

# UNIVERSITÉ D'ORLÉANS

## ÉCOLE DOCTORALE MATHÉMATIQUES, INFORMATIQUE, PHYSIQUE THÉORIQUE ET INGÉNIERIE DES SYSTÈMES

Laboratoire: **PRISME**

**THÈSE** présentée par:

**Mamadou Dian BAH**

soutenue le : **05 mars 2020**

pour obtenir le grade de: **Docteur de l'Université d'Orléans**

Discipline/ Spécialité: **Sciences et Technologies Industrielles**

### Détection des adventices par imagerie aérienne

**THÈSE dirigée par :**

**M. Adel HAFIANE**

**M. Bruno EMILE**

**M. Raphaël CANALS**

MCF HDR, INSA Centre Val de Loire

MCF HDR, Université d'Orléans

MCF, Université d'Orléans

**RAPPORTEURS :**

**Mme Christelle GEE**

**M. Ludovic MACAIRE**

PRU, AgroSup Dijon

PRU, Université de Lille

**JURY :**

**M. Ludovic MACAIRE**

**Mme Christelle GEE**

**M. Stéphane cordier**

**M. Abbas CHEDDAD**

PRU, Université de Lille, President du jury

PRU, AgroSup Dijon

PRU, Université d'Orléans

Associate Professor Blekinge

Institute of Technology, BTH

MCF HDR, INSA Centre Val de Loire

MCF, Université d'Orléans

**M. Adel HAFIANE**

**M. Raphaël CANALS**



---

## Remerciements

“Comme le coureur de fond, le doctorant doit tenir la distance. Mais à la différence du marathonien, personne n’a tracé pour lui de ligne d’arrivée. Le plus dur dans la thèse, c’est de finir.” CAIRN.INFO

Il y a quatre ans la question qui se posait c’était “Thèse ou pas Thèse” et voilà ce jeudi 05 mars 2020 aux alentours de 13h, le jury chargé d’évaluer mes travaux de thèse m’a déclaré «Docteur de l’université d’Orléans».

Il y a tant de personnes à qui je voudrais témoigner officiellement toute ma gratitude. Comme je suis timide je ne vais pas être long.

Tout d’abord, je tiens à remercier Dieu, le tout puissant et miséricordieux, qui m’a donné la force, l’intelligence et la patience d’accomplir ce modeste travail.

Je voudrais remercier mes directeurs de thèse Adel HAFIANE et Bruno EMILE ainsi que mon encadrant Raphaël CANALS, pour leur confiance en m’acceptant comme doctorant, pour leur patience, leur disponibilité et surtout leurs judicieux conseils qui ont contribué à alimenter ma réflexion.

Ce travail n’aurait pas été possible sans le soutien de la Région Centre-Val de Loire que je remercie. Je remercie également les Chambres d’Agriculture de la région Centre-Val de Loire.

Je remercie l’ensemble des membres du jury, qui m’ont fait l’honneur de bien vouloir étudier avec attention mon travail.

Je remercie ensuite tous les enseignants qui m’ont eu en tant qu’élève et étudiant depuis mes premiers pas à l’école maternelle jusqu’au master, sans oublier les professeurs de révision. Vous m’avez inspiré et m’avez permis d’avoir confiance en moi et en l’avenir mais vous avez aussi réussi à me donner l’envie d’apprendre.

Mes remerciements vont pareillement à ma famille et mes amis d’ici et d’ailleurs qui, avec cette question récurrente « quand est-ce que tu la soutiens cette thèse ? », bien qu’angoissante en période fréquente de doutes, m’ont permis de ne jamais dévier de mon objectif final. J’ai également pu compter sur votre amitié et vos soutiens indéfectibles à différents moments de ma vie et surtout en différents lieux.

Je souhaiterais aussi remercier ceux qui ont partagé mon quotidien et grâce à qui ces trois années se sont déroulées dans une ambiance agréable “la team PRISME Galilée” ; vous m’avez accueilli comme un nouveau et nous sommes devenues une famille. Que dire des pauses café et/ou des pauses thé inoubliables ? Des fous rires dans le F214, le bureau de tous les mystères. Une team composée de personnes venues pratiquement de tous les continents avec des cultures différentes mais un objectif commun, la réussite. L’ambiance de travail a été excellente et nous avons pu ainsi allier bons moments et séances de travail. Je remercie Asma Bougrine “l’avocate ...”, la militante “pro café” pour l’ambiance la bonne humeur et les encouragements non seulement dans le bureau F214, mais également lors des conférences et événements (MT180S, ...) auxquels nous avons participé. Tu es une personne exceptionnelle, serviable et toujours au service des autres. Tu as bien mérité ton titre de “meilleure collègue”. Fabrice Atrévie le poto, le doyen, le sage de la team, je tiens également à te remercier tu as été un grand guide, et nous nous sommes beaucoup inspirés de tes expériences car tu as toujours été à l’écoute et tu as su partager. Tu es un “bon”, reste comme tu es et continues à faire ce que tu aimes. IPTA 2019 à Istanbul reste inoubliable car c’était pour nous l’occasion de nous dire “bye-bye” dans une belle

---

ambiance. Koubouratou Idjaton O. "O point" la militante "pro thé de l'équipe", tu as été la dernière à arriver dans la team mais tu as su te créer ta place avec ton sourire et ton ouverture d'esprit. Je tiens à te remercier pour les bons moments passés avec toi, la découverte des repas Béninois et les moments d'échanges sur les murs de ton super château. Je remercie également tous les autres thésards et stagiaires avec qui j'ai pu travailler et passer de bons moments : Omar Zenteno, Asma Rebhi, Azy, Antonio, Rania, Khaouter ...

Je tiens tout particulièrement à remercier, sans nommer individuellement, la team Orleans "Team York", "team vendredi" pour le soutien. Comme on dit souvent, nous sommes arrivés en détail mais l'ouverture d'esprit et la tolérance de chacun ont fait de nous une famille. Avec ce jour (vendredi) de partage, d'amitié et de no-stress, vous m'avez permis de m'évader du monde scientifique.

Je tiens également à remercier l'amour de ma vie Tahirou Bah, cette réussite est en grande partie grâce à toi, et je t'en suis sincèrement reconnaissant et je te témoigne toute ma gratitude. Tu as su me soutenir, me supporter, m'encourager,... pendant ma thèse et plus particulièrement durant les derniers mois de rédaction qui n'ont pas toujours été des plus agréables. Cette thèse et moi te devons beaucoup. Merci.

Je remercie Mamadou Sanoussy Barry, Boubacar Sidy Diallo, Ibrahima Diallo et Pathé Bah qui ont pris le temps de relire une grande partie de mes travaux, et qui se sont arrachés les cheveux pour essayer de comprendre le sens souvent tortueux de mes phrases.

J'aimerais exprimer ma gratitude à tous les chercheurs et spécialistes, trop nombreux pour les citer, qui ont pris le temps de discuter de mon sujet. Chacun de ces échanges m'a aidé à faire avancer mon analyse.

Je tiens aussi à remercier mes autres "team" pour leur soutien : team HS, Dar-es-salam, Gamal, René Lévesque, Montpellier, Toulouse, Strasbourg, AJGF, AEGO, ...

J'exprime encore mes remerciements à ceux qui ont pu faire le déplacement, qui ont posé des congés pour assister à ma soutenance (Conakry, Nice, Toulouse, Angers, Versailles, Paris, Bourges, Orléans, ...), vous êtes les meilleurs.

Enfin, je ne pourrais finir ces remerciements sans penser encore une fois à ma famille dont l'affection, l'amour, le soutien et l'encouragement constants m'ont été d'un grand réconfort et ont contribué à l'aboutissement de ce travail.

# Sommaire

Liste des figures	v
Liste des tables	xi
Introduction	1
<b>1 Problématique des adventices</b>	<b>3</b>
1.1 Bref historique de l'agriculture . . . . .	4
1.2 Adventices . . . . .	6
1.3 Désherbage traditionnel . . . . .	7
1.3.1 Désherbage agronomique . . . . .	7
1.3.2 Désherbage manuel . . . . .	7
1.3.3 Désherbage mécanique . . . . .	7
1.3.4 Désherbage chimique . . . . .	7
1.4 Désherbage localisé . . . . .	8
1.4.1 Imagerie aérienne et satellitaire au service du désherbage . . . . .	9
1.4.2 Désherbage robotisé . . . . .	11
1.4.3 Avantage du drone comme support d'imagerie . . . . .	13
1.5 Conclusion . . . . .	13
<b>2 État de l'art : Détection des adventices par la vision artificielle</b>	<b>15</b>
2.1 Segmentation végétation et non végétation . . . . .	16
2.2 Détection des rangées de culture . . . . .	18
2.2.1 Transformée de Hough . . . . .	18
2.2.2 Régression linéaire . . . . .	19
2.2.3 Analyse de régions (blob analysis) . . . . .	20
2.2.4 Méthode des bandes horizontales . . . . .	20
2.2.5 Approche basée sur la stéréoscopie . . . . .	21
2.3 Détection des adventices par les méthodes d'apprentissage machine . . . . .	21
2.3.1 Apprentissage machine classique . . . . .	21
2.3.2 Apprentissage profond . . . . .	25
2.4 Conclusion . . . . .	29
<b>3 Données et systèmes d'acquisition</b>	<b>31</b>
3.1 Contexte d'acquisition des images . . . . .	32
3.2 Drones . . . . .	34

3.2.1	Aile volante eBee de Sensefly . . . . .	34
3.2.2	DJI Phantom 3 . . . . .	35
3.2.3	DJI Inspire 2 . . . . .	36
3.3	Conclusion . . . . .	37
<b>4</b>	<b>Détection des adventices par utilisation de l'information géométrique du champ</b>	<b>39</b>
4.1	Détection des rangs par la transformée de Hough et le SLIC (CR-Hough-SLIC) . . . . .	40
4.1.1	Transformée de Hough . . . . .	41
4.1.2	Squelettisation . . . . .	43
4.1.3	SLIC . . . . .	44
4.1.4	Algorithme de détection des rangées de culture . . . . .	45
4.2	CRoWNet : approche deep learning pour la détection des rangées de culture	46
4.2.1	S-SegNet . . . . .	48
4.2.2	HoughCNet . . . . .	49
4.3	Détection des adventices . . . . .	52
4.4	Expérience et résultats . . . . .	54
4.4.1	Métriques d'évaluation . . . . .	54
4.4.2	Evaluation du CR-Hough-SLIC . . . . .	55
4.4.3	Évaluation du CRoWNet . . . . .	58
4.4.4	Évaluation qualitative . . . . .	59
4.4.5	Évaluation quantitative . . . . .	60
4.4.6	Détection des adventices . . . . .	63
4.5	Conclusion . . . . .	66
<b>5</b>	<b>Détection des adventices par apprentissage</b>	<b>67</b>
5.1	Labellisation non supervisée des données d'apprentissage . . . . .	68
5.2	Apprentissage machine classique . . . . .	69
5.2.1	SVM ou Séparateurs à Vaste Marge . . . . .	69
5.2.2	Forêts d'arbres aléatoires . . . . .	70
5.2.3	Extraction de caractéristiques . . . . .	71
5.3	Apprentissage profond (deep learning) . . . . .	73
5.4	Expériences et résultats . . . . .	74
5.4.1	Extraction de caractéristiques et apprentissage machine . . . . .	77
5.4.2	ResNet . . . . .	80
5.4.3	Classification d'images . . . . .	83
5.5	Conclusion . . . . .	85
<b>6</b>	<b>Apprentissage avec une seule classe</b>	<b>87</b>
6.1	Procédure de collecte des données . . . . .	88
6.2	Extraction de descripteurs profonds . . . . .	89
6.3	SVM à classe unique ou SVM 1-classe . . . . .	90
6.4	Résultats et discussion . . . . .	90
6.5	Conclusion . . . . .	93
	<b>Conclusion et perspectives</b>	<b>95</b>

<b>Publications</b>	<b>99</b>
<b>A Fonctionnement des réseaux de neurones artificiels</b>	<b>101</b>
A.1 Neurones formels . . . . .	101
A.2 Perceptron multicouche (MLP) . . . . .	102
A.2.1 Descente du gradient . . . . .	103
<b>B Résultats détaillés de l'apprentissage machine</b>	<b>107</b>
<b>Bibliographie</b>	<b>112</b>



# Liste des figures

1.1	Adventices dans un champ de maïs. . . . .	6
1.2	Pulvérisation de pesticides et ses conséquences. . . . .	8
1.3	Absorption spectrale de la végétation souce [22]. . . . .	10
1.4	Recensements agricoles et enquête sur la structure des exploitations agricoles 2016. SAU est une notion normalisée dans la statistique agricole européenne qui désigne la surface agricole utilisée. Selon l'INSEE, elle comprend les terres arables (y compris pâturages temporaires, jachères, cultures sous abri, jardins familiaux, ...), les surfaces toujours en herbe et les cultures permanentes (vignes, vergers, ...). Image issue de [135]. . . . .	12
2.1	Résultats obtenus dans l'article [85]. (a) Image agronomique simulée. (b) Le résultat de la transformée de Hough sur (a). (c) Résultat de la détection des rangées. . . . .	19
2.2	Exemple de réseau de convolutionnel profond [1]. . . . .	26
2.3	Evolution du taux d'erreur en fonction des années, de l'approche et de la profondeur des réseaux [3]. . . . .	27
2.4	Application d'une convolution sur une image [96]. . . . .	28
2.5	Exemple de sous-échantillonnage par max-pooling [54]. . . . .	28
2.6	Exemple de trois modèles de mauvaises herbes obtenus à partir de la texture d'une seule feuille [40]. . . . .	29
3.1	Image du champ de maïs acquise à 100 m avec une aile volante et une caméra couleur. La résolution spatiale des images est de 2,5 cm. . . . .	33
3.2	Proportion de végétation minimales et maximales en pourcentage présentes dans le pixel ayant le plus grand taux de végétation en fonction de la résolution spatiale de l'image et de la taille de l'adventice est répertoriée [109]. . . . .	34
3.3	Dégradation de l'image des taux de végétation de 6 mm à 6 cm de résolution spatiale [109]. . . . .	34
3.4	Images acquises dans les champs de haricot et d'épinard à 20 m d'altitude avec une caméra Sony. La résolution spatiale est inférieur à 1 cm. . . . .	35
3.5	Image du champ de betterave. . . . .	36
3.6	Exemple d'images RGB (a) et CIR (b) provenant de la base publique utilisée [142]. . . . .	36
3.7	L'aile volante eBee de Sensefly (a), les drones DJI Phantom 3 Pro (b) et DJI Inspire 2 (c). . . . .	37

4.1	Organigramme de la méthode de détection des rangées de culture combinant transformée de Hough et superpixels. . . . .	40
4.2	Segmentation d'une image provenant d'un champ de maïs (a) et sa segmentation à l'aide du seuillage de Otsu appliqué sur l'indice ExG (b). . . .	41
4.3	Exemple de fonctionnement de la transformée de Hough pour détecter une droite à partir de deux points situés dans une image. . . . .	42
4.4	(a) Image segmentée et son squelette en bleu. (b) et (c) représentent la transformée de Hough appliquée sur une image segmentée et sur son squelette, respectivement. La durée de traitement a été réduite de 78 %. . . .	43
4.5	(a) Espace colorimétrique CIELAB. La clarté L dérive de la luminance de la surface ; les deux paramètres a et b expriment l'écart de la couleur par rapport à celle d'une surface grise de même clarté. (b) Illustrations des zones de recherche du centre du cluster avec le SLIC. . . . .	44
4.6	Exemple de 1021 superpixels (contours en bleu) générés à partir de l'image couleur (image avant segmentation) (a) et sur son image segmentée (b). . .	45
4.7	Exemple de résultat de détection de rangées de culture. . . . .	46
4.8	Organigramme pour la détection des rangs de culture avec CNN (CRowNet). . . . .	48
4.9	Architecture du SegNet. . . . .	50
4.10	Organigramme du processus HoughCNet. En entrée, nous avons une image segmentée et en sortie, nous obtenons des rangées de culture après la combinaison de la transformée de Hough et du CNN. . . . .	50
4.11	(a) Image binaire en blanc et son squelette en rouge. (b) Transformée de Hough effectuée sur le squelette. (c) Lignes détectées avec la transformée de Hough où chaque ligne a une valeur correspondant au nombre de votes dans l'espace de Hough ( $\mathcal{L}$ ). . . . .	51
4.12	Exemples de trois filtres dilatés. (a) Filtre $f_1$ utilisé dans la convolution dilatée 1 ; chaque élément dans $f_1$ a un champ réceptif de $3 \times 3$ . (b) Filtre $f_2$ produit à partir de $f_1$ par convolution dilatée 2 ; champ récepteur pour chaque élément est $7 \times 7$ . (c) Ce filtre $f_3$ est obtenu de $f_2$ par une convolution dilatée 4 ; éléments dans $f_3$ ont un champ récepteur de $15 \times 15$ . Les couches de convolution dilatées générées par ces filtres ont le même nombre de paramètres. . . . .	51
4.13	Architecture du réseau convolutionnel utilisé pour le HoughCNet. . . . .	52
4.14	(a) Image originale. (b) Masque de rangées de culture en blanc et en rouge les superpixels qui n'intersecte pas les rangées détectées. (c) Résultat de la détection en exploitant la relation entre superpixels et (d) Résultat de la détection d'adventices après avoir utilisé le masque de rangée de culture (en rouge les adventices potentielles et en magenta les adventices inter-rangs). . . . .	53
4.15	Exemple de détection de rangées de culture dans le champ de maïs avec une compacité de 5 et une taille de superpixel variable. (a) correspond à l'image originale. Les 5 autres images (b, c, d, e et f) correspondent respectivement au résultat de la détection des rangées de culture avec $S = 32, 15, 10, 5, 4$ ( $k = 0,1\%N, 0,5\%N, 1\%N, 5\%N, 10\%N$ ). Pour rappel, $N$ correspond au nombre de pixels total de l'image. . . . .	56

- 4.16 Exemple de détection de rangées de culture dans le champ de betterave avec une compacité de 5 et une taille de superpixel variable. (a) correspond à l'image originale. Les 5 autres images (b, c, d, e et f) correspondent respectivement au résultat de la détection des rangées de culture avec  $S = 32, 15, 10, 5, 4$  ( $k = 0,1\%N, 0,5\%N, 1\%N, 5\%N, 10\%N$ ). Pour rappel,  $N$  correspond au nombre de pixels total de l'image. . . . . 57
- 4.17 Exemple de données d'apprentissage. (a) L'image originale  $X$ . (b) Rangées détectées avec la transformée de Hough  $\mathcal{L}$ . (c) Vérité terrain. Dans la deuxième ligne de la figure, nous avons l'image  $X$  après le changement de contraste. (d) Réglage du contraste : les 1 % de pixels sont saturés à basse et haute intensité pour chaque canal de  $X$ . (e) Le canal vert de l'image RGB est remplacé par le résultat de la segmentation de l'arrière-plan effectuée avec l'indice ExG et le seuillage de Otsu (BW). (f) Les canaux bleu et rouge sont réglés sur 0 et le canal vert correspond à BW. . . . . 59
- 4.18 Exemples de détection de rangées de culture sur des images prises dans un champ de maïs. (a) Image originale. (b) Superposition des rangées de culture détectées avec S-SegNet (en bleu) et l'image originale après changement de contraste. (c) Superposition des rangées de culture détectées (en bleu) avec CRowNet et l'image originale après changement de contraste. . . 60
- 4.19 Détection des rangées de culture à l'aide de S-SegNet et de CRowNet. (a) Image originale. (b) et (c) sont les résultats de S-SegNet et CRowNet respectivement. La couleur blanche correspond aux pixels de la vérité terrain qui ont été bien détectés, le magenta aux pixels qui sont sur la rangée prédite et n'appartenant pas à la vérité terrain (surdétection), et le vert les pixels non détectés. . . . . 60
- 4.20 (a) et (b) sont respectivement les résultats de CRowNet après squelettisation puis dilatation. La couleur blanche correspond aux vrais positifs (VPs), le magenta aux faux positifs (FPs), et le vert aux faux négatifs (FNs). (c) Nombre de rangées de culture détectées en fonction de la qualité de la segmentation. CR-Hough est le résultat obtenu avec la transformée de Hough et les informations de champ [85, 59]. CR-Hough-SLIC est le résultat obtenu avec la combinaison de la transformée de Hough et SLIC [13]. . . . . 61
- 4.21 Rangées de culture bien détectées en fonction de la qualité de la segmentation. (a) et (b) présentent les résultats obtenus à partir des données recueillies dans les champs de betterave et de maïs. . . . . 63
- 4.22 Les trois colonnes présentent la vérité terrain et le résultat de la méthode de détection d'adventices sur trois images différentes avec des WIR. La première ligne présente les images originales. La deuxième ligne présentent la vérité terrain (en rouge les adventices, en vert les cultures et en blanc la délimitation des rangées de culture. La dernière ligne présentent le résultat après la combinaison de la relation entre superpixels et le masque représentant les rangées de culture. Le magenta représente les adventices inter-rangs, le rouge les adventices trop proches des rangées de culture (adventices potentielles) et en cyan les indécis. . . . . 65

---

5.1	Illustration du fonctionnement du SVM. . . . .	69
5.2	Fonctionnement du random forest. . . . .	70
5.3	Différentes orientations. . . . .	72
5.4	Fonctionnement d'un réseau résiduel. . . . .	74
5.5	(a) Image originale et rangées de culture détectées (en bleu). (b) Adventices potentielles (en bleu) et inter-rang en rouge. Parties du champ de haricot (c) et du champ d'épinard (d) avec les adventices étiquetées manuellement en rouge. L'étiquetage manuel a pris environ 2 jours et moins de 30 minutes avec la méthode de labellisation non supervisée sur 4 images de tailles $N=7360 \times 4912$ pixels. . . . .	75
5.6	Exemple d'échantillons de culture et d'adventices de taille $64 \times 64$ pixels avec et sans fond. Haricot : échantillons de culture (a et b), échantillons d'adventices (c et d). Épinards : échantillons de culture (e et f) et échantillons d'adventices (g et h). . . . .	76
5.7	Courbes ROC des modèles SVM créés par chaque caractéristique pour chaque champ. La première ligne représente le champ d'épinard et la seconde le champ de haricot. La première et la deuxième colonnes correspondent respectivement à l'apprentissage effectuée sur les données supervisées et non supervisées. TVP : taux de vrais positifs et TFP : taux de faux positifs. . . . .	78
5.8	Courbes ROC des modèles RF créés par chaque caractéristique pour chaque champ. La première ligne représente le champ d'épinard et la seconde le champ de haricot. La première et la deuxième colonnes correspondent respectivement à l'apprentissage effectuée sur les données supervisées et non supervisées. TVP : taux de vrais positifs et TFP : taux de faux positifs. . . . .	79
5.9	Évolution de la fonction perte (fonction de coût) lors de l'apprentissage sur les données avec annotations supervisées et non supervisées dans les parcelles d'épinard et de haricot. train et val représentent respectivement les courbes d'apprentissage et de validation. Les courbes de perte lors de la validation (val) diminuent pendant environ les 80 premières epochs avant d'augmenter puis converger. La première ligne représente le champ d'épinard et la deuxième le champ de haricot. La première et la deuxième colonnes sont respectivement l'apprentissage sur les données supervisées et non supervisées. . . . .	81
5.10	Courbes ROC des données de test avec l'étiquetage non supervisé et supervisé des données. De gauche à droite, les courbes ROC calculées à partir des données de test du haricot (a) et des épinards (b). Dans le champ de haricot, les aires sous la courbe AUC sont de 88,73 % pour les données non supervisées et de 94,84 % pour les données supervisées. Dans le champ d'épinard, les AUC sont de 94,34 % pour les données non supervisées et de 95,70 % pour les données supervisées. Par données supervisées et non supervisées, nous entendons respectivement les données étiquetées par un humain et étiquetées automatiquement. TVP et TVF représentent respectivement les taux de vrais et faux positifs. . . . .	82

5.11	Courbes ROC des données de test avec les données sur les mauvaises herbes du champ de haricot échangées avec celles du champ d'épinard. De gauche à droite, les courbes ROC obtenues à partir des données de test du haricot (a) et des épinards (b). Dans le champ de haricot, les aires sous la courbe AUC sont de 91,37 % pour les données non supervisées et de 93,25 % pour les données supervisées. Dans le champ d'épinard, l'aire sous la courbe (AUC) est de 82,70 % pour les données non supervisées et de 94,34 % pour les données supervisées. Par données supervisées et non supervisées, on entend respectivement les données étiquetées de manière supervisée et non supervisée. . . . .	83
5.12	Exemples de classification d'images de drones avec des modèles créés par des données non supervisées dans deux champs différents. De la première à la deuxième ligne, nous avons des échantillons de culture d'épinard et de haricot. Sur la première colonne, nous avons les échantillons obtenus après l'utilisation de la fenêtre coulissante, sans les lignes de culture et les informations de fond. Les points bleus, rouges et blancs signifient que les plantes sont identifiées respectivement comme des cultures, des mauvaises herbes et des décisions incertaines. Dans la deuxième colonne, nous avons les mauvaises herbes détectées en rouge après l'application des lignes de culture et des informations de base. . . . .	84
6.1	(a) Exemple d'une image en fausse couleur (CIR), dont les 3 canaux RGB sont respectivement le proche infrarouge, le vert et le rouge. Sur l'image (b), la végétation est représentée en blanc. . . . .	88
6.2	(a) et (b) sont respectivement des images avec superpixels (limites blanches) générées avec la méthode SLIC (compacité = 10 et nombre de superpixels = 1000) et la méthode watershed appliquée sur la version dilatée de l'image segmentée avec un élément structurant sphérique de rayon 5. . . . .	89
6.3	Architecture de ResNet18 avec extraction des descripteurs profonds. Les flèches rouges indiquent où les descripteurs ont été extraits. Les rectangles horizontaux indiquent la taille à la sortie des couches. . . . .	90
6.4	Exemple d'échantillons collectés (en RGB). Les rectangles verts et rouges sont respectivement des échantillons de culture et de mauvaises herbes. . .	91
6.5	AUC obtenue avec chaque descripteur en utilisant US0-4 comme base d'apprentissage et S0-4 comme base de test. RGB et CIR signifient que le modèle a été entraîné sur des données en vraies et fausses couleurs, respectivement. . . . .	92
A.1	Exemple de neurones ou de perceptron à une couche cachée. . . . .	101
A.2	Perceptron multicouche à deux couches cachées et une couche de sortie. . .	102
A.3	Exemples de frontières de décision : (a) convexe ouverte ; (b) convexe fermée ; (c) concave ouverte ; et (d) concave fermée. . . . .	103
A.4	Exemple d'un réseau de neurones avec 2 entrées, 1 couche cachée composée de 2 neurones et 1 couche de sortie à 1 neurone. . . . .	104



# Liste des tableaux

2.1	Indices de couleur pour la segmentation végétation et non végétation. . . .	17
4.1	Résultat moyen obtenu avec $k = 0.5\%N$ et $0.1\%N$ respectivement dans les champ de maïs et de betterave avec les différentes métriques. $N$ correspond au nombre de pixels total de l'image. . . . .	57
4.2	Table de pondération de chaque classe et par type de réseau. . . . .	58
4.3	Résultats quantitatifs de la détection des rangs de culture avant leur élargissement dans le champ de betterave. La moyenne est calculée pour chaque rang. . . . .	61
4.4	Moyenne des résultats des méthodes de détection des rangées de culture obtenues après avoir ramenée les rangées de culture à la même largeur. FCN-W et FCN-W-16 signifient respectivement des FCN avec un stride 8 (FCN-8) et 16 (FCN-16) avec des classes rééquilibrées. . . . .	62
4.5	IoU moyenne en fonction du taux d'infestation par les adventices. . . . .	63
5.1	Nombre d'opérations en virgule flottante par seconde (FLOPS) par architecture (en milliard) [71]. . . . .	74
5.2	Données d'apprentissage et de validation dans les champs de haricot et d'épinard. . . . .	77
5.3	Nombre d'échantillons de test utilisés pour chaque champ. . . . .	77
5.4	Résultats des données de test recueillies dans le champ de haricot avec ResNet18, SVM et Random Forest (RF). Pour le SVM et le RF, seuls les résultats des meilleures caractéristiques sélectionnées sont présentés. . . . .	80
5.5	Résultats des données de test recueillies dans le champ d'épinard avec ResNet18, SVM et Random Forest (RF). Pour le SVM et le RF, seul les résultats des meilleures caractéristiques sélectionnées sont présentés. . . . .	81
6.1	Échantillons prélevés pour l'apprentissage et le test. . . . .	91
6.2	AUC des modèles obtenus avec SVM 1-classe et ResNet18. Les Model0000 à Model004 sont les modèles entraînés par SVM 1-classe sur les 5 bases de données non supervisées. Le Model0-4 est le modèle entraîné sur la fusion des 5 bases de données d'apprentissage. ResNet18-000 est le modèle créé avec la base de données S000. . . . .	93
B.1	SVM avec des noyaux gaussien (RBF) et linéaire sur les données supervisées et non supervisées dans le champ de haricot. . . . .	108

## LISTE DES TABLEAUX

---

B.2	Résultat du random forest (RF) avec 100 et 200 arbres sur les données supervisées et non supervisées dans le champ de haricot. . . . .	109
B.3	SVM avec des noyaux gaussien (RBF) et linéaire sur les données supervisées et non supervisées dans le champ d'épinard. . . . .	110
B.4	Résultat du random forest (RF) avec 100 et 200 arbres sur les données supervisées et non supervisées dans le champ d'épinard. . . . .	111

# Introduction

Un des grands enjeux de demain est de renforcer la productivité agricole mais surtout d'améliorer la qualité des productions tout en réduisant fortement l'utilisation des intrants. Relever ce défi, notamment grâce aux progrès scientifiques et technologiques, est un objectif recherché par tous les pays. Aujourd'hui, pour augmenter leurs productivités dans l'objectif de nourrir une population de plus en plus importante dans un environnement souvent hostile (changements climatiques, insectes, adventices, ...), les agriculteurs ont souvent recours aux produits chimiques (engrais, pesticides, ..). Cependant, une pulvérisation non maîtrisée de ces produits est néfaste pour l'homme et l'environnement. Par exemple, en avril 2009, une étude a montré qu'habiter à moins de 500 mètres d'une surface agricole sur laquelle des pesticides avaient été épandus augmentait le risque de maladie de Parkinson de 75% [34]. Pour les agriculteurs, les pertes causées chaque année par les adventices et les insectes peuvent atteindre 40 % des rendements globaux des cultures et ce pourcentage devrait augmenter de manière significative dans les années à venir [48]. Au vu des conséquences des pesticides sur l'environnement, le gouvernement français, à travers son plan EcoPhyto 2+, ambitionne d'ici 2025 de diminuer de 50 % la consommation de pesticides [2]. Ainsi, pour limiter la quantité de produits chimiques tout en continuant d'augmenter la productivité, un certain nombre d'agriculteurs se tournent vers l'agriculture de précision. L'agriculture de précision consiste à utiliser un ensemble de techniques et de pratiques pour mieux comprendre et analyser les besoins, le fonctionnement et l'organisation physique et biochimique des cultures et développer des outils d'aide à la décision (OAD).

Ces dernières années, l'utilisation des nouvelles technologies dans le monde agricole s'est considérablement accrue (capteurs de télédétection, robots, outils de cartographie, matériel agricole, ...). Par rapport aux acquisitions robotisées ou satellitaires, les drones ont été considérés comme plus efficaces car ils permettent une acquisition rapide de la parcelle cultivée avec une très haute résolution spatiale et un faible coût [153, 171]. Cependant, les applications qui en découlent concernent principalement la modulation de la fertilisation azotée, l'irrigation ou la prévision des rendements, mais très peu le désherbage. Ainsi, pour l'étude d'un désherbage localisé, cette thèse a été effectuée dans le cadre d'un projet de recherche financé par la Région Centre-Val de Loire.

Ce manuscrit est structuré en six chapitres. Le premier chapitre présente les problématiques du désherbage. Les enjeux de la gestion des adventices y sont détaillés ainsi que les moyens de lutte actuellement mis en place. Un bref historique des évolutions technologiques dans le secteur agricole ainsi que les avancées liées à l'agriculture de précision y sont décrites.

Le deuxième chapitre présente un état de l'art sur les travaux déjà réalisés autour de

la détection des adventices par traitement d'images. Les méthodes de détection d'adventices basées sur l'analyse spatiale et spectrale ainsi que les méthodes d'apprentissage sont étudiées. Nous entendons par analyse spatiale l'ensemble des méthodes exploitant l'alignement des cultures pour détecter les adventices situées dans l'espace inter-rang. L'analyse spectrale, quant à elle, regroupe les méthodes utilisant le spectre d'imagerie à travers des méthodes de segmentation et d'apprentissage. L'objectif principal de ce chapitre est de faire ressortir les approches développées ou majoritairement utilisées pour la détection des adventices et leur limitation.

Le troisième chapitre décrit les images manipulées dans cette thèse ainsi que les systèmes d'acquisition utilisés. Ces images ont servi à développer les différentes méthodes proposées dans cette thèse.

Deux nouvelles méthodes de détection des rangées de culture ont été développées et sont présentées dans le quatrième chapitre. Ces méthodes sont utilisées pour la détection des adventices en appliquant une analyse spatiale. Cependant, les méthodes utilisant l'approche spatiale sont généralement limitées à la détection des adventices inter-rang. Pour détecter les adventices inter-rang et intra-rang, deux nouvelles approches d'apprentissage ont été explorées : une première approche qui s'appuie sur les méthodes d'apprentissage machine (*machine learning*) classiques (sélection manuelle de descripteurs et de machine d'apprentissage) et une seconde qui utilise l'apprentissage profond (*deep learning*). Pour fonctionner correctement, les algorithmes d'apprentissage supervisé ont besoin d'un volume très important de données d'apprentissage. Mais la création de grands ensembles de données agricoles avec des annotations au niveau des pixels est une tâche qui prend énormément de temps. Compte tenu du problème de collecte de données, une méthode de détection des adventices entièrement automatisée est proposée dans le cinquième chapitre.

Toujours dans le but de réduire le temps d'annotation, une méthode employant un classifieur à classe unique sur des échantillons de culture collectés de façon non supervisée est présentée dans le sixième chapitre.

Une conclusion synthétise l'ensemble de ces travaux ainsi que les perspectives d'applications liées à la production de cartes de localisation d'adventices et au développement d'un service de gestion localisé par drone.

# Chapitre 1

## Problématique des adventices

### Sommaire

---

<b>1.1</b>	<b>Bref historique de l'agriculture . . . . .</b>	<b>4</b>
<b>1.2</b>	<b>Adventices . . . . .</b>	<b>6</b>
<b>1.3</b>	<b>Désherbage traditionnel . . . . .</b>	<b>7</b>
1.3.1	Désherbage agronomique . . . . .	7
1.3.2	Désherbage manuel . . . . .	7
1.3.3	Désherbage mécanique . . . . .	7
1.3.4	Désherbage chimique . . . . .	7
<b>1.4</b>	<b>Désherbage localisé . . . . .</b>	<b>8</b>
1.4.1	Imagerie aérienne et satellitaire au service du désherbage . . . .	9
1.4.2	Désherbage robotisé . . . . .	11
1.4.3	Avantage du drone comme support d'imagerie . . . . .	13
<b>1.5</b>	<b>Conclusion . . . . .</b>	<b>13</b>

---

L'agriculture a commencé au moment où l'homme a décidé de passer de l'économie de cueillette, de la chasse et de la pêche à celle de la culture des terres avec des méthodes qui varient d'une région à une autre. Les premières pratiques agricoles consistaient à cultiver une parcelle jusqu'à l'épuisement du sol, puis à changer de parcelle agricole, et à y revenir lors des prochaines campagnes agricoles. Ce type d'agriculture a connu une évolution lente qui s'est perpétué sur des siècles. Du Moyen âge à nos jours, le secteur agricole a connu des avancées technologiques considérables. De plus avec le développement rapide de l'industrie des capteurs et des systèmes de télécommunication (l'arrivée de la 5G, l'internet des objets, ...) de nouvelles avancées sont donc à prévoir.

Ce chapitre est subdivisé en 5 sections. La première section présente un bref historique des évolutions du secteur agricole. Dans les deuxième et troisième sections, les enjeux de la gestion des adventices y sont détaillés, tout comme les moyens de lutte actuellement mis en place. La quatrième section décrit les principes du désherbage localisé ainsi que les techniques d'imagerie et les outils employés. Enfin la cinquième section conclut ce chapitre.

