, en définissant la segmentation d'images et en mettant en évidence son intérêt et ses difficultés. Ensuite, il donne un aperçu de quelques méthodes récentes proposées dans la littérature dans le même contexte, suivi d'une synthèse et d'un tableau comparatif des méthodes présentées selon certains critères sélectionnés.

2.2 Segmentation d'images

La segmentation d'images est une tâche importante dans tout processus de traitement et analyse d'images, présente dans plusieurs applications de vision par ordinateur. Elle consiste à diviser une image en plusieurs régions ou segments distincts, chacun représentant une partie spécifique de l'image. L'étape de segmentation est obtenue par l'acquisition de l'image, son amélioration, ainsi que sa description et prise de décision finale. Les segments sont généralement définis en fonction de caractéristiques telles que la couleur, la texture, la luminosité et la forme des objets dans l'image. Plusieurs approches de segmentation ont été développées dans le but d'extraire des attributs pour caractériser des objets d'intérêt. La robustesse de chacune de ces approches dépend grandement des caractéristiques de l'image acquise.

2.2.1 Préliminaires et définition

La segmentation d'images est une opération d'analyse d'images qui vise à simplifier ou changer la représentation d'une image en une illustration plus simple et facile à analyser. Considérant son importance, elle s'intègre dans de nombreuses applications importantes telles que l'interprétation automatique d'images biomédicales et satellitaires, le suivi et l'estimation de mouvement dans une séquence vidéo, etc. Dans notre cas, nous nous intéressons à une application pratique de la segmentation d'images dans le domaine d'agriculture, qui est la

segmentation de pommes dans des images en couleurs. La segmentation d'images, telle qu'elle est définie, a pour but de regrouper des pixels entre eux suivant des critères prédéfinis pour former des régions homogènes, où chaque région est un ensemble connexe de pixels ayant des propriétés communes (intensité lumineuse, texture, etc.) qui les différencient des pixels des régions voisines, tels que :

Soient Ω une image segmentée en N régions, notées R_i , où $i = 1, 2, \dots, N$ et $L(\cdot)$ est un prédicat logique qui estime l'homogénéité d'une certaine région par rapport à un ensemble prédéfini de caractéristiques. La définition mathématique de la segmentation d'images est donnée par les quatres conditions suivantes :

$$R_i \cap R_j = \emptyset, \forall i, j \in 1, 2, \dots, N, \quad (2.1)$$

$$\Omega = \bigcup_{i=1}^N R_i, \quad (2.2)$$

$$L(R_i) = True, \forall i, j \in 1, 2, \dots, N, \quad (2.3)$$

$$L(R_i \cup R_j) = False, \forall i, j \in 1, 2, \dots, N, i \neq j. \quad (2.4)$$

Les conditions (2.1) et (2.2) impliquent que la segmentation doit être complète. Spécifiquement, la condition (2.1) indique que chaque pixel doit être affecté à une seule région et la condition (2.2) signifie que l'union de toutes les régions résultante doit former l'image originale. La condition (2.3) exige que les pixels appartenant à la même région doivent être connectés et homogènes. La condition (2.4) indique que les régions résultantes doivent être disjointes et l'union de deux régions casse la relation d'homogénéité entre les pixels.

2.2.2 Intérêt de la segmentation des pommes

La segmentation des pommes présente plusieurs intérêts dans l'industrie alimentaire, notamment pour la sélection, le tri et la classification des pommes en fonction de leur qualité et de leur état de maturité. Cela permet aux producteurs et aux distributeurs de mieux gérer leur stock et d'optimiser leur processus de production en fonction des besoins du marché. La segmentation des pommes peut également être utilisée pour améliorer la qualité des produits dérivés, tels que les jus et les compotes de pommes, en sélectionnant les fruits les plus appropriés pour chaque processus de transformation. Enfin, la segmentation d'image de la pomme peut être utilisée dans la recherche en biologie végétale pour étudier la croissance et le développement des fruits, ainsi que pour détecter et prévenir les maladies et les parasites qui peuvent affecter la production de pommes.

L'émergence des robots de cueillette autonomes, apportera des bénéfices énormes. En premier lieu, ces machines intelligentes soulageront les humains de mener des opérations répétitives, intensives et fatigantes dans des environnements difficiles, tout en prolongeant le temps de travail et la rapidité de production ainsi que la réduction de coûts. En deuxième lieu, le renforcement technologique dans le domaine d'agriculture aidera à une planification de la récolte avec une précision élevée, la protection des arbres, etc. Ces robots sont basés essentiellement sur la vision par ordinateur qui tente d'imiter le système de vision humain pour analyser, traiter et comprendre une ou plusieurs images prises par un système d'acquisition (par exemple : caméras ou capteurs) et dans laquelle la segmentation d'images joue un rôle primordial. En effet, le rôle de la segmentation est double : simplifier une image prise dans un verger en quelque chose de plus significatif et facile à comprendre par un robot et

supprimer les parties non intéressantes d'une image telles que l'arrière-plan et focaliser sur les régions d'intérêts (les pommes).

2.2.3 Challenge de la segmentation des pommes

La segmentation d'image de la pomme présente plusieurs défis techniques. Tout d'abord, les pommes peuvent varier considérablement en apparence en raison de facteurs tels que leur variété, leur état de maturité et leur exposition à la lumière. En conséquence, les algorithmes de segmentation d'image doivent être capables de traiter une grande variété de couleurs, de formes et de textures.

De plus, les images de pommes peuvent contenir des objets en arrière-plan, tels que des feuilles, des branches ou d'autres fruits. Ces éléments peuvent interférer avec la segmentation de la pomme elle-même, nécessitant des techniques de suppression de bruit et de séparation d'objets pour isoler la pomme.

Enfin, la segmentation d'image de la pomme peut être compliquée par des facteurs tels que l'éclairage et les ombres. Les algorithmes doivent être capables de reconnaître les pommes dans des conditions de faible éclairage et de distinguer les zones d'ombre des zones de pomme réelles.

Dans l'ensemble, la segmentation d'image de la pomme est un défi intéressant pour les chercheurs en vision par ordinateur et nécessite des techniques avancées pour obtenir des résultats précis et fiables.

2.3 Quelques méthodes de segmentation de pommes

Dans cette section, nous présentons quelques méthodes récentes qui ont été exhibées dans la littérature s'inscrivant dans le cadre de la segmentation de pommes.

2.3.1 A novel green apple segmentation algorithm based on ensemble U-Net under complex orchard environment

Li et al. [34] ont proposé un algorithme de segmentation de pommes basé sur l'apprentissage en profondeur. Ils ont utilisé le réseau neuronal convolutif U-Net de segmentation sémantique d'images comme architecture de réseau de base, et ont optimisé sa structure et ses paramètres pour mieux extraire les caractéristiques de l'image. Cependant, ils ont constaté que la reconnaissance des pommes vertes dans les images de fruits et de branches avec des caractéristiques de couleur de fond similaires, des occlusions de fruits, un chevauchement et diverses conditions d'éclairage était difficile. Pour remédier à ce problème, les auteurs ont ajouté et intégré une structure Edge sur la base de U-Net pour améliorer la segmentation. En outre, le schéma général pré-entraînement avec ImageNet a été utilisé pour former les modèles d'apprentissage profond de manière plus efficace et plus stable. Le modèle a été testé en utilisant un serveur de laboratoire à l'aide d'un cadre d'apprentissage en profondeur basé sur PyTorch, et la méthode de descente de gradient stochastique a été utilisée pour entraîner le réseau de manière conjointe de bout en bout. Les auteurs ont également enrichi le dataset de tests pour rendre le modèle plus adapté aux différents cas possibles, et ont apporté des améliorations progressives au modèle proposé pour améliorer sa performance.

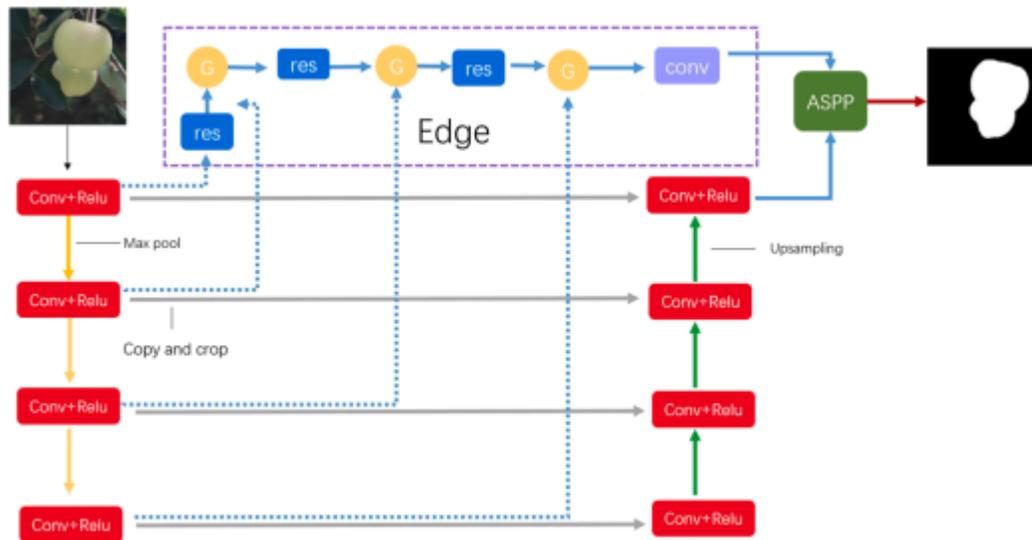


FIGURE 2.1 – Schéma général de la méthode proposée par Li et al. [34].

Validation de la méthode

Le temps de calcul de cette méthode est inférieur à celui des autres méthodes, mais les performances de ségmentation de celle-ci sont meilleures. Bien que le modèle proposé a produit des résultats relativement bons, certains de ses aspects doivent encore améliorer. Essentiellement, il faut augmenter sa vitesse sans sacrifier sa précision de la ségmentation.

2.3.2 A method of apple image segmentation based on color-texture fusion feature and machine learning

Ces dernières années, certains chercheurs ont développé de nombreuses méthodes pour segmenter les fruits de la pomme dans les images. Ils ont utilisé des caméras couleur, des caméras spectrales et des caméras thermiques pour obtenir des images de pommiers dans les vergers. Dans cet article, Zhang et al. [17] ont présenté une étude portée sur des algorithmes de ségmentation des fruits de pommes. Dans cette étude, l'effet de nombreuses caractéristiques de couleur sur la classification des pixels de fruits de pommes par rapport aux autres pixels a été évalué. Trois caractéristiques de couleur ont été sélectionnées. Ces caractéristiques permettent de distinguer efficacement les pixels de fruits de pomme des autres pixels. La matrice de co-occurrence des niveaux de gris (GLCM) a été utilisée pour extraire les caractéristiques de texture. Les meilleurs paramètres de distance et d'orientation pour la GLCM ont été trouvés. Neuf algorithmes d'apprentissage automatique ont été utilisés pour développer des classificateurs de pixels. Le classificateur a été entraîné avec 100 pixels et testé avec 100 pixels. Parallèlement, un classificateur a été utilisé pour segmenter ces images. L'analyse de régression a été effectuée sur les résultats de l'étiquetage artificiel et de la classification du classificateur. Les objectifs de cette recherche étaient donc d'évaluer et optimiser l'adéquation des caractéristiques de couleur et de texture, et de développer un classificateur de pixels basé sur l'apprentissage automatique. La Figure 2.2 exhibe les étapes essentielles de l'algorithme de Zhang et al. [17].

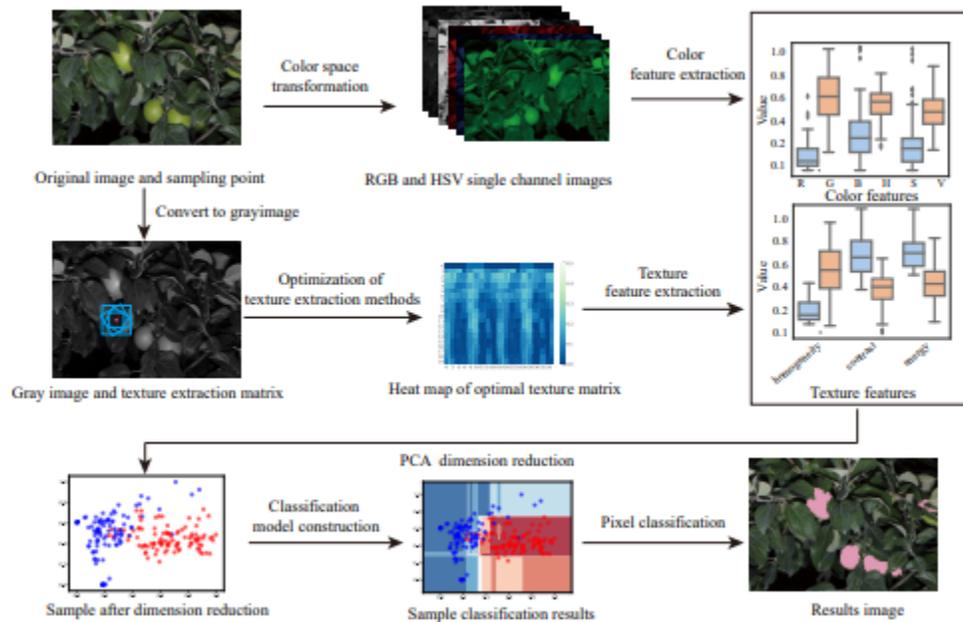


FIGURE 2.2 – Étapes générales de l’algorithme de ségmentation du fruit de pomme de Zhang et al. [17].