### 1.1 Bref historique de l'agriculture

A partir de la fin du Moyen-Age, l'assolement triennal est devenue la méthode la plus pratiquée, c'est-à-dire que 2/3 des terres étaient utilisées et 1/3 étaient laissées au repos (jachère) [39, 32]. A la fin du XIII<sup>e</sup> siècle, suite à une forte croissance démographique qui a vu la population européenne multipliée par deux, voire par trois (70 à 80 millions de personnes), cette pratique a montré ses limites suite à l'incapacité à augmenter les surfaces cultivées et la productivité agricole. Elle se révéla d'ailleurs incapable de nourrir la surpopulation française de 1350 (20 millions d'habitants), ouvrant le champ à un siècle de famines, de guerres et d'épidémies, accentuée par la malnutrition [41, 117].

Au début du XVIII<sup>e</sup> siècle, de grands progrès ont commencé à être enregistrés dans l'agriculture : ce fut la "première révolution agricole". Cette première révolution agricole est parfois divisée en deux révolutions : la révolution fourragère, au XVIII<sup>e</sup> siècle, puis la révolution de la mécanisation, suite à la révolution industrielle, au XIX<sup>e</sup> siècle [63]. Elle a d'abord commencé un siècle plus tôt aux Pays-Bas avant de passer en Angleterre puis d'arriver en France au début du XIX<sup>e</sup> siècle. La révolution fourragère était fondée sur le seul remplacement des jachères par les prairies artificielles [160]. Cette technique permettait d'utiliser toutes les terres tout en alternant les cultures sur les 3 portions de terres d'une année à l'autre. Sur la première portion, on cultivait des céréales ; sur la deuxième portion, on introduisait des pommes de terre, des betteraves, des navets, pour varier l'alimentation et enfin, sur la dernière, on cultivait des légumineuses (trèfle) au lieu de la jachère. Le trèfle fixait l'azote de l'atmosphère en une forme d'engrais et servait également à nourrir les animaux qui, pour leur part, produisaient du fumier permettant de renforcer la fertilisation du sol. C'est entre les années 1830 et 1850 que la science agronomique moderne a pris son envol, aussi bien en France qu'en Angleterre et en Allemagne. En 1834, l'industriel américain Cyrus McCormick breveta la première moissonneuse-batteuse mécanique. Dans les années 1840, l'industriel allemand Justus von Liebig fit des contributions majeures à la chimie organique et à l'agronomie, et créa les premiers engrais chimiques. L'introduction de l'engrais chimique a permis d'augmenter la production végétale et donc

aussi la production animale de 400 % [47]. L'augmentation des rendements a eu pour conséquence la volonté de transformer et d'exporter les produits agricoles. Les progrès agricoles ont joué un rôle important dans la révolution industrielle et seraient même nécessaires à tout progrès industriel. Avec l'amorce de la révolution industrielle à la fin du XIX<sup>e</sup> siècle, de nouvelles technologies ont émergé pour donner naissance à une nouvelle forme d'agriculture : ce fut le début de la deuxième révolution agricole. Cette deuxième révolution agricole reposa sur une utilisation accrue de la technologie. Les agriculteurs ont commencé à produire plus avec moins de main-d'œuvre. Les machines ont remplacé petit à petit les travailleurs, rendant les fermes plus efficaces pour produire de la nourriture. Les fermes qui étaient autrefois communales ont commencé à se privatiser et les gens se sont éloignés de l'agriculture de subsistance. Les chemins de fer installés lors de la révolution industrielle ont permis aux agriculteurs de vendre leurs récoltes dans d'autres villes. On considère que le processus de la révolution agricole a entretenu par ailleurs des liens de causes à effet, réciproques et étroits, avec la révolution industrielle et la transition démographique. Cependant, la modernisation de l'agriculture s'est traduite par une utilisation accrue d'énergie et d'intrants (engrais, phytosanitaires, eau). Une dérogation majeure pour l'environnement commença à se produire à mesure que l'économie passa de l'agriculture durable à l'industrie manufacturière et à d'autres industries polluantes. Dès les années 1980, des études scientifiques ont commencé à interpeller les décideurs sur les conséquences d'une utilisation non contrôlée de ces outils. Wes Jackson [82] a écrit dans un examen des pratiques qui ont poussé l'agriculture américaine au bord du désastre "La charrue pourrait bien avoir détruit plus d'options pour les générations futures que l'épée". Il préconisa l'adoption d'une nouvelle tendance appelée agriculture durable dont l'objectif est de répondre aux besoins alimentaires de la société sans compromettre la capacité des générations futures à satisfaire leurs propres besoins. Les adeptes de l'agriculture durable cherchent à intégrer trois objectifs principaux dans leur travail : un environnement sain, la rentabilité économique et l'équité sociale et économique [79]. L'agriculture durable recommande la sélection des espèces et des variétés bien adaptées aux sites et aux conditions de la ferme, une diversification des cultures (y compris l'élevage) et des pratiques culturales pour améliorer la stabilité biologique et économique de la ferme. Pour atteindre ces objectifs, une nouvelle forme d'agriculture commence à émerger, appelé agriculture de précision. L'agriculture de précision permet de mieux comprendre et d'analyser les besoins, le fonctionnement et l'organisation physique et biochimique des cultures et de développer des outils d'aide à la décision pour les acteurs du domaine agricole. Cette nouvelle approche de l'agriculture s'appuie sur l'exploitation des technologies de l'information et de la communication afin de caractériser le milieu (relief, sol, couvert végétal, ...) dans toute sa variabilité spatiale [10]. L'utilisation des nouvelles technologies dans l'agriculture s'est considérablement développée ces dernières années (capteurs de télédétection, robots, outils de cartographie, matériel agricole, ...). Les applications concernent principalement la modulation de la fertilisation azotée, l'irrigation ou la prévision des rendements, mais très peu le désherbage. Les méthodes utilisées pour la lutte contre les mauvaises herbes, communément appelées adventices, continuent non seulement à coûter cher aux agriculteurs mais représentent aussi une menace pour l'homme et l'environnement.

## 1.2 Adventices

Depuis l'apparition de l'agriculture jusqu'à l'agriculture intensive moderne, la lutte contre les mauvaises herbes a toujours été considérée comme un souci majeur. En effet, il n'y a pas de mauvaises herbes mais seulement des plantes qui poussent dans les champs, qui captent la lumière et les nutriments des plantes cultivées et/ou qui entravent la récolte (Figure 1.1); ces plantes sont communément appelées **adventices** [124]. En 2009, d'après l'organisation de recherche environnementale de la Nouvelle Zélande *Land Care*, les adventices sont à l'origine d'environ 95 milliards de dollars de pertes sur la production vivrière à l'échelle mondiale, contre 85 milliards de dollars pour les agents pathogènes, 46 milliards de dollars pour les insectes et 2.4 milliards pour les vertébrés. Parmi ces 95 milliards de dollars de pertes dues aux adventices, on peut comptabiliser 380 millions de tonnes de blé, soit plus de la moitié de la production mondiale prévue en 2009. Au vu de ces chiffres, l'Organisation des Nations Unies pour l'Alimentation et l'Agriculture (FAO) considère que les adventices devraient être reconnues comme l'ennemi naturel numéro 1 des agriculteurs [51].

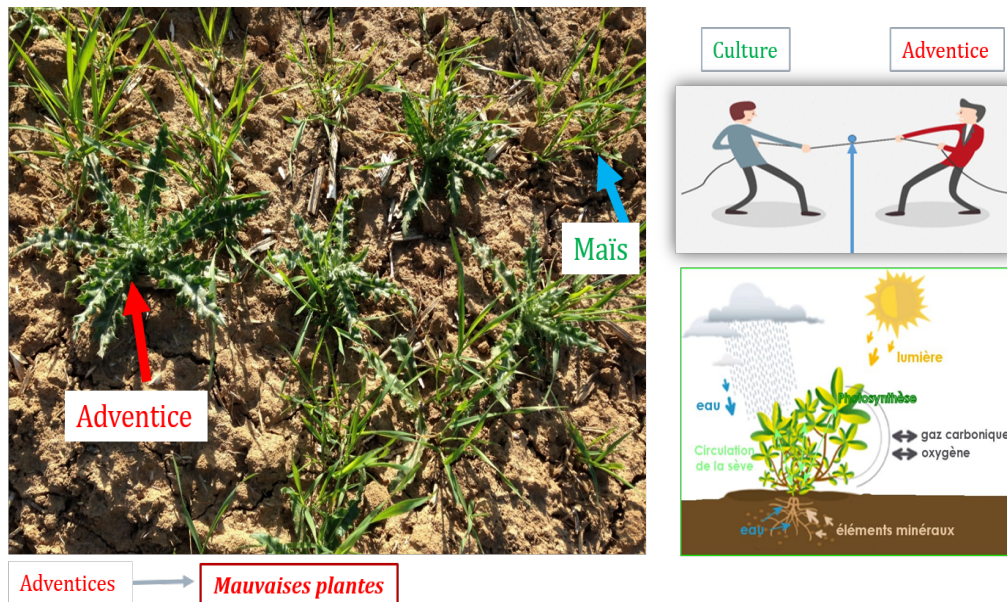


FIGURE 1.1 – Adventices dans un champ de maïs.

Cependant comme dans toutes les batailles, il est primordial de connaître son ennemi. C'est la raison pour laquelle, face aux conséquences néfastes des adventices, des chercheurs se sont penchés très tôt sur l'étude de l'interaction entre les cultures et les adventices, d'où la naissance de **la malherbologie** [18, 154]. Cette science permet de comprendre les facteurs botanique, physiologique et génétique qui influencent le développement des espèces ou des communautés d'adventices. Il devient ainsi possible d'exploiter ces facteurs pour maintenir l'enherbement d'une parcelle en-dessous d'un seuil de nuisibilité globale [27, 114]. Ce seuil de nuisibilité est défini comme étant le niveau d'infestation à partir duquel une baisse de rendement de la culture est mesurée pour une espèce d'adventice donnée. Le seuil de nuisibilité prend en compte non seulement la perte de rendement mais aussi les coûts du désherbage.

### 1.3 Désherbage traditionnel

La gestion des adventices peut s'opérer à différentes échelles de temps et combiner des interventions à des stades de développement différents de la culture ou des adventices. En France, différents types de lutte contre les adventices ont été mis en place.

#### 1.3.1 Désherbage agronomique

Son but est d'utiliser les principes de la nature pour lutter contre les adventices [94, 31]. Nous avons le *mulching* ou paillis qui consiste à étouffer les adventices en mettant une barrière physique entre les adventices et le soleil et l'air. De même, le faux semis accompagné du labour est utilisé pour stimuler la levée des adventices. Avec le labour les plantes nuisibles et leurs graines qui sont enfouies dans le sol sont éliminés. Varier les périodes de semis présente un effet bénéfique en décalant la levée des cultures en dehors des dates de germination privilégiées des adventices. Certaines espèces de plantes sont parfois associées à la culture principale pour renforcer la résistance aux mauvaises herbes. La rotation des cultures est également utilisée pour limiter l'accumulation des populations de mauvaises herbes. Elle modifie donc les conditions de croissance d'une année à l'autre, une situation à laquelle peu d'espèces de mauvaises herbes s'adaptent facilement.

#### 1.3.2 Désherbage manuel

Il consiste à parcourir le champ en éliminant adventice après adventice à la main ou avec un instrument de labour tel que la binette, le couteau désherbeur, ... Cette technique est écologique et très efficace pour éliminer les adventices à des stades précoces. Elle demande cependant une main-d'œuvre importante et n'est pas adaptée quand il s'agit d'une grande surface agricole.

#### 1.3.3 Désherbage mécanique

Le désherbage mécanique est une bonne alternative au désherbage manuel. Il est pratiqué grâce à des engins agricoles tels que la bineuse, la herse stérile, la houe rotative, ... Le désherbage mécanique est appliqué le plus souvent sur les cultures appelées sarclées puisque l'inter-rang est suffisamment grand pour y passer un outil. Parmi les cultures sarclées, nous pouvons citer le maïs, la betterave ou encore la pomme de terre.

#### 1.3.4 Désherbage chimique

Ce type de désherbage se fait avec un herbicide organique de synthèse traditionnellement pulvérisé de façon uniforme dans tout le champ, tout en espérant éliminer les plantes nuisibles [31]. Cette technique est la plus répandue actuellement ; elle est parfois combinée au désherbage mécanique pour réduire la quantité d'herbicide : c'est le *désherbinage*. En 2010, les dépenses de pesticides dans l'agriculture française étaient estimées à 2 475 millions d'euros [25].



FIGURE 1.2 – Pulvérisation de pesticides et ses conséquences.

Une pulvérisation non maîtrisée est cependant néfaste pour l'homme et l'environnement [23, 129], puisque les particules chimiques s'infiltrant dans l'eau, l'air ainsi que dans les cultures (Figure 1.2).

Il est bien connu que les mauvaises herbes ne poussent pas uniformément dans un champ ; il existe en effet une variabilité spatiale significative de la densité des plantes adventices et de leur type [152]. Par exemple, dans [163], Wilson and Brain avaient mené une étude sur le pâturin noir ; ils avaient rapporté que sur une période de 10 ans, le pâturin noir avait poussé dans des zones bien définies et de façon stable. Dans [115], les auteurs avaient fait le même constat sur trois espèces de graminées adventices. Afin de limiter la quantité de produits chimiques tout en continuant d'augmenter la productivité, un certain nombre d'agriculteurs se sont tournés vers l'agriculture de précision.

## 1.4 Désherbage localisé

On parle de désherbage localisé lorsque nous disposons d'un moyen qui permet de désherber d'une manière locale en utilisant des moyens d'identification et de localisation des zones à désherber. L'apparition du GPS (Global Positioning System) dans les années 1990 a été le moteur de cette pratique. Le système de localisation GPS est un système géré par le département américain de la défense et constitué par un réseau de 24 satellites qui permet aux utilisateurs de récepteurs GPS de se localiser en tout point dégagé du globe. L'usage de ce système dans l'agriculture de précision ne s'arrête pas à la cartogra-

phie des parcelles, il est également mis en œuvre pour l'exécution automatique de cartes d'interventions préprogrammées, comme les cartes de fertilisation, en fournissant à chaque instant la position du matériel dans la parcelle [158, 175]. En 1996, Statford *et al.* [148] ont proposé un outil de cartographie des mauvaises herbes en partant de l'hypothèse que l'agriculteur possède un registre mental qualitatif sur la répartition des adventices dans ses champs et le fait qu'il parcourt régulièrement son champ. Cet outil permet à l'agriculteur de faire sa propre carte d'infestation en renseignant l'espèce et la densité d'adventices lors de ses déplacements dans ses champs à l'aide d'un ordinateur portable et d'un GPS qu'il embarque dans un sac à dos. Toutefois, ce système peut facilement atteindre ses limites dans des champs de plusieurs hectares, d'où la nécessité d'utiliser des engins motorisés. Rew *et al.* [139] ont proposé d'utiliser un motorcycle avec deux opérateurs qui sont équipés d'un GPS et d'un ordinateur de bord pour créer la carte d'infestation. Dans [111], Lutzman *et al.* suggère l'utilisation d'un opérateur avec un enregistreur vocal pour nommer les adventices et d'un système de reconnaissance vocale pour créer une carte d'infestation à l'aide d'un GPS.

Certes, l'évaluation visuelle des adventices a montré un intérêt pour la création des cartes d'infestation ; les estimations visuelles sont néanmoins subjectives et peuvent avoir plusieurs sources d'imprécision, à la fois de la vision humaine en soi et du jugement humain. Dans [10, 42, 92], il a été montré que plusieurs facteurs peuvent biaiser les données de l'observateur, notamment le manque d'expérience, le manque d'attention après une période d'évaluations répétitives et l'influence des cas précédemment évalués. Ainsi l'utilisation des méthodes automatisée est devenue une nécessité.

### 1.4.1 Imagerie aérienne et satellitaire au service du désherbage

Depuis le lancement des satellites civils comme Landsat au début des années 70 puis des satellites à basse résolution spatiale comme NOAA-AVHRR au début des années 80, la télédétection est devenue incontournable pour de nombreux domaines comme la climatologie, l'océanographie, l'environnement ou l'agriculture. De manière générale, grâce à la vision synoptique, répétitive et objective qu'elle apporte sur les différents éléments de la surface terrestre, la télédétection est un outil privilégié pour l'étude et le suivi de la végétation sur l'ensemble du globe. Cela passe par la compréhension des variations spatio-temporelles des différents types de végétation, de leurs propriétés structurales et biophysiques. Pour l'étude de la végétation, des séries temporelles d'indices de végétation marqueurs de l'activité photosynthétique sont traditionnellement utilisées. Ces indices de végétation sont définis comme étant « des transformations spectrales de deux ou plusieurs bandes conçues pour améliorer la contribution des propriétés de la végétation et pour permettre des comparaisons spatiales et temporelles fiables de l'activité photosynthétique terrestre et des variations structurelles de la canopée » [75]. Généralement, ces indices sont basés sur les différences de réflectance des bandes spectrales issues du visible et du proche infrarouge, bande de réflectance maximale de la végétation. Dans le visible, nous utilisons habituellement la bande du rouge en raison de l'absorption des pigments chlorophylliens dans cette gamme de longueurs d'onde (Figure 1.3 ). L'indice de surfaces foliaires (Leaf Area Index : LAI) est le premier indice de structuration du couvert végétal à être calculé par combinaison des bandes rouge et proche infrarouge [86]. Plus tard, les études ont montré que, contrairement à la différence des deux bandes qui sont très

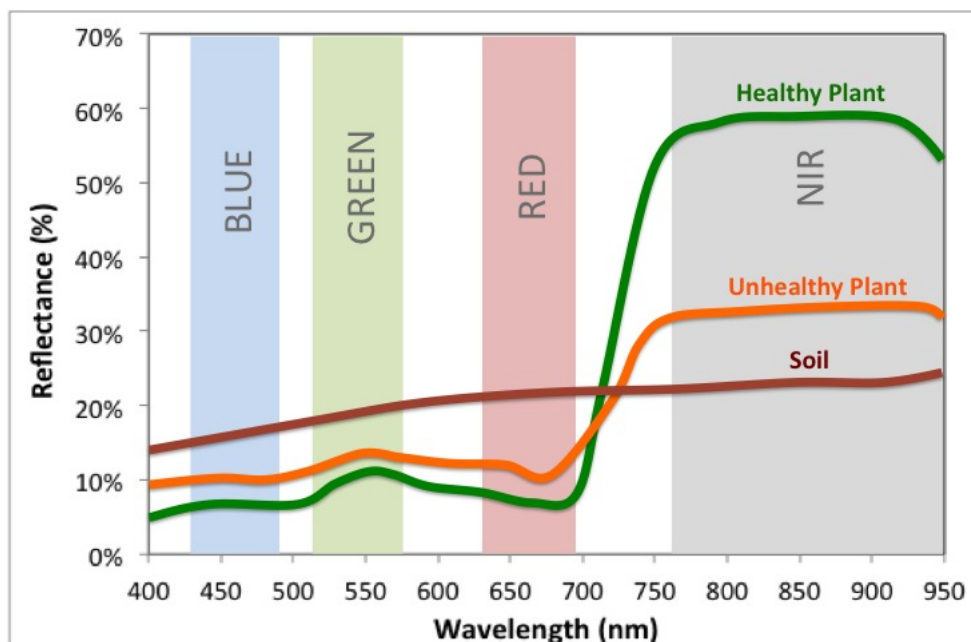


FIGURE 1.3 – Absorption spectrale de la végétation saine [22].

sensibles aux variations d'éclairement, la normalisation par la somme des deux bandes réduit cet effet et garde une valeur constante quelque soit l'éclairement global. Ce nouvel indicateur est appelé indice de végétation par différence normalisée (*NDVI* en anglais) [155]. Le *NDVI* est un indicateur de l'activité photosynthétique (vigueur et quantité de végétation) mais est également sensible à la structure du couvert. Par conséquent, c'est un indicateur qui peut être relié à différentes propriétés biophysiques de la végétation (ex. *LAI* ou production de biomasse) et à sa dynamique saisonnière.

En 1985, pour identifier les mauvaises herbes, Menges *et al.* [118] ont montré qu'avec des photographies aériennes prises à des altitudes comprises entre 610 et 3050 m, il était possible de distinguer certaines adventices dans certaines cultures, notamment : l'asclépiade grimpante dans des orangers, le parthenium de l'herbe à poux dans des carottes, le johnsongrass dans les cultures de coton et sorgho, la roquette londonienne dans les choux, et l'amarante Palmer dans le coton. Dans cette étude les auteurs ont également montré l'intérêt de la réflectance dans le proche infrarouge à 850 nm par rapport à celle du vert à 550 nm. Richardson *et al.* [140] ont utilisé des vidéos acquises à 900 m du sol pour tester la faisabilité de l'utilisation des vidéos aériennes pour distinguer les mauvaises herbes. Pour obtenir des images dans le visible et le proche infrarouge, des filtres à bande étroite visible et proche infrarouge ont été placés sur les objectifs de la caméra. Les résultats obtenus ont montré qu'il était possible de distinguer des zones composées uniquement d'adventices telles que le johnsongrass ou l'amarante par rapport à des zones composées exclusivement de culture de coton et de cantaloup (melon). Dans [49], les auteurs ont traité des images et des vidéos numériques pour détecter le doryphore commun et le doryphore drummond

au cours de leur floraison. En plus des images aériennes, certains auteurs ont exploité les images satellitaires pour la détection des adventices dans les prairies [9, 50, 133]. Par exemple, dans [9], des images Spot de résolution spatiale égale à 20 m ont été utilisées pour détecter les infestations de faux balais (*Ericameria austrotexana*) dans les prairies du Texas. Les images des satellites SPOT, TM et AVHRR ont été utilisées avec succès pour détecter et cartographier à grande échelle de vastes agrégations de mauvaises herbes possédant une signature spectrale extrêmement différente. Ullah *et al.* ont cartographié la floraison de la malédiction de Patterson (*Echium plantagineum*) autour du lac Hume, dans le nord-est de Victoria, à l'aide de données Landsat TM [157]. Peters *et al.* [133] ont travaillé sur la détection des taches de guttierrézie faux-sarothra dans les prairies du Nouveau-Mexique en utilisant le satellite météorologique américain NOAA-10 (National Oceanic and Atmospheric Administration 10). La résolution spatiale de ce dernier est de 1100 m.

Avec les instruments de télédétection une indication de l'emplacement des taches de mauvaises herbes sur une image est obtenue en se basant sur la proportion de signatures spectrales de culture et d'adventices que chaque pixel de l'image enregistre. Cette proportion dépend de la surface occupée par chaque classe par rapport à ce pixel et, dans une moindre mesure, de la densité des mauvaises herbes par rapport au sol. De toute évidence, il existe une relation entre la taille des pixels et la taille des mauvaises herbes en termes de détection et d'identification. En effet, une combinaison de pixels de grande taille et de bandes spectrales relativement larges augmente les risques de moyennage des informations. Par exemple, de petites plantules de mauvaises herbes individuelles ont très peu d'influence sur la réflectance d'un pixel aussi petit que 10 cm × 10 cm et ne serait donc pas détectable [99].

Les limites identifiées pour les applications des systèmes de télédétection en gestion agricole comprennent : la collecte et la livraison d'images en temps opportun, le manque de données à haute résolution spatiale, l'interprétation des images et l'extraction des données, et l'intégration de ces données aux données agronomiques dans les systèmes experts [45, 81].

### 1.4.2 Désherbage robotisé

A partir des années 90, la plupart des méthodes utilisées se sont basées sur les images acquises par des robots ou des caméras montées sur des tracteurs. Ces robots, au premier abord, étaient petits, lents et légers. En effet, l'autonomie des robots leur procurait un atout considérable. L'une des premières démonstrations d'un robot de pulvérisation de précision a été proposée par Lee *et al.* [103] en 1999. Les auteurs ont montré que leur système était capable de traiter 48% des adventices présentes dans un champ de tomates à une vitesse de 0,8 km/h. Dans [159], les auteurs ont proposé un robot de pulvérisation à la demande dont le but était de détecter les adventices dans les rangées de culture puis de pulvériser de l'herbicide uniquement sur leurs feuilles. Les résultats ont démontré la faisabilité de l'utilisation du GPS RTK (Cinématique Temps Réel ou *Real Time Kinematic* en anglais) pour le contrôle automatique de la trajectoire des coupeurs de désherbage mécaniques fonctionnant dans la zone inter-rang. Il y a ainsi des philosophies pour l'automatisation des tracteurs, la construction de robots spécialisés pour chaque tâche et pour la fabrication de robots hautement polyvalents et modulaires.

## 1.4. DÉSHÉRBAGE LOCALISÉ

Dans ce contexte, la société suisse Ecorobotix a par exemple proposé un robot fonctionnant entièrement à l'énergie solaire qui applique une micro-dose d'herbicide grâce à deux bras robotiques parallèles. Lamm *et al.* [100] ont mis au point un système de désherbage robotisé en temps réel. Lors d'essais sur 14 champs de coton, le système a correctement pulvérisé 88,8 % des mauvaises herbes tout en identifiant correctement et sans pulvériser 78,7 % des plantes de coton à une vitesse 0,45 m/s. Downey *et al.* [44] ont opté pour l'utilisation d'une caméra embarquée sur un tracteur et d'un GPS différentiel pour effectuer une cartographie des adventices dans un champ de coton. La précision globale du système était d'environ 85 % et ce résultat était comparable à la précision du désherbage d'une équipe qui a parcouru le champ à pied pour arracher les adventices. Cependant, malgré l'efficacité rencontrée lors des premières expérimentations des robots désherbeurs, un problème demeure sur leur efficacité dans les grandes cultures. Les céréales (blé, orge, maïs, ...), les oléagineux (tournesol, colza, soja, ...) et les protéagineux (pois, féveroles, ...) composent le secteur des grandes cultures. Le secteur des grandes cultures représente en France une place importante dans l'agriculture mondiale. En 2010, près des deux tiers des exploitations agricoles géraient des grandes cultures en France métropolitaine. Selon une étude réalisée en 2016 [135], une exploitation agricole disposait en moyenne de plus de 63 hectares, soit 7 hectares de plus qu'en 2010 et 20 de plus qu'en 2000 (Figure 1.4).

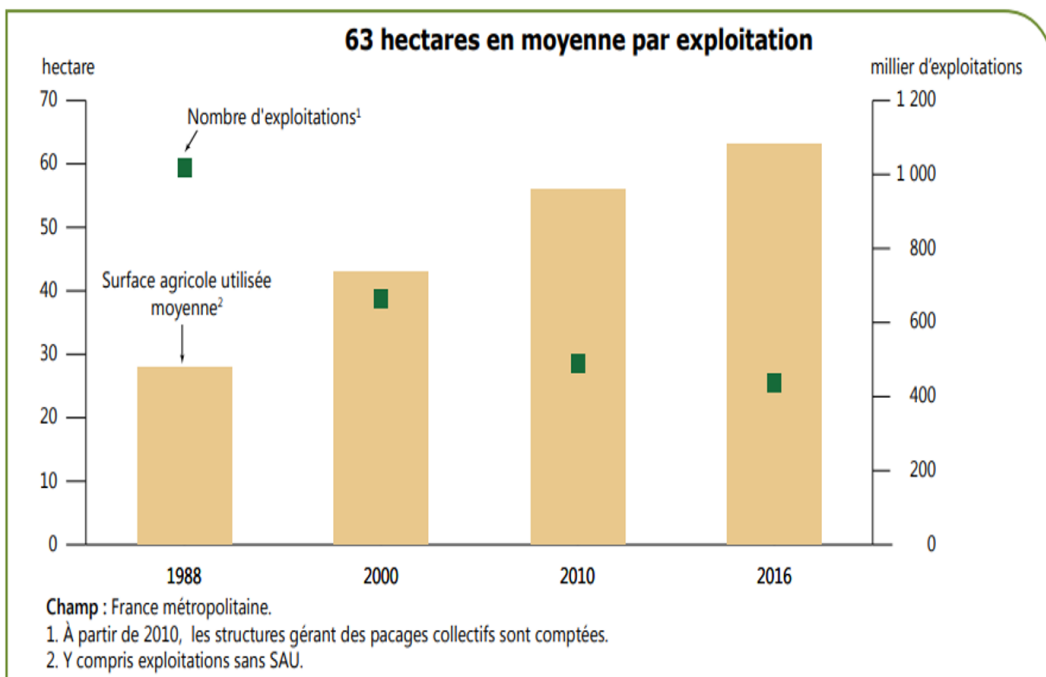


FIGURE 1.4 – Recensements agricoles et enquête sur la structure des exploitations agricoles 2016. SAU est une notion normalisée dans la statistique agricole européenne qui désigne la surface agricole utilisée. Selon l'INSEE, elle comprend les terres arables (y compris pâturages temporaires, jachères, cultures sous abri, jardins familiaux, ...), les surfaces toujours en herbe et les cultures permanentes (vignes, vergers, ...). Image issue de [135].

### 1.4.3 Avantage du drone comme support d'imagerie

Récemment, les drones sont apparus comme un bon compromis entre période d'acquisition, rapidité d'acquisition et résolution spatiale dans les grandes cultures [45, 130, 149, 166]. Contrairement aux satellites et aux avions, les petits drones peuvent opérer à des altitudes plus basses, ce qui permet aux utilisateurs finaux de recueillir des données géoréférencées à plus haute résolution spatiale, essentiel pour la détection et la localisation des adventices [97]. De plus, l'imagerie par drone est généralement moins coûteuse pour les petites surfaces, plus facile à garder dans un local et à utiliser, ce qui rend son utilisation plus accessible aux organisations locales qui gèrent les cultures. Par exemple, le drone eBee de Sensefly (une filiale suisse de la société Parrot) peut voler jusqu'à 80 km/h avec une autonomie moyenne de 45 mn, pour une envergure d'environ 110 cm.

Dans [130], des images de 2,5 cm de résolution spatiale ont été acquises à 30 m d'altitude au dessus d'un champ de maïs de surface  $140 \times 100$  m avec un drone quadricopter MD4-1000 de la société Microdrone dans le but de détecter les adventices à des phases précoces du développement de la culture. Lottes et *al.*, dans leur étude sur la détection des adventices, ont utilisé le drone DJI Phantom 4 de la société DJI pour l'acquisition des images à une altitude de 3 m pour une résolution spatiale de 0.2 mm/pixel [107].

## 1.5 Conclusion

Dans ce chapitre, nous avons constaté que plusieurs avancées technologiques ont été enregistrées dans l'agriculture depuis la première révolution agricole au XVIII<sup>e</sup> siècle. Cependant, malgré les efforts considérables menés dans le développement de nouveaux systèmes d'acquisition (capteurs de télédétection, robots, outils de cartographie, matériel agricole, ...), la détection automatique des adventices dans les images reste un grand défi. Des méthodes précises et robustes de détection et d'identification automatique des mauvaises herbes sont nécessaires, notamment pour répondre aux conditions réelles dans lesquelles les mauvaises herbes et les plantes cultivées s'occultent fréquemment et l'apparence des feuilles est déformée par les insectes ou les conditions météorologiques, le stress hydrique ou nutritionnel ou encore les ombres.

## 1.5. CONCLUSION

---

# Chapitre 2

## État de l'art : Détection des adventices par la vision artificielle

### Sommaire

---

<b>2.1</b>	<b>Segmentation végétation et non végétation . . . . .</b>	<b>16</b>
<b>2.2</b>	<b>Détection des rangées de culture . . . . .</b>	<b>18</b>
2.2.1	Transformée de Hough . . . . .	18
2.2.2	Régression linéaire . . . . .	19
2.2.3	Analyse de régions (blob analysis) . . . . .	20
2.2.4	Méthode des bandes horizontales . . . . .	20
2.2.5	Approche basée sur la stéréoscopie . . . . .	21
<b>2.3</b>	<b>Détection des adventices par les méthodes d'apprentissage machine . . . . .</b>	<b>21</b>
2.3.1	Apprentissage machine classique . . . . .	21
2.3.2	Apprentissage profond . . . . .	25
<b>2.4</b>	<b>Conclusion . . . . .</b>	<b>29</b>

---

Les fortes avancées technologiques dans la conception des capteurs optiques et des calculateurs font du traitement d'images et de la vision par ordinateur des domaines essentiels pour l'agriculture de précision. Si couplés aux informations de localisation fournies par un GPS (par exemple), ils permettraient d'obtenir un aperçu spatialisé des indicateurs phénologiques. La vision artificielle, quant à elle, utilise les techniques de traitement d'images et d'intelligence artificielle pour reproduire des mécanismes de la vision humaine afin de faciliter les tâches de reconnaissance, de détection, ... Son utilisation a augmenté de façon exponentielle au cours des dernières décennies. Ses applications vont de la médecine au divertissement en passant par la télédétection.

Dans ce chapitre, une étude est effectuée sur les techniques de vision artificielle déjà réalisées autour de la détection des adventices. Cette étude prend en compte les méthodes de détection d'adventices basées sur l'analyse spatiale et spectrale ainsi que les méthodes d'apprentissage.

Ce chapitre est divisé en 4 sections. La première section décrit les méthodes employées pour la séparation de la végétation du reste des éléments contenus dans l'image (sol, ombres, ...). Dans la deuxième section, les méthodes de détection des rangées de culture pour un guidage précis des engins agricoles ainsi que la détection des adventices inter-rang sont présentées. La troisième section décrit les méthodes de détection d'adventices par apprentissage machine. La dernière section correspond à la conclusion du chapitre.

## 2.1 Segmentation végétation et non végétation

Pendant cette dernière décennie, les techniques de traitement d'images ont été utilisées avec succès dans la mesure de la fertilisation azotée ainsi que dans d'autres applications, mais la détection des adventices restent un défi de taille.

Cependant, un prétraitement robuste est nécessaire pour éliminer les perturbations indésirables afin d'établir une distinction entre la végétation et le sol [67]. Les approches courantes de segmentation utilisent la couleur et l'information multispectrale pour séparer la végétation de l'arrière-plan (sol et résidus). Des indices spécifiques sont calculés à partir des spectres de l'image pour segmenter efficacement la végétation. L'avantage d'utiliser des indices de couleur est qu'ils permettent d'accentuer une couleur particulière de l'image, comme c'est le cas de la couleur verte ou du proche infrarouge pour les plantes.

Parmi ces indices, nous pouvons citer l'indice d'excès de verdure de la végétation (*Excess Green vegetation index*) : ExG [164], l'indice de végétation normalisé : NDI [131], l'indice de couleur pour l'extraction de la végétation (*Colour Index of Vegetation Extraction*) : CIVE [91], le *VEGetative index* : VEG [65], l'indice de rougeur de la végétation (*Excess Red vegetative index*) : ExR [120] et le ExGR qui représente la différence entre ExG et ExR [119]. La Table 2.1 présente les formules de ces indices. NIR correspond au proche infra rouge.

où  $r$ ,  $g$  et  $b$  sont les coordonnées chromatiques :

$$g = \frac{R^*}{R^* + G^* + B^*}, r = \frac{G^*}{R^* + G^* + B^*}$$

$$b = \frac{B^*}{R^* + G^* + B^*}$$

TABLE 2.1 – Indices de couleur pour la segmentation végétation et non végétation.

Méthodes	Formules
ExG [164]	$ExG = 2g - r - b$ (2.1)
NDI [131]	$NDI = \frac{G - R}{G + R}$ (2.2)
ExR [120]	$ExR = 1.4r - g$ (2.3)
ExGR [119]	$ExGR = ExG - ExR$ (2.4)
CIVE [91]	$CIVE = 0.441r - 0.811g + 0.385b + 18.78745$ (2.5)
VEG [65]	$VEG = \frac{g}{rb^{1-a}} ; a = 0.667$ (2.6)
NDVI [155]	$NDVI = \frac{NIR - R}{NIR + R}$ (2.7)

et  $R^*$ ,  $G^*$  et  $B^*$  sont les valeurs RGB normalisées.

$$R^* = \frac{R}{255}, G^* = \frac{G}{255}, B^* = \frac{B}{255}$$

Dans [119], après avoir comparé les indices ExG, NDI et ExGR sur des plantes de soja ayant différentes couvertures de sol (sol nu, paille de blé et tiges de maïs), les auteurs ont constaté que ExGR avec un seuillage fixé à 0 fournissait les meilleurs résultats de segmentation pour les différents arrière-plans ou types de sol et que les deux autres indices avec un seuillage adaptatif sur les valeurs de type Otsu [127] ne montraient leurs efficacités que lorsqu'il s'agissait d'un sol nu. Pour se doter d'un outil de segmentation efficace permettant l'identification de la végétation, d'autres auteurs ont proposé une méthode combinant quatre indices différents de végétation, à savoir ExG, ExGR, CIVE et VEG [64]. Ces indices ont alors été pondérés en fonction de leurs efficacités. Dans cette étude, CIVE s'est avéré le plus efficace et VEG le moins pertinent. Toutefois, pour les images multispectrales, le NDVI reste le meilleur indicateur [30, 52, 128].