Validation de la méthode

Les caractéristiques de couleur permettent de distinguer efficacement les pixels de la pomme des autres. Tandis que les caractéristiques de texture n’étaient pas très performantes. En outre, l’algorithme de classification basé sur la forêt aléatoire a permis de classer efficacement les pixels des fruits de la pomme, avec une précision de 94 %.

2.3.3 An accurate detection and segmentation model of obscured green fruits

La détection et la ségmentation précises des fruits verts obscurcis est un facteur clé de l’efficacité du robot de cueillette. Dans l’environnement réel et complexe des vergers, il existe de nombreux facteurs d’influence tels que les changements d’éclairage, l’occultation des branches et des feuilles, les conditions météorologiques, etc. qui compliquent le travail des robots de cueillette. Liu et al. [36] ont proposé une méthode de détection et de ségmentation des pommes vertes obscurcies, basée sur un modèle de détection d’objets à convolution complète en une étape (FCOS). Dans leur méthode, le réseau pyramidal de caractéristiques (FPN) dans FCOS est remplacé par un réseau pyramidal de caractéristiques résiduelles (RFPN), ce qui améliore la précision de détection des fruits verts de différentes tailles. De plus, comme les limites des fruits flous ne sont souvent pas identifiées avec précision, un réseau d’attention par blocs convolutionnel à deux couches est connecté à un FCOS modifié pour récupérer les bords des fruits verts incomplets.

Les auteurs ont évalué leur méthode sur deux datasets contenant des images de pommes vertes. Essentiellement, un dataset contenant des images collectées à la base de démonstration agricole de la ville de Yantai à l’aide de l’appareil d’acquisition Sony Alpha 7 II caméra avec une résolution de 6000×4000 . Les images du dataset ont été capturées dans différentes

conditions d'éclairage, y compris les cas d'éclairage frontal, rétro-éclairé, obscurci, après la pluie et la nuit.

Validation de la méthode

Les résultats expérimentaux montrent que les précisions de détection et de segmentation des pommes vertes ont atteint 81,2 % et 85,3 % respectivement sur le dataset utilisé. Par contre, sur le deuxième dataset, appelé Apple-ape, la précision de détection et celle de segmentation ont atteint 77,2 % et 79,7 % respectivement.

2.3.4 Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot

Jia et al. [28] ont présenté un algorithme de reconnaissance de profondeur pour le système de vision du robot cueilleur de pommes. Il étend avec succès l'application de l'apprentissage profond à la direction de la cueillette des robots agricoles. Afin de mieux appliquer les bonnes performances de l'extraction de caractéristiques et de la détection de cibles utilisées dans l'apprentissage profond à la détection de fruits dans les vergers, un modèle de détecteur de vision de robot de récolte basé sur le réseau neuronal convolutionnel de région de masque (Mask R-CNN) est proposé. Le modèle a été amélioré pour le rendre plus adapté à la reconnaissance et à la segmentation des pommes qui se chevauchent. Spécifiquement, la combinaison des deux réseaux ResNet et DenseNet, qui peut réduire considérablement les paramètres d'entrée, est utilisée comme réseau de base pour l'extraction des caractéristiques. Les cartes de caractéristiques obtenues sont par la suite introduites au réseau de proposition de régions (RPN) afin de générer la région d'intérêt. Et enfin, le masque est généré par le réseau de convolution complet (FCN) afin d'obtenir la région où se trouve la pomme. La méthode est testée sur un ensemble aléatoire de 120 images.

Jia et al. [28] ont fait recours à la technique de transfert d'apprentissage pour entraîner leur algorithme. Spécifiquement, l'algorithme proposé a été tout d'abord pré-entraîné en utilisant la base ImageNet [42] en raison de manque de datasets et entraîné entièrement par la suite en utilisant des images acquises à l'aide d'une caméra d'une résolution de 6000×4000 pixels dans différentes conditions de lumière naturelle.

Validation de la méthode

Les évaluations ont abouti à un taux de précision de 97,31% et un taux de rappel de 95,70%. Les auteurs ont aussi noté la rapidité de reconnaissance de leur approche par rapports à ces méthodes concurrentes.

2.3.5 An apple image segmentation method based on a color index obtained by a genetic algorithm

Zou et al. [67] ont présenté une approche de segmentation d'images à base de seul utilisant une nouvelle technique de calcul des indices de couleurs, qui sont des caractéristiques importantes pour la segmentation d'images en couleurs. L'approche proposée a été utilisée avec succès pour segmenter des fruits de pommes dans des images en couleurs. Le point fort de la méthode de Zou et al. [67] consiste en l'utilisation de l'algorithme génétique afin optimiser les

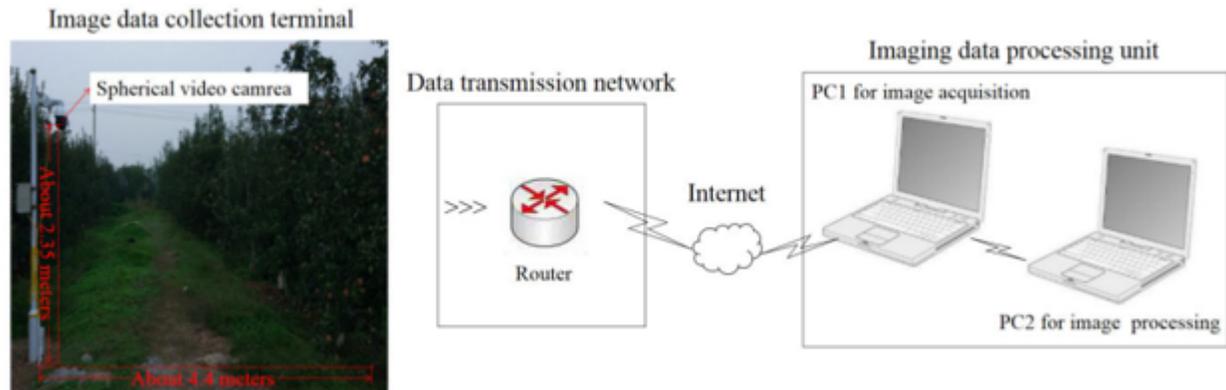


FIGURE 2.3 – Architecture matérielle du système de surveillance à distance de la croissance des pommes [59].

paramètres de leur approche. Ainsi, l'extraction de caractéristiques de couleurs pertinentes qui peuvent être utilisées dans une variété de techniques de segmentation d'images, en partant des plus simples, comme le seuillage, aux plus complexes techniques de classification de pixels de machine learning et de réseaux de neurones profonds de la segmentation sémantique.

Validation de la méthode

Les résultats expérimentaux montrent que l'approche introduite permet de segmenter efficacement et rapidement les images de pommiers dans les vergers. Spécifiquement, la méthode proposée a obtenu une efficacité moyenne de 90%, une intersection sur l'union moyenne de 81%, une moyenne du coefficient de Dice de 98% et enfin une vitesse de segmentation moyenne de 20 ms.

2.3.6 Deep learning approach for aApple edge detection to remotely monitor apple growth in orchards

L'acquisition automatique et intelligente d'informations sur la croissance des pommes à long terme est un avantage prometteur pour les producteurs, qui peuvent ainsi planifier l'application de nutriments et de pesticides pendant la maturation des fruits. Le but principal de cette étude, menée par Wang et al. [59] était de développer un système de surveillance de la croissance des pommes dans les vergers en utilisant un réseau d'apprentissage en profondeur de détection de bord. Cela permettrait une estimation à distance de la taille des pommes tout au long de leur période de croissance. Pour construire ce système, les chercheurs ont utilisé une caméra vidéo sphérique ainsi que deux ordinateurs personnels pour acquérir régulièrement des images de pommes.

En ce qui concerne le logiciel, les chercheurs ont mis au point un réseau de détection de bords qui fusionne des caractéristiques convolutionnelles (FCF). Ce réseau a été utilisé pour segmenter les images de pommes. Pour éliminer les pommes non pertinentes dans les images, des points sur les pommes à surveiller ont été sélectionnés manuellement dans les images en tant que points de semence. Ensuite, la méthode de croissance régionale a été effectuée sur ces points de semence pour suivre la croissance des pommes.

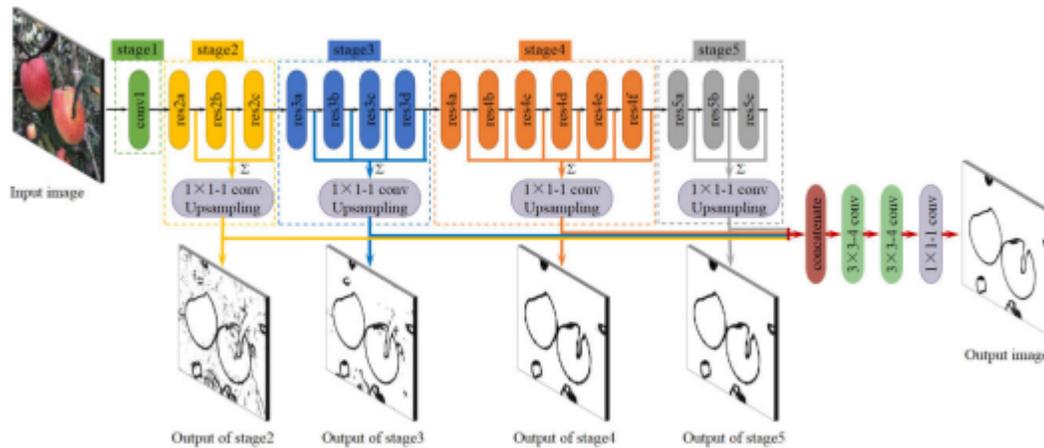


FIGURE 2.4 – Architecture du réseau de neurones profond de détection de bords de pommes [59].

Pour suivre la croissance des pommes, les chercheurs ont appliqué la méthode de croissance régionale aux cartes de bord extraites. Ils ont ensuite calculé les diamètres horizontaux des pommes. Ce système de suivi de la croissance des pommes a été utilisé depuis la date suivant l'éclaircissage jusqu'à la maturation des pommes.

Les résultats de cette étude peuvent être utilisés pour optimiser la gestion des vergers et peuvent également servir de référence pour le suivi de la croissance d'autres fruits pendant leur période de croissance. En somme, cette étude présente une méthode prometteuse pour la surveillance à distance de la croissance des fruits, qui peut être utilisée dans les applications de gestion de vergers (voir la Figure 2.3 pour illustration).

Validation de la méthode

Les résultats expérimentaux ont montré que le score F1 de la méthode FCF était de 53,1% Sur l'ensemble de test des pommes, et que le temps d'exécution moyen était de 0,075 s par image, ce qui était meilleur que les cinq autres méthodes comparées. L'erreur absolue moyenne des diamètres horizontaux des pommes détectées par ce système était de 0,90 mm, et elle a diminué de 67,9 % par rapport à la méthode basée sur l'ajustement du cercle (2,8 mm).

2.3.7 Fast segmentation of colour apple image under all-weather natural conditions for vision recognition of picking robots

Ji et al. [27] ont proposé une méthode pour la reconnaissance en temps réel des pommes du type Fuji par des robots de cueillette, et cela, en utilisant la méthode de la coupe normalisée. La méthode proposée se décompose en quatre étapes :

- Dans la première étape, la méthode du décalage moyen adaptatif (Adaptive Mean shift) est utilisée pour produire une segmentation initiale de l'image originale. Les pixels sont regroupés en régions selon leurs coordonnées spatiales, ainsi que leurs coordonnées chromatiques en RVB.
- Dans la deuxième étape, la segmentation initiale issue de la première étape est complexe, ce qui engendre la difficulté de séparation des régions de pommes des régions

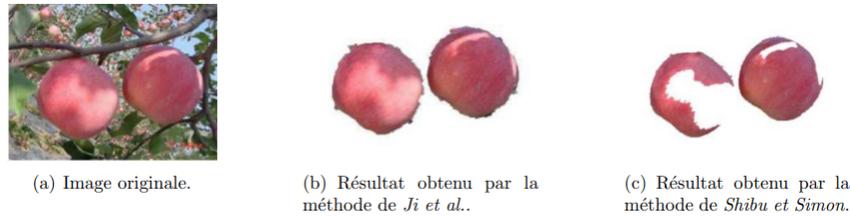


FIGURE 2.5 – Exemples de segmentation d'une image prise en un jour ensoleillé [27].

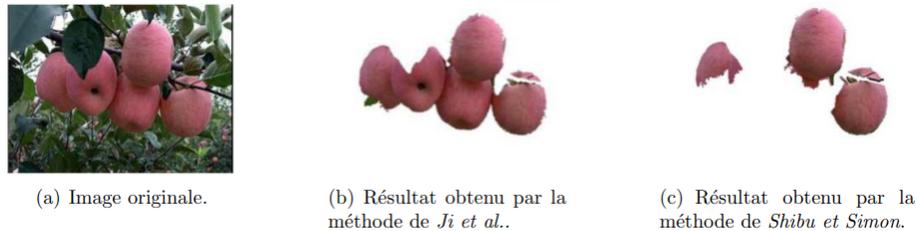


FIGURE 2.6 – Exemples de segmentation d'une image prise en un jour nuageux [27].

de l'arrière-plan de l'image segmentée. Afin de produire une image plus simple, les caractéristiques des couleurs rouge et bleue sont extraites de l'image segmentée.

- Dans la troisième étape, la segmentation initiale est raffinée par la méthode de la coupe normalisée traditionnelle (Ncut).
- Enfin, dans la dernière étape, les régions qui représentent des pommes sont extraites en utilisant un modèle de correspondance du contour entre l'image segmentée et l'image originale.

Validation de la méthode

Afin de valider la méthode proposée, les auteurs ont testé leur méthode sous différentes conditions d'éclairage, à savoir : jours ensoleillés, jours nuageux et nuit, et ont comparé ces résultats avec ceux de la méthode proposée par Shibu et Simon [49]. La Figure 2.5 illustre des résultats de segmentation par la méthode de Ji et al. [27] et la méthode de Shibu et Simon [49] d'une image prise en un jour ensoleillé. La Figure 2.6 illustre des résultats de segmentation par la méthode de Ji et al. [27] et la méthode de Shibu et Simon d'une image prise en un jour nuageux. La Figure 2.7 illustre des résultats de segmentation par la méthode de Ji et al. et la méthode de Shibu et Simon d'une image prise en nuit.

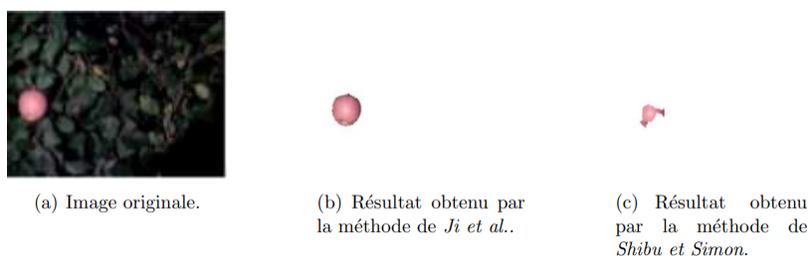


FIGURE 2.7 – Exemples de segmentation d'une image prise en nuit [27].

Les tests de validation ont montré que la méthode proposée par Ji et al. est plus résistante à l'éclairage fort et l'ombre. Aussi, elle se caractérise par un temps d'exécution acceptable.

2.4 Synthèse des travaux connexes

Dans cette section, nous présentons une synthèse des méthodes que nous avons étudiées dans la section précédente. Pour ce faire, nous établissons un tableau comparatif de ces méthodes selon quelques critères de comparaisons. Ensuite, nous présentons une discussion de la comparaison faite.

2.4.1 Tableau comparatif et discussions

Le Tableau 2.1 présente différentes méthodes utilisées pour la détection et la segmentation de pommes dans des images en couleur comparées selon quelques critères de comparaison. Spécifiquement, pour chaque étude, nous donnons le type de pomme étudié, la technique de segmentation utilisée, les caractéristiques, et l'objectif de la méthode.

Méthode	Type de pomme	Technique de segmentation	Caractéristique	Objectif
Li et al. [34] (2021)	Granny Smith	U-Net + structure edge	2D / Automatique	Extraction de pommes
Zhang et al. [17] (2020)	Golden et Fuji	Algorithmes de machine learning	2D / Automatique	Extraction de régions + Classification
Liu et al. [36] (2022)	Granny Smith	Model FCOS + RFPN	2D / Automatique	Extraction de pommes
Jia et al. [28] (2020)	Granny Smith	R-CNN + ResNet et DenseNet	2D / Automatique	Extraction de régions
Zou et al. [67] (2022)	Golden et Fuji et Granny Smith	Extraction d'indices de couleur + Seuillage	2D / Automatique	Extraction de pommes
Wang et al. [59] (2020)	Fuji	Réseau de détections de bords + FCF	2D / Automatique	Extraction de pommes et suivi de la croissance des régions
Ji et al. [27] (2016)	Granny Smith	K-means + Seuillage	2D / Automatique	Extraction de pommes

TABLEAU 2.1 – Comparaison des méthodes de segmentation d'images étudiées.

La première étude (Li et al, 2021) utilise un réseau de neurones convolutifs de type U-Net, combiné à une structure de détection de bords, pour extraire les pommes Granny Smith à partir d'images 2D.

La deuxième étude (Zhang et al, 2020) utilise un algorithme d'apprentissage automatique de classification pour extraire des régions de pommes Golden et Fuji à partir d'images 2D.

La troisième étude (Liu et al, 2021) utilise un modèle FCOS et RFPN pour extraire des pommes Granny Smith à partir d'images 2D.

La quatrième étude (Jia et al, 2020) utilise un réseau de neurones de type R-CNN combiné à des modèles ResNet et DenseNet pour extraire des régions de pommes Granny Smith à partir d'images 2D.

La cinquième étude (Zou et al, 2022) utilise une méthode d'extraction d'indices de couleur et de seuillage pour extraire des pommes Golden, Fuji et Granny Smith à partir d'images 2D.

La sixième étude (Wang et al, 2020) utilise un réseau de détection de bords combiné à une méthode de croissance de régions (FCF) pour extraire et suivre la croissance des régions de pommes Fuji à partir d'images 2D.

Enfin, la septième étude (Ji et al, 2017) utilise une méthode de clustering k-means combinée à une méthode de seuillage pour extraire des pommes Granny Smith à partir d'images 2D.

Pour récapituler cette étude comparative que nous avons menée, la majorité des méthodes récentes analysées ci-dessus sont automatiques et visent à traiter des images 2D. Et ce, avec diverses techniques de machine learning, particulièrement l'utilisation des réseaux de neurones convolutionnels avec des approches telles que U-Net, R-CNN, et FCOS, visant principalement à extraire des pommes de types Granny Smith et Fuji pour la plupart.

2.5 Conclusion

En conclusion de ce chapitre d'état de l'art, nous avons exploré activement le domaine de la segmentation des pommes dans le contexte de la vision par ordinateur. Nous avons mis en évidence l'ampleur des recherches dans ce domaine, avec de nombreuses méthodes développées pour extraire et suivre la croissance des pommes dans les vergers.

Nous avons constaté que les réseaux de neurones convolutionnels, tels que U-Net, R-CNN et FCOS, ont été largement utilisés pour la segmentation des pommes. Ces approches ont montré leur efficacité en combinant des techniques de détection d'objets et de segmentation pour une délimitation précise des pommes.

De plus, d'autres méthodes telles que l'algorithme de classification, la détection de bords et la combinaison de l'extraction de couleur et de seuillage ont également été explorées pour segmenter les images de pommes. Ces approches complémentaires offrent des alternatives et des possibilités de combinaison pour améliorer la précision de la segmentation.

En termes de perspectives d'amélioration, l'utilisation de données supplémentaires pour l'entraînement des modèles pourrait contribuer à améliorer leurs performances. La combinaison de différentes méthodes de segmentation peut également permettre d'obtenir une segmentation plus précise et robuste des pommes dans les images. De plus, l'intégration de la segmentation avec des robots pour une récolte autonome des fruits est une direction prometteuse à explorer.

En conclusion, ce chapitre a mis en évidence la diversité des approches de segmentation des pommes, avec une prédominance des réseaux de neurones convolutionnels. Ces avancées dans le domaine de la vision par ordinateur ouvrent la voie à des applications pratiques pour la surveillance et la gestion des cultures de pommes. Dans le chapitre suivant, nous présenterons notre contribution personnelle, basée sur les réseaux de neurones et l'apprentissage automatique, pour la segmentation des images de pommes.

Nouvelle approche de segmentation de pommes : YOLOv8+SAM

3.1 Introduction

Dans le chapitre précédent, nous avons examiné en détail l'état de l'art en matière de segmentation d'images, mettant en évidence les différentes techniques et modèles existants. Dans ce troisième chapitre, nous allons présenter notre contribution personnelle qui introduit une toute nouvelle approche pour la segmentation d'images de pommes. Notre approche novatrice repose sur la combinaison de deux modèles puissants : YOLOv8 pour la détection précise des pommes sur l'image, et SAM (Segment Anything Model) qui se distingue par son excellente capacité de segmentation.

La détection précise des objets est une étape cruciale dans la segmentation d'images, car elle permet de localiser les pommes présentes dans une scène donnée. Pour ce faire, nous avons utilisé le modèle YOLOv8, connu pour sa capacité à identifier et à localiser différents objets dans une image avec une grande précision. En utilisant YOLOv8, nous avons pu détecter les pommes de manière fiable et obtenir une localisation précise de leur position sur l'image.

Cependant, la simple détection des pommes ne suffit pas pour réaliser une segmentation complète et précise. C'est là que notre deuxième modèle, SAM, entre en jeu. SAM est un modèle de segmentation polyvalent qui peut segmenter pratiquement n'importe quoi dans une image. En combinant les capacités de détection de YOLOv8 avec la puissance de segmentation de SAM, nous avons pu développer une approche novatrice pour la segmentation d'images de pommes.

Notre approche consiste à utiliser YOLOv8 pour détecter les pommes sur l'image, puis à utiliser les résultats de la détection comme entrée pour SAM, qui effectue ensuite une segmentation fine des pommes détectées. Cette combinaison permet d'obtenir une segmentation précise et détaillée des pommes présentes dans l'image, en fournissant des contours nets et en distinguant clairement les pommes des autres éléments de la scène.

Dans ce chapitre, nous décrirons en détail notre approche de combinaison des modèles YOLOv8 et SAM pour la segmentation d'images de pommes. Nous présenterons également les résultats obtenus grâce à notre approche, en mettant en évidence sa performance par rapport aux méthodes existantes. Enfin, nous discuterons des avantages, des limitations et des perspectives d'amélioration de notre approche, en ouvrant la voie à de futures recherches dans le domaine de la segmentation d'images de fruits.

En résumé, ce chapitre présente notre contribution personnelle dans le domaine de la segmentation d'images de pommes, en introduisant une approche novatrice qui combine les

modèles YOLOv8 et SAM. Grâce à cette combinaison, nous avons réussi à obtenir des résultats prometteurs en termes de détection et de segmentation précise des pommes. Les détails de notre approche et les résultats obtenus seront présentés de manière exhaustive, suivis d'une discussion approfondie sur les perspectives dans ce domaine de recherche passionnant.

3.2 You Only Look Once (YOLO)

3.2.1 Présentation de YOLO

Le modèle de vision par ordinateur YOLO (You Only Look Once) est un réseau neuronal qui prédit la position des rectangles englobants et les probabilités de classification pour une image en un seul passage. En termes simples, il s'agit d'une méthode permettant d'identifier et de reconnaître des objets dans des images en temps réel.

Le modèle est devenu largement connu après que ses développeurs, Joseph Redmon, Santosh Divvala, Ross Girshik et Ali Farhadi, ont présenté la nouvelle architecture lors de la conférence CVPR (Computer Vision and Pattern Recognition) en 2016 [9], et ont même remporté le prix OpenCV People Choice Awards pour cela [10].

Le YOLO original (You Only Look Once) a été écrit par Joseph Redmon dans un framework personnalisé appelé Darknet [44], qui est un framework de recherche très flexible écrit dans des langages de bas niveau. Darknet a produit une série des meilleurs détecteurs d'objets en temps réel en vision par ordinateur : YOLO, YOLOv2, YOLOv3, YOLOv4, YOLOv5, YOLOv6, YOLOv7, et tout récemment YOLOv8. Bien que d'autres modèles de détection en une étape existent, les modèles YOLO sont rapides et compacts, ce qui les rend faciles à entraîner et à déployer, notamment sur des appareils avec des ressources de calcul limitées. La version de base de YOLO prédit les images à 45 IPS (images per second), tandis que la version plus légère, Fast YOLO, les traite à 155 IPS. La Figure 3.1 illustre une brève comparaison entre les versions de YOLO.