Certes, les indices de végétation ont prouvé leurs efficacités pour la distinction entre végétation et non végétation, mais il reste cependant difficile à séparer les cultures des adventices uniquement à l'aide d'informations spectrales en raison de leurs fortes similitudes. En partant de l'hypothèse que la plupart des cultures sont semées en rangs, différentes méthodes de détection des rangées de culture ont été proposées dans la littérature. L'objectif de ces méthodes est de localiser les rangées de culture pour ensuite détecter les adventices qui se situent dans l'espace inter-rang. Généralement, ces méthodes se répartissent en quelques catégories selon leur principe de détection, comme la transformée de Hough, la régression linéaire, l'analyse de régions (*blob analysis* en anglais), l'analyse des bandes horizontales et la vision stéréoscopique. Dans la section suivante, ces approches sont détaillées.

## 2.2 Détection des rangées de culture

### 2.2.1 Transformée de Hough

La transformée de Hough est une technique de reconnaissance de formes inventée en 1962 par Paul Hough [73] et brevetée par IBM. Cette technique développée il y a plus d'un demi-siècle continue à faire ses preuves dans le domaine de la vision artificielle à tel point qu'elle est considérée comme un outil standard pour la détection de formes. Elle est utilisée dans presque tous les domaines, par exemple dans le milieu médical pour la reconnaissance de profils osseux compacts dans des images cliniques de tomographie à rayons X [26] ou encore dans les voitures autonomes pour la détection des voies sur les routes [172]. Elle est aussi employée pour la détection des immeubles sur des images aériennes [143]. Dans l'agriculture, cette méthode est l'une des plus utilisées pour la détection des rangées de culture. Elle est souvent intégrée dans les outils de guidage des engins agricoles puisqu'elle a une forte capacité à recoller les rangées discontinues dues généralement à des problèmes de semis ou à la faible croissance de certaines plantes. Elle a par exemple été mise en oeuvre pour guider des robots mobiles dans des champs de betterave et de colza dans [176]. Avec des tests sur le terrain, les auteurs ont montré que le système était suffisamment précis et rapide pour contrôler un désherbeur et un robot mobile en boucle fermée avec un écart-type de positionnement de 2,7 et 2,3 cm, respectivement. Baker *et al.* [16] ont appliqué la transformée de Hough pour extraire les rangées de culture dans des images acquises par un robot en utilisant un intervalle d'angle bien précis. Ainsi la transformée de Hough a été calculée sur un intervalle d'angle allant de  $45^\circ$  à  $135^\circ$  avec un pas de  $1^\circ$ . Dans [59], l'indice ExG [164] et le seuillage de Otsu [127] ont aidé à éliminer le fond (sol, résidus, paille, ...) avant de réaliser une double transformée de Hough [73] pour identifier les rangées de cultures dans des images en perspective. De plus, toute région de l'image avec au moins un pixel appartenant aux rangées détectées était considérée comme de la culture, ou adventice dans le cas contraire. Cette technique a souffert de l'impact des adventices trop proches des rangées de culture. Étant données les difficultés rencontrées dans l'évaluation des algorithmes de détection des adventices dans [85], les auteurs ont simulé des champs agricoles pour évaluer les performances de leurs algorithmes. Ils ont considéré que le taux d'infestation pouvait être obtenu facilement sur les données simulées, contrairement aux vraies données terrain qui exigeaient un énorme travail manuel. Au total, deux méthodes de détection de mauvaises herbes ont été proposées. La première a utilisé le filtre de Gabor pour détecter les rangées de culture tandis que la seconde a employé la transformée de Hough. Les images étant sans effet de perspective, les auteurs ont considéré que dans l'espace de Hough, tous les maxima locaux étaient alignés sur le maximum global et avaient le même angle. En termes de résultats, la transformée de Hough s'est avérée plus efficace que le filtre de Gabor. Dans [132], pour la détection des rangées de culture, les paramètres associés à la transformée de Hough (angle, distance minimale entre les rangs et nombre de lignes à détecter) ont été fixés manuellement en fonction de la connaissance du champ. Ji *et al.* [84] ont proposé d'utiliser la Random Hough Transform (RHT) [168] pour détecter les rangées de culture. La particularité de RHT est que la transformée de Hough est appliquée sur certains pixels de la végétation aléatoirement sélectionnés. Comparé à la transformée de Hough traditionnelle, la RHT a permis de gagner environ 45% en temps de traitement. Après le calcul de la transformée de Hough, le plus grand défi est

de sélectionner les lignes qui traversent le mieux le centre des rangées de culture à cause des mauvaises herbes entre les rangs et de la largeur de ces rangées. Pour relever ce défi, Jones *et al.* [85] ont proposé d’exploiter le parallélisme des rangées de culture ainsi que les connaissances préalables sur la distance entre les rangées. Ce parallélisme a permis d’émettre l’hypothèse que si les rangées de culture sont parallèles dans l’espace image, les votes de ces lignes seront alignés autour de l’angle correspondant à leur orientation globale. De plus, ces votes seront séparés par une distance équivalente à la distance inter-rang. Pour extraire les lignes principales de chaque rangée de culture, ils ont donc identifié le premier maximum, puis avec un pas correspondant à la distance inter-rang théorique, ils se sont déplacés de chaque côté de ce maximum pour récupérer les autres maxima. Cette méthode a prouvé une certaine efficacité pour détecter les rangées de culture avec une distance entre les rangs de culture et une orientation constantes (Figure 2.1).

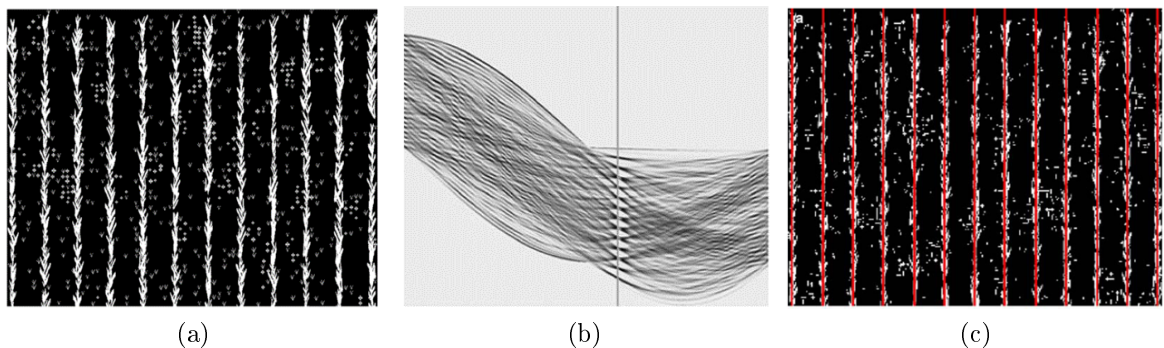


FIGURE 2.1 – Résultats obtenus dans l’article [85]. (a) Image agronomique simulée. (b) Le résultat de la transformée de Hough sur (a). (c) Résultat de la détection des rangées.

Mais dans la réalité des parcelles agricoles, l’imperfection du sol et la présence de chemins de tracteurs ne permettent pas toujours d’avoir des rangs droits, parallèles et avec la même distance inter-rang. Il est donc indispensable de passer à d’autres méthodes de détection des rangées de culture.

### 2.2.2 Régression linéaire

En partant de l’hypothèse que chaque rangée de culture est un nuage de points qui peut être représenté par une droite, plusieurs auteurs se sont tournés vers la régression linéaire. Sogaard et Olsen [147] ont localisé les rangées de culture dans un champ d’orge en utilisant une régression linéaire pondérée. Dans leur expérience, ils ont montré qu’en fonction du niveau de croissance de la culture, il était possible de détecter les rangées de culture avec une précision allant de  $\pm 6$  mm à  $\pm 12$  mm. Hague *et al.* [65] ont déterminé la position et l’orientation des rangées de culture en appliquant la version non linéaire du filtre de Kalman appelée filtre de Kalman étendu [88]. Montalvo *et al.* [122] ont proposé d’utiliser une régression linéaire pour la détection des rangées de culture dans un champ de maïs avec un taux d’infestation élevé. La méthode consistait à faire un double seuillage de l’indice ExG, le premier seuillage pour isoler la végétation et le second sur la végétation pour séparer la culture de la non culture. Pour une bonne identification des rangées de culture, ils se sont servis des informations connues de la parcelle (disposition des rangs,

espacement, nombre de rangs, ...) et les informations sur la géométrie du système de caméra monté sur un tracteur. Ces informations concernant la parcelle et la caméra ont permis de construire un modèle du champ de culture. Le décalage maximal que pouvait avoir une rangée de culture à gauche et à droite a été estimé par observation de 3 séquences vidéo de 1400 images.

Le modèle du champ de culture est une image binaire avec le même nombre de rangées que dans les images acquises. Ce modèle est utilisé comme masque sur le résultat du second seuillage. Enfin, une régression linéaire est utilisée pour ajuster les rangées de culture. Avec la régression linéaire, l'objectif est d'obtenir le meilleur alignement de pixels dans chaque rang de culture identifiée.

### 2.2.3 Analyse de régions (blob analysis)

L'analyse des taches ou blob est une technique fondamentale de la vision par ordinateur basée sur l'analyse de régions d'images qui présentent certaines cohérences. Fontaine *et al.* [53] se sont basés sur la direction et le centre de gravité des différentes régions pour proposer une méthode de détection de rangées. En associant le centre de gravité et la direction de ces régions à la direction de déplacement d'un tracteur sur lequel était montée une caméra, Burgos-Artizzu *et al.* [24] ont proposé une méthode pour détecter les rangées de culture. Dans [130], une procédure d'analyse d'images par objet (OBIA) a été développée sur des séries d'images aériennes pour la discrimination automatique entre des rangées de culture et les adventices dans les champs de maïs. Les images manipulées ont été acquises, en début de saison, par drone à 30 m de hauteur ; la résolution spatiale des images était de 2 cm. Les images ont été segmentées en objets multi-pixels homogènes en utilisant un algorithme multi-échelle [20]. La segmentation à grande échelle a ainsi mis en évidence les structures des rangées de culture (orientation) et celle à petite échelle a accentué les objets se trouvant à l'intérieur des rangées de culture (séparation entre végétation et sol nu). Puisque les images ont été subdivisées en différents objets homogènes, après l'élimination du sol nu, les régions restantes ont été fusionnées en fonction de l'orientation et de la valeur du rapport entre la longueur et la largeur. Enfin est considérée comme culture tout objet assez grand et ayant une direction proche de celle des rangées de culture. Cette procédure a été appliquée rangée par rangée. Pour tester leur méthode, 49 cadres (échantillons) de 1×1 m ont été placés à différents endroits du champ. Les résultats ont montré que les bords des rangées de culture ont été détectés avec une efficacité de 90% et que le processus était fortement influencé par la présence des adventices très proches ou situées dans les rangées de culture.

### 2.2.4 Méthode des bandes horizontales

Cette méthode est majoritairement appliquée sur des images acquises par une caméra montée sur tracteur. L'objectif est de considérer l'image comme un tableau dans lequel chaque rangée correspond à une colonne. Dans [24], les images sont acquises à une fréquence de 25 images (720 x 576 pixels) par seconde (vidéos) avec deux caméras montées à 2.15 m du sol, sur le toit d'un tracteur circulant à une vitesse de 7 km/h. Pour isoler la végétation du reste, une version modifiée de l'indice ExG a été proposée nommée ExG-Modify. Les coefficients utilisés pour le calcul de l'indicateur de couleurs ont été obtenus

en utilisant un algorithme génétique. Pour le seuillage, les auteurs ont préféré employer la valeur moyenne de ExGModify par rapport à la méthode de Otsu qu'ils trouvaient moins rapide. Une méthode appelée "traitement d'images rapide" (*Fast image processing*) a été mise en œuvre. Cette méthode consiste à diviser les images en bandes verticales. Pour chaque bande, un vecteur de même colonne est créé. Dans chaque case de ce vecteur est stocké le nombre de pixels blancs correspondants dans la colonne du même numéro. Un seuil équivalent à  $3/4$  de la hauteur de la bande est utilisé sur les valeurs contenues dans le vecteur. Les colonnes du vecteur non nulles à l'issue du seuillage sont considérées comme étant les rangées de culture. Cependant, dans certains cas, il est difficile de savoir quand commence la rangée et où elle se termine, sans oublier l'impact des adventices inter-rang.

### 2.2.5 Approche basée sur la stéréoscopie

La stéréovision permet de retrouver la profondeur des points d'une scène à partir de deux projections de cette scène (points de vue différents). Certains auteurs proposent d'utiliser cette approche pour identifier les rangées de culture. Le but est de créer une carte altimétrique du champ qui va permettre, par la suite, de localiser les emplacements des rangées de culture. Mais ces méthodes sont généralement utilisées dans le cas où les hauteurs de plantes sont suffisamment importantes [93].

Indépendamment des efforts considérables consacrés à la détection des mauvaises herbes, un algorithme capable de détecter efficacement les mauvaises herbes inter-rang et intra-rang est nécessaire. Il a été constaté que celles basées sur l'alignement et la fréquence des rangées de culture sont moins efficaces si le taux d'infestation est relativement élevé.

## 2.3 Détection des adventices par les méthodes d'apprentissage machine

En plus des méthodes exploitant la structure géométrique du champ, des auteurs se sont focalisés sur la classification des images à partir d'algorithmes d'apprentissage. Parmi les méthodes proposées, nous discernons celles qui utilisent les méthodes d'apprentissage classiques et celles qui appliquent un apprentissage profond.

### 2.3.1 Apprentissage machine classique

Dans [131], la couleur et la forme de la végétation ont été employées pour discriminer les mauvaises herbes et les cultures, sachant que la couleur a principalement servi à éliminer l'arrière-plan (sol, cailloux, ...). La méthode consistait à calculer la différence normalisée entre les bandes verte et rouge et à appliquer un seuil. En s'appuyant sur la périodicité de l'alignement des cultures, les auteurs ont choisi de classer toute végétation localisée entre les rangées de culture comme étant des mauvaises herbes. Pour réduire l'erreur de détection, ils se sont orientés sur l'utilisation des attributs de formes. L'idée était de caractériser les objets isolés avec un certain nombre de descripteurs géométriques dont la longueur des axes principaux, le rapport de forme, la surface, etc. Ainsi chaque objet a été

### 2.3. DÉTECTION DES ADVENTICES PAR LES MÉTHODES D'APPRENTISSAGE MACHINE

---

représenté comme un vecteur dans l'espace de caractéristiques. Le critère de Fisher [150] a été appliqué pour la sélection des meilleurs descripteurs. Les plantes qui se trouvaient à l'intérieur des rangées ont été reclassées par trois classifieurs différents : la méthode des  $k$  plus proches voisins (k-NN), la règle de Bayes et une approche heuristique. Après comparaison des résultats obtenus par les trois méthodes, il en est ressorti que la méthode heuristique était la moins efficace des trois. Ils ont aussi montré que l'utilisation des caractéristiques géométriques améliorerait la détection de 10%.

Dans [151], des filtres de Gabor 2D ont été appliqués pour extraire les caractéristiques et un réseau neuronal artificiel (ANN) a été employé pour la classification des mauvaises herbes à feuilles larges et des graminées. Dans [162], un système de caméra bi-spectrale équipée d'un GPS différentiel a été utilisé pour la détection des adventices. Pour récupérer la végétation, la différence entre le proche infrarouge et la bande rouge a été calculée. Un seuillage a permis d'isoler la végétation. Le bruit résiduel après la segmentation a été éliminé à l'aide d'une approche région qui consistait à éliminer des groupements de pixels dont le nombre était inférieur à un seuil minimum fixé. Mais cette méthode s'est avérée insuffisante pour éliminer complètement le bruit résiduel ; des filtres morphologiques ont alors été employés. Une opération de morphologie (fermeture) a été appliquée avec un élément structurant circulaire. Différents descripteurs ont ensuite été calculés : des descripteurs de régions, des descripteurs de contours et des descripteurs basés sur le squelette des régions. L'utilisation des descripteurs basés sur le squelette a été motivée par leur invariance au changement de forme global. Une centaine d'attributs ont été calculés au total. En utilisant une méthode de réduction de dimension, le nombre de descripteurs a été ramené à 15. Parmi les 15 attributs exploités, nous pouvons citer l'excentricité, la compacité et le moment de Hue. Pour la classification, au final, 37 classifieurs incluant Bayes, le plus proche voisin, les arbres de décision, etc., ont été appliqués. Les résultats obtenus ont permis de constater que tous les classifieurs avaient réussi à plus de 95 %.

D'autres auteurs se sont appuyés sur la variation morphologique et ont utilisé des réseaux de neurones artificiels pour séparer les mauvaises herbes de la culture de maïs [83]. Dans [6], la méthode des  $k$  plus proches voisins (k-NN) a été employée sur la transformée en ondelettes de Haar afin de reconnaître et de différencier les feuilles des adventices étroites et larges pour la mise en place d'un système automatique de contrôle en temps réel de pulvérisateur. La précision moyenne de l'algorithme proposé était de 94 % et il était moins robuste aux conditions d'éclairage, au vent et à d'autres paramètres environnementaux naturels. Dans [8], la méthode du support à vaste marge (SVM) et les caractéristiques de forme ont été proposées pour la classification efficace des cultures et des mauvaises herbes dans les images numériques. Dans le cadre de leur expérience, quatorze caractéristiques de cultures et de mauvaises herbes ont été testées pour trouver la combinaison optimale de caractéristiques qui fournissait le taux de classification le plus élevé. L'analyse des résultats a révélé que le SVM atteignait une précision de 97 % sur un ensemble de 224 images de test. Un réseau de neurones artificiel et 17 attributs géométriques ont été utilisés pour classer 20 espèces de plantes dans [69]. Au total, 400 images de feuilles ont servi pour l'apprentissage et 137 pour le test. L'efficacité globale obtenue était de 92 %. Latha *et al.* [101] ont proposé une méthode de détection d'adventices en partant de l'hypothèse que dans les images, les fréquences de bordure et les nervures de la culture et des mauvaises herbes ont des propriétés de densité différentes (bords forts et faibles). Dans [76], les méthodes de classification et de reconnaissance de forme ont été déployées sur des images

### 2.3. DÉTECTION DES ADVENTICES PAR LES MÉTHODES D'APPRENTISSAGE MACHINE

---

aériennes en vue de la reconnaissance de trois espèces de mauvaises herbes : jacinthe d'eau ; solanum viarum (pomme à soude tropicale) ; stipe à feuilles dentées.

Une méthode semi-supervisée a été proposée dans [136] pour distinguer les mauvaises herbes et la culture dans deux champs : tournesol et maïs. Le seuillage de Otsu a été appliqué deux fois sur l'indice ExG. Le premier seuillage a permis de supprimer l'arrière-plan et le second a été utilisé pour créer deux classes censées représenter la culture et les mauvaises herbes. La méthode des k-means a été mise en œuvre pour sélectionner cent échantillons de chaque classe (culture, adventices et sol) pour l'apprentissage. Les caractéristiques utilisées pour les k-means sont la moyenne et l'écart type des trois canaux de l'image couleur et du ExG. Le classifieur SVM a été employé avec des caractéristiques géométriques, spatiales, des statistiques de premier et de second ordres extraits sur les bandes rouge, bleue, verte et l'indice ExG.

Les attributs calculés étaient :

- Attributs géométriques : le nombre total de pixels, la largeur maximale et la hauteur maximale ;
- Attributs spatiaux : pour les caractéristiques spatiales, deux approches ont été utilisées. La première consistait à utiliser les rangées de culture obtenues par la transformée de Hough. Elle calculait la moyenne et l'écart type de la distance entre les pixels formant un objet et la rangée de culture la plus proche. La seconde approche utilisait un voisinage de 5 pixels autour de l'objet en question et calculait ensuite la moyenne et l'écart type des pixels qui se trouvaient dans ce voisinage ;
- Statistiques du premier ordre : la moyenne, la variance, le kurtosis et la dissymétrie ;
- Statistiques du second ordre : la matrice de co-occurrence (GLCM) a été calculée pour chaque canal, y compris le ExG. De cette matrice, quatre attributs ont été calculés : l'énergie, le contraste, l'homogénéité et la corrélation.

Après avoir effectué une sélection de caractéristiques, il a été remarqué que les caractéristiques géométriques et la distance à la ligne de culture n'apportaient pas beaucoup d'informations. Le modèle formé s'était avéré efficace dans les champs de tournesol, mais moins robuste dans les champs de maïs en raison de l'ombre des plantes de maïs.

Dans [15], les auteurs ont exploité des attributs de texture extraits sur des ondelettes pour détecter et caractériser quatre types de mauvaises herbes dans un champ de betterave sucrière. Des réseaux de neurones artificiels ont été mis à contribution comme classifieur. L'utilisation des ondelettes s'est révélée efficace pour la détection des mauvaises herbes, même à un stade de croissance de la betterave supérieur à 6 feuilles, c'est-à-dire lorsque les plantes apparaissent avec des occlusions considérables dans les images et les caractéristiques de la forme des plantes ne peuvent être utilisées. Cependant, la méthode présentée n'est utile que dans les cas où il existe une différence significative entre les fréquences de texture des feuillages. Les résultats ont montré que les caractéristiques de texture des ondelettes permettaient de distinguer les mauvaises herbes entre les cultures avec un taux de détection correcte de 96 %, alors qu'au plus 4 % des betteraves ont été classées comme mauvaises herbes à tort. Dans [107], une méthode de détection et de reconnaissance des mauvaises herbes dans des images acquises par drone volant très proche du sol (3 à 15 m au-dessus du sol) a été proposée. La méthode utilisait des caractéristiques statistiques et géométriques (distance de la ligne de culture et relation spatiale entre chaque objet et ses voisins) extraites sur les objets segmentés et un certain nombre de points clés créés à partir d'une grille du masque de l'image de végétation. Une fois les attributs extraits,

### 2.3. DÉTECTION DES ADVENTICES PAR LES MÉTHODES D'APPRENTISSAGE MACHINE

---

la méthode de la forêt d'arbres aléatoires (*random forest* : RF en anglais) a été employée pour la classification. Une précision de 96 % a été obtenue pour la détection des mauvaises herbes et de 86 % pour leurs identifications. Bakhshipour et Jafari [14] ont évalué la détection des mauvaises herbes à l'aide du SVM et des réseaux de neurones artificiels dans quatre espèces de mauvaises herbes communs des champs de betterave à sucre, en se basant sur leurs formes. Dans [57], une procédure semi-automatique d'analyse d'images basée sur des objets (OBIA) a été développée avec les forêts aléatoires (RF) combinée à des techniques de sélection de caractéristiques pour classer le sol, les mauvaises herbes et le maïs.

Une autre méthode de traitement d'images automatique a été développée pour discriminer les cultures et les pixels des mauvaises herbes en combinant des informations spatiales et spectrales extraites sur des images multi-spectrales à quatre bandes [110]. Les images manipulées avaient été acquises à l'aide d'une perche dans deux champs de betterave et de maïs sur des surfaces délimitées de taille 5×5 m. Le capteur se situait à 3 m au-dessus du sol avec une résolution spatiale de 6 mm. Au total, 14 images ont servi à évaluer la méthode composée de 2 approches successives.

- Approche basée sur l'information spatiale : les cultures étant semées en rangs parallèles, l'orientation globale des rangées de culture a été déterminée grâce à la transformée de Fourier. A l'aide de la transformée de Hough, les rangées de culture ont été identifiées. Une régression linéaire a été effectuée sur les pixels du même côté pour délimiter les rangées de culture. Enfin les plantes qui se trouvaient dans l'espace inter-rang étaient considérées comme des mauvaises herbes.
- Approche basée sur l'information spectrale : des pixels ont été sélectionnés sur les deux classes formées à l'issue de l'approche spatiale pour former la base d'apprentissage. Un SVM avec un noyau gaussien radial (RBF) a servi pour la classification.

Sur les 14 images évaluées (7 dans chaque champ), le taux moyen de bonne détection était de 88 % pour les cultures et 79 % pour les adventices avec l'approche spectrale et de 85 % pour la culture et 75 % pour les adventices avec l'approche spatiale. La combinaison spatiale et spectrale a permis d'avoir une amélioration de la détection d'adventices de 10 à 15%. Les erreurs de classification se situaient majoritairement au niveau des plantes localisées sur les bords des rangées ; cette erreur était causée par le mélange spectral du sol et de la végétation.

Dans [7], une méthode de classification de deux types d'adventices (monocotylédone et dycotylédone) a été proposée par utilisation du méta-classifieur AdaBoost [55]. Plusieurs classifieurs faibles ont été testés pour le AdaBoost, dont le Naïve Bayes [116], le BayesNet [56], la régression logistique simple [72], l'arbre décisionnel [112] et l'arbre aléatoire [104]. Parmi ces classifieurs, les meilleurs résultats ont été obtenus avec le Naïve Bayes pour lequel une précision globale de 98,16% pour les deux types d'adventices a été observée.

Une combinaison d'algorithmes de détection de rangées de culture et une approche basée sur l'analyse d'objet (OBIA) a aussi été proposée pour obtenir une méthode de détection d'adventices précise et robuste dans un champ de maïs [57]. Pour ce faire, les auteurs ont sélectionné une zone de travail sur laquelle des images très haute résolution ont été acquises. La transformée de Hough a alors été appliquée pour la détection des rangées de culture. Les plantes se trouvant à l'intérieur des rangées ont été masquées, ce qui a permis d'identifier les adventices inter-rang. La base d'apprentissage a été formée à partir des adventices inter-rang. Au total, 16 attributs de textures, 15 attributs géométriques et

3 attributs spectraux ont été calculés sur cette base. La méthode du random forest a été utilisée comme machine d'apprentissage. L'efficacité moyenne obtenue était de 95%.

Dans [58], le potentiel d'une nouvelle caméra mosaïque hyperspectrale instantanée pour la classification des mauvaises herbes et du maïs a été testé. Dans cet article, 3 types d'adventices habituellement rencontrées dans les champs de maïs, le souci des champs (*calendula arvensis*), le rumex (patience) et le cirse des champs (*cirsium arvense*) ainsi que le maïs ont été observés durant leur période de développement. Un dispositif comportant une caméra hyperspectrale a été conçu dans le but de faire des acquisitions de ces 4 types de plantes 2 fois par semaine (les lundis et vendredis). En combinant les bandes spectrales du visible et du proche infrarouge, un total de 180 caractéristiques a été calculé. De plus, avec l'analyse en composantes principales (ACP) et la random forest, le 10 et 30 meilleures composantes ont été sélectionnées, respectivement. Les résultats obtenus ont montré que les 30 composantes du *Random Forest* ont fourni les informations les plus pertinentes. Il est montré que la culture (*Z. mays*) peut être reconnue avec une très grande précision (94 %). Les valeurs de précision pour les trois types de mauvaises herbes sont respectivement de 95,9 %, 70,3 % et 65,9 %. Les auteurs de [110] ont développé un traitement automatique de l'image pour discriminer entre les pixels de culture et des mauvaises herbes en combinant les informations spatiales et spectrales extraites des images multispectrales à quatre bandes. Les données d'image ont été capturées à 3 m au-dessus du sol avec une caméra montée sur un poteau maintenu manuellement. L'approche spatiale (Hough Transform) a été utilisée pour détecter les rangées de culture et construire un ensemble de données d'apprentissage. Le SVM a ensuite été appliqué sur l'information spectrale pour faire la classification.

En exploitant l'information spectrale, cette dernière méthode tout comme d'autres de la littérature supposent que les cultures et les adventices ont des informations spectrales différentes, ce qui n'est pas toujours le cas dans les parcelles agricoles. Le succès de ce type de méthode repose en fait sur une meilleure sélection des caractéristiques qui implique une analyse humaine en fonction du domaine agricole. Nous pouvons constater que le point commun entre tous ces articles est que les caractéristiques sélectionnées changent en général d'un type de culture à un autre ou d'un type de données à un autre. C'est la raison pour laquelle d'autres auteurs ont décidé d'utiliser les méthodes d'apprentissage profond qui ne nécessitent pas de sélection manuelle des caractéristiques.

### 2.3.2 Apprentissage profond

Récemment, les réseaux neuronaux convolutionnels (CNN) ont émergé comme une approche intéressante et puissante pour des problèmes rencontrés dans différents domaines tels que la vision par ordinateur, la reconnaissance vocale et le traitement du langage naturel [102, 95, 29]. Le premier succès majeur a été atteint sur ImageNet Large Scale Vision Recognition Challenge 2012 (ILSCVR12) avec le réseau AlexNet [95]. AlexNet a montré qu'un grand réseau neuronal convolutionnel profond est capable d'atteindre des résultats record sur un ensemble de données très complexe en utilisant un apprentissage purement supervisé.

Généralement, dans les réseaux de neurones artificiels classiques, 3 couches de neurones sont visibles : les couches neuronales d'entrée, intermédiaires et de sortie. Plus nous avons de couches intermédiaires, plus le réseau est profond et complexe pour l'appren-

## 2.3. DÉTECTION DES ADVENTICES PAR LES MÉTHODES D'APPRENTISSAGE MACHINE

tissage. En traitement d'image, pour apprendre à un réseau à reconnaître deux classes à partir d'échantillons d'images, nous préférons utiliser une liste de descripteurs ou caractéristiques (*features* en anglais) car utiliser chaque pixel de l'image augmenterait le nombre de neurones de la couche d'entrée et par conséquent le nombre de poids à entraîner. Pour un bon apprentissage, la phase d'extraction de caractéristiques est effectuée par un expert car le résultat en dépend. En d'autres termes, le concepteur doit réaliser une grande partie du travail. C'est lors de cette phase qu'intervient le concept d'apprentissage profond (deep learning) dont le but est de se passer de la phase d'intervention de l'expert en la remplaçant par des couches de convolution, d'où l'appellation de réseau profond ou réseaux neuronaux convolutionnels (CNN).

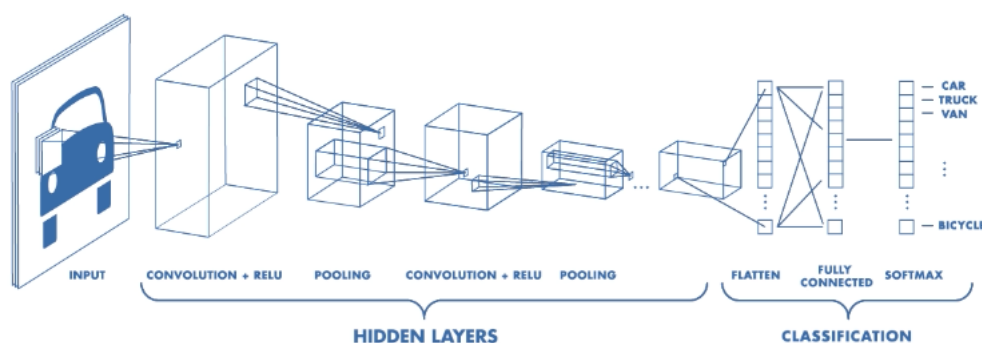


FIGURE 2.2 – Exemple de réseau de convolutionnel profond [1].

Les CNN sont constitués de deux types de couches, les couches convolutionnelles ou convolutive qui extraient des caractéristiques différentes dans les images et les couches entièrement connectées basées sur un réseau de neurones classique multicouche pour la classification (Figure 2.2). Le nombre de couches de convolution dépend de la tâche de classification ainsi que du nombre et de la taille des données d'entraînement. Le deep learning a la particularité de disposer d'une multitude de couches de convolution pour une bonne efficacité, ce qui implique plusieurs coefficients de pondération à régler. Par conséquent, une quantité très importante de données d'apprentissage est requise pour un apprentissage efficace. Ceci explique la réussite du réseau AlexNet [95] à la compétition ImageNet. Cette compétition consiste à entraîner un algorithme d'apprentissage sur une grande base de plus de 1 million d'images et 1000 objets différents (classes). Tous les algorithmes entraînés sont testés sur 100 000 images.

Sur la Figure 2.3, l'impact du résultat des méthodes de deep learning de 2012 à 2017 sur la base ImageNet est notable. En 2015, des méthodes plus performantes qu'un être humain sur cette même base ont été développées. Nous pouvons également noter que c'est durant cette année qu'une architecture utilisant plus de 100 couches a été proposée [71].

Une couche de convolution est composée de plusieurs filtres ou noyaux de taille identique qui transforment l'image en entrée en une sortie dont la valeur des pixels dépend de chaque filtre, tel que représenté sur la Figure 2.4. De plus, les coefficients de ces filtres sont ajustés lors de l'apprentissage. Tout comme dans les réseaux neuronaux classiques, des **fonctions d'activation** sont utilisées pour rendre les couches de convolution non

### 2.3. DÉTECTION DES ADVENTICES PAR LES MÉTHODES D'APPRENTISSAGE MACHINE

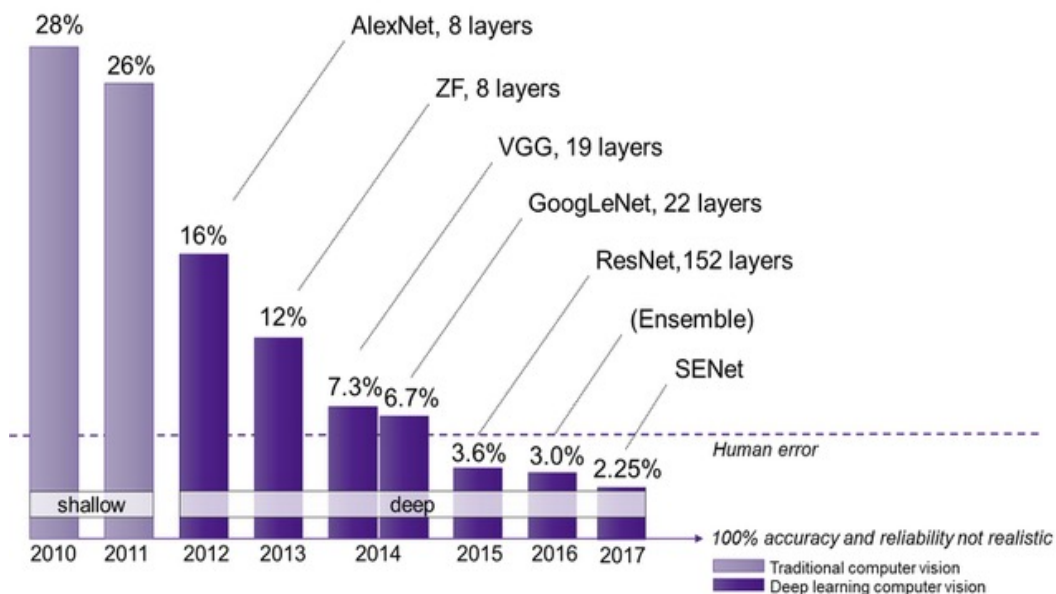


FIGURE 2.3 – Evolution du taux d’erreur en fonction des années, de l’approche et de la profondeur des réseaux [3].

linéaires. En fonction des applications et des architectures des réseaux, le pas de déplacement des filtres après chaque convolution est spécifié pour éviter le recouvrement entre fenêtre : c’est le *stride*. Par contre, empiler des couches de convolution entraîne indirectement une augmentation des opérations de calcul ; des couches de *pooling* sont par conséquent rajoutées dans le but de réduire continuellement la dimension des images et donc le nombre de paramètres à traiter dans le réseau : c’est une forme de sous-échantillonnage spatiale de l’image. Cette couche applique une fonction permettant de sélectionner par exemple un pixel sur quatre. La fonction couramment mise en œuvre est la fonction maximum, d’où son appellation *max-pooling* (Figure 2.5). Le *max-pooling* permet de réduire la taille des images tout en conservant les informations importantes. Pour gérer les pixels situés sur les bords de l’image, en fonction de la taille du filtre, le *padding* est utilisé. Dans le cas de la classification, après la phase d’extraction de caractéristiques opérée, le dernier résultat qui est sous forme de matrice est converti en un vecteur pour être placée en entrée de couches entièrement connectées (*fully connected layer* ou *FCL* en anglais). Les FCL ont le même principe de fonctionnement qu’un perceptron multicouche. Enfin la dernière couche du réseau est une généralisation de fonction logistique appelée fonction *softmax* ou fonction exponentielle normalisée ; elle normalise les K sorties de la dernière couche entièrement connectée en une distribution de probabilités dont la somme totale est égale à 1.