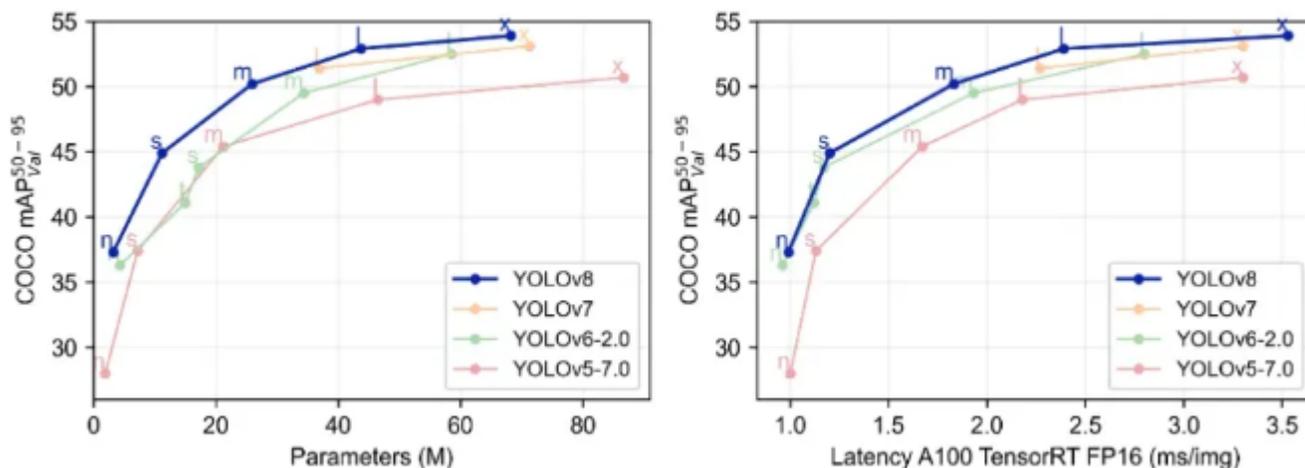


FIGURE 3.1 – Comparaison entre les versions YOLO [22]

3.2.2 Architecture des modèles YOLO

L'architecture du modèle YOLO (You Only Look Once) se compose de plusieurs couches qui permettent la détection d'objets dans une image de manière efficace et en temps réel. Voici une description générale de l'architecture du modèle YOLOv3, qui est l'une des versions les plus couramment utilisées :

Couche d'entrée (Input Layer) :

Cette couche reçoit l'image d'entrée et la redimensionne à une taille compatible avec le réseau. Couches de traitement (Feature Extraction Layers) :

Ces couches sont généralement basées sur des architectures de réseau convolutionnel (par exemple, Darknet-53) et sont responsables de l'extraction des caractéristiques de l'image. Elles sont composées de multiples blocs de convolution, suivis de couches d'activation (généralement ReLU) et de regroupement (max pooling). Couches de détection (Detection Layers) :

Ces couches sont responsables de la détection des objets dans l'image à différentes échelles et niveaux de précision. Elles sont situées à des niveaux spécifiques du réseau, et chaque couche de détection est associée à des paramètres spécifiques pour prédire les boîtes englobantes (bounding boxes) et les classes des objets détectés. Chaque couche de détection est généralement suivie d'une couche d'adaptation pour ajuster les prédictions. Couche de regroupement (Downsampling Layer) :

Cette couche est utilisée pour réduire la résolution spatiale de la sortie précédente et agréger les informations à une échelle supérieure. Elle permet de détecter des objets de plus grande taille et de réduire le coût de calcul pour les prochaines couches. Couches de détection supplémentaires :

Selon l'architecture spécifique, il peut y avoir plusieurs couches de détection à différentes échelles et résolutions, permettant ainsi la détection d'objets de différentes tailles. Couche de sortie (Output Layer) :

Cette couche fournit les prédictions finales du modèle, comprenant les boîtes englobantes prédites, les scores de confiance associés à chaque boîte et les classes prédites pour les objets détectés. Des techniques telles que l'algorithme de non-maximum suppression sont souvent utilisées pour supprimer les détections redondantes et améliorer la précision du modèle. L'architecture du modèle YOLO est conçue pour être rapide et efficace, permettant une détection en temps réel sur des images de haute résolution. Elle offre une approche "end-to-end" pour la détection d'objets, où la détection et la classification sont effectuées simultanément dans un seul réseau neuronal, ce qui distingue YOLO des approches de détection traditionnelles basées sur des régions d'intérêt (region-based).

La Figure 3.2 illustre un exemple de cette architecture.

3.2.3 YOLOv8

Le 10 janvier 2023, la dernière version de YOLO, YOLOv8, a été lancée. Développé par Ultralytics, Ce modèle, à la pointe de la technologie, s'appuie sur les succès des versions précédentes de YOLO et intègre de nouvelles avancées pour améliorer ses performances et sa polyvalence. En résumé, c'est un outil robuste et adaptable pour la détection d'objets qui offre le meilleur de la technologie de pointe la plus récente et la capacité d'utiliser et de comparer toutes les versions précédentes de YOLO. [22]

La Figure 3.6 illustre une détection par défaut par YOLO sur des images contenant des pommes. On peut tout de suite remarquer que la précision est quelque-peu médiocre compte

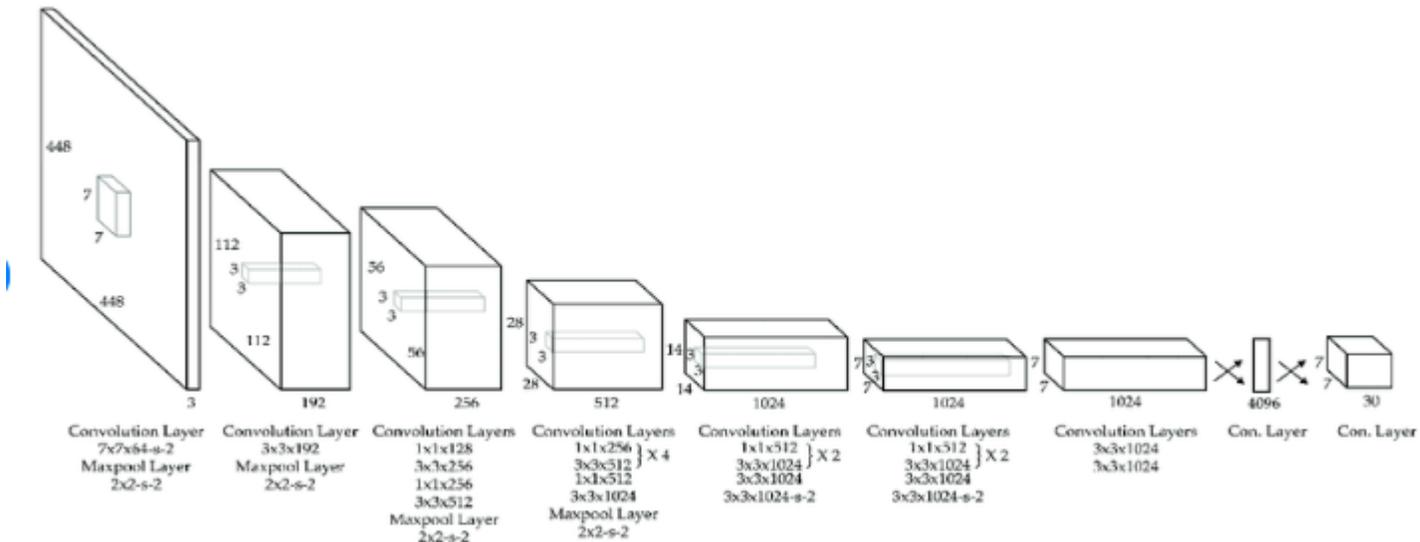


FIGURE 3.2 – Architecture YOLO [54]

tenu de nos objectifs.

3.3 Segment Anything Model (SAM)

Le modèle Segment anything (SAM), développé par Meta AI Research [12], a récemment attiré une attention significative. Entraîné sur un vaste ensemble de données de segmentation comprenant plus d'un milliard de masques, SAM est capable de segmenter n'importe quel objet dans une image donnée. Dans l'étude originale sur SAM, les auteurs se sont tournés vers des tâches de transfert immédiat (comme la détection des contours) pour évaluer les performances de SAM [18].

Il est important de mentionner que SAM effectue une segmentation rapide, ce qui diffère de la segmentation sémantique de deux manières : (1) les masques générés par SAM n'ont pas d'étiquettes ; (2) SAM repose sur des indications spécifiques. En d'autres termes, SAM découpe uniquement les objets de l'image sans leur attribuer d'étiquettes, et les objets découpés dépendent des indications fournies. Grâce à ce qu'on appelle l'ingénierie des indications, SAM a démontré une remarquable performance de transfert immédiat sans avoir besoin de réglages fins, ce qui a conduit beaucoup à penser que SAM est l'équivalent, pour la vision par ordinateur, du modèle GPT-3 de Brown et al. [55]. SAM est entraîné sur SA-1B, qui contient plus d'un milliard de masques provenant de 11 millions d'images, ce qui en fait le plus grand ensemble de données de segmentation jamais publié [12].

3.3.1 Architecture du modèle SAM

L'architecture du modèle Segment anything (SAM) repose sur un réseau de neurones convolutifs profonds et utilise une approche de segmentation basée sur des indications. Voici une description générale de l'architecture du modèle SAM :

Couche d'entrée (Input Layer) :

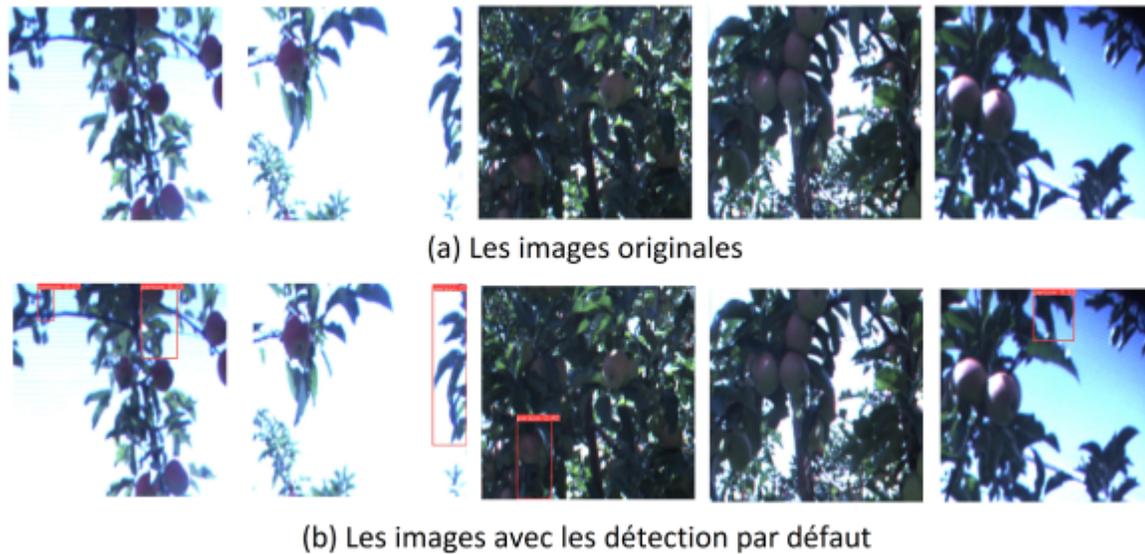


FIGURE 3.3 – Exemples de détection par YOLOv8 par défaut

Cette couche reçoit l'image d'entrée et la prépare pour le traitement ultérieur. Couches de traitement (Feature Extraction Layers) :

Ces couches sont responsables de l'extraction des caractéristiques de l'image à différentes échelles et niveaux d'abstraction. Elles peuvent être basées sur des architectures convolutionnelles pré-entraînées telles que ResNet, VGG, ou d'autres architectures spécifiques. Couche d'indication (Prompt Layer) :

Cette couche permet de fournir des indications spécifiques au modèle pour guider le processus de segmentation. Les indications peuvent être sous la forme d'un texte, d'une image de référence ou de tout autre format approprié. Couches de déconvolution (Deconvolution Layers) :

Ces couches effectuent la déconvolution des caractéristiques extraites et des indications fournies afin de générer des masques de segmentation. Couche de sortie (Output Layer) :

Cette couche fournit les masques de segmentation résultants, qui représentent les régions découpées des objets dans l'image. L'architecture du modèle SAM est conçue pour permettre une segmentation précise et rapide des objets dans une image, en utilisant des indications pour guider le processus de découpage. Les détails spécifiques de l'architecture, tels que le nombre de couches, les hyperparamètres et les détails de l'apprentissage, peuvent varier en fonction de l'implémentation spécifique de SAM et des objectifs de recherche.

La Figure 3.4 illustre une segmentation par default du modèle SAM sur une image de pommes.

En tenant compte des forces et des faiblesses de SAM, nous adaptons SAM au domaine de notre étude avec des modifications significatives. Nous introduisons SAM+YOLOv8, un cadre unifié qui permet aux utilisateurs d'effectuer la détection de la pomme et sa segmentation.