Dans le domaine agricole, les algorithmes d’apprentissage profond sont employés dans plusieurs applications [90]. Dans la détection des adventices, cette approche a été utilisée pour classer les parcelles de jacinthe d’eau, les touffes dentelées et les pommes de soude tropicales dans [76]. Mortensen *et al.* [123] ont appliqué l’apprentissage profond pour la segmentation sémantique dans le contexte de cultures mixtes sur des images d’une parcelle d’essais de radis oléagineux avec orge, herbe, mauvaises herbes, souches et sol. Milioto *et al.* [121] ont fourni une classification précise des mauvaises herbes dans des champs

## 2.3. DÉTECTION DES ADVENTICES PAR LES MÉTHODES D'APPRENTISSAGE MACHINE

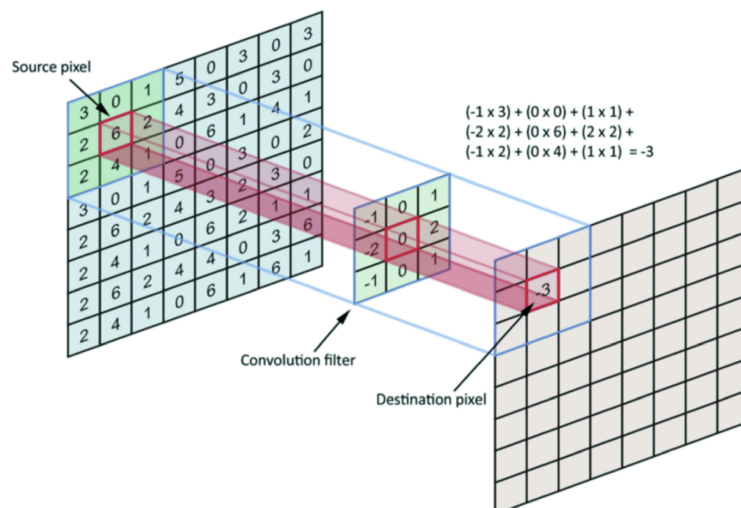


FIGURE 2.4 – Application d’une convolution sur une image [96].

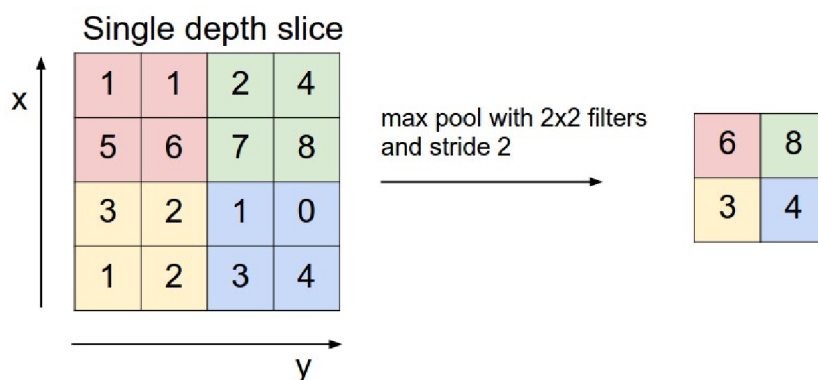


FIGURE 2.5 – Exemple de sous-échantillonnage par max-pooling [54].

de betteraves sucrière avec des robots agricoles mobiles. Dans [134], Potena *et al.* ont mis en oeuvre deux architectures d’apprentissage profond pour détecter les adventices à partir d’images multispectrales acquises par un robot. La première a servi à séparer la végétation du sol et la seconde à réaliser la classification culture/adventices. Les résultats obtenus étaient comparables à ceux de [121] sur la même base de test. Dos Santos Ferreira *et al.* [43] ont appliqué une méthode appelée ConvNet pour la détection des mauvaises herbes dans les cultures de soja. Les résultats ont été comparés à ceux du SVM, Adaboost et le Random Forest appliqués sur 218 attributs. 15000 échantillons ont été utilisés pour l’apprentissage. Le ConvNet s’est montré plus efficace et a atteint une précision supérieure à 98 % pour la détection des mauvaises herbes à feuilles larges et des graminées par rapport au sol et à la culture de soja. Dans [141], les auteurs ont employé un réseau entièrement convolutif pour détecter les adventices dans des images multispectrales acquises par un drone à 2 m d’altitude dans un champ de betterave. L’une des contributions principales de cet article est l’utilisation de la segmentation sémantique en temps réel (2Hz) sur des images de drone. Pour annoter leurs images, ils ont sélectionné des zones de  $40 \times 40$  m et il a fallu environ 60 mn pour annoter une image, en sachant qu’ils avaient choisi au total 132, 243 et 90 images respectivement de cultures, d’adventices et de cultures et adventices mélangées. L’indice NDVI a été utilisé pour séparer la végétation et le sol.

Pour ce faire, un flou gaussien a été appliqué sur le NDVI accompagné d'un seuillage de Otsu (seuil=1.2). Seules 3 bandes ont été traitées : le proche infrarouge, le red-edge et le NDVI. Leur méthode s'est cependant confrontée au problème récurrent des données : ils ont constaté qu'un modèle créé avec des données issues d'un champ A n'était pas efficace dans un champ B.

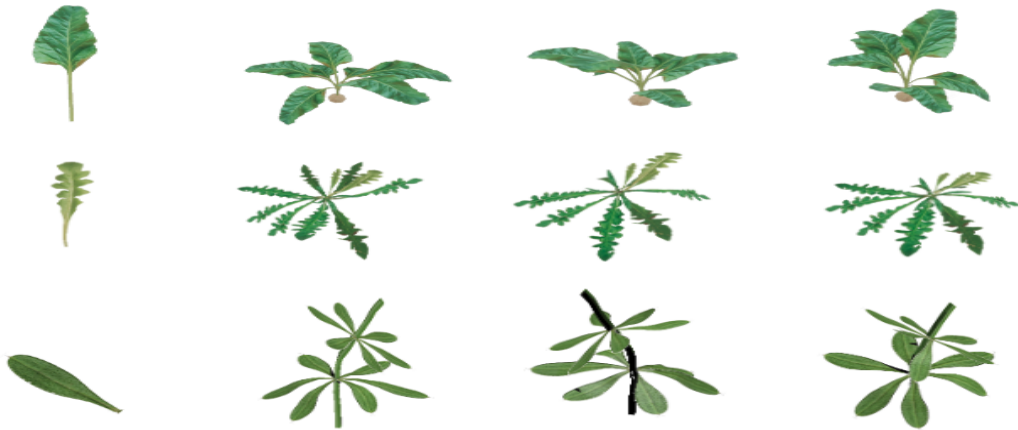


FIGURE 2.6 – Exemple de trois modèles de mauvaises herbes obtenus à partir de la texture d'une seule feuille [40].

Le principal point commun entre les algorithmes d'apprentissage machine est le besoin en données. Pour une bonne optimisation des modèles d'apprentissage profond, il est nécessaire d'avoir à disposition un certain nombre de données étiquetées. Mais comme nous l'avons mentionné précédemment, la création de grands ensembles de données agricoles avec des annotations au niveau des pixels est une tâche qui demande beaucoup de temps. Peu d'efforts ont été faits pour développer des systèmes entièrement automatiques d'apprentissage et d'identification des mauvaises herbes dans les champs agricoles. Dans [40], Di Cicco *et al.* suggéraient l'utilisation d'ensembles de données synthétiques pour l'apprentissage (Figure 2.6). Cette technique nécessite cependant une modélisation précise en termes de texture, de modèles 3D et de conditions d'éclairage.

## 2.4 Conclusion

Dans ce chapitre, des méthodes de l'état de l'art appliquées pour la détection des adventices ont été présentées. Ces méthodes se subdivisent en deux groupes : les méthodes exploitant la géométrie du champ (l'alignement des rangées de culture) pour détecter les adventices situées dans l'inter-rang et les méthodes utilisant l'apprentissage machine pour détecter les adventices situées à l'intérieur et à l'extérieur des rangées de culture. Les premières sont généralement utilisées dans le cas où les images acquises ne possèdent pas une résolution spatiale permettant d'identifier individuellement les plantes. Ainsi seule l'hypothèse de l'alignement des plantes est exploitable. Cependant, même si les secondes méthodes sont parfois jugées efficaces, elles souffrent du manque et de la complexité de la collecte et surtout de l'annotation des données d'apprentissage.

## 2.4. CONCLUSION

---

# Chapitre 3

## Données et systèmes d'acquisition

### Sommaire

---

<b>3.1</b>	<b>Contexte d'acquisition des images</b>	<b>32</b>
<b>3.2</b>	<b>Drones</b>	<b>34</b>
3.2.1	Aile volante eBee de Sensefly	34
3.2.2	DJI Phantom 3	35
3.2.3	DJI Inspire 2	36
<b>3.3</b>	<b>Conclusion</b>	<b>37</b>

---

La détection des adventices est un processus complexe qui dépend souvent de différents facteurs :

- Le système d'acquisition (drones, robot ou satellites et capteurs) ;
- L'éclairage (soleil, ...);
- La structure géométrique du champ, notamment existence ou non de rangées de culture ;
- Le niveau de croissance des cultures et des adventices à détecter.

Ce chapitre est divisé en quatre sections. La première décrit le contexte d'acquisition des images manipulées dans cette thèse et la seconde présente les drones utilisés.

## 3.1 Contexte d'acquisition des images

Au démarrage de cette thèse, en collaboration avec la chambre régionale d'agriculture Centre-Val de Loire et les chambres départementales d'agriculture, le but était de développer une méthode de détection d'adventices sur des données acquises à l'aide d'une aile volante (section 3.2.1) à une altitude de 100 m et qui embarquait une caméra canon de 12 mégapixels (MP). Ces données devaient être acquises pour initialement effectuer de la préconisation azotée en grande culture mais l'objectif a été revu pour que, en un seul passage, on puisse faire non seulement de la préconisation en engrais mais également proposer une carte d'infestation en adventices. Pour réaliser cette étude, les images ont été acquises dans un champ de maïs.

A l'obtention de ces images une analyse visuelle a été effectuée. Il a été constaté que les images présentaient des rangées de culture et, visuellement, avec une résolution spatiale de 2,5 cm, il était parfois difficile, même pour les experts d'y repérer les adventices (Figure 3.1). En échangeant avec les experts des chambres d'agriculture, il est apparu que pour repérer les adventices, il fallait observer l'alignement des plantes pour s'apercevoir que les adventices poussaient de façon désordonnée. Ainsi, à travers cet alignement, l'hypothèse consistait à ce que toutes les plantes qui étaient hors ou qui perturbaient cet alignement pouvaient être considérées comme des adventices.

Dans le but de proposer une méthode capable de reproduire au mieux l'approche utilisée par les experts, des méthodes ont été développées voir le Chapitre 4.

Cependant, avec cette approche, on ne peut procéder que par zonage, à savoir identifier des zones d'agrégation de plantes plutôt que des plantes individuellement. Ceci peut s'expliquer en observant les résultats de la simulation effectuée dans la thèse de Marine Louargant [109] sur la relation entre la surface des plantes et la résolution de la caméra.

Il peut être constaté que sur les Figures 3.2 et 3.3, qu'il y a un grand risque de mélange spectrale avec une résolution spatiale de 2.5 cm. De plus, le nombre de pixels contenant de la végétation pure peut être faible si la surface des plantes est petite. Par conséquent, il a fallu effectuer des acquisitions à une altitude plus proche du champ tout en sachant qu'un compromis en a découlé, à savoir le rallongement du temps d'acquisition.

Face à la limitation de l'aile volante qui ne peut réaliser des acquisitions en dessous de 50 m, et pour tenir compte du compromis cité ci-avant, il a été décidé d'exploiter de nouvelles images acquises dans des champs d'épinard et de haricot à une altitude de 20 m avec un capteur Sony embarqué sur un drone DJI Phantom 3 pro (section 3.2.2). Le capteur utilisé était un appareil photo couleur Sony ILCE-7R de 36 MP.

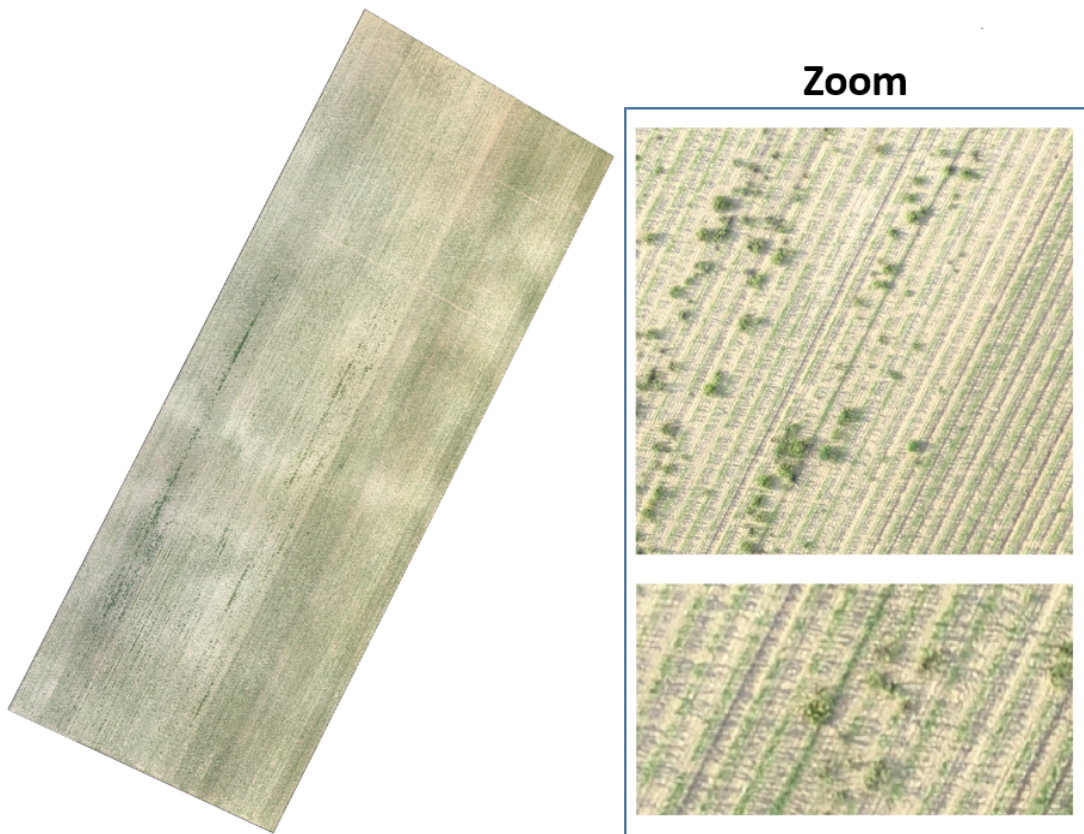


FIGURE 3.1 – Image du champ de maïs acquise à 100 m avec une aile volante et une caméra couleur. La résolution spatiale des images est de 2,5 cm.

La résolution spectrale des images à disposition était subcentimétrique (Figure 3.4). Avec ces données comme dans le cas des images acquises à une altitude de 100 m une analyse visuelle a été effectuée. Les images présentaient des rangées de culture bien identifiables et il était possible de repérer visuellement les plantes et de séparer les adventices des cultures.

Pour ces images, nous avons opté pour une approche par apprentissage machine dont le but est de former des modèles de détection des adventices inter-rangs et intra-rangs robustes et facilement reproductibles dans d'autres champs. Pour cette approche une question s'est posée, à savoir : comment réaliser la collecte des données et le choix des descripteurs ? Car dans les champs, d'une semaine à une autre, la croissance des plantes et les conditions d'acquisition peuvent être différentes. L'extraction et la sélection des caractéristiques est également un grand problème puisque, dans la littérature on constate que différentes caractéristiques sont utilisées sur la même culture pour des hauteurs et des croissances de plantes différentes. Dans le chapitre 5 une méthode est proposée pour répondre à cette question.

En plus des images de betterave acquises dans un champ de betteraves à 10 m d'altitude avec le capteur RedEdge-M embarqué sur un drone DJI Inspire2 (section 3.2.3) ont été manipulées. Ces dernières proviennent d'une base de données publique publiée dans [142]. Les images ont une résolution spatiale de 1 cm (Figure 3.5). Le capteur RedEdge-M

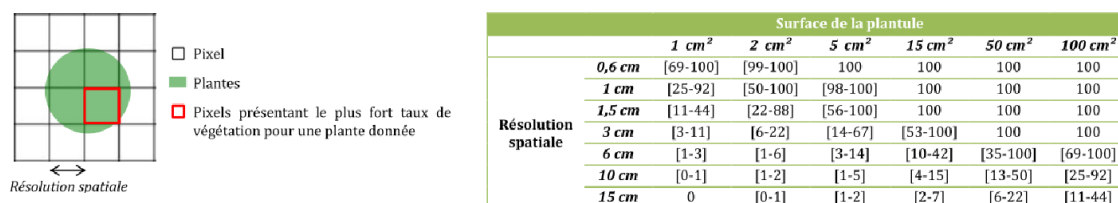


FIGURE 3.2 – Proportion de végétation minimales et maximales en pourcentage présentes dans le pixel ayant le plus grand taux de végétation en fonction de la résolution spatiale de l'image et de la taille de l'adventice est répertoriée [109].

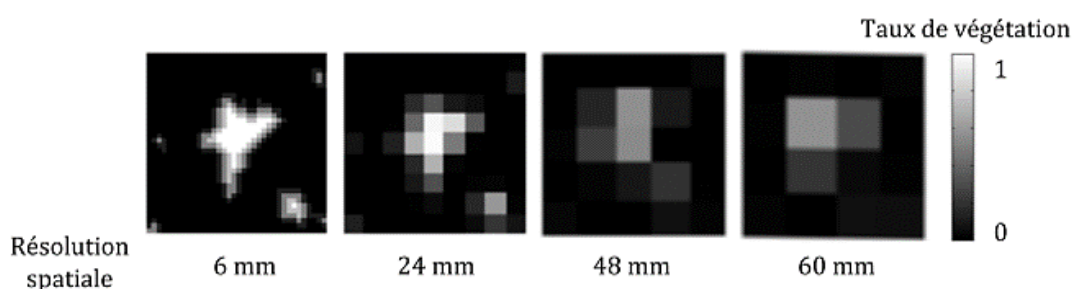


FIGURE 3.3 – Dégradation de l'image des taux de végétation de 6 mm à 6 cm de résolution spatiale [109].

est une caméra multispectrale de la compagnie MicaSense®, conçue pour une intégration facile et flexible sur une large variété de drones. Cette caméra est capable, de manière simultanée de capter des images sur 5 bandes spectrales différentes dont les trois bandes du visible (RGB), le red-edge et le proche infrarouge. Avec un poids de 173 g, le RedEdge-M a une résolution spatiale de 8 cm à 120 m d'altitude. Il dispose de plusieurs modules, dont un module GPS et un capteur de lumière. Le GPS permet de géolocaliser les images et également à automatiser les acquisitions. Quant au capteur de lumière, il permet de corriger le changement d'éclairage lors du post-traitement des images. C'est un capteur de lumière incidente à 5 bandes et il mesure la lumière ambiante pendant le vol pour chacune des 5 bandes.

La base de données contient cinq images orthomosaïquées (000, 001, 002, 003 et 004) et leurs vérités terrain correspondantes. Chaque image orthomosaïquée est une partie du champ. La taille des cultures et des adventices est de 15-20 cm et de 5-10 cm respectivement. Dans cette base en plus des bandes brutes provenant du capteur, deux autres types d'images en bande composite sont disponibles, les images en couleur (RGB) et les images en fausses couleurs (CIR) formées en empilant les canaux R, V et le proche infra rouge.

## 3.2 Drones

### 3.2.1 Aile volante eBee de Sensefly

Le drone utilisé est une aile volante communément appelé l'eBee de la société SenseFly. Ce drone très léger (700 g, charge comprise) est équipé de capteurs de position tels qu'une

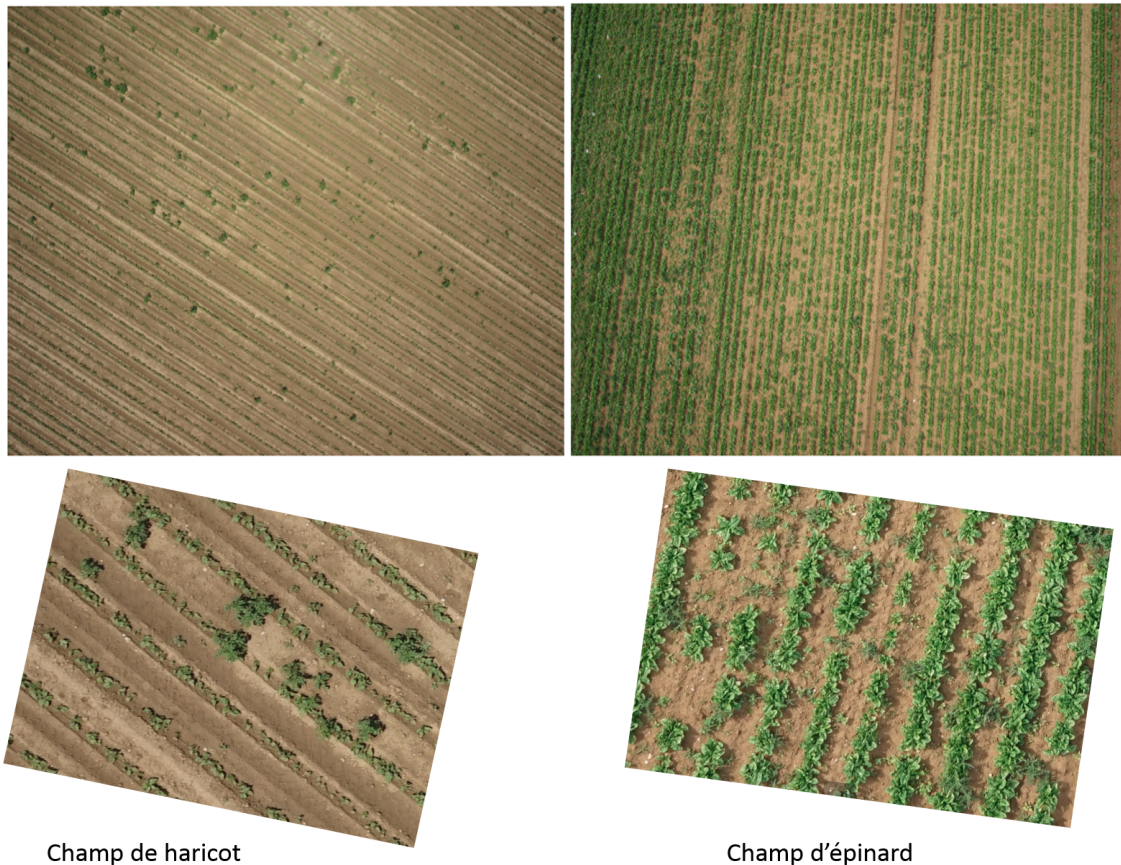


FIGURE 3.4 – Images acquises dans les champs de haricot et d'épinard à 20 m d'altitude avec une caméra Sony. La résolution spatiale est inférieure à 1 cm.

centrale inertielle et un GPS, permettant ainsi de géoréférencer les données acquises lors du vol ; il a également la réputation d'être résistant au vent (jusqu'à 50 km/h). Grâce à ce système, les vols sont effectués automatiquement à une hauteur de vol configurable comprise entre 50 et 150 m avec une erreur moyenne de positionnement absolue du GPS qui est de 2 à 5 m, ce qui peut induire une erreur moyenne de mosaïquage de 2,06 m d'erreur à moins de 2 cm. Il peut couvrir jusqu'à 150 ha en un seul vol à 150 m d'altitude.

### 3.2.2 DJI Phantom 3

La gamme de drone multicoptère Phantom de DJI® possède une technologie avec un rapport qualité/prix sur le marché considéré abordable. Actuellement, 1 drone sur 2 vendu dans le monde est conçu par DJI (Chine). Le drone DJI Phantom 3 pro possède un poids en vol de 1,28 kg, batterie et hélices autobloquantes comprises. En tenant compte de ses hélices, il a une envergure de 59 cm et peut atteindre une altitude maximale de 6000 m. Cependant, la réglementation en France ne permet pas de voler à plus de 150 m. Le Phantom 3 pro dispose d'une autonomie en vol d'environ 23 mn. Ses vitesses d'ascension et de descente maximales sont respectivement de 5 m/s et de 3 m/s ; de plus, il peut atteindre une vitesse de 16 m/s en vol. La plage de précision du vol stationnaire est de

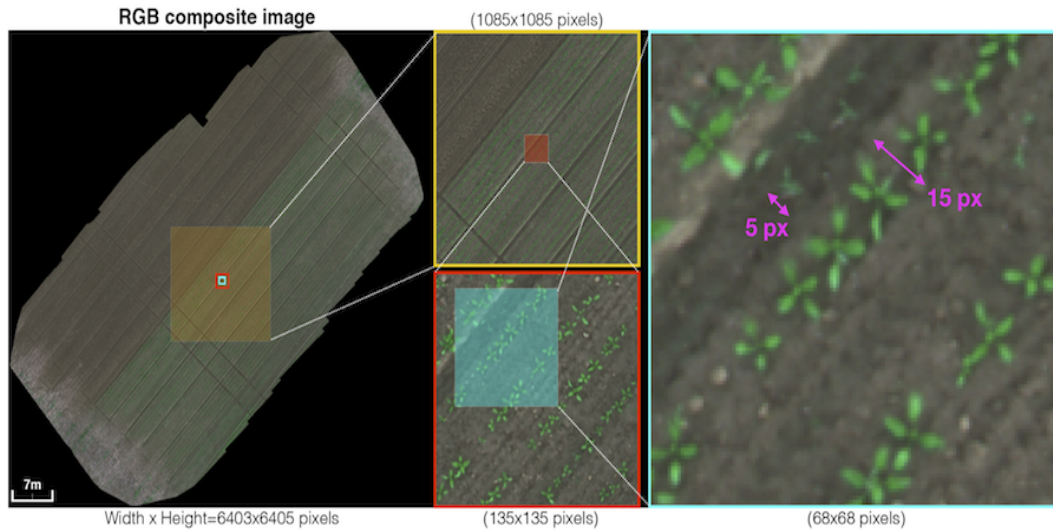
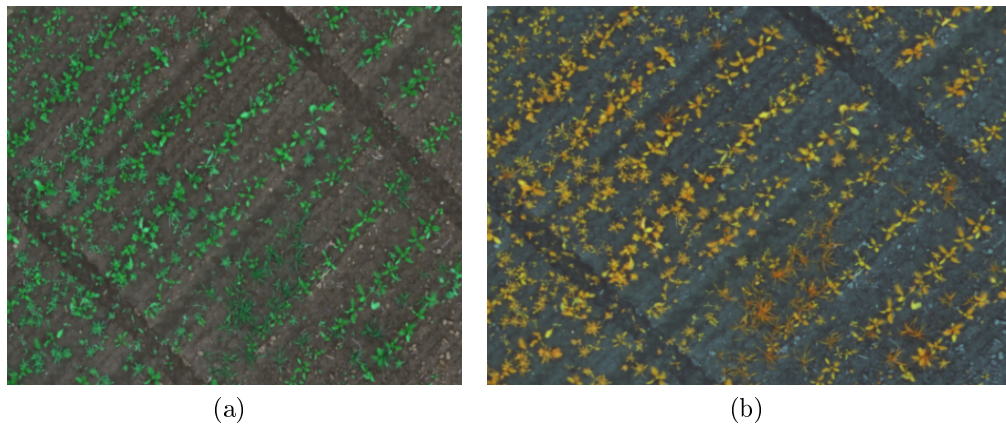


FIGURE 3.5 – Image du champ de betterave.



(a)

(b)

FIGURE 3.6 – Exemple d’images RGB (a) et CIR (b) provenant de la base publique utilisée [142].

$\pm 10$  cm en vertical et  $\pm 1.5$  m en horizontal ; pour améliorer son positionnement, il se sert d’une technologie qui exploite le système de positionnement satellitaire GPS/GLONASS (GLONASS est la forme abrégée de Global Navigation Satellite System qui est la version du GPS créé par la Russie). Tout comme le GPS, une flotte de 24 satellites est disponible pour localiser un point sur le globe. Le but d’utiliser le système GPS/GLONASS est qu’il améliore le positionnement en sélectionnant le système qui a le plus de satellites visibles par le drone lors de l’acquisition. Il dispose en plus d’une technologie de positionnement par caméra (*vision positioning* en anglais) qui couple les performances d’un sonar et de la caméra pour affiner son positionnement lors de vols en intérieur.

### 3.2.3 DJI Inspire 2

Le DJI Inspire 2 est un quadcoptère, qui pèse 3440 g sans nacelle et sans caméra. Il peut embarquer jusqu’à 810 g de matériel au décollage, ce qui peut être très utile. Il

### 3.3. CONCLUSION

---

dispose d'une technologie double batterie qui lui procure une autonomie de 27 mn. Avec des vitesses de pointe à 94 km/h (26 m/s), une vitesse de 6 m/s en ascension et 4 m/s en descente, ce drone est capable de passer de 0 à 80 km/h en 5 s. Le DJI Inspire 2 est robuste face au vent dont la vitesse maximale ne dépasse pas 10 m/s et peut être utilisé dans des zones à basse température, jusqu'à  $-20^{\circ}\text{C}$ , grâce à la technologie auto-chauffante dont il est équipé.

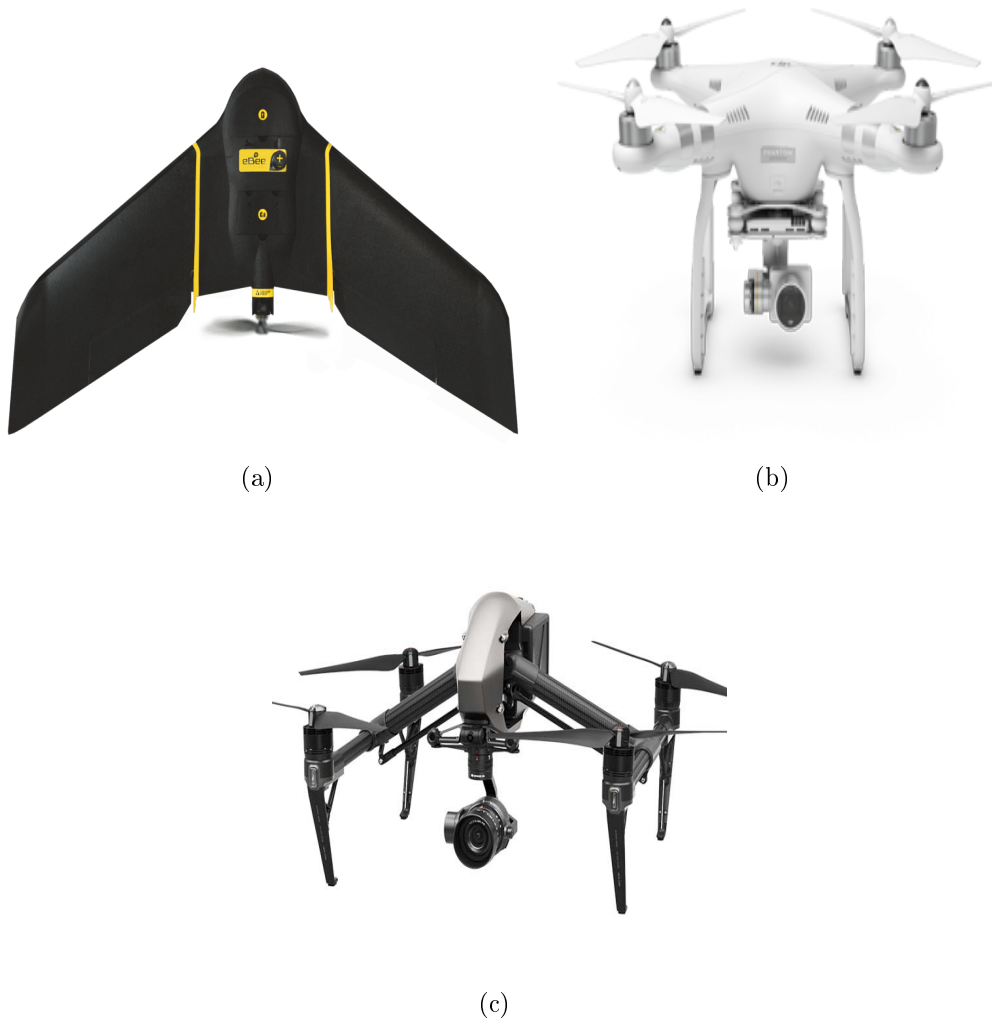


FIGURE 3.7 – L'aile volante eBee de Sensefly (a), les drones DJI Phantom 3 Pro (b) et DJI Inspire 2 (c).

### 3.3 Conclusion

Dans ce chapitre les systèmes d'acquisition et les types de données ont été présentés. Dans la suite de ce manuscrit ces données seront utilisées pour évaluer les méthodes proposées.

### 3.3. CONCLUSION

---

# Chapitre 4

## Détection des adventices par utilisation de l'information géométrique du champ

### Sommaire

---

<b>4.1</b>	<b>Détection des rangs par la transformée de Hough et le SLIC (CR-Hough-SLIC)</b>	<b>40</b>
4.1.1	Transformée de Hough	41
4.1.2	Squelettisation	43
4.1.3	SLIC	44
4.1.4	Algorithme de détection des rangées de culture	45
<b>4.2</b>	<b>CRowNet : approche deep learning pour la détection des rangées de culture</b>	<b>46</b>
4.2.1	S-SegNet	48
4.2.2	HoughCNet	49
<b>4.3</b>	<b>Détection des adventices</b>	<b>52</b>
	Utilisation des relations spatiales entre superpixels	52
<b>4.4</b>	<b>Expérience et résultats</b>	<b>54</b>
4.4.1	Métriques d'évaluation	54
4.4.2	Evaluation du CR-Hough-SLIC	55
4.4.3	Évaluation du CRowNet	58
4.4.4	Évaluation qualitative	59
4.4.5	Évaluation quantitative	60
4.4.6	Détection des adventices	63
<b>4.5</b>	<b>Conclusion</b>	<b>66</b>

---

## 4.1. DÉTECTION DES RANGS PAR LA TRANSFORMÉE DE HOUGH ET LE SLIC (CR-HOUGH-SLIC)

Plusieurs études ont montré l'intérêt de l'automatisation de la détection des rangées de culture pour la navigation robotisée, tout comme pour la détection des mauvaises herbes situées entre les rangées. Généralement, ces méthodes se répartissent en quelques catégories selon l'approche utilisée, notamment la transformée de Hough, la régression linéaire, l'analyse de blob, la vision stéréoscopique ou l'analyse des bandes horizontales comme mentionné dans le Chapitre 2.

Dans ce chapitre, deux nouvelles méthodes de détection de rangées de culture ainsi qu'une méthode de détection d'adventices sont présentées. La première méthode de détection des rangées de culture proposée combine des approches de traitement d'images classiques, à savoir la transformée de Hough et les superpixels formés par simple regroupement (clustering) itératif linéaire, communément appelé SLIC [4]. Quant à la deuxième méthode de détection de rangées de culture, elle utilise une approche totalement nouvelle et est motivée par l'efficacité récente des algorithmes d'apprentissage profond. Cette méthode utilise les réseaux entièrement convolutionnels pour la détection des rangées de culture et s'appuie sur deux architectures différentes de réseaux convolutionnels : le S-SegNet basée sur le SegNet et le HoughCNet qui utilise la transformée de Hough. Enfin la méthode de détection d'adventices consiste à utiliser les rangées de culture détectées pour localiser les adventices.

Ce chapitre est divisé en cinq sections : les deux premières sections présentent les méthodes de détection de rangées de culture. La méthode de détection des adventices est présentée dans la troisième section. Les résultats sont commentés et discutés dans la quatrième section. La cinquième section conclut ce chapitre.

### 4.1 Détection des rangs par la transformée de Hough et le SLIC (CR-Hough-SLIC)

Comme présenté dans l'organigramme de la Figure 4.1, cette méthode prend comme entrée une image qui est ensuite segmentée pour éliminer les éléments indésirables tels que le sol, les cailloux et la paille. Dans ce chapitre l'indice ExG suivi d'un seuillage de Otsu est choisi comme méthode de segmentation (section 2.1). La Figure 4.2 montre un exemple de segmentation sur une image provenant d'un champ de maïs.

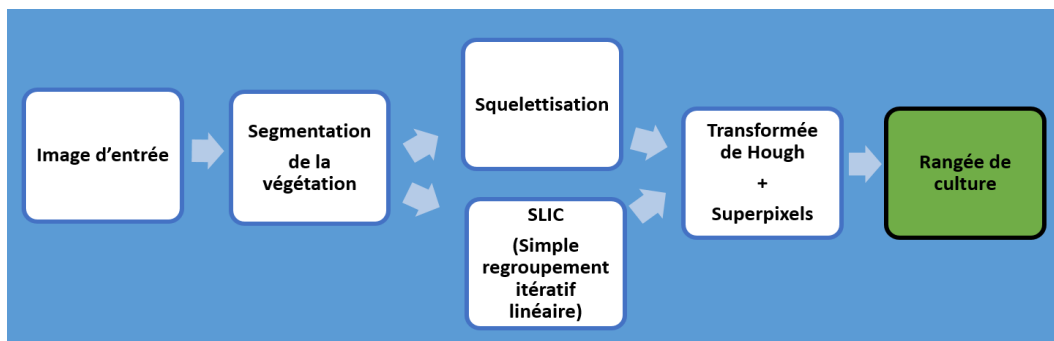


FIGURE 4.1 – Organigramme de la méthode de détection des rangées de culture combinant transformée de Hough et superpixels.

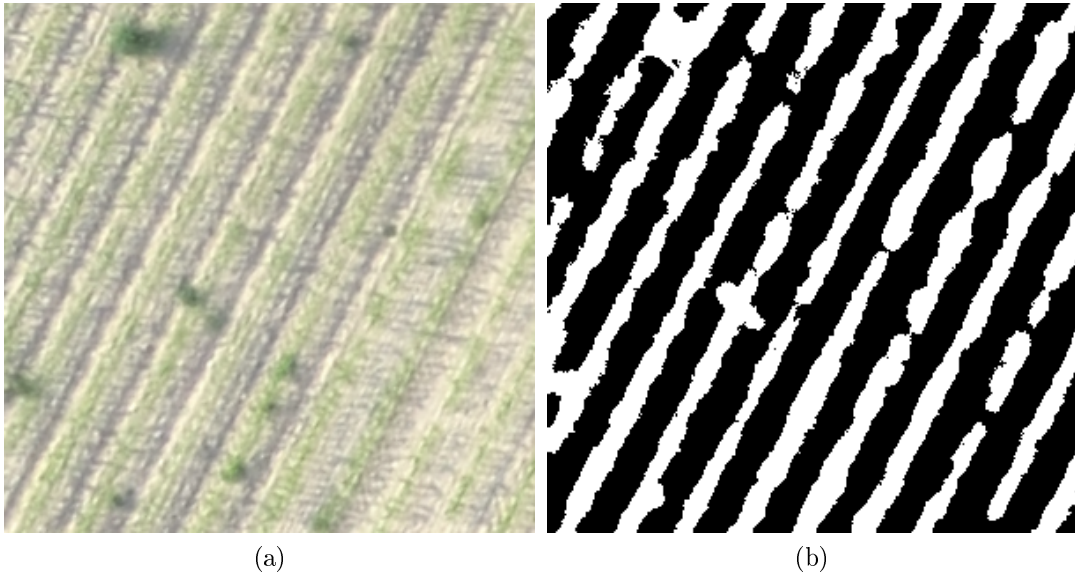


FIGURE 4.2 – Segmentation d’une image provenant d’un champ de maïs (a) et sa segmentation à l’aide du seuillage de Otsu appliqué sur l’indice ExG (b).

Après la segmentation de la végétation, une squelettisation est appliquée (voir section 4.1.2) dans le but de réduire le temps de traitement de la transformée de Hough qui est couplée aux superpixels formés par le SLIC pour détecter les rangées de culture.

### 4.1.1 Transformée de Hough

Le fonctionnement de la transformée de Hough se base sur un mécanisme d’accumulation (vote) de l’espace image (pixels) à un espace de paramètres multidimensionnel appelé espace de Hough. Cet espace permet de simplifier le problème complexe de détection globale de formes dans l’espace image. Elle est utile pour la détection de contours descriptibles par peu de paramètres, notamment les droites, les cercles, les ellipses, etc.

Par exemple, pour détecter une droite passant par deux points dans une image (espace  $(x, y)$ ), nous pouvons nous servir des paramètres cartésiens (Équations 4.1) composés de la pente  $a$  et l’ordonnée à l’origine  $b$ . Dans ce cas, l’espace de Hough aura un accumulateur à deux dimensions ou matrice  $M(a,b)$  dont chaque case correspond aux votes pour une droite particulière. Les cases qui ont les plus grands votes correspondent aux principales droites présentes dans l’image.

$$y = ax + b \tag{4.1}$$

Pour un point  $(x_0, y_0)$  de l’espace image, toutes les droites qui passent par ce point ont un couple  $(a, b)$  qui satisfait les Équations 4.2 et 4.3.

Dans l’espace image :

$$y_0 = ax_0 + b \tag{4.2}$$

Dans l’espace de Hough :

$$b = -x_0a + y_0 \tag{4.3}$$

#### 4.1. DÉTECTION DES RANGS PAR LA TRANSFORMÉE DE HOUGH ET LE SLIC (CR-HOUGH-SLIC)

Ainsi, pour un point dans l'espace image, il y a une multitude de droites dans l'espace de Hough, et pour tous les points appartenant à une même droite dans l'espace image, on a une intersection de droites au point  $(a, b)$  : le nombre de droites qui s'intersectent au point  $(a, b)$  correspond au nombre de votes.

L'un des inconvénients d'utiliser les paramètres cartésiens est qu'ils ne sont pas bornés et peuvent varier de  $-\infty$  à  $+\infty$ . Par conséquent, dans l'espace de Hough, l'accumulateur peut avoir une infinité de case, ce qui est moins pratique et risque d'être très coûteux en temps de calcul. C'est la raison pour laquelle, en 1972, Duade *et al.* [46] ont proposé, pour la première application de la transformée de Hough dans le contexte de la détection des lignes dans des images, d'utiliser la formulation dite « normale » de la droite ou coordonnées polaires (Équation 4.4) puisque, contrairement à l'espace  $(a, b)$ , l'espace paramétrique  $(\rho, \theta)$  est homogène et surtout borné ( $\rho$  est borné par la taille de l'image et  $\theta$  est borné dans l'intervalle  $[-\pi/2, \pi/2]$ ).  $\rho$  est la distance de la droite à l'origine du repère et  $\theta$  est l'angle que fait la perpendiculaire à la droite avec l'axe x.

$$\rho = x \cos(\theta) + y \sin(\theta) \quad (4.4)$$

Avec les coordonnées polaires, chaque point  $(x_i, y_i)$  du plan image donne une sinusoïde dans l'espace de paramètres  $(\rho, \theta)$ . Par conséquent,  $N$  points colinéaires appartenant à la même droite donneront des courbes qui s'intersectent à un point  $(\rho_j, \theta_j)$  de l'espace de paramètres (Equation 4.5 et Figure 4.3). Dans l'espace de Hough, le nombre de votes au point  $(\rho_j, \theta_j)$  sera de  $N$  (en référence aux  $N$  points de la droite). Plus une ligne est longue dans l'espace image, plus le nombre de vote en sa faveur est important dans l'espace de Hough.

$$\rho = x_i \cos(\theta) + y_i \sin(\theta) \quad (4.5)$$

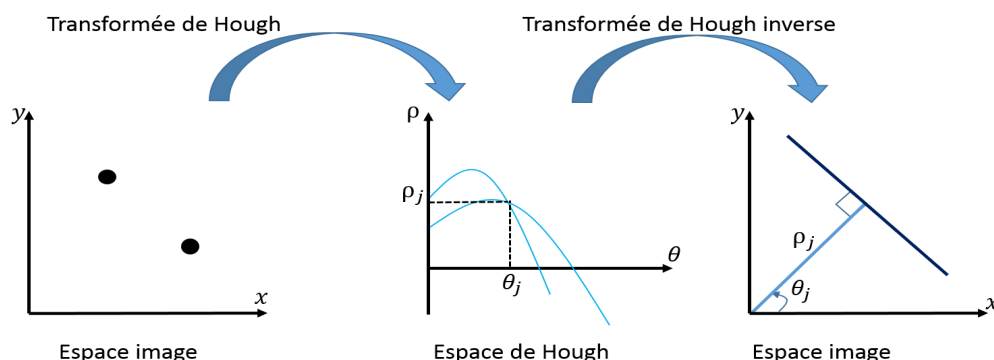


FIGURE 4.3 – Exemple de fonctionnement de la transformée de Hough pour détecter une droite à partir de deux points situés dans une image.

Cependant, malgré l'efficacité de la transformée de Hough, cette méthode présente une faiblesse due à son temps de calcul. Pour surmonter ce problème, de nombreuses variantes de la transformée de Hough ont été proposées dans la littérature [78, 89, 169].

De notre côté, pour cette méthode, nous avons appliqué la transformée de Hough sur le squelette des rangées de culture puisqu'elle permet de maintenir la structure globale du

champ avec peu de pixels (Figure 4.4). Avec cette opération, la durée du traitement peut être réduite de plus de 75 %.

##### 4.1.2 Squelettisation

En principe, une rangée de culture est une succession de plusieurs lignes très rapprochées, et en plus non uniforme sur les rangées et entre les rangées. Ainsi, plus il y a de plantes dans les rangées, plus il faut de temps pour calculer la transformée de Hough. Pour résoudre le problème du temps de calcul, nous proposons d'utiliser les squelettes des rangées. En vision par ordinateur, il est devenu très important de réduire au minimum nécessaire la quantité d'informations à traiter car cela permet de réduire le temps de calcul, la distorsion du contour et certains bruits locaux tout en conservant des propriétés topologiques et géométriques importantes. Par exemple, la représentation en lignes minces (squelettes) des rangées de culture serait plus proche de la conception humaine de ces motifs dans les champs agricoles. En effet, la squelettisation est un processus qui permet de représenter un motif par une collection d'arcs et de courbes minces (ou presque minces) [98]. Selon la méthode de squelettisation, le résultat peut être appelé squelette ou amincissement. Le terme squelette ou *skeleton* est utilisé lorsque le résultat ne dépend pas du motif original et est plus générique, alors que l'amincissement ou *thinning* est une représentation en ligne du motif original. Les résultats des deux méthodes peuvent être assez similaires en apparence, c'est pourquoi le terme général utilisé est *skeleton*. Dans ce travail, c'est malgré tout le *thinning* qui est utilisé dans le but de trouver les lignes médianes des rangées de culture. L'amincissement itératif est la méthode la plus utilisée ; elle est basée sur la suppression ou le maintien de couches successives de pixels sur les bords jusqu'à ce qu'il ne reste qu'un squelette [98, 17]. Il donne avec peu de pixels les mêmes informations sur la géométrie du champ : direction des rangées, distance inter-rangées et la périodicité des rangées de culture (Figure 4.4).

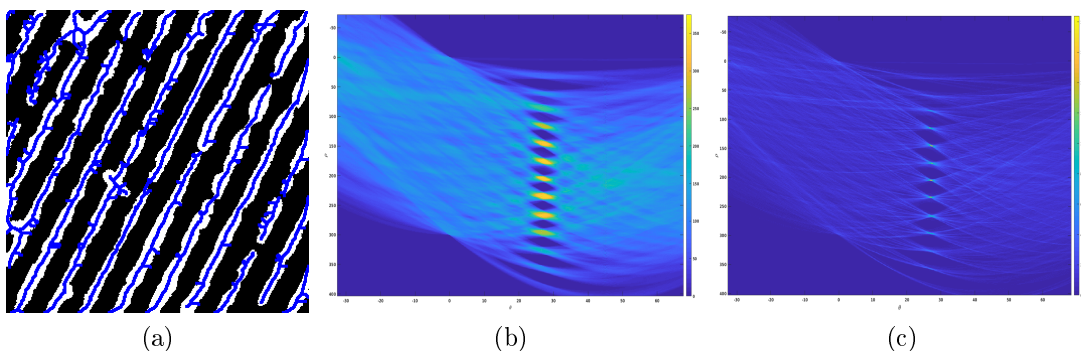


FIGURE 4.4 – (a) Image segmentée et son squelette en bleu. (b) et (c) représentent la transformée de Hough appliquée sur une image segmentée et sur son squelette, respectivement. La durée de traitement a été réduite de 78 %.

### 4.1.3 SLIC

Le SLIC [4] est choisi car il est simple et efficace en termes de qualité des résultats et de temps de calcul. C'est une adaptation de la méthode des k-moyennes (k-means) pour la génération de superpixels avec un contrôle sur la taille et la compacité des superpixels. Sa particularité est qu'il évite de considérer l'ensemble des pixels de l'image lors du calcul de la distance euclidienne. L'espace considéré est proportionnel à la taille fixée du superpixel. La distance pondérée calculée combine l'information de couleur et la proximité spatiale tout en contrôlant la taille et la compacité des superpixels. Le SLIC regroupe les pixels dans un espace plan couleur et image combiné en cinq dimensions  $[labxy]$  pour générer efficacement des superpixels compacts et presque uniformes.  $[lab]$  est le vecteur couleur des pixels dans l'espace colorimétrique CIELAB (Figure 4.5). Cet espace couleur est largement utilisé pour caractériser les couleurs de surface. Il essaye de prendre en compte la réponse logarithmique de l'œil. Il possède le grand avantage d'être uniforme sur le plan perceptif pour de petites distances de couleurs.  $[xy]$  correspond à la position des pixels dans l'image.

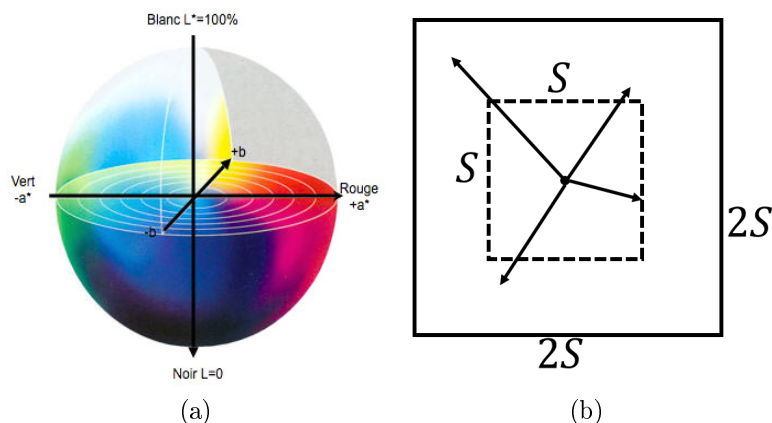


FIGURE 4.5 – (a) Espace colorimétrique CIELAB. La clarté  $L$  dérive de la luminance de la surface ; les deux paramètres  $a$  et  $b$  expriment l'écart de la couleur par rapport à celle d'une surface grise de même clarté. (b) Illustrations des zones de recherche du centre du cluster avec le SLIC.

Si nous souhaitons obtenir  $k$  superpixels, l'image sera divisée en régions quasi identiques de surface  $S \times S$  avec  $S = \sqrt{N/k}$  où  $N$  correspond au nombre de pixels dans l'image. Dans chaque région ou superpixel, un pixel est choisi aléatoirement : c'est le seed. Pour éviter de choisir un pixel se trouvant sur un contour ou sur un pixel de bruit, le pixel choisi est remplacé par son pixel voisin dans une fenêtre  $(3 \times 3)$  qui a le plus faible gradient. Pour mettre à jour le pixel central choisi, la distance est calculée dans un espace de surface  $(2 \times S)^2$  autour du superpixel. Chaque pixel de l'image est associé au centre du cluster le plus proche que la zone de recherche chevauche. Une fois que tous les pixels sont associés au centre du faisceau le plus proche, un nouveau centre est calculé comme vecteur  $[labxy]$  moyen de tous les pixels appartenant au cluster. Ensuite, itérativement, les processus d'association des pixels avec le centre du cluster le plus proche et le recalcul du centre du cluster sont répétés jusqu'à convergence. A noter que le SLIC peut aussi être appliqué sur des images niveau de gris.

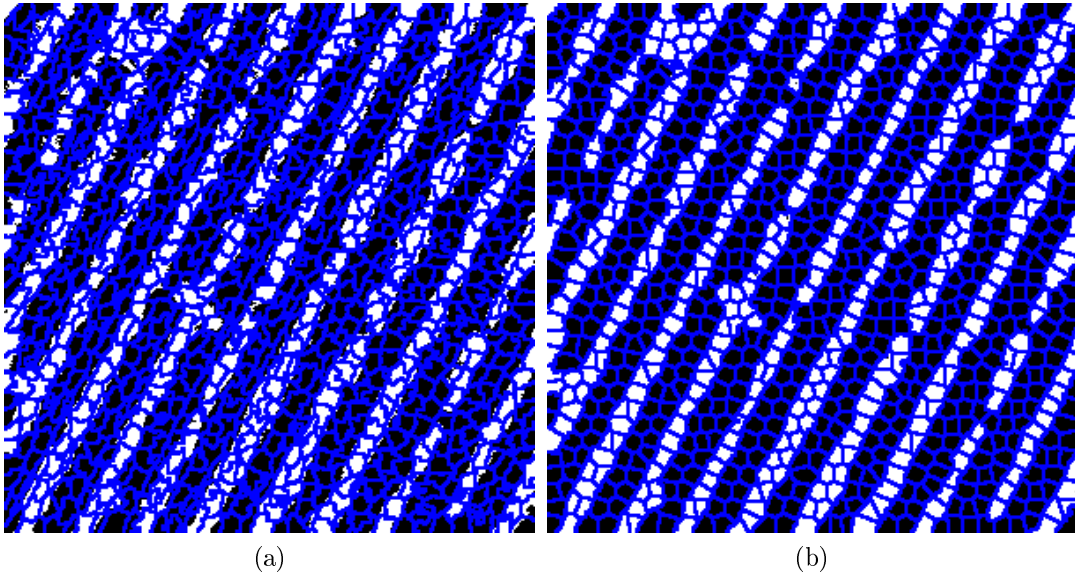


FIGURE 4.6 – Exemple de 1021 superpixels (contours en bleu) générés à partir de l’image couleur (image avant segmentation) (a) et sur son image segmentée (b).

La Figure 4.6 présente un exemple de segmentation en superpixels d’une image à l’aide du SLIC. Sur cet exemple, nous pouvons constater que les superpixels formés sur l’image originale sont moins réguliers, ce qui est normal d’ailleurs car la couleur et l’hétérogénéité des plantes sont prises en compte, contrairement à ceux formés sur l’image segmentée où seul le nombre de superpixels compte vraiment.

#### 4.1.4 Algorithme de détection des rangées de culture

Cette section présente l’algorithme proposé pour la détection des rangées de culture. La transformée de Hough  $H(\theta, \rho)$  est calculée sur le squelette avec une résolution  $\theta$  égale à  $0,1^\circ$ , laissant  $\theta$  prendre des valeurs dans la plage  $] -90^\circ; 90^\circ]$  et  $\rho$  égale à 1. A partir d’un histogramme de directions des squelettes, l’angle le plus représenté est choisi comme orientation principale  $\theta_{lines}$  des rangées de culture.  $H(\theta, \rho)$  a été normalisée  $H_{norm}(\theta, \rho)$  afin de donner le même poids à toutes les rangées, en particulier les rangées courtes proches des bords de l’image [59].  $H_{norm}(\theta, \rho)$  est définie comme le rapport entre l’accumulateur de l’image de végétation et l’accumulateur d’une image totalement blanche de même taille  $H_{ones}(\theta, \rho)$ . Pour ne pas tenir compte des petites rangées créées par l’agrégation des mauvaises herbes entre les rangs, un seuil de 0,1 a été appliqué à la transformée de Hough normalisée. De plus, dans l’agriculture moderne, les cultures sont généralement semées en rangs parallèles avec la même distance inter-rang, ce qui implique que les pics principaux correspondant aux rangées de culture sont alignés autour d’un angle dans l’espace de Hough avec les mêmes intervalles. Malheureusement, en raison des réalités du domaine agricole, les rangées ne sont pas parfaitement parallèles, de sorte que les pics dans l’espace Hough ont des angles proches mais différents et la distance inter-rang n’est pas toujours constante. Afin de ne sauter aucune rangée pendant la détection, toutes les rangées ayant un pic dans l’espace de Hough dont l’angle par rapport à l’orientation globale ( $\theta_{lines}$ ) ne

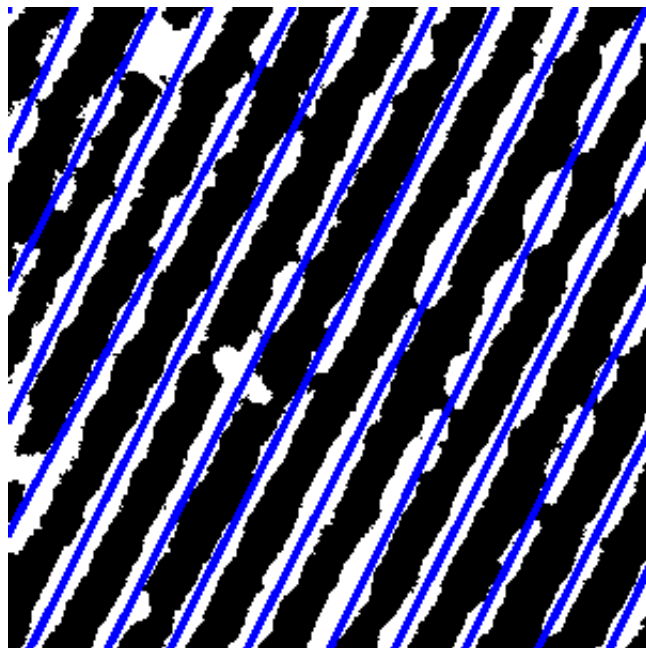


FIGURE 4.7 – Exemple de résultat de détection de rangées de culture.

dépasse pas  $20^\circ$  sont retenues. Cependant, pour éviter de détecter plus d'un pic dans une agrégation, à chaque fois que nous repérons le pic d'une rangée de culture dans  $H_{norm}(\theta, \rho)$ , nous identifions le squelette correspondant et supprimons les votes de ce squelette dans  $H_{norm}(\theta, \rho)$  avant de continuer. Toutes les étapes sont résumées dans l'algorithme 1. Sur la Figure 4.7, nous montrons un exemple de résultat de détection de rangées.

## 4.2 CRowNet : approche deep learning pour la détection des rangées de culture

La segmentation d'images est l'une des tâches fondamentales de la vision par ordinateur. La segmentation sémantique s'apparente à la compréhension d'une image à l'échelle du pixel, c'est-à-dire l'assignement à chaque pixel d'un label correspondant à sa classe d'objet. Cette méthode est adaptée aux objets réguliers et irréguliers contrairement aux méthodes de détection d'objets où les objets doivent s'insérer dans un rectangle englobant. Dans la littérature, nous pouvons rencontrer des méthodes qui se basent sur la reconnaissance d'objet pour extraire puis classer des régions de forme libre de l'image. Le R-CNN [62], par exemple, sélectionne 2000 objets dans l'image. Les attributs (caractéristiques) de ces objets sont extraits à partir d'un CNN avant d'être classés à l'aide d'un SVM. Il peut être construit avec des CNN jugés performants, notamment le AlexNet, VGG, GoogleNet, ... Cette approche a cependant été jugée très coûteuse en temps de calcul et ces caractéristiques ne contiennent pas suffisamment d'informations spatiales pour bien délimiter les régions. A cause de ces limites, différents travaux de recherches ont été menés [70, 137, 138].

Les progrès récents de la segmentation sémantique tendent à convertir les réseaux de CNN classiques (VGG, AlexNet, ...) en des réseaux entièrement convolutionnels (FCN ou

---

**Algorithme 1** : Détection des rangées de culture.

---

```

input   : squelette, superpixels
output : rangées de culture
2 Calcul de l'orientation de tous les squelettes ;
4 Calcul de l'orientation globale  $\theta_{lines}$  des rangées ;
6 Transformée de Hough sur les squelettes  $H(\theta, \rho)$  ;
8  $H_{norm}(\theta, \rho) = H(\theta, \rho) / Hones(\theta, \rho)$  ;
10  $H_p(\theta, \rho) = H_{norm}(\theta, \rho) > 0,1$  ;
12 while maximum de  $H_p(\theta, \rho) > 0,1$  do
14   Calcul de la valeur maximale de  $H_p(\theta, \rho)$  et de son angle  $\theta_m$  ;
16   Récupération des superpixels (sup) formant la valeur maximale ;
18   Transformée de Hough ( $H_{sup}(\theta, \rho)$ ) sur les superpixels de l'étape précédente
   (sup) ;
20    $H_p(\theta, \rho) = H_p(\theta, \rho) - H_{sup}(\theta, \rho)$  ;
22   if  $\theta_m > \theta_{lines} - 20^\circ$  and  $\theta_m < \theta_{lines} + 20^\circ$  then
24      $\lfloor$  une rangée de culture est détectée ;

```

---

*Fully Convolutional Networks* en anglais). Un FCN [106] ne dispose que de couches de convolution et de pooling qui lui permettent de faire des prédictions sur des entrées de tailles arbitraires. Les hiérarchies des caractéristiques profondes (*deep features* en anglais) codent la localisation et la sémantique dans une pyramide non linéaire du local vers le global. L'un des problèmes avec le réseau FCN est qu'en faisant de la propagation à travers plusieurs couches de convolution et de pooling alternées, la résolution des images en sortie est sous-échantillonnée (contours de régions relativement flous). Différentes versions améliorées du FCN ont été proposées, à savoir le SegNet [12], le DeepLab [28], ... Bien que les études existantes se soient concentrées sur l'utilisation du SegNet et du FCN pour la segmentation des cultures, des adventices et de l'arrière-plan, aucune, à notre connaissance, ne les avait employés pour détecter et accentuer l'alignement des pixels de plantes situées à l'intérieur des rangées.

Dans cette section, l'approche proposée utilise des réseaux entièrement convolutionnels pour la détection des rangées de culture. Elle s'appuie sur deux réseaux différents, le S-SegNet et le HoughCNet. La structure globale de la méthode est présentée dans la Figure 4.8. En entrée, la méthode récupère une image RGB et à l'aide du réseau S-SegNet, des bandes correspondant aux rangées de culture sont extraites. Le réseau S-SegNet a donc pour but d'extraire des bandes avec des contours bien lisses et de réduire l'impact des erreurs de semis en raccordant les rangées discontinues. Cependant, pour renforcer sa robustesse face à la présence importante d'adventices dans l'inter-rang, le réseau S-SegNet est suivi par le réseau HoughCNet dont l'objectif est d'ajuster les rangées impactées par les adventices. En effet, le HoughCNet utilise la transformée de Hough et un réseau convolutionnel profond pour extraire les pixels qui forment l'alignement des rangées de culture dans l'image. Pour s'assurer que le HoughCNet n'a pas ajouté de bruit, une intersection est appliquée entre sa sortie et celle du S-Segnet. A l'issue de cette intersection, nous observons généralement qu'il y a des discontinuités ou que les bandes obtenues ne sont pas lisses. Le S-SegNet est alors à nouveau appliqué pour ajuster ces

## 4.2. CROWNET : APPROCHE DEEP LEARNING POUR LA DÉTECTION DES RANGÉES DE CULTURE

bandes. Selon la densité d'adventices, cette opération est appliquée une ou plusieurs fois. Une boucle dont la condition d'arrêt est basée sur le score du IoU (intersection over union) est appliquée. La boucle est arrêtée lorsque l'IoU entre le S-SegNet et la sortie ajustée (HoughCNet + Intersection + S-SegNet) est élevée ( $\text{IoU} > \varepsilon$ ).

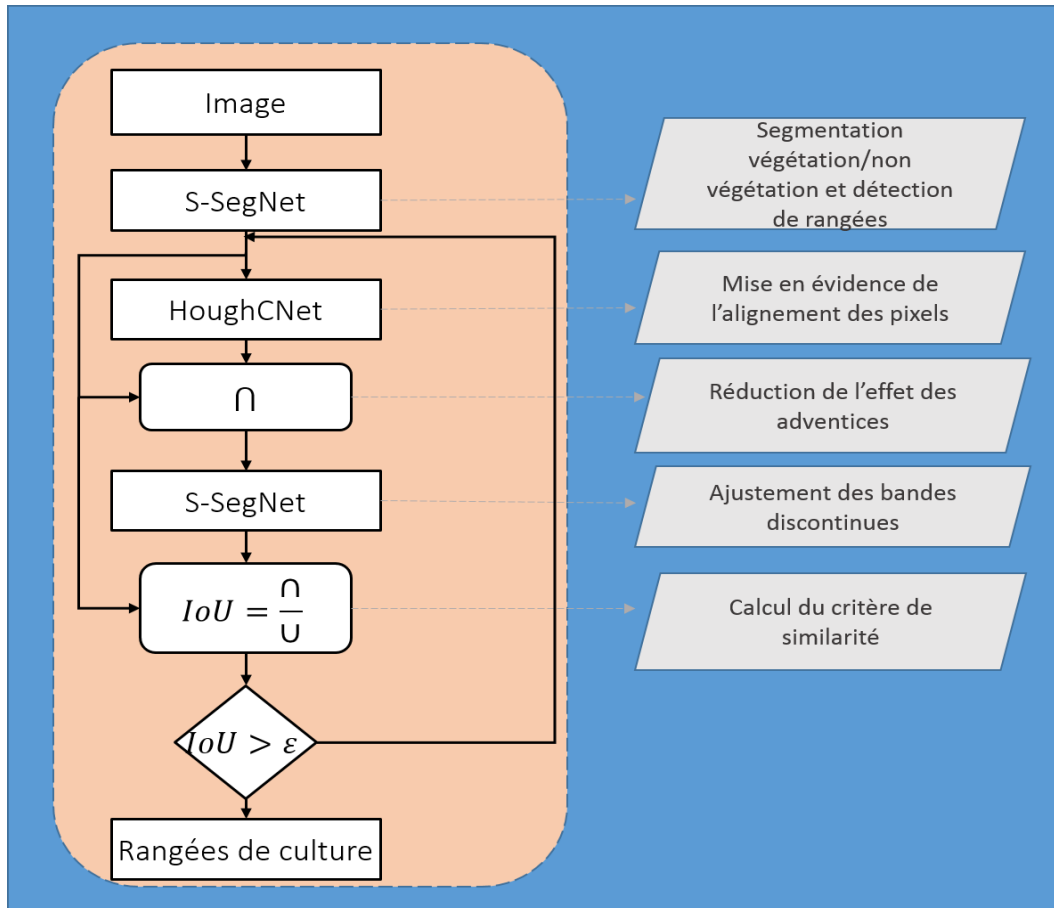


FIGURE 4.8 – Organigramme pour la détection des rangs de culture avec CNN (CRowNet).

### 4.2.1 S-SegNet

Dans l'état de l'art présenté dans le Chapitre 2, nous avons vu que récemment, diverses études ont été proposées pour segmenter la végétation (cultures et adventices) et le sol. Toutes ces études visent à segmenter les plantes individuellement afin d'obtenir le résultat qui correspond le mieux aux contours de chacune de ces plantes. Cette segmentation est ensuite utilisée dans d'autres applications, comme l'extraction des rangées de culture pour guider les robots ou la détection des mauvaises herbes inter-rang. Dans la présente étude, la combinaison de la segmentation de la végétation et de la détection des rangées de culture est proposée en une seule étape (Figure 4.9) en se basant sur l'architecture SegNet : c'est le S-SegNet.

SegNet [12] est un réseau de deep learning qui s'appuie sur le FCN pour réaliser une segmentation sémantique. A la base, il a été conçu pour des applications de compréhension de scène en temps réel. Il est donc conçu pour être à la fois efficace en termes de mémoire

et de temps de calcul lors de l'inférence (phase de segmentation). De plus, le nombre de paramètres à entraîner est très faible par rapport à un CNN classique. Son architecture est composée d'encodeurs, de décodeurs et d'une couche de classification par pixel. Sa particularité est que chaque encodeur a son décodeur correspondant. Les encodeurs sont généralement des couches de convolution comme celles présentes dans le réseau VGG16 [146]. Le décodeur consiste à projeter sémantiquement les caractéristiques discriminantes apprises par l'encodeur à une résolution inférieure vers une résolution supérieure : le décodage n'est rien d'autre que du sur-échantillonnage. SegNet tire son originalité de la façon dont le décodeur est appliqué pour passer d'une résolution inférieure à une résolution supérieure. Après qu'une opération de sous-échantillonnage ait été effectuée (max-pooling) à la sortie d'un encodeur, les indices correspondant au pooling sont conservés. En d'autres termes, l'emplacement de la valeur maximale dans chaque fenêtre de pooling est mémorisé pour chaque carte de caractéristiques de l'encodeur. Ainsi ces indices sont utilisés par le décodeur correspondant pour effectuer un sur-échantillonnage non linéaire. L'avantage de cette technique est qu'elle permet de réduire le nombre de paramètres puisqu'elle n'a pas besoin de faire de l'apprentissage pour sur-échantillonner.

Au total, l'architecture du SegNet est composée de 13 couches de convolution (encodeurs) qui correspondent aux 13 premières couches du VGG16. Chaque couche de convolution est suivie d'une normalisation par lot (*batch normalisation* en anglais) [80] et d'une fonction d'activation non linéaire, le ReLu (Rectified Linear Unit), suivie d'un max-pooling avec une fenêtre de taille  $2 \times 2$  et un stride 2 (pas de recouvrement). Cette dernière opération permet de sous-échantillonner la carte de caractéristiques en entrée par un facteur 2. Quant au max-pooling, il permet de renforcer l'invariance à la translation sur de petits décalages spatiaux de l'image en entrée. Cependant, même si l'utilisation du max-pooling est bénéfique pour la robustesse face à la translation, il cause une perte de la résolution spatiale, ce qui peut être moins bénéfique dans des applications où la délimitation des frontières est essentielle. Avec le réseau SegNet, la conservation des positions des valeurs max lors du max-pooling est un bon moyen de faire face à ce problème.

Comme mentionné plus haut sur la particularité du SegNet, chacune des 13 couches de convolution a un décodeur correspondant. Le décodeur sur-échantillonne l'encodeur à partir des indices mémorisés et conserve le nombre de bandes. Un encodeur avec 3 canaux (RGB) aura 3 canaux après encodage, ce qui est le cas pour le premier encodeur. Les résultats du sur-échantillonnage sont épars, c'est la raison pour laquelle une couche de convolution suivie d'une normalisation par lot est rajoutée pour produire une carte de caractéristiques plus dense.

Enfin la sortie du dernier décodeur est mise en entrée d'une couche softmax pour la classification. Le softmax classe chaque pixel indépendamment. En sortie du softmax, nous obtenons une image de probabilités avec  $K$  classes. L'image de segmentation prédite va correspondre à la classe qui a la plus grande probabilité pour chaque pixel.

### 4.2.2 HoughCNet

Au lieu d'utiliser des superpixels ou des informations de distance inter-rang pour identifier les votes représentant les rangs de culture, le HoughCNet combine la transformée de Hough et un réseau convolutionnel (Figure 4.10).

Dans un premier temps, la transformée de Hough normalisée  $H_{norm}(\rho, \theta)$  est calculée

## 4.2. CROWNET : APPROCHE DEEP LEARNING POUR LA DÉTECTION DES RANGÉES DE CULTURE

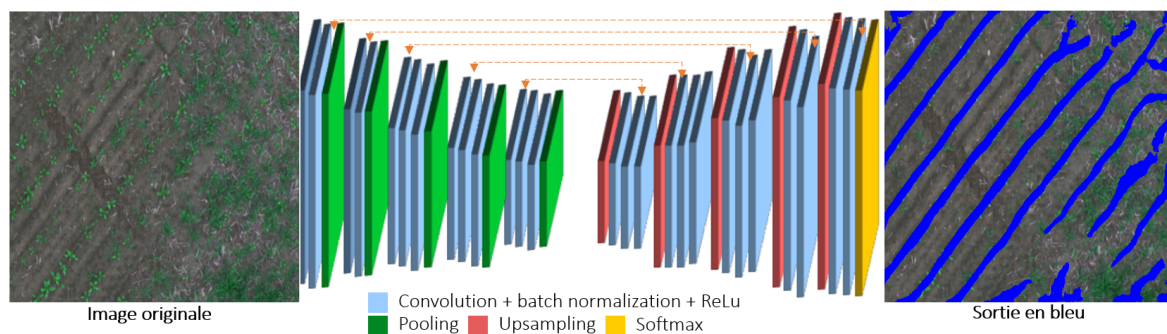


FIGURE 4.9 – Architecture du SegNet.

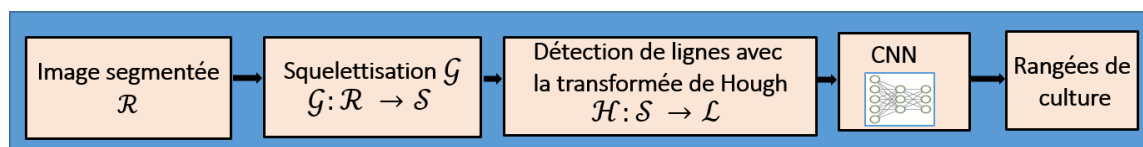


FIGURE 4.10 – Organigramme du processus HoughCNN. En entrée, nous avons une image segmentée et en sortie, nous obtenons des rangées de culture après la combinaison de la transformée de Hough et du CNN.

afin de donner le même poids à toutes les rangées de culture, en particulier les plus courtes, près des bords [59]. Toutes les lignes sont ensuite détectées pour obtenir une image  $\mathcal{L}$ .