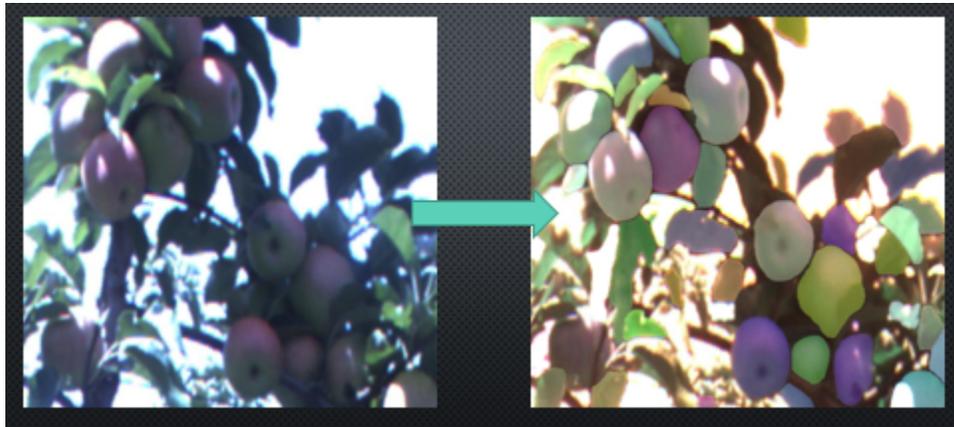


FIGURE 3.4 – Segmentation par défaut par le modèle SAM

3.4 L'approche proposée : Combinaison de YOLOv8 et SAM (YOLOv8+SAM)

Malgré les avantages mentionnés ci-dessus, l'application directe de SAM produit des résultats non-flexible et ne permet pas l'exécution d'autres tâches nécessitant une compréhension sémantique. Par conséquent, SAM n'est pas adapté aux tâches que nous souhaitons appliquées.

La Figure 3.5 illustre des exemple de segmentations qui donnent pas les résultat que nous attendant :

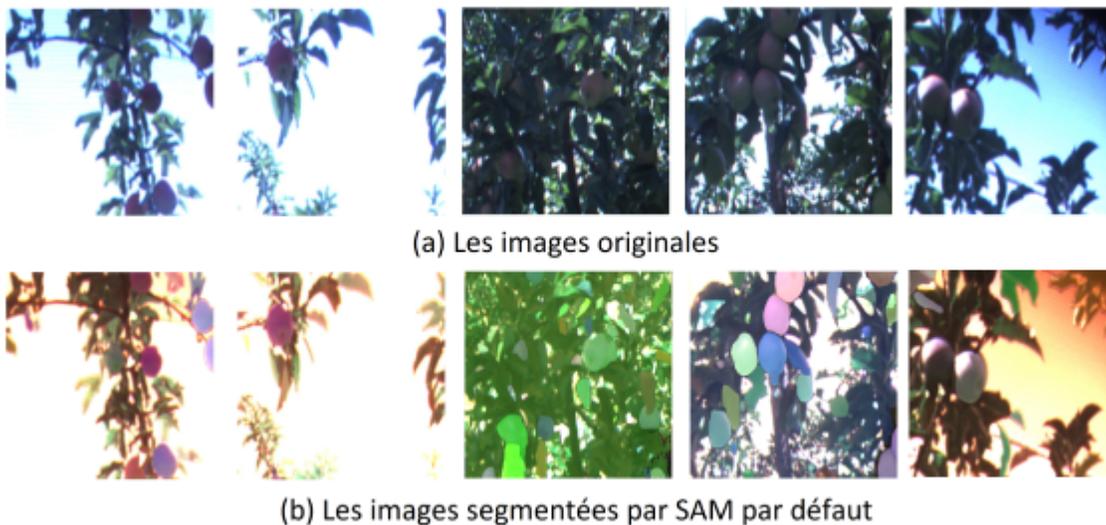


FIGURE 3.5 – Exemples de segmentation de SAM par défaut

La Figure 3.6 illustre la chaîne de traitement globale de l'approche que nous proposons. La



FIGURE 3.6 – Chaîne de traitement globale de YOLOv8+Sam

La première image est l'image en entrée, la seconde l'application du modèle YOLOv8 pré-entraîné, en troisième et en final, l'application du modèle de segmentation, et donc le résultat.

3.4.1 Architecture du modèle de segmentation proposé

Avant de passer à l'utilisation de YOLOv8+SAM, nous avons effectué un transfert d'apprentissage sur le modèle YOLOv8. Nous avons modifié la liste des classes pour inclure une nouvelle classe spécifique aux pommes. Par conséquent, nous avons supprimé toutes les autres classes qui n'étaient pas pertinentes pour notre objectif de détection de pommes.

Pour entraîner le modèle, nous avons utilisé notre dataset, qui a été préalablement décrit dans un fichier au format "yaml". Cet ensemble de données est spécifiquement conçu pour l'entraînement du modèle de détection de pommes.

Après l'entraînement, le modèle est maintenant bien préparé pour effectuer des opérations de détection de pommes dans l'environnement où les images ont été prises. Il est capable d'identifier et de localiser avec précision (0,846) les pommes dans des images en se basant sur les connaissances acquises lors de l'entraînement.

Ces étapes de transfert d'apprentissage et d'entraînement sur notre dataset ont permis d'adapter le modèle YOLOv8 à la tâche spécifique de détection de pommes. Grâce à ces efforts, nous disposons maintenant d'un modèle performant et spécialisé qui peut être utilisé pour la suite du traitement.

Le succès de SAM pour "segmenter n'importe quoi" a motivé la communauté à explorer "X n'importe quoi". En particulier, SAM s'est avéré polyvalent dans de nombreux projets lorsqu'il est combiné avec d'autres modèles pour obtenir des performances impressionnantes. Le tableau 3.1 exhibe une liste des projets de projets ayant combinés SAM avec d'autres modèles[52].

Inspirés des travaux présentés dans le tableau 3.1, nous présentons une nouvelle approche, et ce, en combinant SAM avec le modèle de détection YOLOv8, nommée par conséquent 'YOLOv8+SAM'. Spécifiquement, la nouvelle approche conçoit une chaîne de traitement de notre approche pour résoudre les problèmes liés à la segmentation de régions précises (les pommes dans notre cas), en combinant les avantages de SAM et les performances de détection de YOLOv8. La chaîne de traitement de notre approche est illustrée à la Figure 3.6.

Projet	Type
Grounded Segment Anything IDEA-Research [2023]	Grounded SAM
Grounded SAM : zero shot anomaly detection Cao [2023]	Grounded SAM
Grounded SAM : From Objects to Parts Cheems [2023]	Grounded SAM
SEEM Zou et al. [2023]	SAM+Label
Semantic Segment Anything (SSA) Chen et al. [2023b]	SAM+Label
Segment Anything with Clip Park [2023]	SAM+Label
Caption-Anything Ttengwang [2023]	SAM+Label
Label-Anything-Pipeline Yu et al. [2023a]	SAM+Label
AnyLabeling Vietanhdev [2023]	SAM+Label
Magic Copy Kevmo [2023]	Image editing
IEA : Image Editing Anything Feizc [2023]	Image editing
Edit Anything by Segment-Anything Gasvn [2023]	Image editing
Transfer-Any-Style Huage [2023]	Image editing
Inpaint Anything Yu et al. [2023b]	Image editing
MOTRv2 Yan [2023]	Tracking in Video
Track-Anything Gaomingqi [2023]	Tracking in Video
Segment and Track Anything (SAM-Track) Zxyang [2023]	Tracking in Video
Anything-3D Adamdad [2023]	3D
3D-Box via Segment Anything Chen [2023]	3D
Segment Anything Model (SAM) in Napari Karol [2023]	Medical
SAM Medical Imaging Amine [2023]	Medical
napari-segment-anything Okuma [2023]	Medical

TABLEAU 3.1 – Liste de projets ayant combinés SAM avec d’autres modèles[63]

En résumé, nous avons chargé un modèle YOLOv8 pré-entraîné sur le dataset COCO [23]. Ensuite, nous l’avons ré-entraîné sur notre propre dataset, bénéficiant ainsi d’un Transfer Learning, en remplaçant la dernière couche de classification par une nouvelle couche adaptée à notre tâche spécifique. On entraîne ensuite seulement une ou plusieurs de ses dernières couches sur nos propres données d’entraînement. Une fois l’entraînement terminé, nous extrayons les boîtes englobantes générées par le détecteur, celles-ci sont enfin insérées comme indication (Prompt) dans le modèle SAM pour générer un masque pour la zone de la pomme.

3.5 Tests et évaluations

3.5.1 Environnement de travail

Pour réaliser notre travail, nous avons mis en œuvre la méthode proposée ci-dessus en utilisant Google Colab doté de GPU à 12Go de RAM. Les tests ont été effectués sur un ordinateur portable Lenovo ThinkPad équipé d’un processeur Intel(R) Core(TM) i5-8250U @ 1.60GHz 1.80GHz, 8 Go de RAM, système d’exploitation 64-bits fonctionnant sous Windows 10.

Afin de valider notre méthode de segmentation, nous avons utilisé le dataset acfr-multifruit que nous avons customisé au début de la conception, le quel contient 223 images dédiées à

la phase de validation, soit tout juste 20% du nombre total d'images.

3.5.2 Dataset

Dans le cadre de l'entraînement de notre modèle, nous avons utilisé le dataset acfr-multifruit, duquel nous avons pris une partie des images que nous avons nous-même étiquetées en utilisant un outil d'étiquetage en ligne très efficace qui est Roboflow.

3.5.2.1 acfr-multifruit

Le dataset acfr-multifruit est la propriété de l'équipe d'agriculture au Centre Australien de Robotique de Terrain de l'Université de Sydney en Australie. Il est disponible en téléchargement sur le site (<http://data.acfr.usyd.edu.au/ag/treecrops/2016-multifruit/>). Les images de ce dataset ont été capturées par un véhicule terrestre équipé d'une caméra vidéo sphérique numérique Point Grey Ladybug trois, comprenant six caméras 2MP orientées pour capturer une vue panoramique complète à 360 degrés[50].

Il convient de noter que les images ont été prises en préservant les conditions naturelles des pommes telles qu'elles se trouvent dans le verger. Par conséquent, il y a des variations significatives intra-régions (au sein d'une même région) et inter-régions (entre différentes régions) en raison de la variabilité des conditions d'éclairage, de la distance entre les fruits et la caméra, du regroupement des fruits, de l'angle de capture de la caméra, etc[51].

Dans l'ensemble, ce dataset comprend plus de 8000 images, dont 1120 images contiennent des pommes des variétés Kanzi et Pink Lady, (celles précisément que nous avons utilisées pour générer notre propre dataset). La couleur des pommes varie du rouge vif à un mélange de rose-vert et de vert, comme illustré dans les exemples de la Figure 3.7.



FIGURE 3.7 – Quelques images dans acfr-multifruit.

3.5.2.2 Roboflow

Roboflow est une plateforme de vision par ordinateur qui fournit des outils et une infrastructure pour gérer et annoter des ensembles de données d'images, entraîner des modèles d'apprentissage automatique et les déployer en production. Son objectif est de simplifier le processus de création et de déploiement d'applications de vision par ordinateur en offrant une gamme de fonctionnalités et de services.

Nous avons tout d'abord téléchargé notre ensemble de données d'images et ensuite utilisé leurs outils d'annotation pour étiqueter et annoter les objets d'intérêt dans les images, les pommes dans notre cas. La plateforme prend en charge différents types d'annotations tels que les boîtes englobantes, les polygones et les points clés. Roboflow propose également des techniques d'augmentation de données pour augmenter la diversité des données d'entraînement et améliorer les performances du modèle. Nous avons donc pu étiqueter l'ensemble des 1120 images en englobant tout les pommes contenues sur les images.

Une fois notre ensemble de données prêt, Roboflow permet également d'entraîner des modèles d'apprentissage automatique à l'aide de frameworks populaires tels que TensorFlow, PyTorch et YOLO. Il offre une intégration transparente avec ces frameworks, en fournissant des pipelines préconfigurés et des options de personnalisation du modèle.

De plus, Roboflow propose des options de déploiement pour nos modèles entraînés, notamment l'exportation vers des formats populaires tels que TensorFlow SavedModel, ONNX et Core ML. Cela nous permet d'intégrer notre modèle de vision par ordinateur dans différentes applications et plates-formes, tel que google colab.

En résumé, Roboflow est conçu pour simplifier le processus de bout en bout de développement de modèles de vision par ordinateur, de la gestion des ensembles de données au déploiement du modèle, ce qui facilite le travail des chercheurs et des développeurs sur des tâches de vision par ordinateur.

Une fois notre dataset prêt, nous n'avons qu'à l'importer sur notre notebook colab pour pouvoir l'utiliser.

3.5.2.3 Répartition

Nous avons réparti les images comme suit : 70% pour l'entraînement soit 783 images. 20% pour la validation soit 223 images. 10% pour le test, soit 114 images.

La figure 3.8 résume le processus de création de notre dataset personnalisé sous roboflow, de l'étiquetage des images à la génération du fichier de configuration `.yaml` en passant par la répartition des images.