Soit  $\Psi_i$  une image de la même taille que l'image d'entrée et  $(X_i, Y_i)$  le couple de coordonnées de tous les pixels de  $\Psi_i$  (Équation 4.6) appartenant à la ligne représentée dans  $H_{norm}(\rho, \theta)$  par le pic  $P_i$  de valeur  $\alpha_i$ .

$$\Psi_i(x, y) = \alpha_i, (x, y) \in (X_i, Y_i) \quad (4.6)$$

$$\mathcal{L} = \arg \max_{i \in [1, N]} (\Psi_i) \quad (4.7)$$

L'image  $\mathcal{L}$  contenant les lignes des  $N$  pics de  $H_{norm}(\rho, \theta)$  est obtenue par l'Équation 4.7. La Figure 4.11 présente un exemple de résultats. Une fois que  $\mathcal{L}$  est obtenue, elle est utilisée comme entrée du CNN pour détecter les rangées de culture. Tout le processus est présenté dans la Figure 4.10.

L'architecture de CNN proposée est composée de 5 couches de convolution à trou ou convolution dilatée. Une couche de convolution dilatée est définie comme une couche dans laquelle la convolution est réalisée avec des filtres dilatés (Figure 4.12). Chaque couche de convolution est composée de 32 filtres. La normalisation par lot et le ReLU sont appliqués après chaque couche de convolution. Les convolutions dilatées sont appliquées en raison de leur capacité à étendre le champ réceptif sans perte de résolution et à augmenter les paramètres [170, 66]. Le champ réceptif croît de façon exponentielle tandis que le nombre de paramètres croît linéairement.

## 4.2. CROWNET : APPROCHE DEEP LEARNING POUR LA DÉTECTION DES RANGÉES DE CULTURE

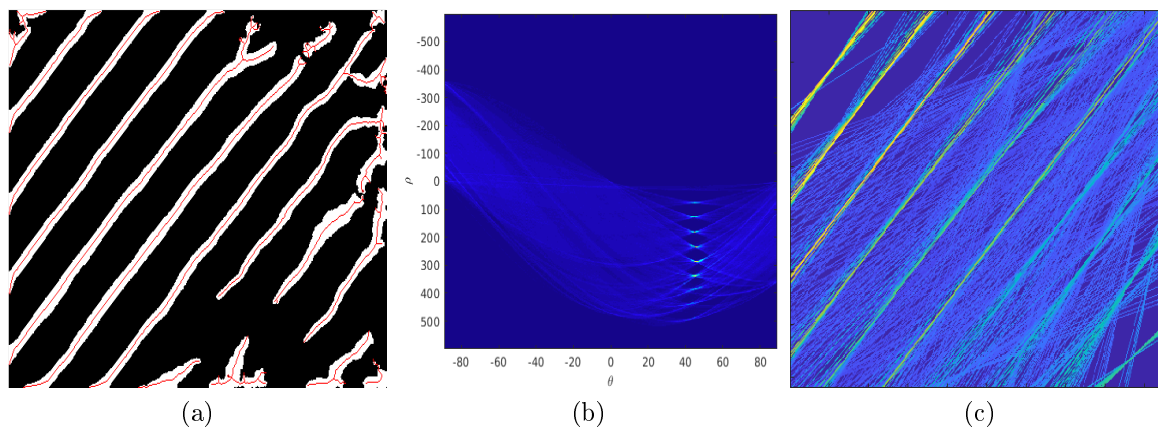


FIGURE 4.11 – (a) Image binaire en blanc et son squelette en rouge. (b) Transformée de Hough effectuée sur le squelette. (c) Lignes détectées avec la transformée de Hough où chaque ligne a une valeur correspondant au nombre de votes dans l'espace de Hough ( $\mathcal{L}$ ).

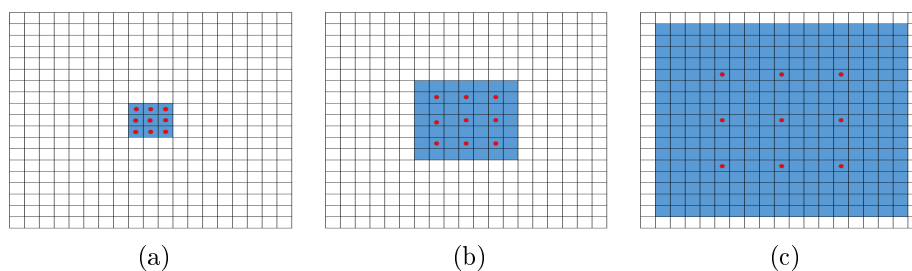


FIGURE 4.12 – Exemples de trois filtres dilatés. (a) Filtre  $f_1$  utilisé dans la convolution dilatée 1; chaque élément dans  $f_1$  a un champ réceptif de  $3 \times 3$ . (b) Filtre  $f_2$  produit à partir de  $f_1$  par convolution dilatée 2; champ récepteur pour chaque élément est  $7 \times 7$ . (c) Ce filtre  $f_3$  est obtenu de  $f_2$  par une convolution dilatée 4; éléments dans  $f_3$  ont un champ récepteur de  $15 \times 15$ . Les couches de convolution dilatées générées par ces filtres ont le même nombre de paramètres.

L'architecture du CNN appliquée sur la transformée de Hough est montrée dans la Figure 4.13. Généralement, dans  $\mathcal{L}$ , chaque rangée de culture est représentée par plus d'une ligne à cause de son épaisseur, en sachant que nous aimerions identifier celle qui passe le mieux par le centre. Ainsi, les deux premières couches du réseau utilisent 4 comme valeur de dilatation, ce qui permet de mettre en évidence la ligne qui est au centre et inversement de réduire l'influence des lignes éloignées du centre des rangées de culture. Ces couches sont suivies d'un sous-échantillonnage qui conserve la valeur maximale dans un voisinage de  $2 \times 2$  (pooling maximum) et d'un sur-échantillonnage qui remet l'image à sa taille originale. Le processus est ensuite répété avec 2 autres couches de convolution de valeur de dilatation 2. La dernière couche de convolution ( $1 \times 1 \times 2$ ) est appliquée, où 2 est le nombre de classes. Enfin le softmax est appliqué pour normaliser les valeurs de chaque pixel (Figure 4.13).

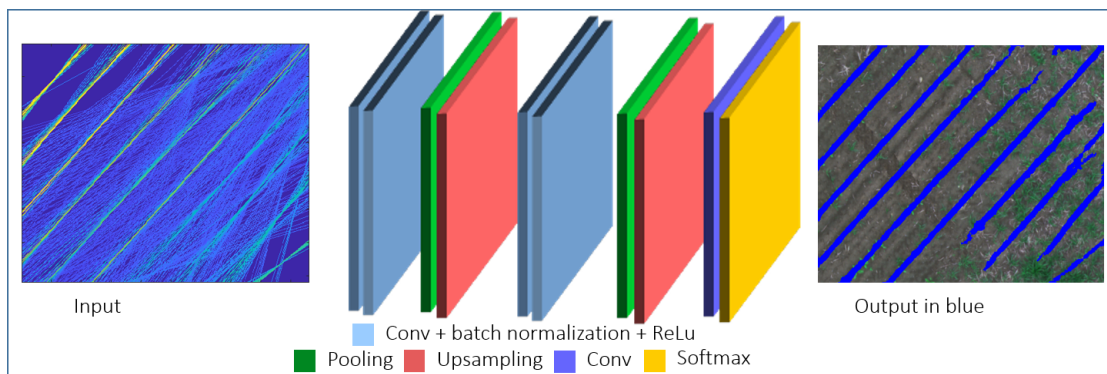


FIGURE 4.13 – Architecture du réseau convolutionnel utilisé pour le HoughCNet.

### 4.3 Détection des adventices

Une fois les rangées de culture détectées, les adventices inter-rang sont détectées en étudiant la relation spatiale de ces rangées avec tous les composants connexes (les régions) de l'image segmentée. Ainsi chaque région intersectant une rangée de culture est automatiquement considérée comme de la culture, de l'adventice dans le cas contraire. Cette approche ne permet cependant de ne détecter que les agrégations de plantes isolées hors des rangées de culture et ne permet donc pas de détecter les adventices qui poussent dans l'intra-rang.

A défaut de pouvoir détecter les adventices situées à l'intérieur des rangées de culture, nous proposons d'utiliser une méthode de détection des adventices *potentielles*, c'est-à-dire des régions de végétation qui sont en contact avec les rangées de culture détectées et se distinguent beaucoup plus dans l'inter-rang par rapport à leurs voisines.

#### Utilisation des relations spatiales entre superpixels

---

**Algorithme 2** : Détection d'adventices potentielles.

---

```

input :  $Sup, \theta$ 
output : Adventices potentielles
2 for tous les superpixels ( $Sup$ ) do
4    $MajAx$  = calculer le grand axe de  $Sup$  ;
6    $Sup_N$  = Identifier les voisins de  $Sup$ ;
7   for tous les voisins de  $Sup$  ( $Sup_N$ ) do
9      $mergeSup$  = fusionner  $Sup$  et  $Sup_N$ ;
11     $mergeMajAx$  = calculer le grand axe de  $mergeSup$ ;
13     $mergeOrientation$  = orientation de  $mergeSup$ ;
15    if  $mergeMajAx > MajAx$  et  $mergeOrientation < \theta - 40^\circ$  then
17       $Sup$  et  $Sup_N$  sont des adventices potentielles;

```

---

Nous partons de l'hypothèse que la présence d'une adventice à une position  $p$  à l'intérieur d'une rangée de culture va entraîner une perturbation de l'alignement des plantes

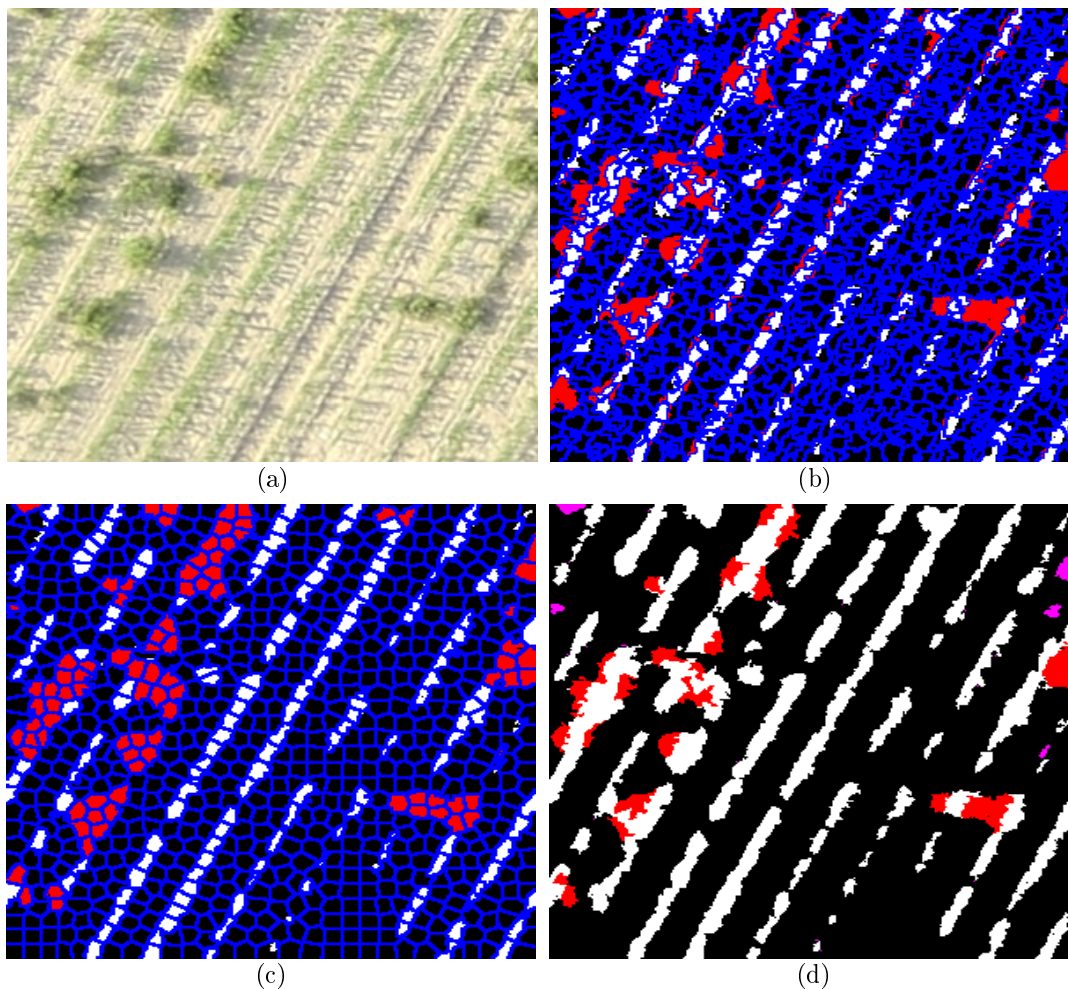


FIGURE 4.14 – (a) Image originale. (b) Masque de rangées de culture en blanc et en rouge les superpixels qui n’intersecte pas les rangées détectées. (c) Résultat de la détection en exploitant la relation entre superpixels et (d) Résultat de la détection d’adventices après avoir utilisé le masque de rangée de culture (en rouge les adventices potentielles et en magenta les adventices inter-rangs).

à cette position. Ceci implique que l’angle formé par la fusion de deux superpixels dont l’un est de la culture et l’autre une adventice sera différente de l’orientation globale de la rangée d’une valeur  $d$ . En testant différentes valeurs de  $d$  allant de  $10^\circ$  à  $90^\circ$ , la valeur de  $d = 40^\circ$  a montré de meilleurs résultats. Afin de nous concentrer uniquement sur les fusions impliquant des superpixels alignés, nous avons également comparé le grand axe des 2 superpixels avant fusion et celui après fusion. L’algorithme 2 explique la méthode.

Dans le but de réduire la surdétection d’adventices, à la sortie de l’algorithme 2 les adventices détectées sont croisées avec le masque des rangées détectées. Ainsi tous les pixels n’appartenant pas à ce masque sont considérés comme des adventices : ce sont les adventices potentielles.

## 4.4 Expérience et résultats

Les méthodes développées dans ce chapitre ont été évaluées dans deux champs différents : betterave et maïs. De plus, les données ont été acquises avec des altitudes de vol et des capteurs d'images différents. Ces deux types de données aux caractéristiques différentes ont été choisis dans le but de comparer la robustesse des méthodes de détection des rangées proposées. Les données sont présentées dans le Chapitre 3.

Les images de betterave acquises grâce à un capteur RedEdge-M sont subdivisées en 5 sous-ensembles d'images numérotés de 000 à 004. Chaque sous-ensemble correspond à une partie du champ. Pour l'apprentissage des réseaux dans la méthode CRowNet, seules les images du sous-ensemble 003 ont été utilisées et l'évaluation a été effectuée sur les sous-ensembles d'images 000-002. Les images du sous-ensemble 004 n'ont pas été utilisées à cause du flou présent sur certaines parties des images.

Pour évaluer les méthodes dans le champ de betterave, 154 images de taille  $480 \times 360$  pixels ont été sélectionnées à partir des jeux de données 000, 001 et 002. Afin de tenir compte des irrégularités des rangs de culture et d'éviter l'impact de l'infestation de mauvaises herbes dans les données d'apprentissage, la vérité terrain a été construite manuellement avec l'outil d'étiquetage de Matlab (Image labeler). L'étiquetage consistait à créer un masque autour de chaque rangée. Le taux d'infestation en mauvaises herbes (WIR ou *Weed Infestation Rate* en anglais) sur ces images varie de faible à élevé. Le WIR (Equation 4.8) est la proportion de pixels de végétation considérés comme des mauvaises herbes dans une image. Selon la classification WIR, nous avons 45 images sans adventices (WIR < 5%), 21 images avec une faible infestation ([5% - 15%]), 50 images avec un taux d'infestation modéré ([15% - 35%]) et 38 images fortement infestées (WIR > 35%). Ces taux d'infestation ont été calculés à partir de la vérité terrain des images disponibles dans la base de données publique. Au total, 1325 rangées de culture sont à détecter.

Dans le champ de maïs, 80 échantillons ont été sélectionnés ; la taille de l'échantillon est de  $300 \times 300$  pixels. Le nombre total de rangées de culture est de  $N_{CR} = 1081$ . Sur la Figure 4.18a, nous pouvons voir que l'image est affectée par les ombres de la végétation et que certaines rangées de culture sont difficilement localisables, même à l'œil nu.

$$WIR(\%) = \frac{Adventice(pixels) \times 100}{Adventice(pixels) + Culture(pixels)} \quad (4.8)$$

### 4.4.1 Métriques d'évaluation

Quatre indices ont été calculés : le rappel (*Rappel*), la précision (*Précision*), le *F-score* et l'intersection sur l'union (*IoU*). Le *Rappel* reflète la capacité à détecter les vraies rangées (Équation 4.9), la *Précision* (Équation 4.10) indique l'exactitude des rangées détectées et le *F-score* indique l'équilibre entre *Précision* et *Rappel* (équation 4.11). Ces indices ont été calculés pour chaque rangée de culture. De plus, pour évaluer l'efficacité de la détection dans chaque image, l'IoU a été calculée (Équation 4.12). L'IoU, aussi appelée coefficient de similarité de Jaccard, est le rapport entre le nombre de pixels des rangées de culture correctement détectés (l'intersection) et le nombre total de pixels des rangées de culture de la vérité terrain (GT) et ceux détectés (P) (l'union). L'IoU prend en compte à la fois les fausses alarmes et les pixels non détectés pour chaque classe.

$$Rappel = \frac{VP}{VP + VN} \quad (4.9)$$

$$Précision = \frac{VP}{VP + FP} \quad (4.10)$$

$$F - score = \frac{2 \times Rappel \times Précision}{Précision + Rappel} \quad (4.11)$$

$$IoU = \frac{GT \cap P}{GT \cup P} \quad (4.12)$$

Les vrais positifs (VP) sont des pixels qui sont correctement étiquetés comme une classe positive (rangées de culture), alors que les faux positifs (FP) sont des pixels incorrectement étiquetés comme une classe positive. Enfin, les vrais négatifs (VN) sont des pixels correctement étiquetés comme une classe négative (arrière-plan et/ou végétation ne se trouvant pas dans les rangées de culture), et les faux négatifs (FN) sont des pixels incorrectement étiquetés comme une classe négative.

Pour évaluer efficacement les résultats, les rangées ont été squelettisées. En supposant que l'opération de squelettisation retourne des squelettes qui représentent l'alignement central de chaque rang, ces squelettes ont été dilatés pour obtenir la même largeur que celle des rangées de culture mesurée sur la parcelle. En effet, sur la base des mesures effectuées au sol, dans le champ de betterave, la largeur d'une plante de culture varie de 15 à 20 pixels. Dans le champ de maïs, la largeur des rangs de culture est d'environ 12 pixels. Nous avons donc dilaté le squelette des rangées détectées avec un élément structurant carré de  $15 \times 15$  pixels sur les images de betterave et  $11 \times 11$  pixels sur les images de maïs.

#### 4.4.2 Evaluation du CR-Hough-SLIC

Comme nous pouvons le constater dans l'algorithme 1 de la section 4.1.4, cette méthode prend en entrée le squelette et une image segmentée en superpixels. Pour la squelettisation, nous avons utilisé l'amincissement itératif proposé par la fonction *thin* de Matlab<sup>®</sup>. Puisque la génération de superpixel par la méthode du SLIC demande le contrôle de la compacité  $C$  et du nombre de superpixels  $k$ , une étude a été effectuée sur l'impact de ces paramètres lors de la détection des rangées de culture. La fonction *superpixels* de Matlab<sup>®</sup> a été utilisée pour la génération des superpixels. Cette fonction prend en entrée une image RGB, le nombre de superpixels  $k$  et la compacité  $C$ . La compacité varie de 1 à 20 ; elle permet de contrôler la forme des superpixels. Une valeur de compacité très grande rend la forme des superpixels plus régulière, proche d'un carré ; une valeur faible permet aux superpixels de mieux adhérer aux contours des objets de forme irrégulière. Étant donné que le SLIC sera appliqué sur les images segmentées, le paramètre de compacité aura donc peu d'effet ; la taille des superpixels sera ainsi le paramètre déterminant.

Sur les Figures 4.15 et 4.16, des exemples de résultats sont présentés pour une compacité constante ( $C=5$ ) et des tailles de superpixels variables. On peut voir que la détection est mieux réussie sur les Figures 4.15c et 4.15d pour le champ de maïs et les Figures 4.16c et 4.16d du champ de betterave. Sur les Figures 4.15e, 4.15f, 4.16e et 4.16f, la surdétec-

tion est bien visible au niveau des rangées de culture, surtout celles qui ont le plus de végétation. De plus, la présence d'adventices inter-rang accentue cette surdétection. Par exemple, dans le champ de maïs, une taille de superpixel de taille  $32 \times 32$  pixels correspond à une surface au sol d'environ  $80 \times 80$  cm, ce qui est supérieur à la largeur des rangées de culture. Les meilleurs résultats sont obtenus pour des valeurs de  $S$  égal à 15 dans le champ de maïs et 32 dans celui de betterave. Nous constatons que plus  $S$  s'approche de la largeur des rangées de culture (qui est d'environ 12 pixels pour le champ de maïs et entre 15-20 pixels pour celui de betterave) meilleur est le résultat. De plus, nous pouvons remarquer que la qualité du résultat et la taille des superpixels sont liées puisque plus la résolution est importante, plus la taille des superpixels augmente. Ceci est dû au fait qu'une bonne résolution spatiale engendre plus d'hétérogénéité dans l'image parce que beaucoup de détails sont visibles aussi bien au niveau des plantes qu'au niveau du sol.

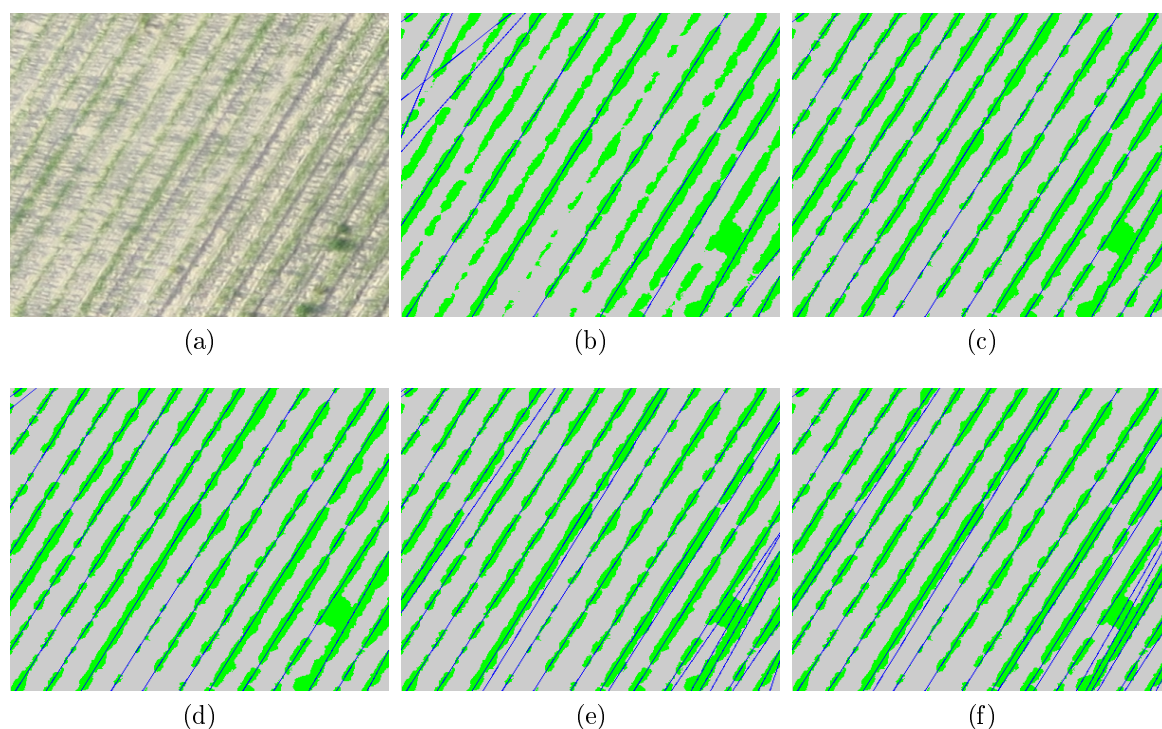


FIGURE 4.15 – Exemple de détection de rangées de culture dans le champ de maïs avec une compacité de 5 et une taille de superpixel variable. (a) correspond à l'image originale. Les 5 autres images (b, c, d, e et f) correspondent respectivement au résultat de la détection des rangées de culture avec  $S = 32, 15, 10, 5, 4$  ( $k = 0,1\%N, 0,5\%N, 1\%N, 5\%N, 10\%N$ ). Pour rappel,  $N$  correspond au nombre de pixels total de l'image.

La Table 4.1 présente le résultat moyen obtenu avec  $k = 0.5\%N$  et  $0.1\%N$  respectivement dans les champ de maïs et de betterave avec les différentes métriques. Nous constatons que en termes de *Précision* et d'IoU les résultats de la méthode dans les deux champs sont comparables avec un écart de moins de 1.5 %. Toutefois le *Rappel* obtenu dans le champ de betterave, est de 6 % supérieur à celui dans le champ de maïs et que l'IoU moyenne par image dans le champ de maïs est de 8 % plus grande que celle dans le champ de betterave. L'écart d'efficacité entre les deux dernières métriques signifient que dans le champ de betterave plus de rangées de culture sont détectées par rapport au

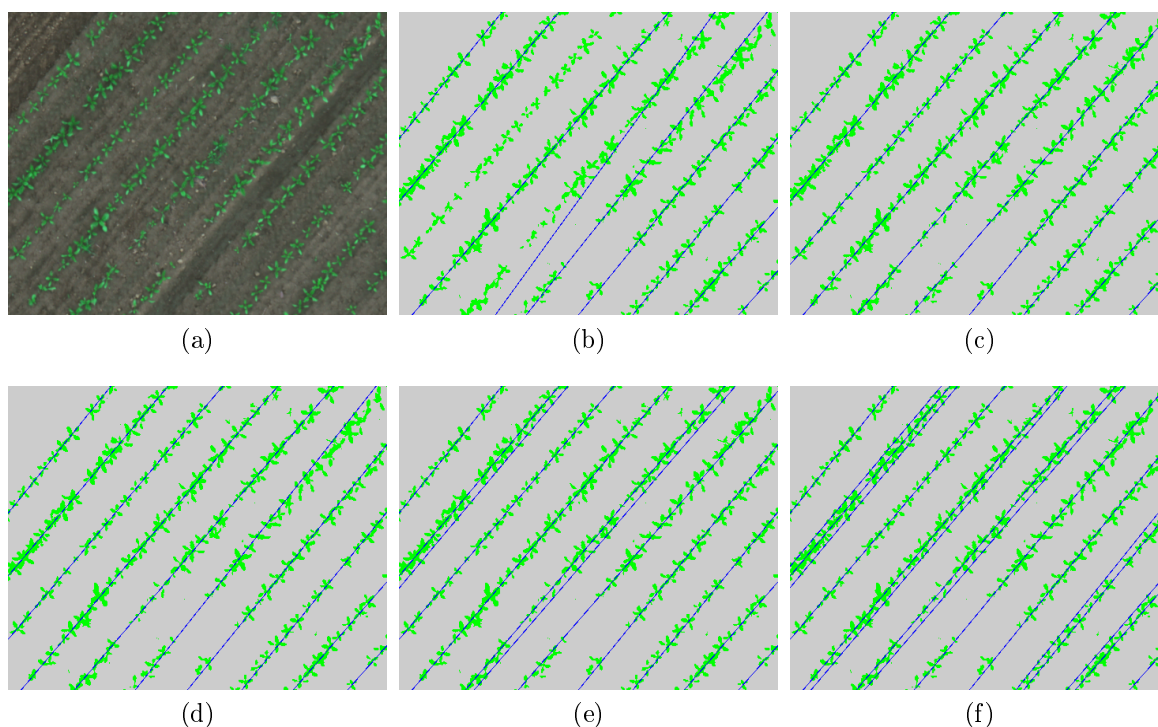


FIGURE 4.16 – Exemple de détection de rangées de culture dans le champ de betterave avec une compacité de 5 et une taille de superpixel variable. (a) correspond à l'image originale. Les 5 autres images (b, c, d, e et f) correspondent respectivement au résultat de la détection des rangées de culture avec  $S = 32, 15, 10, 5, 4$  ( $k = 0,1\%N, 0,5\%N, 1\%N, 5\%N, 10\%N$ ). Pour rappel,  $N$  correspond au nombre de pixels total de l'image.

champ de maïs maïs avec une surdétection plus élevée.

Cette différence de performance peut s'expliquer par le fait que certains échantillons provenant du champ de betterave contiennent, en plus des rangées de culture, des zones entièrement couvertes d'adventices, ce qui perturbe fortement la méthode. Pour améliorer ce résultat, il faudrait trouver un bon compromis entre la taille des rangées de culture à retenir et celles à éliminer. Toutes les plantes ne poussent pas de la même façon dans les différentes rangées, ce qui peut causer un problème si un seuil de longueur trop grand est choisi. Dans l'autre sens, choisir un seuil trop petit reviendrait à augmenter la surdétection causée par les adventices inter-rang.

TABLE 4.1 – Résultat moyen obtenu avec  $k = 0,5\%N$  et  $0,1\%N$  respectivement dans les champ de maïs et de betterave avec les différentes métriques.  $N$  correspond au nombre de pixels total de l'image.

Données	k	Rappel	Précision	F-score	IoU	IoU/image
Maïs	0,5%N	0,7974	0,8514	0,8213	0,7516	0,8290
Betterave	0,1%N	0,8567	0,8658	0,8540	0,7660	0,7492

### 4.4.3 Évaluation du CRowNet

L'annotation manuelle d'une image prend énormément de temps car elle nécessite non seulement la possibilité de tracer une ligne continue par rangée de culture, mais aussi de répéter le même processus pour chaque image. Afin de simplifier le processus d'annotation, nous avons procédé de manière semi-supervisée. L'image orthomosaïquée 003 a été divisée en patches de  $250 \times 250$  pixels. Cette taille a été choisie pour obtenir des images montrant des rangées de culture avec des dispositions différentes (discontinuité, contours irréguliers, mauvaises herbes entre les rangs, ...) et aussi pour faciliter l'apprentissage CNN. Nous avons ensuite appliqué la méthode du CR-Hough-SLIC (section 4.1) pour détecter les rangées de culture dans tous les échantillons. Les échantillons dans lesquels toutes les rangées de culture sont correctement détectées ont été sélectionnés pour construire la base de données d'apprentissage. Le nombre d'échantillons d'apprentissage recueillis est de 134. En appliquant une augmentation de données grâce à trois rotations ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  et  $135^\circ$ ) et trois changements de contraste, la taille de la base d'apprentissage a été multipliée par sept. La Figure 4.17 présente un exemple de la base de données utilisée pour l'apprentissage. Le S-SegNet a été entraîné avec toutes les données d'apprentissage, tandis que le réseau HoughCNet n'a été entraîné qu'avec seulement les images originales et les données augmentées par rotation puisque les échantillons avec le changement de contraste ont la même orientation que les images originales.

TABLE 4.2 – Table de pondération de chaque classe et par type de réseau.

Architecture	Classe	Nombre de pixels	Fréquence	Poids
S-SegNet	Rangée	7441648	0,1269	7,8780
	Arrière-plan	51183352	0,8731	1,1454
HoughCNet	Rangée	2214900	0,0661	15,1248
	Arrière-plan	31285100	0,9339	1,0708

La vérité terrain montre que les classes sont déséquilibrées : la majorité des pixels appartiennent à l'arrière-plan et cela peut affecter le processus d'apprentissage. A cause de ce déséquilibre de classe, le modèle entraîné peut devenir plus sensible à l'identification de la classe majoritaire par rapport à la classe minoritaire. Ce problème de déséquilibre de classe est résolu en pondérant à nouveau chaque classe par l'inverse de sa fréquence lors du calcul de la fonction de perte. Ce processus a l'avantage d'augmenter le poids des classes sous-représentées et donc d'influencer la fonction de perte en attribuant un poids relativement plus élevé aux classes mineures (Table 4.2). La différence de poids entre le S-SegNet et le HoughCNet s'explique par le fait que le nombre d'échantillons utilisés n'est pas le même et que les images annotées employées pour entraîner le S-SegNet sont des versions dilatées de celles utilisées pour le HoughCNet.

L'algorithme de descente de gradient stochastique avec un terme d'inertie (momentum) ou (SGDM : *Stochastic Gradient Descent with Momentum* en anglais) (Annexe A) est appliqué comme procédé d'optimisation pour entraîner les réseaux. Le taux d'apprentissage est fixé à 0.001 et la valeur du momentum à 0.9. La fonction perte a été calculée en utilisant l'entropie croisée pondérée. Pour le S-SegNet, le nombre d'epochs et le mini-lot (mini-batch) sont réglés sur 1000 et 4, respectivement. En ce qui concerne le HoughCNet, le nombre d'epochs et la taille du mini-batch sont fixés respectivement à 200 et 20.

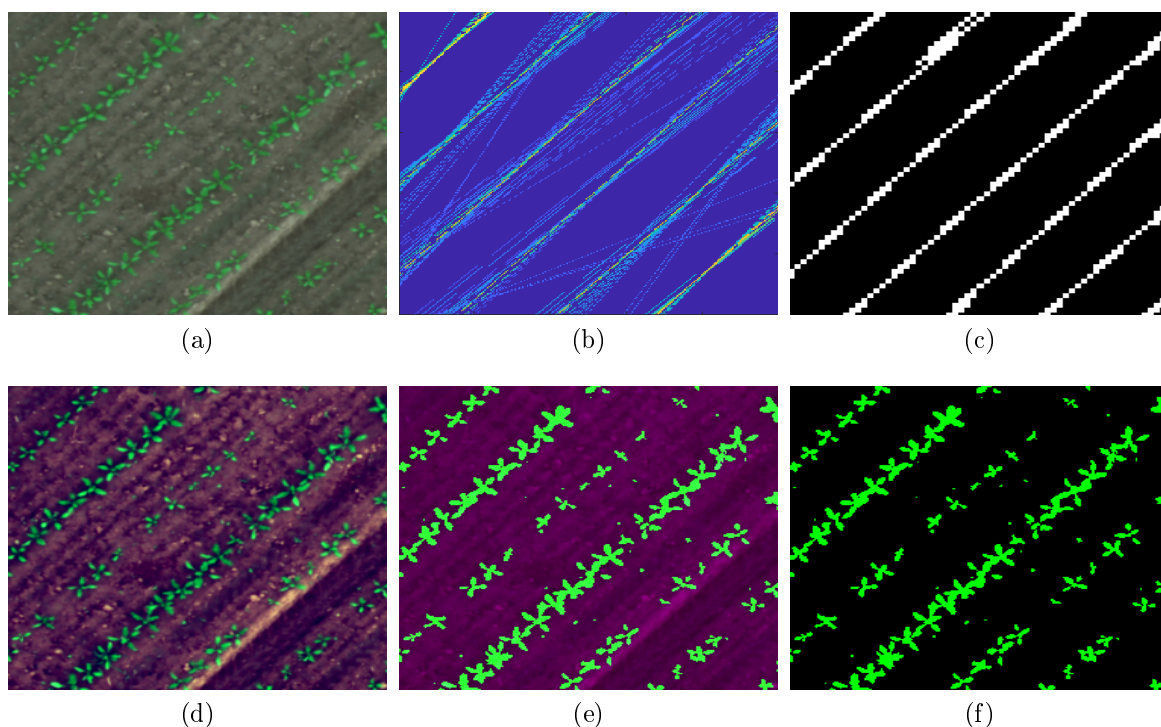


FIGURE 4.17 – Exemple de données d’apprentissage. (a) L’image originale  $X$ . (b) Rangées détectées avec la transformée de Hough  $\mathcal{L}$ . (c) Vérité terrain. Dans la deuxième ligne de la figure, nous avons l’image  $X$  après le changement de contraste. (d) Réglage du contraste : les 1 % de pixels sont saturés à basse et haute intensité pour chaque canal de  $X$ . (e) Le canal vert de l’image RGB est remplacé par le résultat de la segmentation de l’arrière-plan effectuée avec l’indice ExG et le seuillage de Otsu (BW). (f) Les canaux bleu et rouge sont réglés sur 0 et le canal vert correspond à BW.

La transformée de Hough  $H(\theta, \rho)$  est calculée sur  $\mathcal{S}$  avec une résolution  $\theta$  égale à  $0.1^\circ$ , laissant  $\theta$  prendre des valeurs dans la plage  $] -90^\circ; 89^\circ]$  et une résolution  $\rho$  égale à 1.  $\mathcal{S}$  est obtenu en utilisant la fonction Matlab *thin* avec  $n = Inf$ ; avec cette option, les pixels du contour sont supprimés selon la configuration locale jusqu’à ce que l’image ne change plus. La valeur du critère d’arrêt  $\epsilon$  choisi pour le CRowNet est 0,85. Cette valeur de  $\epsilon$  a été considérée comme un bon compromis entre une bonne détection et le nombre d’itérations.

#### 4.4.4 Évaluation qualitative

Sur la base du résultat visuel, une analyse qualitative a été effectuée. Dans les Figures 4.18 et 4.19, nous pouvons voir que le S-SegNet est capable de localiser et de représenter des rangées de culture avec des bandes même pour des rangées discontinues. Nous notons également qu’avec S-SegNet, les rangées de culture avec un faible taux d’infestation en adventices dans le voisinage sont bien détectées, tandis que celles qui sont discontinues et avec des adventices à proximité immédiate sont détectées mais avec des perturbations. En utilisant le CRowNet, la détection a été améliorée et l’impact des adventices a été réduit. L’autre remarque qui peut être traitée est que les bandes détectées sont principalement

situées au centre des rangées de culture et ne couvrent pas tous les pixels de la vérité terrain. Ceci est normal puisque les modèles sont entraînés pour détecter l’alignement central. De plus, la Figure 4.18 montre que même si toutes les plantes ont la même valeur, la méthode réussit à détecter les rangées. Ceci s’explique en partie par le fait que le modèle a appris à reconnaître un alignement lors de son apprentissage.

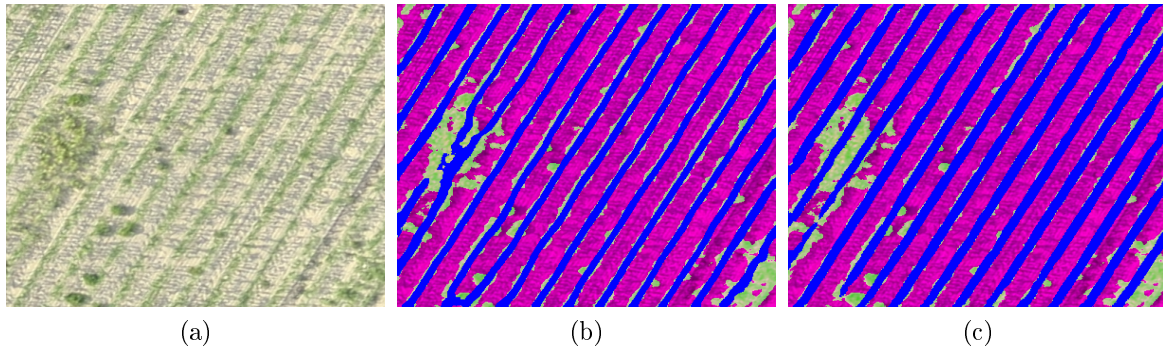


FIGURE 4.18 – Exemples de détection de rangées de culture sur des images prises dans un champ de maïs. (a) Image originale. (b) Superposition des rangées de culture détectées avec S-SegNet (en bleu) et l’image originale après changement de contraste. (c) Superposition des rangées de culture détectées (en bleu) avec CRowNet et l’image originale après changement de contraste.

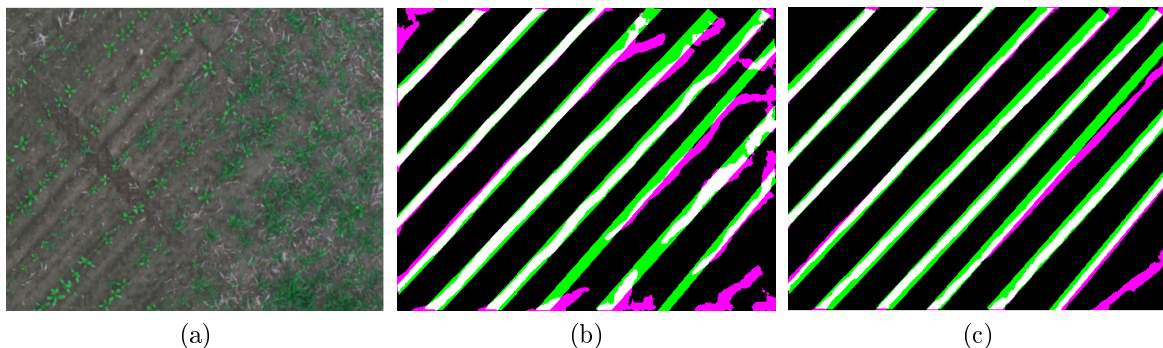


FIGURE 4.19 – Détection des rangées de culture à l’aide de S-SegNet et de CRowNet. (a) Image originale. (b) et (c) sont les résultats de S-SegNet et CRowNet respectivement. La couleur blanche correspond aux pixels de la vérité terrain qui ont été bien détectés, le magenta aux pixels qui sont sur la rangée prédite et n’appartenant pas à la vérité terrain (surdétection), et le vert les pixels non détectés.

#### 4.4.5 Évaluation quantitative

La Table 4.3 montre que les résultats de S-SegNet et de CRowNet sont comparables. Cependant le CRowNet est plus robuste lorsqu’il s’agit d’images avec un WIR élevé. Avec le CRowNet, chaque rangée est détectée en moyenne avec un *Rappel* de 70,72 % et une *Précision* de 90,10 %. L’évaluation qualitative effectuée est bien reflétée dans l’évaluation

#### 4.4. EXPÉRIENCE ET RÉSULTATS

quantitative. Les résultats s’expliquent par la position des rangs détectés en fonction de la vérité terrain.

TABLE 4.3 – Résultats quantitatifs de la détection des rangs de culture avant leur élargissement dans le champ de betterave. La moyenne est calculée pour chaque rang.

Méthode	Rappel moyen	Précision moyenne	$F - score$ moyen	IoU moyenne
S-SegNet	0,6614	0,8981	0,7472	0,60081
CRowNet	0,7072	0,9010	0,7794	0,6486

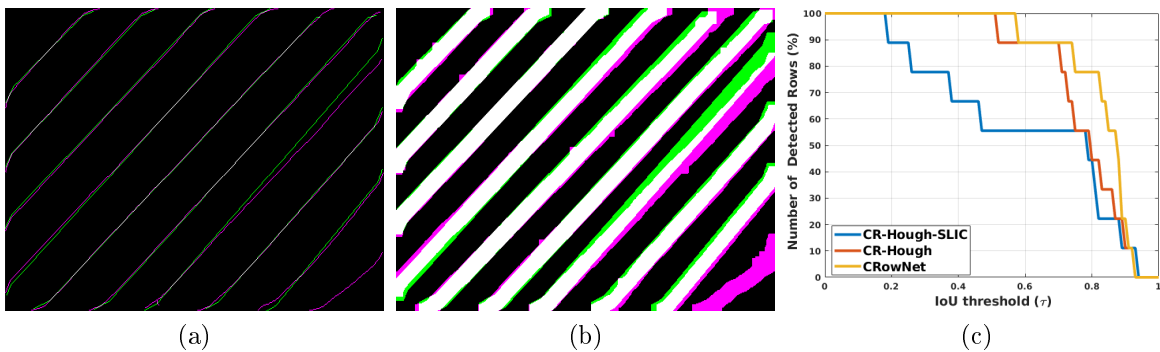


FIGURE 4.20 – (a) et (b) sont respectivement les résultats de CRowNet après squelettisation puis dilatation. La couleur blanche correspond aux vrais positifs (VPs), le magenta aux faux positifs (FPs), et le vert aux faux négatifs (FNs). (c) Nombre de rangées de culture détectées en fonction de la qualité de la segmentation. CR-Hough est le résultat obtenu avec la transformée de Hough et les informations de champ [85, 59]. CR-Hough-SLIC est le résultat obtenu avec la combinaison de la transformée de Hough et SLIC [13].

Les Figures 4.20a et 4.20b présentent les résultats après l’amincissement et ceux après la dilatation. Le CRowNet a été comparé à différentes méthodes de segmentation sémantique dont le SegNet avec des classes déséquilibrées, FCN [106], FCN-W et FCN-W-16. FCN-W et FCN-W-16 signifient respectivement FCN avec un *stride* 8 (FCN-8) et 16 (FCN-16) avec des classes rééquilibrées. Un *stride* correspond à la quantité de mouvement qu’un filtre de convolution doit effectuer pour passer d’une position à une autre.

Outre les méthodes de segmentation sémantique, le CRowNet a été comparé à CR-Hough-SLIC (section 4.1) et à l’utilisation traditionnelle de la transformée de Hough pour détecter les rangées de culture. En effet, les méthodes les plus couramment utilisées pour détecter les rangées de culture dans les images aériennes utilisent la transformée de Hough et les informations géométriques du champ (distance entre les rangées, orientation globale, etc.). Selon plusieurs travaux de recherches, la transformée de Hough est appliquée à tous les pixels de la végétation [85, 59], ou aux centres des plantes [84], mais le principe fondamental reste le même : on calcule d’abord la transformée de Hough puis on identifie le pic du vote majoritaire, à partir de la position de ce pic et de la distance estimée entre les rangs, d’autres pics sont identifiés. Cette méthode suppose que dans les images sans perspectives, les rangées de culture sont parallèles avec une distance constante entre les rangées. Dans la suite de cette section, cette approche sera appelée CR-Hough. Pour le

TABLE 4.4 – Moyenne des résultats des méthodes de détection des rangées de culture obtenues après avoir ramenée les rangées de culture à la même largeur. FCN-W et FCN-W-16 signifient respectivement des FCN avec un stride 8 (FCN-8) et 16 (FCN-16) avec des classes rééquilibrées.

Données	Méthode	Rappel	Précision	F-score	IoU	IoU/image
Betterave	SegNet	0,6740	0,8024	0,7179	0,6350	0,7650
	FCN	0,2625	0,4969	0,3215	0,2135	0,2839
	FCN-W	0,6512	0,6442	0,6318	0,4924	0,4984
	FCN-W-16	0,6544	0,6635	0,6442	0,5063	0,5111
	<b>CRowNet</b>	<b>0,9056</b>	0,9037	<b>0,9039</b>	<b>0,8319</b>	<b>0,7896</b>
	CR-Hough-SLIC	0,8567	0,8658	0,8540	0,7660	0,7492
	CR-Hough	0,8687	<b>0,9094</b>	0,8876	0,8027	0,7299
Maïs	<b>CRowNet</b>	<b>0,8093</b>	0,8457	<b>0,8250</b>	<b>0,7572</b>	0,8280
	CR-Hough-SLIC	0,7974	<b>0,8514</b>	0,8213	0,7516	<b>0,8290</b>
	CR-Hough	0,7907	0,8438	0,8149	0,7407	0,8154

calcul de CR-Hough, les valeurs théoriques utilisées pour les images du champ de betterave est de  $45^\circ$  pour l'orientation globale ; la distance entre les rangs est estimée à 50 pixels et la largeur de rang à 20. Sur les images du champ de maïs, l'orientation globale, la distance entre les rangs et la largeur des rangs sont respectivement de  $29^\circ$ ,  $22^\circ$  et  $12^\circ$ . Le nombre de rangées de culture dans un champ peut parfois être considéré comme un indicateur de bonne croissance des plantes et une prémonition d'un bon rendement. Par conséquent, une courbe du taux de bonne détection (GDR) des rangées de culture selon la qualité de la segmentation requise pour l'IoU ( $\tau$ ) est tracée (Equation 4.13).  $\tau$  est une valeur de seuil entre 0 et 1.

$$GDR = \frac{Card(IoU \geq \tau)}{N_{CR}} \quad (4.13)$$

Les Figures 4.20c et 4.21 montrent la performance des méthodes selon l'IoU obtenue pour chaque rangée de culture. Nous notons qu'en raison de l'irrégularité de la forme des rangs de culture, il est difficile d'obtenir un recouvrement de 100 % entre la vérité terrain et la détection. De plus, nous pouvons observer, dans la Figure 4.21a, que pour une valeur de  $\tau = 0.70$ , le nombre de lignes bien détectées avec CRowNet est de 93.58%. Dans le champ de maïs, la qualité de la segmentation présente des résultats proche et ceci est mis en évidence dans la Figure 4.21b et la Table 4.4. Nous pouvons également voir que la méthode CR-Hough-SLIC surpasse le CRowNet pour les valeurs élevées du seuil de l'IoU.

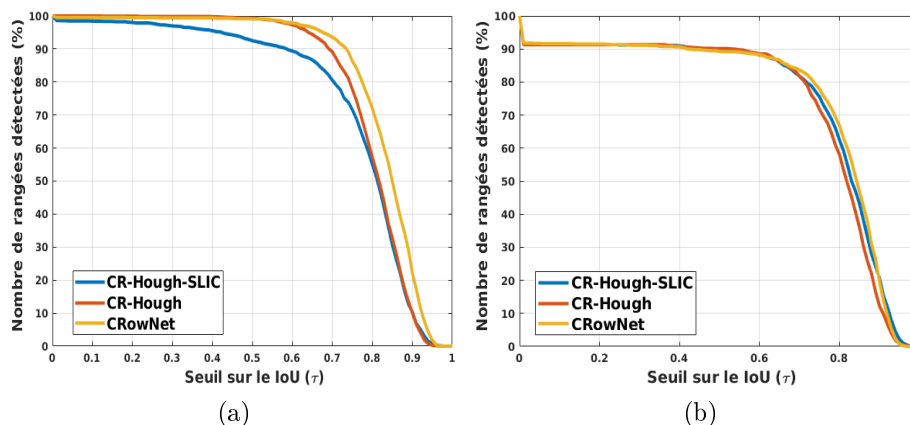


FIGURE 4.21 – Rangées de culture bien détectées en fonction de la qualité de la segmentation. (a) et (b) présentent les résultats obtenus à partir des données recueillies dans les champs de betterave et de maïs.

TABLE 4.5 – IoU moyenne en fonction du taux d’infestation par les adventices.

Méthode	Taux d’infestation		
	Faible	Modéré	Élevé
CRowNet	<b>0,8449</b>	<b>0,8026</b>	<b>0,7239</b>
CR-Hough-SLIC	0,7986	0,7971	0,6387
CR-Hough	0,7903	0,7474	0,6513

#### 4.4.6 Détection des adventices

Pour évaluer la méthode, les 80 échantillons du champ de maïs ont été utilisés. La labellisation a été effectuée manuellement en s’appuyant sur la vérité terrain réalisée lors de l’évaluation des méthodes de détection des rangées de culture. Toutes les agrégations de pixels en contact avec les rangées et qui en ressortent beaucoup trop dans l’inter-rang sont labellisées comme adventices (Figure 4.22).

Les adventices inter-rang ont été détectées en utilisant les composantes connexes (4 connexités). De plus, pour la détection des adventices potentielles, le nombre de superpixels générés est égal à environ 1% du nombre total de pixels dans l’image et la compacité a été fixée à 10. En effectuant une analyse qualitative sur la Figure 4.22, nous constatons que la qualité du résultat dépend fortement de la qualité de la détection des rangées de culture. Les images dans lesquelles les rangées de culture ont été bien détectées, les adventices inter-rang ont toutes été détectées. Le grand défi des méthodes exploitant l’information spatiale reste la détection des adventices intra-rang et celles juxtaposées aux rangées de culture. Avec les superpixels, nous constatons qu’il est possible de détecter les adventices inter-rangs et potentielles mais avec une certaine surdétection. Cette surdétection est généralement causée par des superpixels de cultures qui ne sont pas en contact avec la rangée de culture et qui ont les mêmes caractéristiques que les adventices potentielles. Cependant, pour une pulvérisation efficace, les agriculteurs préfèrent traiter les zones exemptes de mauvaises herbes plutôt que de manquer des mauvaises herbes qui pourraient concurrencer les cultures.

Une analyse quantitative a été effectuée en comparant les taux d'infestation en mauvaises herbes théorique ( $WIR_T$ ) et estimé ( $WIR_E$ ). Seules les adventices inter-rang et potentielles ont été utilisées. Ainsi, sur les 80 échantillons, les taux d'infestation moyen sont  $WIR_T = 0.0388$  et  $WIR_E = 0.0442$ . L'écart moyen entre le  $WIR_T$  et  $WIR_E$  par image est de 1.65 %. Nous remarquons que le taux d'infestation estimé, est en moyenne supérieur à celui théorique ce qui rejoint les remarques de l'analyse visuelle.

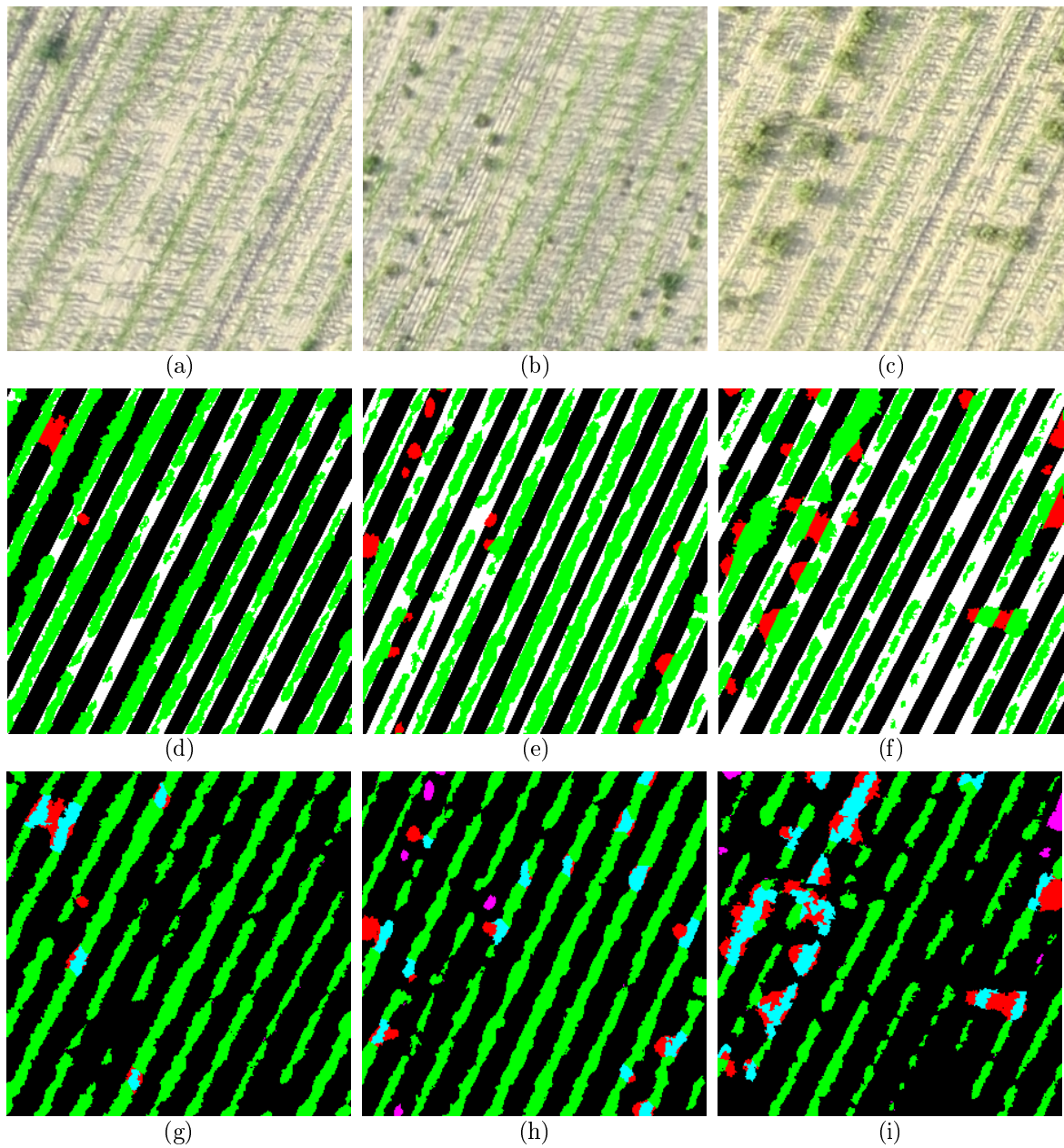


FIGURE 4.22 – Les trois colonnes présentent la vérité terrain et le résultat de la méthode de détection d’adventices sur trois images différentes avec des WIR. La première ligne présente les images originales. La deuxième ligne présentent la vérité terrain (en rouge les adventices, en vert les cultures et en blanc la délimitation des rangées de culture. La dernière ligne présentent le résultat après la combinaison de la relation entre superpixels et le masque représentant les rangées de culture. Le magenta représente les adventices inter-rangs, le rouge les adventices trop proches des rangées de culture (adventices potentielles) et en cyan les indécis.

## 4.5 Conclusion

Dans ce chapitre, nous avons proposé deux méthodes de détection de rangées de culture. La première, appelée dans cette thèse CR-Hough-SLIC, combine l'utilisation de la transformée de Hough et un algorithme de génération de superpixels, le SLIC. La seconde, appelée CRowNet, utilise la transformée de Hough et un algorithme d'apprentissage profond. Ces deux méthodes ont été évaluées dans deux champs différents : un champ de betterave pour lequel une partie des images a été utilisée pour entraîner le CRowNet et un champ de maïs avec une résolution spatiale différente pour tester la robustesse. Les résultats obtenus ont montré qu'il est possible de détecter les rangées de culture en couplant la transformée de Hough et un CNN même avec une forte pression des mauvaises herbes. En plus des méthodes de segmentation sémantique, le CRowNet et le CR-Hough-SLIC ont également été comparés avec le CR-Hough (une méthode utilisant l'information géométrique connue du champ). L' $IoU$  obtenus avec le CRowNet est supérieur à celui de CR-Hough et de CR-Hough-SLIC dans le champ de betterave. Cependant, les résultats restent très proches dans le champ de maïs. Sur les images acquises dans le champ de betterave, 93,58 % des rangées de culture ont été détectées avec un score  $IoU$  supérieur à 70 %. CRowNet a également démontré sa robustesse dans un champ de maïs avec une forte présence d'ombres après un changement de contraste.

Même si le CR-Hough-SLIC a donné des résultats comparables au CRowNet, le choix des paramètres reste déterminant. C'est la raison pour laquelle dans les perspectives, nous prévoyons de renforcer la robustesse de CRowNet en testant d'autres architectures de réseau au niveau de HoughCNet ou même créer un seul réseau capable de fusionner les deux traitements. Dans les perspective nous souhaitons également adapter les méthodes pour la détection des rangées non linéaires.

Après la détection des rangées de culture, une méthode de détection d'adventices inter-rang et potentielles a été appliquée. Même si la méthode de détection d'adventices s'est montrée efficace dans le cas où les rangées de culture sont bien identifiables, cette approche reste beaucoup plus utilisée dans le cadre des grandes cultures et avec des images dont la résolution est faible. Dans la suite de cette thèse, des approches exploitant l'apprentissage machine sont proposées (Chapitre 5) pour la détection des adventices inter-rang et intra-rang.

# Chapitre 5

## Détection des adventices par apprentissage

### Sommaire

---

<b>5.1</b>	<b>Labellisation non supervisée des données d'apprentissage . . .</b>	<b>68</b>
<b>5.2</b>	<b>Apprentissage machine classique . . . . .</b>	<b>69</b>
5.2.1	SVM ou Séparateurs à Vaste Marge . . . . .	69
5.2.2	Forêts d'arbres aléatoires . . . . .	70
5.2.3	Extraction de caractéristiques . . . . .	71
	Information couleurs . . . . .	71
	Descripteurs géométriques . . . . .	71
	Densité de contours . . . . .	72
	Histogramme des gradients orientés (HOG) . . . . .	72
	Descripteurs de Haralick . . . . .	72
	Les ondelettes de Gabor . . . . .	73
<b>5.3</b>	<b>Apprentissage profond (deep learning) . . . . .</b>	<b>73</b>
<b>5.4</b>	<b>Expériences et résultats . . . . .</b>	<b>74</b>
5.4.1	Extraction de caractéristiques et apprentissage machine . . . . .	77
5.4.2	ResNet . . . . .	80
5.4.3	Classification d'images . . . . .	83
<b>5.5</b>	<b>Conclusion . . . . .</b>	<b>85</b>

---

L'apprentissage machine (*machine learning* en anglais) est un domaine qui étudie comment un algorithme pourrait générer des modèles de classification ou de régression à partir d'un ensemble de données. En d'autres termes, faire de l'apprentissage consiste à traiter des données pour en extraire des informations particulières et enfin trouver des liens ou générer un modèle. Dans le domaine agricole, des études exploitent cette méthode pour la séparation des adventices et des cultures. Ces méthodes vont de l'utilisation des méthodes d'apprentissage machine classique aux méthodes d'apprentissage profond. Ces dernières se sont avérées performantes, mais nécessite de disposer d'un grand nombre de données. Comme mentionné dans la section 2.3, la création de grands ensembles de données agricoles avec des annotations au niveau des pixels est une tâche fastidieuse. C'est la raison pour laquelle nous avons développé une méthode de détection d'adventices entièrement automatique. L'automatisation s'étend de la phase de collecte des données à la phase d'apprentissage : les données d'apprentissage sont collectées à partir du résultat de la méthode de détection d'adventices de la section 4.3 ; pour l'apprentissage les algorithmes d'apprentissage profond ou *deep learning* ont été préférés aux méthodes d'apprentissage machine classique (choix manuel des descripteurs et de la machine d'apprentissage).

Ce chapitre est divisé en cinq sections : dans la première section la méthode de labellisation des données d'apprentissage est présentée. Dans le but d'accentuer l'intérêt du *deep learning* par rapport aux méthodes d'apprentissage classique les deux approches ont été étudiées et comparées : la deuxième section décrit les descripteurs et les algorithmes d'apprentissage classique testés ; la troisième section présente le réseau d'apprentissage profond ainsi que la technique appliquée pour l'entraîner. Les résultats sont commentés et discutés dans la quatrième section avant de conclure dans la dernière section.

### 5.1 Labellisation non supervisée des données d'apprentissage

Compte tenu des difficultés rencontrées dans la phase de collecte et de labellisation des données, une méthode entièrement automatisée a été proposée. Cette méthode consiste à utiliser les méthodes de détection d'adventices et des rangées de culture développées dans la section 4.3.

Les adventices ont été collectées en utilisant une fenêtre englobante sur chaque adventice inter-rang. Pour les échantillons de culture, une fenêtre coulissante est appliquée sur l'image en utilisant les positions relatives aux rangées de culture. Ainsi, à chaque position de la fenêtre, si celle-ci croise le masque de rangée de culture et qu'elle ne contient pas de pixels d'adventices inter-rang, l'échantillon est classé comme culture. En général, la classe culture contient beaucoup plus d'échantillons que la classe adventices. Dans le cas où il y a moins d'échantillons d'adventices inter-rang, si plusieurs adventices potentielles ont été détectées, nous avons proposé de rajouter les échantillons de ces adventices potentielles à la classe adventices. Ainsi, la fenêtre qui contient majoritairement des adventices potentielles est étiquetée comme étant de l'adventice.

## 5.2 Apprentissage machine classique

L'idéal pour une bonne classification est d'avoir un classifieur rapide, qui évite le sur-apprentissage, capable de répondre à des problèmes multi-classes, de séparer les classes avec un grand écart ou une marge importante, de gérer des attributs de grandes dimensions. Pour répondre à ces contraintes, les séparateurs à vaste marge (*support vector machine* : SVM) et les forêts d'arbres aléatoires (*random forest*) ont été utilisés.

### 5.2.1 SVM ou Séparateurs à Vaste Marge

Les machines à vecteurs de support ou séparateurs à vaste marge (*support vector machine* : SVM) [33], sont initialement définies pour la discrimination de variables binaires (Figure 5.1). L'idée est de rechercher une règle de décision basée sur une séparation par *hyperplan de marge optimale*. Intuitivement, une bonne séparation est obtenue par l'hyperplan qui a la plus grande distance par rapport à l'échantillon d'apprentissage le plus proche de n'importe quelle classe (appelée marge fonctionnelle) car en général, plus la marge est grande, plus l'erreur de généralisation du classifieur est faible. La popularité de ce type de classifieur vient non seulement du fait qu'il fournit une très grande marge de séparation de classes binaires, mais aussi parce qu'il est adapté aux cas linéaires et non linéaires [36]. Les SVM reposent sur deux idées clés, la notion de marge maximale et la notion de fonction noyau. La première consiste à définir l'hyperplan comme solution d'un problème d'optimisation. La seconde utilise une fonction noyau (ou *kernel*) pour trouver des surfaces séparatrices non linéaires. Le noyau gaussien est le plus populaire [161].

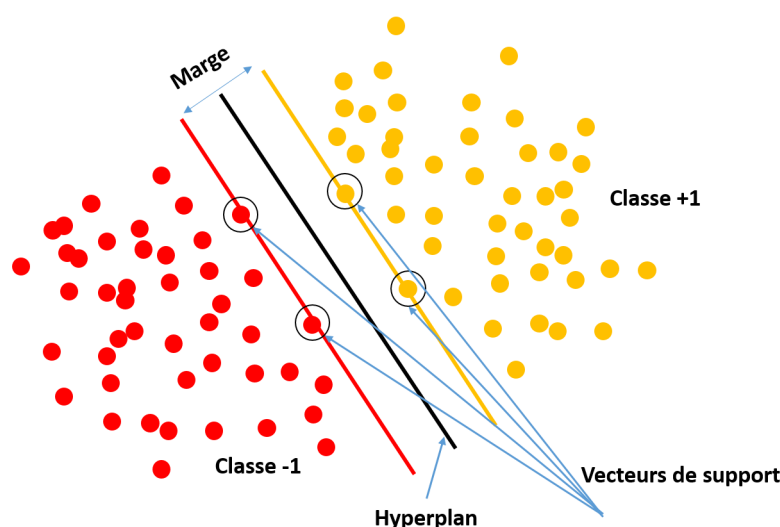


FIGURE 5.1 – Illustration du fonctionnement du SVM.

### 5.2.2 Forêts d'arbres aléatoires

Une forêt d'arbres aléatoires (RF) [104], est un méta-classifieur qui combine plusieurs classifieurs faibles (*weak learners* en anglais) pour en former un qui est fort. RF traite facilement les problèmes multi-classes, il est robuste et a un risque très faible de sur-apprentissage. Il est utilisé dans plusieurs applications telles que le suivi de points en vidéosurveillance, l'imagerie médicale, ou encore les jeux (la Kinect de Microsoft) [145]. Comme son nom l'indique, il est structuré sur le principe d'une forêt avec des arbres, où chaque arbre a des racines, des branches et des feuilles. Les arbres correspondent aux différents classifieurs, le premier nœud correspond à la racine de l'arbre (le point d'entrée des données), chaque nœud est ensuite séparé en nœuds intermédiaires et chaque feuille correspond à un nœud terminal où la décision finale est enregistrée. Les arbres de la forêt sont construits en utilisant l'ensachage ou l'agrégation *bootstrap* [61]. Le principe de l'agrégation bootstrap consiste à construire chaque arbre en sélectionnant un sous-ensemble de  $n$  observations parmi les  $N$  données d'apprentissage ( $n < N$ ) obtenues par échantillonnage aléatoire avec remise. L'objectif est d'obtenir des arbres aussi différents que possible, ou, en d'autres termes, d'obtenir des arbres non corrélés car plus les arbres sont différents, plus la forêt est robuste et donc sa capacité de généralisation est élevée.

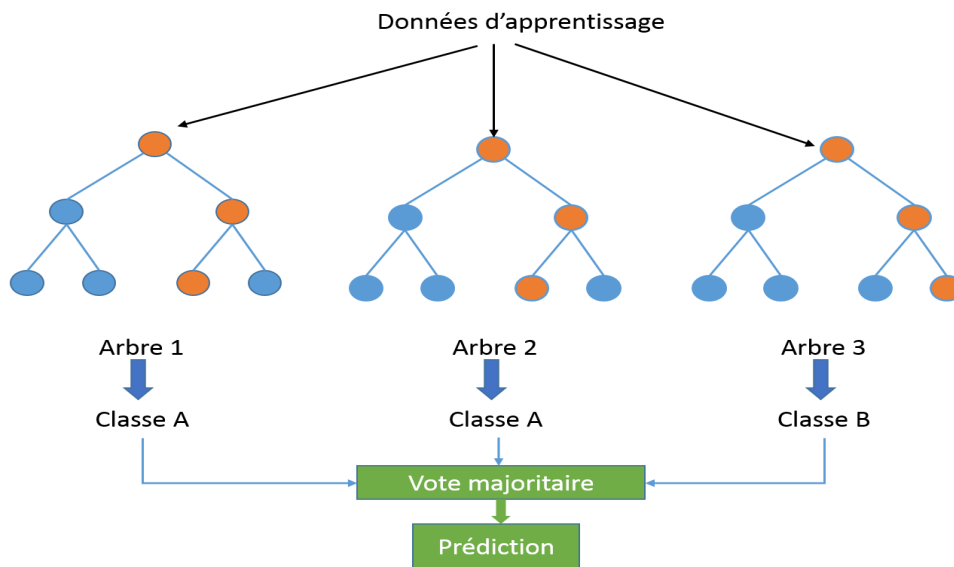


FIGURE 5.2 – Fonctionnement du random forest.

Pour augmenter la vitesse de construction des différents arbres de la forêt, Leo Breiman [104] propose, au lieu d'utiliser tous les descripteurs à chaque nœud, de faire une sélection aléatoire d'un certain nombre  $K$  de descripteurs puis de garder le meilleur en utilisant une fonction de séparation ou classifieur faible (*split function* ou *weak learner* en anglais). Ces classifieurs faibles peuvent être simples (critère d'impureté de Gini, critère d'impureté de l'entropie, etc.) ou complexes (SVM, boosting, etc.) [35, 11, 60]. Le meilleur descripteur est celui qui fournit le gain d'information le plus élevé. Cette méthode de séparation est utilisée sur tous les nœuds de l'arbre jusqu'à atteindre un critère d'arrêt qui peut être la feuille ou lorsque le nœud ne contient que des éléments de la même classe.

Après la construction de chaque arbre, chaque feuille de l'arbre donne la probabilité

qu'à une classe d'appartenir à cette feuille (score). C'est la moyenne des scores se trouvant dans chaque feuille qui donne le vote de l'arbre ou la classification de l'arbre. Cependant, c'est la combinaison des votes de tous les arbres qui donne le résultat de la classification de la forêt c'est le principe de "l'union fait la force" (Figure 5.2).

### 5.2.3 Extraction de caractéristiques

Tout comme pour le choix des machines d'apprentissage, le choix des descripteurs est primordial pour une bonne classification. Par exemple, lors de l'interprétation humaine des images, quatre principaux éléments sont observés : les caractéristiques spectrales, texturales, contextuelles et temporelles. Les caractéristiques spectrales décrivent l'ensemble des variations de niveaux de gris entre les différentes bandes spectrales d'un ensemble d'images multibandes. La texture est considérée comme une distribution de motifs locaux de niveau gris. Les caractéristiques contextuelles contiennent des informations sur la disposition relative des régions appartenant à différentes catégories et les caractéristiques temporelles décrivent les changements dans les attributs de l'image en fonction du temps [87]. En nous appuyant sur des travaux de la littérature (section 2.3), plusieurs caractéristiques ont été calculées, notamment des descripteurs statistiques, de forme et de texture [15, 101, 108, 132].

#### Information couleurs

Les caractéristiques de couleur utilisées sont les moyennes et les écarts types des trois canaux d'image RGB et de l'indice ExG. Afin de rendre les caractéristiques de couleur cohérentes avec les différents niveaux d'éclairage, chaque canal de couleur a été normalisé par la somme des trois canaux de couleur.

#### Descripteurs géométriques

Basé sur [8], trois paramètres à savoir le facteur de forme noté  $F$  (Equation 5.1), l'allongement (Equation 5.2) et la solidité (Equation 5.3) ont été calculés en tant que caractéristiques géométriques. L'ensemble de ces trois descripteurs est appelé *Geo3* dans la suite du manuscrit.

$$F = \frac{4 * \pi * S}{P^2} \quad (5.1)$$

$$Allongement = \frac{S}{E^2} \quad (5.2)$$

$$Solidité = \frac{S}{S_{convexe}} \quad (5.3)$$

$E$  : correspond à l'épaisseur. La surface ( $S$ ) est définie comme le nombre de pixels avec une valeur '1' dans l'image binaire. Le périmètre ( $P$ ) est défini comme le nombre de pixels ayant une valeur '1' dont au moins un des huit pixels voisins a la valeur '0', ce qui implique que le périmètre est le nombre de pixels de bordure.  $S_{convexe}$  est la surface de la plus petite région convexe.