3.5.3 Métriques d'évaluation

Généralement, l'évaluation de la qualité d'une méthode de segmentation implique la comparaison de ses résultats avec des segmentations idéales, souvent créées manuellement par un expert. Cette comparaison est réalisée à l'aide de certaines mesures d'évaluation. Dans notre cas, nous avons sélectionné trois mesures couramment utilisées dans la littérature : la précision, le rappel et le score F1. Ces notions de rappel, précision et F-mesure ont été introduites par Van Rijsbergen [16], et elles reposent sur le principe qu'un système de recherche est efficace s'il parvient à identifier correctement le plus grand nombre possible de pixels. Dans ce qui suit, nous présentons ces trois mesures en détail.

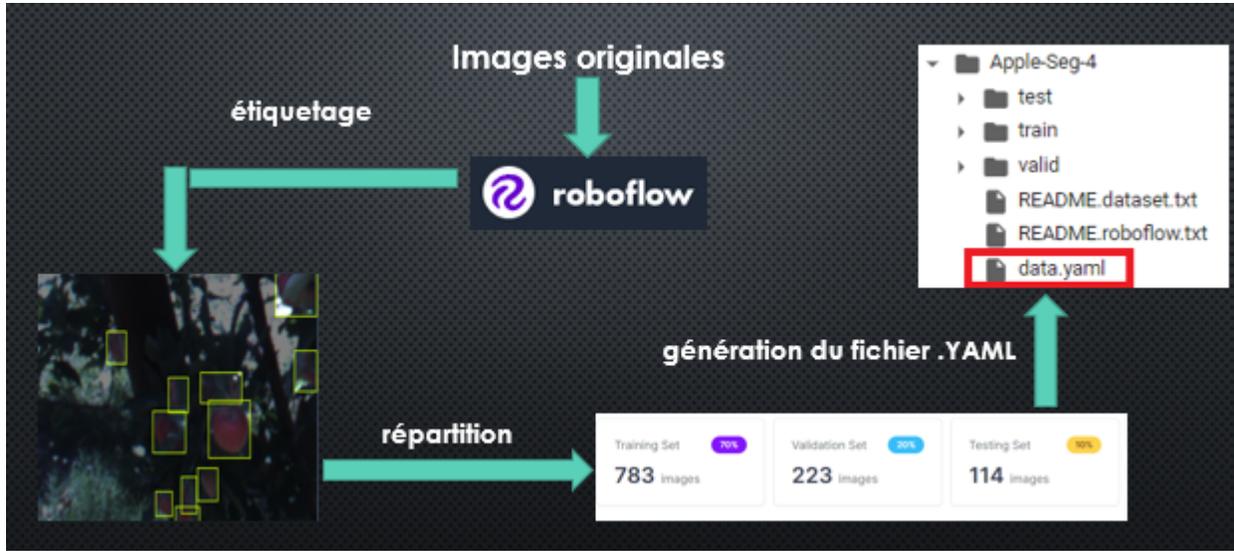


FIGURE 3.8 – Processus d’annotation et de génération du dataset avec Roboflow

3.5.3.1 Précision

La mesure de précision évalue la capacité d’un système à identifier et à restituer uniquement les pixels prédits correctement [39]. Elle est calculée en utilisant les vrais positifs (VP) et les faux positifs (FP). La formule de précision est la suivante :

$$Precision = \frac{VP}{VP + FP} \quad (3.1)$$

3.5.3.2 Rappel

La mesure de rappel évalue la capacité d’un système à trouver toutes les instances pertinentes dans un ensemble de données [39]. En d’autres termes, il s’agit du rapport entre les données correctement prédites et le nombre total de données pertinentes. Le rappel est calculé en utilisant les vrais positifs (VP) et les faux négatifs (FN). La formule du rappel est la suivante :

$$Rappel = \frac{VP}{VP + FN} \quad (3.2)$$

3.5.3.3 F1-score

Pour évaluer l’exactitude des résultats obtenus en appliquant la méthode subRW-A à différentes images, nous avons choisi d’utiliser la mesure d’évaluation F1-score. Cette mesure est également utilisée par Bargoti et Underwood dans leurs travaux sur la segmentation des pommes [53][52]. Notre objectif est de maximiser le F1-score, qui est la moyenne harmonique de la précision et du rappel, en maximisant simultanément ces deux derniers. Le F1-score atteint sa valeur maximale de 1 lorsque la précision est parfaite, sinon il est égal à 0. La formule correspondante est la suivante :

$$F1 - score = \frac{2RappelPrecision}{Rappel + Precision}. \quad (3.3)$$

3.5.4 Résultats de segmentation

Dans cette section, nous nous concentrons sur l'extraction des objets d'intérêt, qui sont des pommes dans notre cas, en utilisant le modèle YOLOv8+SAM. Nous appliquons ce modèle sur un ensemble de test d'images provenant du dataset mentionné précédemment. L'objectif est de détecter et segmenter précisément les pommes présentes dans les images.

Une fois que le modèle est appliqué sur les images de test, nous obtenons des résultats de segmentation. Ces résultats sont illustrés ci-dessous, montrant les pommes extraites avec leurs contours bien définis. Les exemples de segmentation démontrent l'efficacité du modèle YOLOv8+SAM dans la tâche d'extraction d'objets d'intérêt, en l'occurrence les pommes.

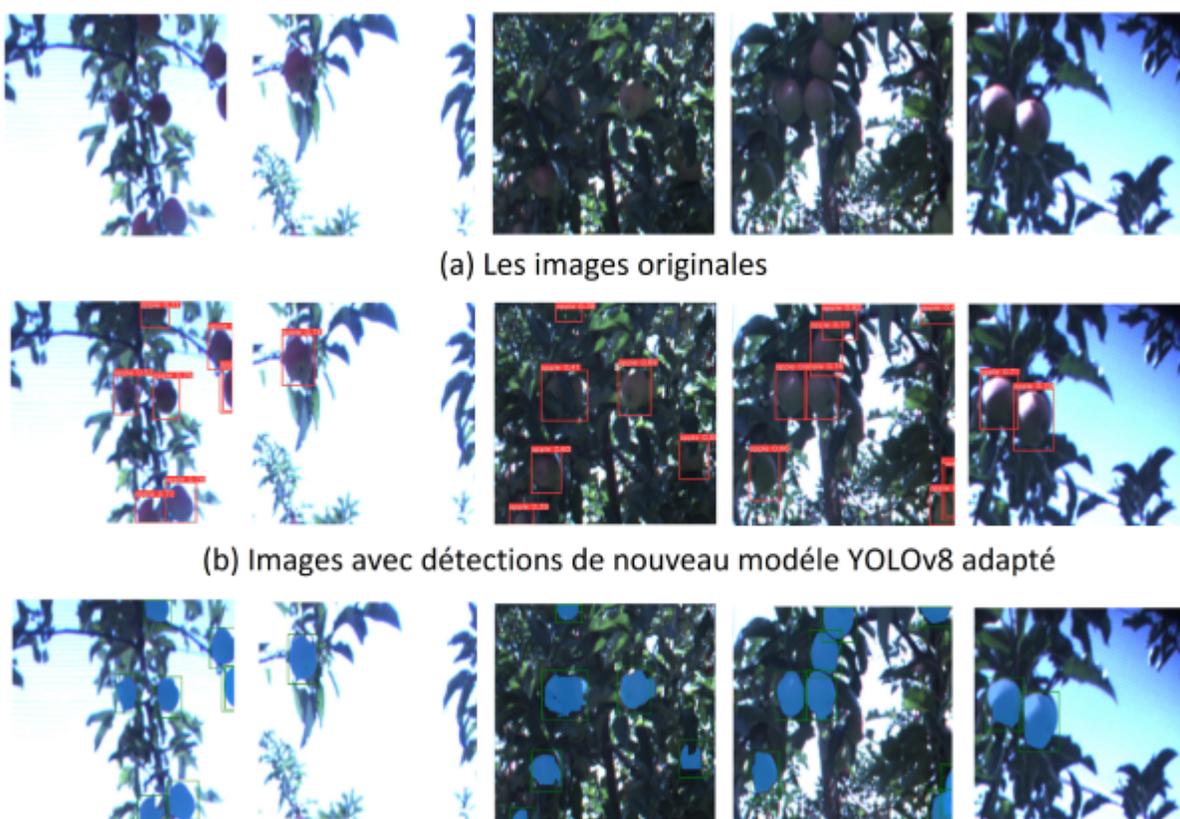


FIGURE 3.9 – Résultats de la segmentation en appliquant YOLOv8+SAM

3.5.5 Évaluation du modèle YOLOv8+SAM

En tant qu'expérimentation, nous avons testé notre méthode sur l'ensemble de test du dataset acfr-multifruit-2016 afin d'évaluer ses performances en présentant la matrice de confusion et en utilisant les métriques de précision, de rappel et de F1-score. Les résultats de l'évaluation et la matrice de confusion sont présentés ci-dessous.

La Figure 3.10 illustre La matrice de confusion générée par notre approche.

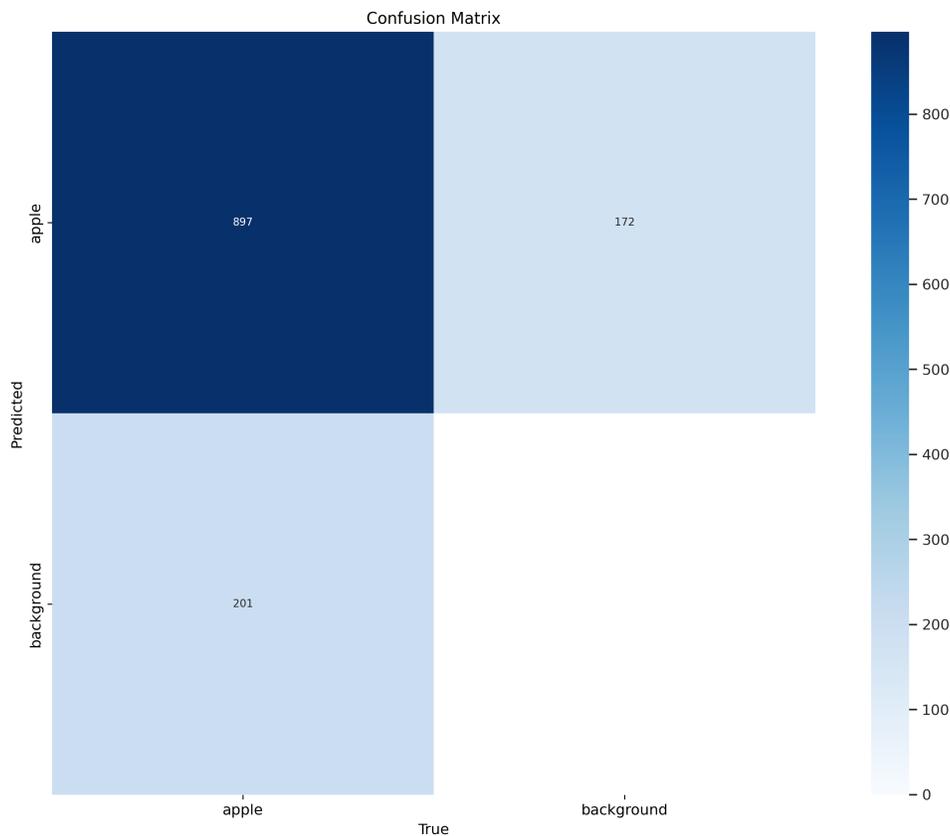


FIGURE 3.10 – Matrice de confusion

3.5.6 Discussion des résultats de la matrice de confusion

Dans notre évaluation, nous avons utilisé une matrice de confusion pour analyser en détail les performances de notre modèle de segmentation d'images de pommes. Cette matrice nous permet de comprendre la qualité de nos prédictions en comparant les résultats réels avec ceux prédits par le modèle. Voici une discussion plus détaillée des résultats obtenus à partir de la matrice de confusion :

Vrais positifs (VP) : Nous avons obtenu un total de 897 vrais positifs, ce qui correspond aux instances de pommes correctement détectées et segmentées par notre modèle. Ces résultats reflètent la capacité de notre modèle à identifier précisément les pommes dans les images.

Faux positifs (FP) : Nous avons observé 172 faux positifs, ce qui signifie que notre modèle a incorrectement prédit la présence de pommes dans certaines régions de l'image où il n'y en avait pas réellement. Ces erreurs peuvent être dues à des similitudes visuelles entre les pommes et d'autres objets ou à des régions de l'image pouvant prêter à confusion.

Faux négatifs (FN) : Nous avons identifié 201 faux négatifs, ce qui indique que notre modèle n'a pas réussi à détecter et segmenter correctement certaines pommes présentes dans les images réelles. Ces erreurs peuvent être attribuées à des facteurs tels que des occlusions partielles, des variations de couleur ou de forme des pommes, ou des problèmes liés à la complexité de l'arrière-plan de l'image.

Vrais négatifs (VN) : Dans notre cas, nous avons obtenu un nombre de vrais négatifs de 0. Cela signifie qu'il n'y avait pas de régions erronément identifiées comme ne contenant pas

de pommes dans nos données d'évaluation.

En examinant ces résultats, nous constatons que notre modèle a réussi à détecter et segmenter un grand nombre de pommes avec 897 vrais positifs. Cependant, nous avons également identifié quelques erreurs, notamment les 172 faux positifs et les 201 faux négatifs. Il est important de comprendre les raisons de ces erreurs afin de les corriger et d'améliorer les performances de notre modèle.

Les faux positifs pourraient être dus à des similitudes visuelles entre les pommes et d'autres objets présents dans l'image, ce qui peut être un défi dans la tâche de segmentation. D'autre part, les faux négatifs peuvent être causés par des facteurs tels que des variations de couleur ou de forme des pommes, des occlusions ou des difficultés à distinguer les pommes du reste de l'image.

Pour améliorer les performances de notre modèle, nous pourrions envisager d'explorer des techniques de prétraitement supplémentaires, telles que l'amélioration du contraste, la normalisation des couleurs ou l'utilisation de filtres adaptatifs, pour améliorer la robustesse de notre modèle face à ces défis.

En conclusion, l'analyse de la matrice de confusion met en évidence à la fois les points forts et les points faibles de notre modèle de segmentation d'images de pommes. En comprenant les sources d'erreurs, nous sommes en mesure de proposer des améliorations potentielles pour accroître la précision et la fiabilité de notre modèle, contribuant ainsi à l'avancement de la segmentation des images de pommes à l'aide de techniques de deep learning.

Précision	Rappel	F1-score
0.846	0.796	0.877

TABLEAU 3.2 – Résultats d'évaluation

Le Tableau 3.2 illustre les résultats d'évaluation du modèle proposé de part le taux de précision, rappel et score F1.

3.5.7 Discussion des résultats

Dans la discussion des résultats, nous tenons à souligner les performances prometteuses de notre approche "YOLOv8+SAM" pour la segmentation des images de pommes. Les mesures de précision, rappel et F1-score sont des indicateurs clés pour évaluer la qualité de notre modèle. Dans notre étude, nous avons obtenu des valeurs respectives de 0,846, 0,796 et 0,877, démontrant ainsi la capacité de notre approche à détecter et segmenter précisément les pommes dans les images agricoles.

La précision, qui mesure la proportion de vrais positifs parmi les prédictions positives, indique que notre modèle est capable d'identifier correctement la plupart des régions d'intérêt contenant des pommes. Le rappel, qui mesure la proportion de vrais positifs détectés parmi tous les exemples réels, souligne la capacité de notre modèle à trouver la plupart des pommes présentes dans les images. Enfin, le F1-score, qui est une mesure harmonique de la précision et du rappel, combine ces deux aspects pour évaluer globalement la performance de notre modèle.

Les résultats obtenus démontrent la robustesse de notre approche "YOLOv8+SAM" pour la segmentation des pommes, avec des scores élevés dans toutes les mesures évaluées. Cela confirme l'efficacité de notre modèle dans la détection précise des pommes et la segmentation

de leurs contours. Ces performances sont encourageantes et démontrent le potentiel de notre méthode pour des applications pratiques telles que le comptage précis des fruits, l'évaluation de leur maturité et la détection de maladies ou d'anomalies.

Néanmoins, il convient de noter qu'il existe toujours des marges d'amélioration. L'amélioration des performances pourrait être réalisée en continuant à enrichir notre jeu de données avec une plus grande variété d'images de pommes provenant de différents contextes et conditions. De plus, des techniques avancées de pré-traitement des images pourraient être explorées afin d'améliorer la qualité des données d'entrée et de réduire les éventuels artefacts.

En conclusion, nos résultats démontrent la pertinence et l'efficacité de notre approche "YOLOv8+SAM" pour la segmentation des images de pommes. Ces résultats encourageants ouvrent de nouvelles perspectives pour l'optimisation des processus agricoles, la gestion des cultures et la surveillance des récoltes. Nous espérons que notre contribution sera utile à la communauté scientifique et à l'industrie agricole, et stimulera de nouvelles recherches visant à améliorer encore davantage la précision et la performance des méthodes de segmentation d'images appliquées à l'agriculture.

3.6 Conclusion

Dans cette étude, nous avons développé un modèle de segmentation d'image novateur pour détecter et segmenter précisément les pommes dans les images agricoles. Notre approche combine les forces de deux modèles, YOLOv8 pour la détection des pommes et SAM pour la segmentation prompte des objets. Cette combinaison a permis d'obtenir des résultats améliorés par rapport aux méthodes traditionnelles de segmentation et de détection.

Les résultats de notre modèle sont prometteurs, avec des mesures de précision, de rappel et de F1-score atteignant respectivement 0,846, 0,796 et 0,877. Ces scores élevés témoignent de la capacité de notre modèle à détecter et segmenter avec précision les pommes dans les images. Cela ouvre de nouvelles perspectives pour des applications pratiques telles que le comptage précis des fruits, l'évaluation de leur maturité et la détection de maladies ou d'anomalies.

Cependant, il est important de souligner que des améliorations supplémentaires peuvent être apportées à notre approche. Par exemple, l'enrichissement de notre jeu de données avec une plus grande variété d'images de pommes provenant de différents contextes et conditions pourrait permettre d'améliorer la précision de la détection après l'entraînement. De plus, l'exploration de techniques de pré-traitement plus avancées des images pourrait contribuer à améliorer la qualité des données d'entrée et à réduire les éventuels artefacts.

En conclusion, notre modèle de segmentation d'image YOLOv8+SAM représente une avancée significative dans la détection et la segmentation des pommes dans les images agricoles. Les résultats encourageants obtenus démontrent son potentiel pour optimiser les processus agricoles, faciliter la gestion des cultures et améliorer la qualité des récoltes. Nous espérons que notre contribution ouvrira la voie à de nouvelles recherches visant à améliorer encore davantage la précision et la performance des méthodes de segmentation d'images appliquées à l'agriculture.

Conclusion générale et perspectives

Ce travail a été réalisé dans le cadre notre projet de fin de cycle master en Intelligence artificielle.

Dans le cadre de cette étude, nous avons abordé le défi de la segmentation d'images appliquée à l'industrie agricole, en mettant l'accent sur la détection et la segmentation précises des pommes. Notre objectif principal était de développer une méthode novatrice et performante qui combine les forces de deux modèles, YOLOv8 pour la détection des pommes et SAM pour la segmentation précise des objets. Cette approche avait pour but d'optimiser les processus de classification, de surveillance et de contrôle de qualité dans l'industrie de la pomme.

L'industrie de la pomme joue un rôle économique crucial au sein du secteur agricole. Cependant, les processus de classification, de surveillance et de contrôle de qualité des pommes peuvent s'avérer complexes et exigeants. La segmentation d'images se présente comme une solution puissante pour détecter, identifier et extraire précisément les pommes dans les images agricoles. Cependant, les méthodes existantes présentent des limitations dans la détection et la segmentation des pommes. C'est pourquoi notre étude visait à proposer une approche novatrice qui combine les forces de la détection et de la segmentation pour améliorer la précision et l'efficacité de ces processus.

Nous avons adopté une approche méthodologique rigoureuse pour réaliser notre étude. Tout d'abord, nous avons effectué une revue approfondie de la littérature pour comprendre les bases théoriques de la segmentation d'images, de la détection d'objets et de la vision par ordinateur. Ensuite, nous avons réalisé un état de l'art des méthodes de segmentation d'images appliquées spécifiquement aux images de pommes, en analysant leurs avantages et leurs limites.

Sur la base de ces analyses, nous avons développé notre méthode novatrice "YOLOv8+SAM". Nous avons utilisé YOLOv8, un modèle de détection d'objets reconnu pour sa précision, comme base pour la détection des pommes. Ensuite, nous avons exploité les coordonnées des boîtes englobantes générées par YOLOv8 pour introduire les régions d'intérêt dans le modèle de segmentation SAM, connu pour sa précision dans la segmentation des objets. Nous avons entraîné et optimisé ces modèles sur notre propre jeu de données spécifique aux pommes.

Les résultats obtenus grâce à notre approche "YOLOv8+SAM" sont très encourageants. Les mesures de précision, de rappel et de F1-score atteignent respectivement 0,846, 0,796 et 0,877, démontrant ainsi la capacité de notre modèle à détecter et segmenter avec précision les pommes dans les images agricoles. Ces résultats témoignent de l'efficacité de notre approche qui combine les capacités de détection de YOLOv8 avec la précision de segmentation de SAM.

Notre contribution se situe dans l'amélioration des processus de classification, de sur-

veillance et de contrôle de qualité dans l'industrie de la pomme. Notre méthode permet une évaluation précise de la qualité des pommes en identifiant les défauts, les maladies ou les anomalies présentes dans les images. Cela facilite la sélection, le tri et la commercialisation des pommes de haute qualité, répondant ainsi aux normes les plus strictes de l'industrie et garantissant une satisfaction maximale des consommateurs. De plus, notre approche offre des perspectives pratiques pour optimiser les processus de tri et de conditionnement, réduisant ainsi les pertes et les déchets.

En conclusion, notre modèle de segmentation d'image "YOLOv8+SAM" représente une avancée significative dans la détection et la segmentation des pommes dans les images agricoles. En intégrant les forces de YOLOv8 et de SAM, nous avons surmonté les limitations des méthodes traditionnelles et obtenu des résultats prometteurs. Notre approche contribue à l'amélioration des processus de classification, de surveillance et de contrôle de qualité dans l'industrie de la pomme, offrant des solutions pratiques pour la gestion agricole, la planification logistique et le contrôle de la qualité des produits. Nous espérons que notre recherche ouvrira la voie à de nouvelles investigations visant à améliorer encore davantage la précision et la performance des méthodes de segmentation d'images appliquées à l'agriculture, favorisant ainsi le développement durable de l'industrie de la pomme et répondant aux exigences croissantes du marché.

Bibliographie

- [1] <https://www.lanutrition.fr/>. Consulté le [02/05/2023].
- [2] <http://www.stargrow.co.za/>. Consulté le [02/05/2023].
- [3] <https://www.gastronomiac.com//>. Consulté le [02/05/2023].
- [4] <http://www.specialtyproduce.com/produce/Kanzi/Apples>. Consulté le [29/05/2023].
- [5] <https://jardinage.ooreka.fr/astuce/voir/410359/pomme/pinklady>,. Consulté le [29/05/2023].
- [6] <https://www.passeportsante.net/fr/Nutrition/EncyclopedieAliments/Fiche.aspx>. Consulté le [29/05/2023].
- [7] Site de la fao. <http://www.fao.org/home/fr/>. Consulté le [02/05/2023].
- [8] Site de la fao. <http://www.fao.org/home/fr/>. Consulté le [02/05/2023].
- [9] In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [10] People choice awards, 2106.
- [11] H. Dianne A. A comprehensive review of apples and apple components and their relationship to human health. Advances in nutrition, 2(5) :408–420, 2011.
- [12] K. Alexander, M. Eric, R. Nikhila, M. Hanzi, R. Chloe, G. Laura, X. Tete, Wh. Spencer, B. Alexander C, L. Wan-Yen, et al. Segment anything. arXiv preprint arXiv :2304.02643, 2023.
- [13] M. AZZI. Contribution à l'étude des propriétés globales et locales des semi-conducteurs en vue d'application photovoltaïque et oscillateur paramétrique optique. PhD thesis, 15-05-2022, 2021.
- [14] JA Bondy and M. USR. Graph theory with applications north, 1979.
- [15] G. Rafael C. Digital image processing. Pearson education india, 2009.
- [16] V-R. C. Information retrieval : theory and practice. In Proceedings of the joint IBM/University of Newcastle upon tyne seminar on data base systems, volume 79, 1979.
- [17] Z. Kunlin C. Zhang and P. Yue. A method of apple image segmentation based on color-texture fusion feature and machine learning. Agronomy, 10(7) :972, 2020.
- [18] Zh. Chaoning, Zh. Sheng, L. Chenghao, Q. Yu, K. Taegoo, Sh. Xinru, Zh. Chenshuang, Q. Caiyan, R. Francois, B. Sung-Ho, et al. A survey on segment anything model (sam) : Vision foundation model meets prompt engineering. arXiv preprint arXiv :2306.06211, 2023.

- [19] T. Chuanqi, S. Fuchun, K. Tao, Zh. Wenchang, Y. Chao, and L. Chunfang. A survey on deep transfer learning. In Artificial Neural Networks and Machine Learning–ICANN 2018 : 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27, pages 270–279. Springer, 2018.
- [20] B. Claude. Graphes et hypergraphes : 2e ed. Dunod, 1973.
- [21] W. Dandan, L. Changying, S. Huaibo, X. Hongting, L. Chang, and H. Dongjian. Deep learning approach for apple edge detection to remotely monitor apple growth in orchards. IEEE access, 8 :26911–26925, 2020.
- [22] DataPhoenix. A guide to the yolo family of computer vision models, Year.
- [23] DataPhoenix. A guide to the yolo family of computer vision models, Year.
- [24] S. Denis, R. Jurij, B. Michael, et al. Modelling apple fruit yield using image analysis for fruit colour, shape and texture. European journal of horticultural science, 74(6) :260, 2009.
- [25] J. Michael I and M. Tom M. Machine learning : Trends, perspectives, and prospects. Science, 349(6245) :255–260, 2015.
- [26] G. Ian, B. Yoshua, and C. Aaron. Deep learning. MIT press, 2016.
- [27] W. Ji, X. Meng, Y. Tao, B. Xu, and D. Zhao. Fast segmentation of colour apple image under all-weather natural conditions for vision recognition of picking robots. International Journal of Advanced Robotic Systems, 13(1) :24, 2016.
- [28] W. Jia, Y. Tian, R. Luo, Z. Zhang, J. Lian, and Y. Zheng. Detection and segmentation of overlapped fruits based on optimized mask r-cnn application in apple harvesting robot. Computers and Electronics in Agriculture, 172 :105380, 2020.
- [29] L. Jonathan, Sh. Evan, and D. Trevor. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3431–3440, 2015.
- [30] J. Anil K. Fundamentals of digital image processing. Prentice-Hall, Inc., 1989.
- [31] H. Kaiming, G. Georgia, D. Piotr, and G. Ross. Mask r-cnn. In Proceedings of the IEEE international conference on computer vision, pages 2961–2969, 2017.
- [32] Z. Kunlin, G. Luzhen, Z. Hang, Z. Chunlong, and L. Wei. An apple image segmentation method based on a color index obtained by a genetic algorithm. Multimedia Tools and Applications, 81(6) :8139–8153, 2022.
- [33] D. LAÏCHE. Sur la coloration de packing de quelques classes de graphes : Cas orienté et non orienté. PhD thesis, Université de Lyon, 2017.
- [34] Q. Li, W. Jia, M. Sun, S. Hou, and Y. Zheng. A novel green apple segmentation algorithm based on ensemble u-net under complex orchard environment. Computers and Electronics in Agriculture, 180 :105900, 2021.
- [35] Qianwen Li, Weikuan Jia, Meili Sun, Sujuan Hou, and Yuanjie Zheng. A novel green apple segmentation algorithm based on ensemble u-net under complex orchard environment. Computers and Electronics in Agriculture, 180 :105900, 2021.
- [36] M. Liu, W. Jia, Z. Wang, Y. Niu, X. Yang, and C. Ruan. An accurate detection and segmentation model of obscured green fruits. Computers and Electronics in Agriculture, 197 :106984, 2022.

- [37] B. Christopher M and N. Nasser M. Pattern recognition and machine learning, volume 4. Springer, 2006.
- [38] R. Janet M, Ng. Quynhanh V, W. Juhi R, D. Larry W, P. Kristine Y, H. Juliette C, B. J Chance, Th. Leslie D, W. Dale R, E. Terry E, et al. Protein, fat, moisture and cooking yields from a us study of retail beef cuts. Journal of Food Composition and Analysis, 43 :131–139, 2015.
- [39] Gh. Maha, C. Pascal, L. Jean-Charles, and L. Alain. Mesures de qualité de clustering de documents : Prise en compte de la distribution des mots clés. In 10ième Conférence Internationale Francophone sur l’Extraction et la Gestion des Connaissances-EGC 2010, 2010.
- [40] L. Mengyuan, J. Weikuan, W. Zhifen, N. Yi, Y. Xinbo, and R. Chengzhi. An accurate detection and segmentation model of obscured green fruits. Computers and Electronics in Agriculture, 197 :106984, 2022.
- [41] Andrew Ng. Machine learning yearning. URL : [http://www.mlyearning.org/\(96\)](http://www.mlyearning.org/(96)), 139, 2017.
- [42] H. Su J. Krause S. Satheesh S. Ma Z. Huang A. Karpathy A. Khosla M. Bernstein A. C. Berg O. Russakovsky, J. Deng and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV), 115(3) :211–252, 2015.
- [43] R. Olaf, F. Philipp, and B. Thomas. U-net : Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015 : 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18, pages 234–241. Springer, 2015.
- [44] J. Peiyuan, E. Daji, L. Fangyao, C. Ying, and M. Bo. A review of yolo algorithm developments. Procedia Computer Science, 199 :1066–1073, 2022.
- [45] H. Rachid. Segmentation non-supervisee d’images couleur par sur-segmentation markovienne en regions et procedure de regroupement de regions par graphes ponderes. 2008.
- [46] D. Shiv Ram, D. Pushkar, S. Nishant, and G. Jay Prakash. Infected fruit part detection using k-means clustering segmentation technique. 2013.
- [47] Z. Rong, D. Lutz, S. Yurui, and B. Michael M. Using colour features of cv.‘gala’apple fruits in an orchard in image processing to predict yield. Precision Agriculture, 13 :568–580, 2012.
- [48] Sh-Sh. Shai and B-D. Shai. Understanding machine learning : From theory to algorithms. Cambridge university press, 2014.
- [49] V.S. Shibu and P. Simon. An efficient method for color image segmentation using adaptive mean shift and normalized cuts. In 2011 Third International Conference on Advanced Computing, pages 309–314. IEEE, 2011.
- [50] B. Suchet and U. James. Deep fruit detection in orchards. arXiv preprint arXiv :1610.03677, 2016.
- [51] B. Suchet and U. James. Image segmentation for fruit detection and yield estimation in apple orchards. To Appear in Journal of Field Robotics, 2016.
- [52] B. Suchet and U. James. Deep fruit detection in orchards. In 2017 IEEE international conference on robotics and automation (ICRA), pages 3626–3633. IEEE, 2017.

- [53] B. Suchet and U. James P. Image segmentation for fruit detection and yield estimation in apple orchards. Journal of Field Robotics, 34(6) :1039–1060, 2017.
- [54] Juan Terven and Diana Cordova-Esparza. A comprehensive review of yolo : From yolov1 to yolov8 and beyond. arXiv preprint arXiv :2304.00501, 2023.
- [55] B. Tom, M. Benjamin, R. Nick, S. Melanie, K. Jared D, D. Prafulla, N. Arvind, Sh. Pranav, S. Girish, A. Amanda, et al. Language models are few-shot learners. Advances in neural information processing systems, 33 :1877–1901, 2020.
- [56] H. Trevor, T. Robert, H. Jerome Friedman, and H. Jerome Friedman. The elements of statistical learning : data mining, inference, and prediction, volume 2. Springer, 2009.
- [57] H. Trevor, Robert Tibshirani, F. Jerome H, and F. Jerome H. The elements of statistical learning : data mining, inference, and prediction, volume 2. Springer, 2009.
- [58] S. VS and S. Philomina. An efficient method for color image segmentation using adaptive mean shift and normalized cuts. In 2011 Third International Conference on Advanced Computing, pages 309–314. IEEE, 2011.
- [59] D. Wang, C. Li, H. Song, H. Xiong, C. Liu, and D. He. Deep learning approach for apple edge detection to remotely monitor apple growth in orchards. IEEE access, 8 :26911–26925, 2020.
- [60] J. Wei, M. Xiangli, T. Yun, X. Bo, and Z. Dean. Fast segmentation of colour apple image under all-weather natural conditions for vision recognition of picking robots. International Journal of Advanced Robotic Systems, 13(1) :24, 2016.
- [61] J. Wei, M. Xiangli, T. Yun, X. Bo, and Z. Dean. Fast segmentation of colour apple image under all-weather natural conditions for vision recognition of picking robots. International Journal of Advanced Robotic Systems, 13(1) :24, 2016.
- [62] J. Weikuan, T. Yuyu, L. Rong, Z. Zhonghua, L. Jian, and Z. Yuanjie. Detection and segmentation of overlapped fruits based on optimized mask r-cnn application in apple harvesting robot. Computers and Electronics in Agriculture, 172 :105380, 2020.
- [63] S. Xin and Y. Lei. The study of adaptive multi threshold segmentation method for apple fruit based on the fractal characteristics. In 2015 8th International Symposium on Computational Intelligence and Design (ISCID), volume 2, pages 168–171. IEEE, 2015.
- [64] L. Yann, B. Yoshua, H. Geoffrey, et al. Deep learning. nature, 521 (7553), 436-444. Google Scholar Google Scholar Cross Ref Cross Ref, page 25, 2015.
- [65] E. Youssef. Segmentation d’images couleur par combinaison LPE-régions/LPE-contours et fusion de régions. Application à la segmentation de toitures à partir d’orthophotoplans. PhD thesis, Université de Technologie de Belfort-Montbeliard ; Université Ibn Tofail . . . , 2013.
- [66] Chunlong Zhang, Kunlin Zou, and Yue Pan. A method of apple image segmentation based on color-texture fusion feature and machine learning. Agronomy, 10(7) :972, 2020.
- [67] K. Zou, L. Ge, H. Zhou, C. Zhang, and W. Li. An apple image segmentation method based on a color index obtained by a genetic algorithm. Multimedia Tools and Applications, 81(6) :8139–8153, 2022.

RÉSUMÉ

Nous sommes fiers de présenter notre travail sur la segmentation d'images de pommes. Notre projet repose sur une approche intelligente en combinant deux modèles pré-entraînés pour obtenir une segmentation précise.

Dans un premier temps, nous avons utilisé le modèle pré-entraîné YOLOv8, reconnu pour sa capacité à détecter et encadrer les objets dans une image. En ajoutant une couche supplémentaire et en ré-entraînant le modèle sur notre propre jeu de données contenant 1120 images de pommes que nous avons soigneusement étiquetées, nous avons adapté le modèle pour se spécialiser dans la détection précise des pommes.

Une fois le modèle entraîné, nous avons extrait les coordonnées des boîtes englobantes qui encadrent les pommes détectées. En utilisant ces coordonnées comme indication, nous avons appliqué un autre modèle pré-entraîné appelé SAM pour effectuer la segmentation du contenu des boîtes englobantes. Cette approche nous a permis d'obtenir une segmentation précise des pommes dans les images.

Notre projet démontre une excellente combinaison de transfert d'apprentissage et d'utilisation de modèles pré-entraînés pour résoudre un problème spécifique de segmentation d'images de pommes. En adaptant et en ré-entraînant le modèle YOLOv8 sur notre propre jeu de données, nous avons réussi à obtenir des résultats de détection précis. Ensuite, l'utilisation du modèle SAM pour segmenter le contenu des boîtes englobantes nous a permis d'obtenir une segmentation détaillée des pommes.

Ce projet ouvre des perspectives intéressantes pour l'automatisation de la segmentation des fruits, en particulier des pommes. Les résultats que nous avons obtenus pourraient être utilisés dans diverses applications, telles que le tri automatique des fruits ou l'analyse de la qualité des récoltes. En conclusion, notre travail présente une approche prometteuse pour la segmentation d'images de pommes en combinant habilement des modèles pré-entraînés, le transfert d'apprentissage et l'entraînement sur notre propre jeu de données. Les résultats obtenus sont encourageants et ouvrent de nouvelles perspectives dans le domaine de la segmentation des fruits. Ces avancées pourraient trouver des applications concrètes, telles que le tri automatique des fruits ou l'analyse de la qualité des récoltes. Nous sommes ravis des résultats que nous avons obtenus jusqu'à présent et nous sommes impatients de voir comment ces travaux pourront être utilisés pour améliorer les processus liés à la segmentation des pommes et potentiellement d'autres fruits.

Mots clés : Ségmentation, Détection, Transfert-Learning

ABSTRACT

We are proud to present our work on apple image segmentation. Our project is based on an intelligent approach that combines two pre-trained models to achieve precise segmentation.

In the first step, we utilized the pre-trained YOLOv8 model, known for its capability to detect and enclose objects in an image. By adding an additional layer and retraining the model on our own dataset consisting of 1120 carefully labeled apple images, we tailored the model to specialize in accurate apple detection.

Once the model was trained, we extracted the coordinates of the bounding boxes encompassing the detected apples. Using these coordinates as guidance, we applied another pre-trained model called SAM to perform segmentation of the contents within the bounding boxes. This approach enabled us to obtain precise segmentation of apples in the images.

Our project showcases an excellent combination of transfer learning and utilization of pre-trained models to address the specific task of apple image segmentation. Through the adaptation and retraining of the YOLOv8 model on our own dataset, we achieved precise detection results. Furthermore, leveraging the SAM model to segment the contents within the bounding boxes yielded detailed apple segmentation.

This project opens up exciting prospects for automating fruit segmentation, particularly for apples. The results we obtained can have various applications, such as automated fruit sorting or crop quality analysis.

In conclusion, our work presents a promising approach to apple image segmentation by skillfully combining pre-trained models, transfer learning, and training on our own dataset. The results obtained are encouraging and offer new perspectives in the field of fruit segmentation. These advancements can find concrete applications in areas like automated fruit sorting and crop quality analysis. We are delighted with the results achieved thus far and look forward to exploring how these findings can enhance apple segmentation processes and potentially extend to other fruits.

Key words : Ségmentation, Détection, Transfert-Learning