### Densité de contours

La densité de contours ( $DC$ ) est une méthode de segmentation d'image qui peut exploiter la fréquence des contours entre les cultures et les adventices [101]. Elle est définie comme la surface des contours de l'image calculés par le filtre de Sobel ( $S_c$ ) par rapport à la surface totale (Equation 5.4).

$$DC = \frac{S_c}{S} \quad (5.4)$$

### Histogramme des gradients orientés (HOG)

Les attributs ou descripteurs de contours correspondent généralement à l'histogramme de l'orientation du gradient. Le HOG compte les occurrences d'orientation de gradient dans les régions localisées d'une image. Le principe du HOG est de diviser l'image en cellules uniformément espacées et se chevauchant. Pour chaque cellule, les normes et les orientations du gradient sont calculées. Un histogramme pondéré par la norme selon des intervalles d'angles spécifiés est calculé puis normalisé. Grâce à la normalisation du contraste local, il devient invariant aux conditions d'éclairage. A la fin, les histogrammes des différentes cellules sont concaténés. Le HOG a été initialement utilisé pour la détection des piétons [38], mais il a prouvé sa robustesse dans le domaine agricole pour identifier les feuilles des plantes dans [167, 77]. Le HOG est calculé en divisant l'histogramme des orientations en 9 classes, comme suggéré dans [38].

### Descripteurs de Haralick

La matrice de cooccurrence permet d'obtenir la fréquence d'occurrence du niveau de gris de deux pixels séparés par une distance  $d$  suivant une direction  $\theta$  (Figure 5.3). Dans [68], les auteurs proposent 14 caractéristiques pouvant être calculées sur cette matrice : descripteurs de Haralick. Ces caractéristiques ont pour but de mettre en évidence certaines caractéristiques visuelles, statistiques, le caractère aléatoire de la distribution des niveaux de gris et la dépendance linéaire des niveaux de gris sur un voisinage de pixels (homogénéité, grossièreté, périodicité, lissé, ...).

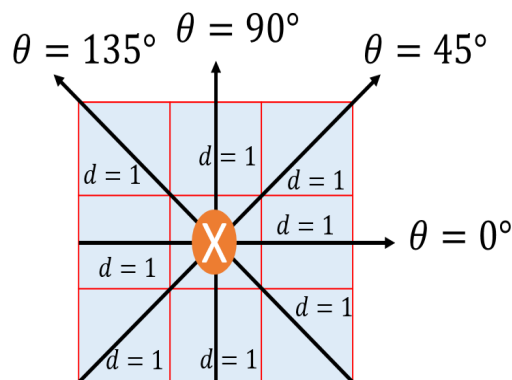


FIGURE 5.3 – Différentes orientations.

En 2012, la méthode Haralick a été appliquée pour extraire les caractéristiques de texture dans la classification des espèces végétales [5]. Au total, 6 attributs de Haralick ont été calculés sur 4 orientations différentes ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  et  $135^\circ$ ). Ces attributs sont l'autocorrélation, le contraste, la corrélation, la dissimilarité, l'énergie et l'entropie. Les mêmes attributs ont été sélectionnés dans ce travail.

### Les ondelettes de Gabor

La transformée de Fourier à courte durée avec une fenêtre gaussienne s'appelle la transformée de Gabor. Elle est capable de préserver les informations locales et globales dans l'image et est particulièrement utile pour analyser des images texturées contenant des caractéristiques de fréquence ou d'orientation très spécifiques [21, 126]. En 2003, Tang *et al.* ont effectué la classification des adventices à feuilles larges et étroites avec une orientation de filtre fixée à  $90^\circ$  et 4 fréquences de filtres (4, 5, 6 et 7) [151]. Les mêmes valeurs de fréquence ont été choisies, mais avec des orientations différentes  $0^\circ$ ,  $45^\circ$  et  $90^\circ$ . Au total 12 attributs ont été générés.

## 5.3 Apprentissage profond (deep learning)

Le réseau d'apprentissage profond choisi dans ce travail est un réseau résiduel (ResNet). Ce réseau a été introduit en 2015 [71]. Il a remporté le défi de reconnaissance d'images sur la base ImageNet la même année. Le ResNet consiste à propager le signal d'entrée dans différents niveaux du réseau afin d'améliorer ses performances et éviter les pertes d'efficacité si le réseau augmente en profondeur (Figure 5.4). La particularité des réseaux résiduels est qu'ils peuvent être extrêmement profonds avec moins de complexité lors de la phase d'apprentissage. Si, avec les réseaux classiques, le but est de pondérer les poids des couches pour obtenir une fonction d'approximation de l'entrée, le but des réseaux résiduels est d'ajuster les poids des différentes couches pour que la variable en entrée de la première couche soit identique à la sortie des différentes couches. En d'autres termes, les réseaux résiduels essaient de ramener les pondérations des couches non linéaires multiples vers 0.

De manière formelle, en empilant les couches de convolution nous nous attendons à obtenir des caractéristiques de l'image placée en entrée  $X$  à la sortie  $\mathcal{H}(x)$ . En effet, en rajoutant  $X$  au bout des couches de convolution, un résultat  $\mathcal{H}(x)$  égal à  $\mathcal{F}(x) + X$  est obtenu. Ceci implique que la fonction  $\mathcal{F}(x) = \mathcal{H}(x) - X$  est la sortie des couches de convolution empilées et correspond à une carte de résidus. Ainsi entraîner un réseau résiduel revient à ramener la valeur de  $\mathcal{F}(x)$  à zéro.

En 2015, avec un réseau résiduel de 152 couches, ResNet152 a remporté la compétition ImageNet. Son architecture a été jugée moins complexe que celle de VGG16 ou VGG19 (Table 5.1).

Cependant, vis à vis de la taille des données, le ResNet avec 18 couches (ResNet18) a été choisi car il réalise un bon compromis entre performance et temps de calcul. Aussi il a montré un meilleur résultat que AlexNet et VGG13 [146] dans le défi ImageNet. Étant donné le nombre de paramètres à mettre à jour dans le réseau ResNet18 et le nombre de données dont nous disposons, il a été décidé d'utiliser l'apprentissage par transfert. L'ap-

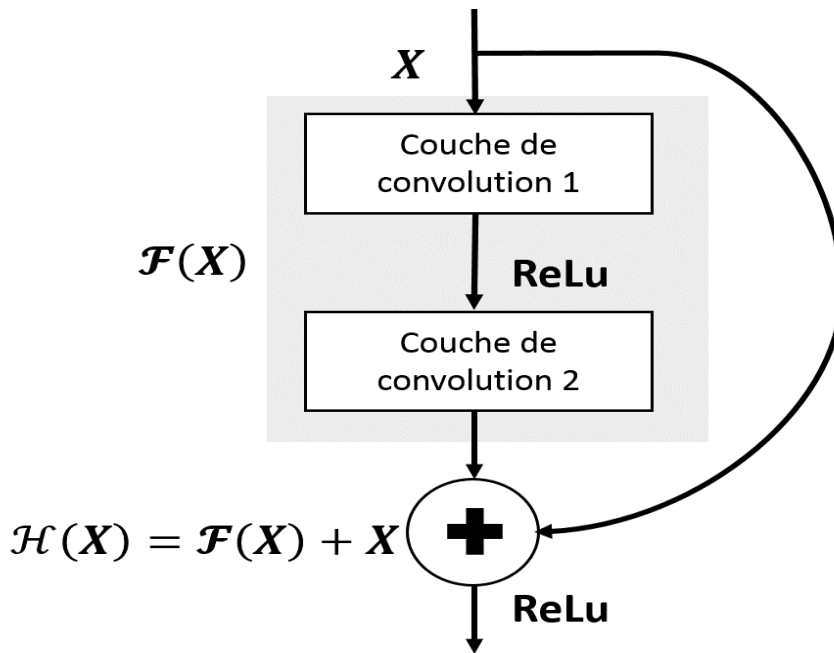


FIGURE 5.4 – Fonctionnement d’un réseau résiduel.

TABLE 5.1 – Nombre d’opérations en virgule flottante par seconde (FLOPS) par architecture (en milliard) [71].

Réseau	VGG16	VGG19	ResNet18	ResNet34	ResNet101	ResNet152
FLOPS	15.3	19.6	1.8	3.6	7.6	11.3

prentissage par transfert est l’approche la plus populaire dans l’apprentissage profond : des modèles pré-entraînés sur un ensemble de données comme ImageNet sont utilisés comme point de départ pour résoudre un autre problème de vision par ordinateur. En raison de l’abondance des catégories et du nombre important d’images dans ImageNet, certaines études ont révélé la performance de la transférabilité des réseaux entraînés avec le jeu de données ImageNet [74, 123]. La technique de transfert learning appliquée est appelée *fine-tuning*. Le *fine-tuning* consiste à utiliser le réseau déjà entraîné sur la base ImageNet et de tronquer la dernière couche du réseau en la remplaçant par une nouvelle couche pertinente pour le problème de classification traité. Dans notre cas, les mille classes d’ImageNet ont été remplacées par seulement deux classes : culture et adventices.

## 5.4 Expériences et résultats

Les expériences ont été menées sur des images acquises dans les champs de haricot et d’épinard (Figure 3.4) avec le drone DJI Phantom 3 Pro (section 3.2.2) embarquant une caméra RGB de 36 MP à une altitude de 20 m. Pour constituer la base de données d’apprentissage non supervisée, deux parties différentes de chaque champ ont été choisies. La première (Partie 1) sert à recueillir les données d’apprentissage et l’autre (Partie 2) pour recueillir les données de test.

Le masque binaire de culture a été créé en intersectant les rangées détectées et les superpixels créés par la méthode SLIC (section 4.1.3). La compacité des superpixels a été fixée à 20 et le nombre de superpixels est égal à  $0,1 \% \times N$ , où  $N = 7360 \times 4912$  pixels. La Figure 5.5 montre un exemple de détection d’adventices dans le champ de haricot. Dans cette expérience, une fenêtre de  $64 \times 64$  pixels a été employée pour collecter les données d’apprentissage avec un recouvrement de 50%. Cette taille de fenêtre offre un bon compromis entre le type de plante et l’information globale sur son voisinage. En effet, une petite fenêtre n’est pas suffisante pour capturer la plante entière et peut conduire à confondre culture et non culture car, dans certaines conditions, la culture et les feuilles d’adventices ont les mêmes caractéristiques visuelles. Par contre, une taille trop grande présente le risque d’avoir des cultures et des mauvaises herbes dans la même fenêtre.

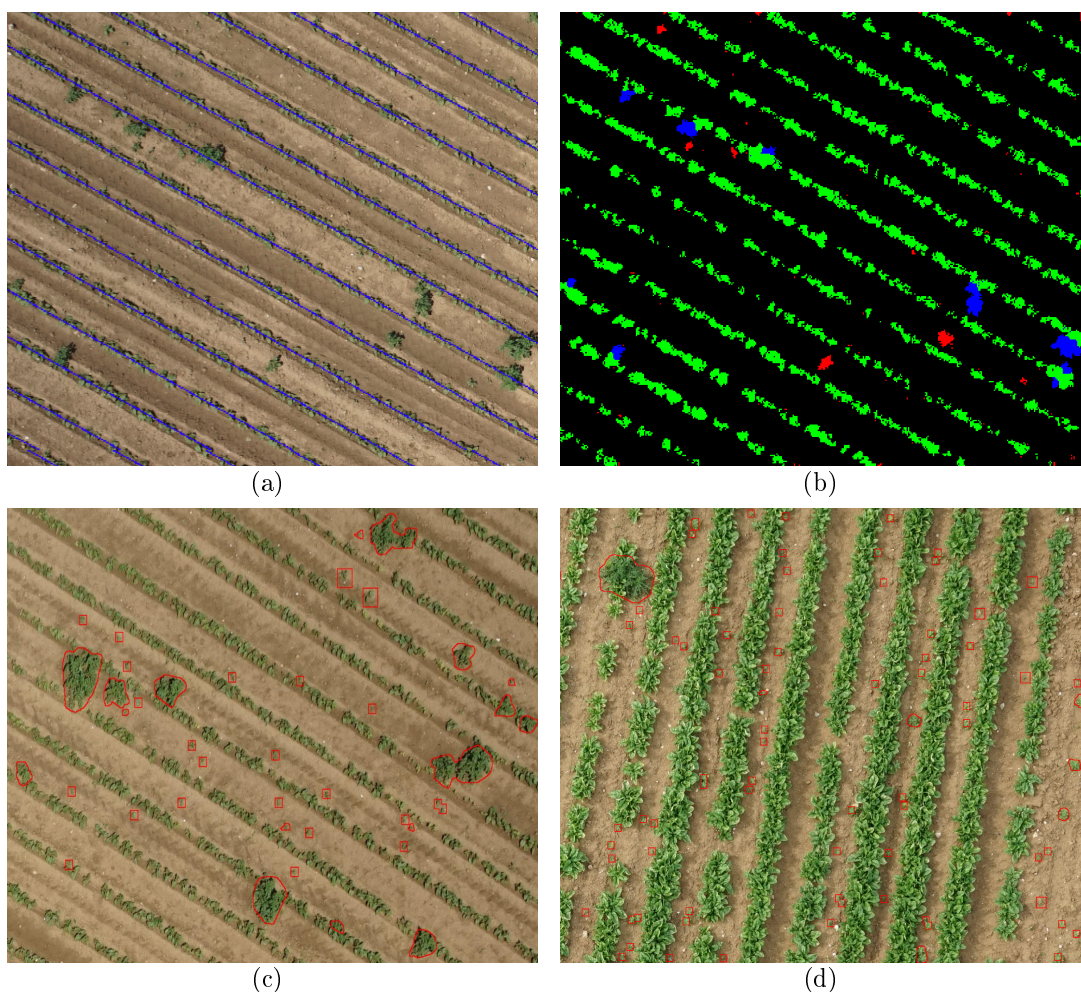


FIGURE 5.5 – (a) Image originale et rangées de culture détectées (en bleu). (b) Adventices potentielles (en bleu) et inter-rang en rouge. Parties du champ de haricot (c) et du champ d’épinard (d) avec les adventices étiquetées manuellement en rouge. L’étiquetage manuel a pris environ 2 jours et moins de 30 minutes avec la méthode de labellisation non supervisée sur 4 images de tailles  $N=7360 \times 4912$  pixels.

Dans le champ de haricot, les adventices étaient majoritairement des chardons et des jeunes pousses de pommes de terre provenant de semis antérieurs sur le même champ

et il y avait peu d'adventices inter-rang (Figure 5.5). Les adventices potentielles ont été incluses dans les échantillons d'adventices ; une augmentation de données a également été effectuée. Le champ d'épinard était plus infecté que celui de haricot, il y avait surtout des chardons. Au total, deux changements de contraste, un lissage avec un filtre gaussien et trois rotations ( $90^\circ$ ,  $180^\circ$ ,  $270^\circ$ ) ont été effectués. Dans les parcelles agricoles, une forte hétérogénéité est parfois rencontrée d'une zone à une autre. Cette hétérogénéité peut correspondre à une différence d'humidité du sol, de présence de paille, etc. Afin de rendre les modèles robustes à l'arrière-plan, des échantillons avec et sans arrière-plan ont été utilisés. Les échantillons sans arrière-plan ont été obtenus en appliquant l'indice ExG suivi du seuillage d'Otsu sur les échantillons précédemment créés (Figure 5.6).

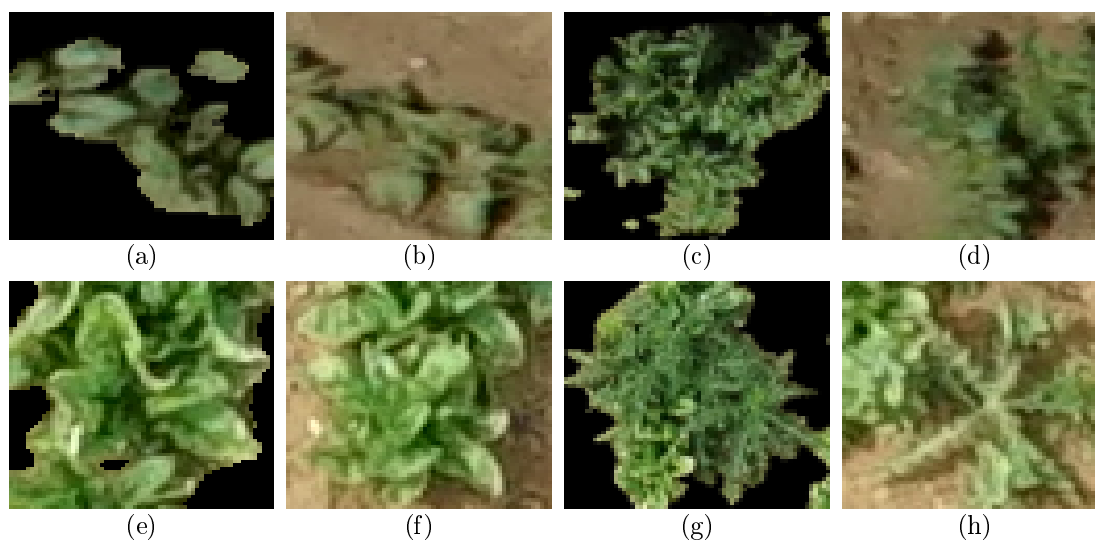


FIGURE 5.6 – Exemple d'échantillons de culture et d'adventices de taille  $64 \times 64$  pixels avec et sans fond. Haricot : échantillons de culture (a et b), échantillons d'adventices (c et d). Épinards : échantillons de culture (e et f) et échantillons d'adventices (g et h).

Les données d'apprentissage supervisées ont été étiquetées manuellement. La Figure 5.5 présente les mauvaises herbes délimitées en rouge. Les données supervisées recueillies étaient également déséquilibrées, de sorte que la même procédure d'augmentation de données que celle effectuée sur les données non supervisées a été appliquée.

Dans le champ d'épinard, un total de 4303 échantillons de culture et 3626 échantillons d'adventices ont été étiquetés de façon non supervisée. Contrairement au champ de haricot, les données étaient moins déséquilibrées. Par conséquent, la seule augmentation de données appliquée est l'ajout d'échantillons sans arrière-plan. Le même traitement a été appliqué aux données supervisées. La Table 5.2 présente les échantillons utilisés par classe et par champ. Nous constatons que le nombre d'échantillons collectés de manière supervisée dans le champ de haricot est 3 fois plus grand que celui des données labellisées de manière non supervisée. Dans le champ d'épinard le nombre d'échantillons des données supervisées est plus du double de celles non supervisées.

Les performances des modèles ont été évaluées sur la base de tests collectée dans la Partie 2 de manière supervisée sur chaque champ ; la Table 5.3 présente les échantillons. Les performances des résultats de la classification sont illustrées par des courbes ROC (receiver operating characteristic). Une courbe ROC est une courbe de probabilité repré-

TABLE 5.2 – Données d'apprentissage et de validation dans les champs de haricot et d'épinard.

<b>Champ de haricot</b>				
<b>Données</b>	<b>Classe</b>	<b>Apprentissage</b>	<b>Validation</b>	<b>Total</b>
Supervisées	Culture	17192	11694	28886
	Adventices	17076	9060	26136
Total		34268	20754	55022
Non supervisées	Culture	7688	1928	9616
	Adventices	5935	1493	7428
Total		13623	3421	17044

<b>Champ d'épinard</b>				
<b>Données</b>	<b>Classe</b>	<b>Apprentissage</b>	<b>Validation</b>	<b>Total</b>
Supervisées	Culture	11350	2838	14188
	Adventices	8234	2058	10292
Total		19584	4896	24480
Non supervisées	Culture	6884	1722	8606
	Adventices	5800	1452	7252
Total		12684	3174	15858

TABLE 5.3 – Nombre d'échantillons de test utilisés pour chaque champ.

<b>Champ</b>	<b>Culture</b>	<b>Adventices</b>
Haricot	2139	1852
Epinard	1523	1825

sentant les performances d'un modèle de classification pour tous les seuils de classification. C'est la courbe du taux de vrais positifs (TVP) en fonction du taux de faux positifs (TFP). En calculant l'aire sous la courbe (AUC ou *area under the curve* en anglais) nous indiquons dans quelle mesure le modèle est capable de distinguer les deux classes : cultures et adventices.

#### 5.4.1 Extraction de caractéristiques et apprentissage machine

Le SVM et le RF ont été appliqués aux caractéristiques extraites des données d'apprentissage (Table 5.2). Pour éviter le déséquilibre entre les classes lors de l'apprentissage dans le champ de haricot, 26000 et 8000 échantillons ont été sélectionnés respectivement pour les données supervisées et non supervisées. Dans le champ d'épinard, le nombre d'échantillons est respectivement de 10000 et 7000 pour les données supervisées et non supervisées. Pour le SVM, le noyau étant un paramètre important pour une bonne clas-

sification, deux noyaux ont été testés : les noyaux gaussien et linéaire. Le RF a été testé avec deux tailles d'arbres différentes : 100 et 200. Tous les résultats sont présentés dans l'Annexe B. Le SVM avec un noyau gaussien a montré des résultats bien meilleurs qu'avec le noyau linéaire dans le champ de haricot et proches dans le champ d'épinard. Pour ce qui est du RF l'augmentation du nombre d'arbres a eu peu d'effet. Dans ce chapitre seul les résultats de SVM avec un noyau gaussien et le RF avec 200 arbres sont présentés.

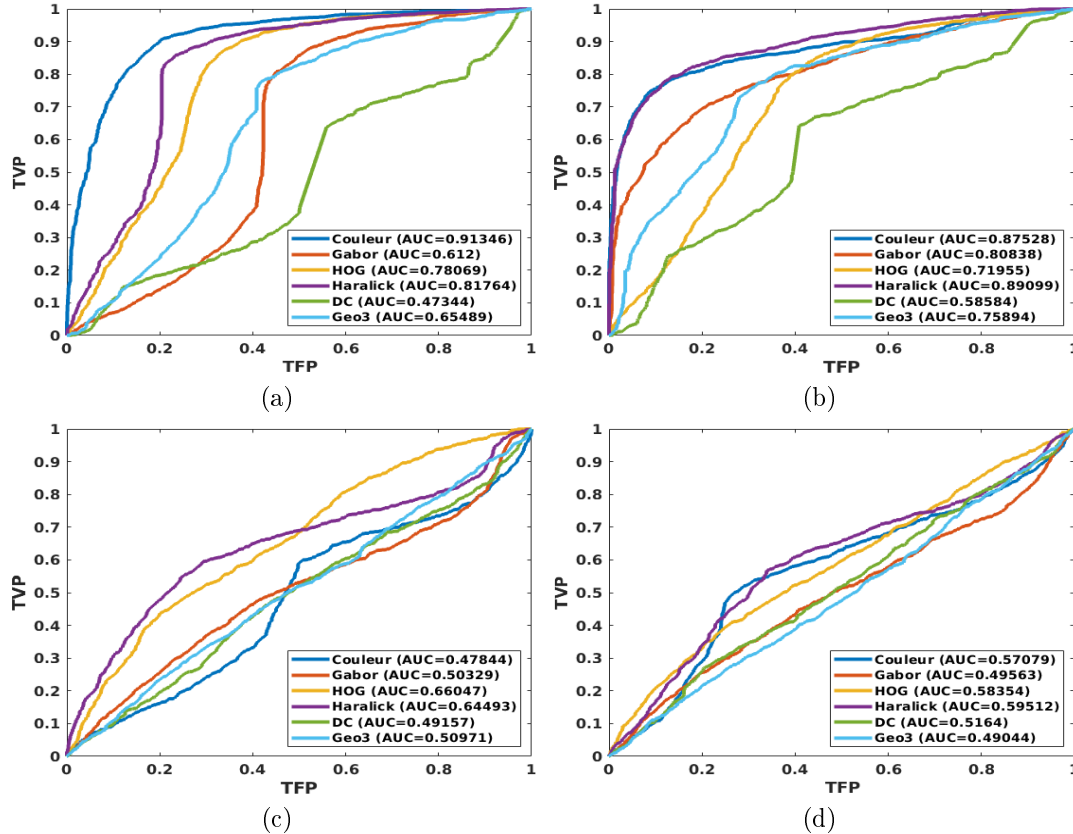


FIGURE 5.7 – Courbes ROC des modèles SVM créés par chaque caractéristique pour chaque champ. La première ligne représente le champ d'épinard et la seconde le champ de haricot. La première et la deuxième colonnes correspondent respectivement à l'apprentissage effectuée sur les données supervisées et non supervisées. TVP : taux de vrais positifs et TFP : taux de faux positifs.

Dans les Figures 5.7 et 5.8 nous pouvons remarquer que les descripteurs de couleur, de Haralick et ceux du HOG fournissent les meilleurs résultats. Dans le champ d'épinard, la forte présence de chardons dont la couleur des feuilles est différente de celle des épinards à un certain niveau de croissance explique l'efficacité des caractéristiques de la couleur. Dans le champ de haricot, les caractéristiques de couleur sont moins efficaces que les caractéristiques de texture (Haralick et HOG) dans les deux types de données (supervisées et non supervisées).

Les Tables 5.4 et 5.5 présentent les résultats de SVM et RF avec les meilleures caractéristiques sélectionnées. En utilisant le SVM avec les caractéristiques combinées, l'amélioration est d'environ 10 % pour les données supervisées et 12 % pour les données non supervisées dans le champ de haricot. Dans le champ d'épinard, l'amélioration est infé-

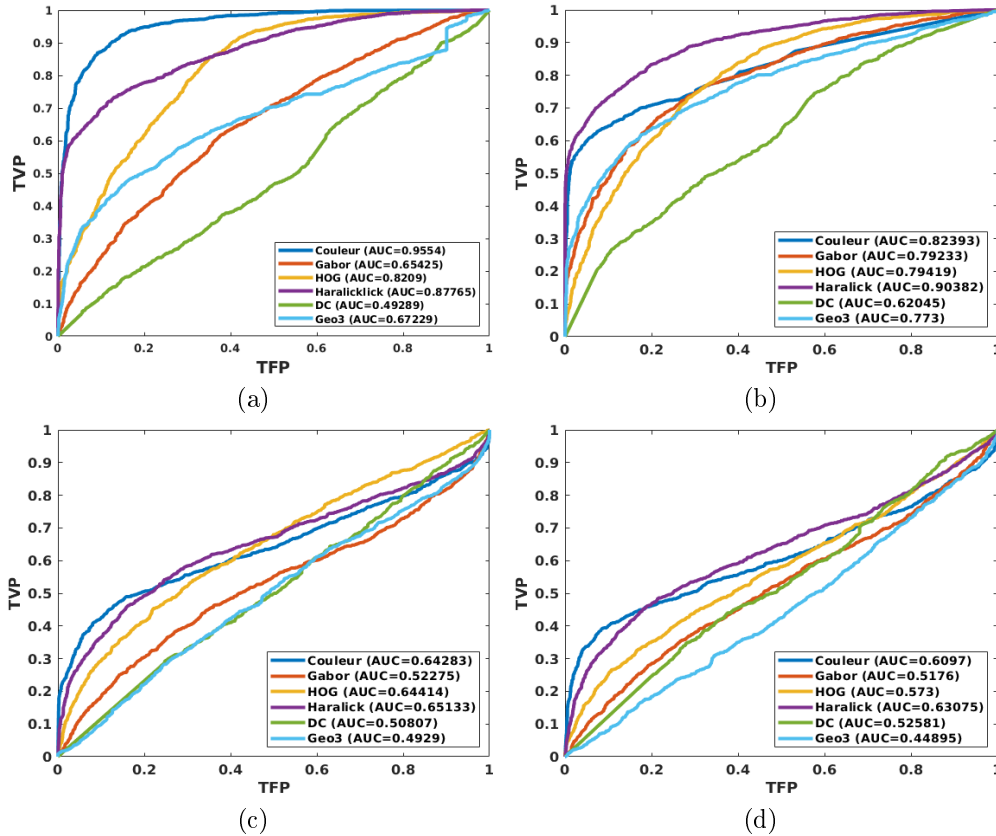


FIGURE 5.8 – Courbes ROC des modèles RF créés par chaque caractéristique pour chaque champ. La première ligne représente le champ d'épinard et la seconde le champ de haricot. La première et la deuxième colonnes correspondent respectivement à l'apprentissage effectuée sur les données supervisées et non supervisées. TVP : taux de vrais positifs et TFP : taux de faux positifs.

rieure à 5 %. Une autre remarque qui peut être fait ici est que d'un type de données à un autre, les meilleures caractéristiques ne sont pas toujours les mêmes. Nous avons également remarqué que les caractéristiques sélectionnées ne permettent pas de détecter les mauvaises herbes présentes dans le champ de haricot. Avec le RF, la procédure de sélection des caractéristiques n'a augmenté les performances que d'environ 1% pour les deux ensembles de données issus du champ d'épinard. Dans le champ de haricot, une amélioration d'environ 1% a été observée sur les données non supervisées et d'environ 5% sur les données supervisées.

De ces résultats, nous pouvons déduire que même si nous parvenons à sélectionner les caractéristiques les plus appropriées pour identifier les mauvaises herbes dans un champ, il est possible que ces caractéristiques ne soient pas adaptées à un autre champ avec un type de culture différent. Ils montrent également que les caractéristiques considérées comme meilleures par un classifieur ne sont pas nécessairement les meilleures pour un autre classifieur. Cependant, dans les parcelles agricoles, d'une année à l'autre, le type d'adventices n'est pas forcément le même ou d'une semaine à l'autre, les croissances des plantes et des adventices ne sont pas les mêmes, ce qui conduit à une nouvelle collecte de données et à une nouvelle sélection de caractéristiques. Ainsi, pour une classification

efficace, il est intéressant d'utiliser un outil capable de générer automatiquement des échantillons et des caractéristiques pertinentes, d'où l'intérêt d'utiliser un apprentissage profond avec un étiquetage de données non supervisé.

### 5.4.2 ResNet

Après la création des classes de culture et d'adventices, 80 % des échantillons ont été sélectionnés de manière aléatoire pour l'apprentissage et la validation de ResNet18, respectivement. La Table 5.2 présente les données d'apprentissage et de validation pour chaque champ.

Pour affiner les réglages, différentes valeurs de taux d'apprentissage ont été testées. Le taux d'apprentissage initial est fixé à 0,01 et mis à jour tous les 200 epochs (nombre de cycles d'apprentissage). La mise à jour se fait en divisant le taux d'apprentissage par 10. La Figure 5.9 montre l'évolution de la fonction de perte (fonction de coût ou *loss function* en anglais) pendant l'entraînement pour les ensembles de données supervisés et non supervisés. À partir de ces chiffres, nous pouvons remarquer que les courbes de perte de validation diminuent durant les 80 premières epochs avant d'augmenter et de converger (comportement proche du phénomène de sur-apprentissage). Ce phénomène de sur-apprentissage est moins mis en évidence dans les données de la parcelle de haricot étiquetées de manière supervisée. Les meilleurs modèles ont été obtenus lors de la première phase d'apprentissage avec un taux d'apprentissage de 0,01.

Dans la Table 5.4, on peut remarquer que ResNet18 donne de bien meilleurs résultats que SVM et RF dans le champ de haricot, avec une différence de performance supérieure à 20 %. Cependant, dans la parcelle d'épinard, les résultats obtenus sont comparables et parfois même les résultats de ResNet18 sont inférieurs à ceux de SVM et RF (Table 5.5). Les performances de ResNet18 dans le champ d'épinard peuvent s'expliquer par la faible quantité de données utilisées pour l'entraînement dans ce champ. Pour les algorithmes d'apprentissage profond, plus nous avons de données, mieux l'algorithme apprend. Nous pouvons également noter que les performances des modèles générés par les deux types de données (supervisées et non supervisées) sont comparables pour les trois méthodes de classification. La différence maximale est d'environ 6% dans les deux champs.

Les deux champs étant infestés principalement par des chardons, nous avons testé la robustesse de nos modèles en échangeant les échantillons d'adventices du champ de haricot avec ceux du champ d'épinard. Dans la Figure 5.11, les résultats obtenus montrent que malgré les petits échantillons récoltés dans le champ de haricot, ces données conviennent

TABLE 5.4 – Résultats des données de test recueillies dans le champ de haricot avec ResNet18, SVM et Random Forest (RF). Pour le SVM et le RF, seuls les résultats des meilleures caractéristiques sélectionnées sont présentés.

Descripteurs	Données	SVM		RF		ResNet18	
		Supervisées	Non	Supervisées	Non	Supervisées	Non
Tous les descripteurs (AUC%)		<b>78.76</b>	<b>68.92</b>	<b>69.76</b>	64.00	-	-
Haralick + Color (AUC%)		71.90	67.66	68.11	<b>64.95</b>	-	-
- (AUC%)		-	-	-	-	<b>94.84</b>	<b>88.73</b>

TABLE 5.5 – Résultats des données de test recueillies dans le champ d'épinard avec ResNet18, SVM et Random Forest (RF). Pour le SVM et le RF, seul les résultats des meilleures caractéristiques sélectionnées sont présentés.

Descripteurs	Données	SVM		RF		ResNet18	
		Supervisées Oui	Non	Supervisées Oui	Non	Supervisées Oui	Non
Tous les descripteurs (AUC%)		<b>96.78</b>	88.56	<b>96.86</b>	95.46	-	-
D1=Haralick + Couleur (AUC%)		95.90	<b>91.20</b>	92.69	95.90	-	-
D1 + HOG + Gabor (AUC%)		96.25	89.05	95.05	<b>96.33</b>	-	-
- (AUC%)		-	-	-	-	<b>95.70</b>	<b>94.34</b>

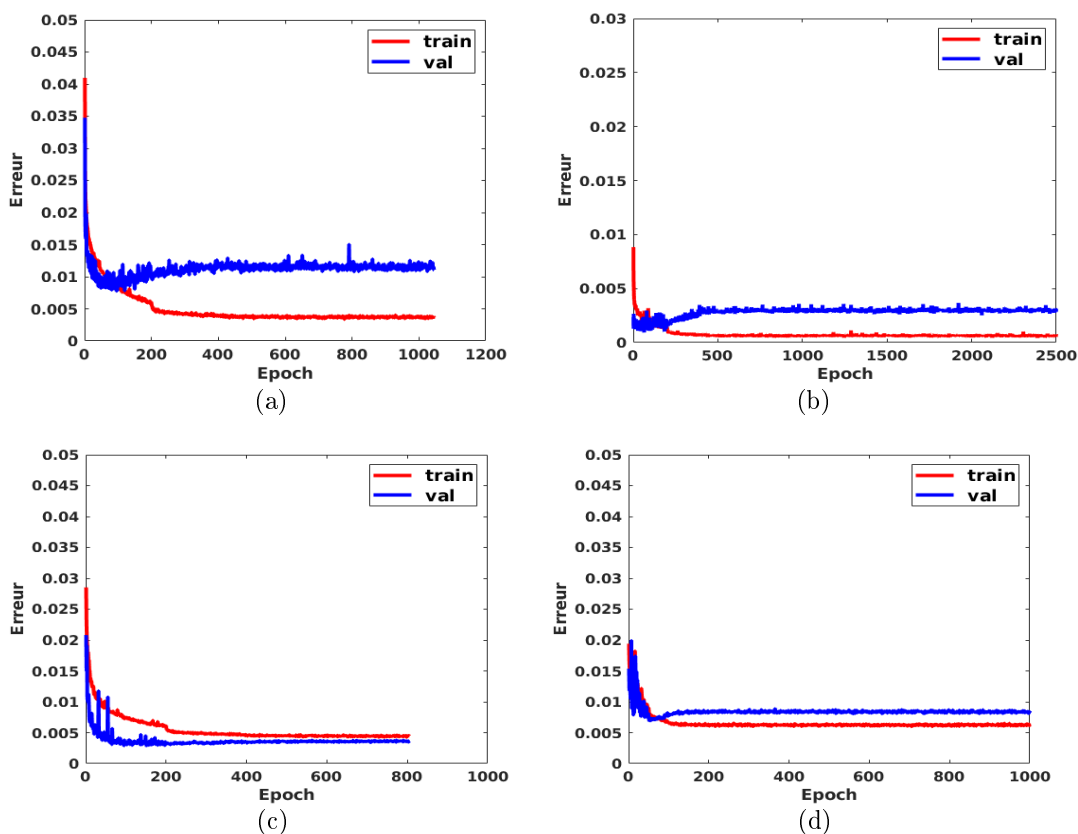


FIGURE 5.9 – Évolution de la fonction perte (fonction de coût) lors de l'apprentissage sur les données avec annotations supervisées et non supervisées dans les parcelles d'épinard et de haricot. train et val représentent respectivement les courbes d'apprentissage et de validation. Les courbes de perte lors de la validation (val) diminuent pendant environ les 80 premières epochs avant d'augmenter puis converger. La première ligne représente le champ d'épinard et la deuxième le champ de haricot. La première et la deuxième colonnes sont respectivement l'apprentissage sur les données supervisées et non supervisées.

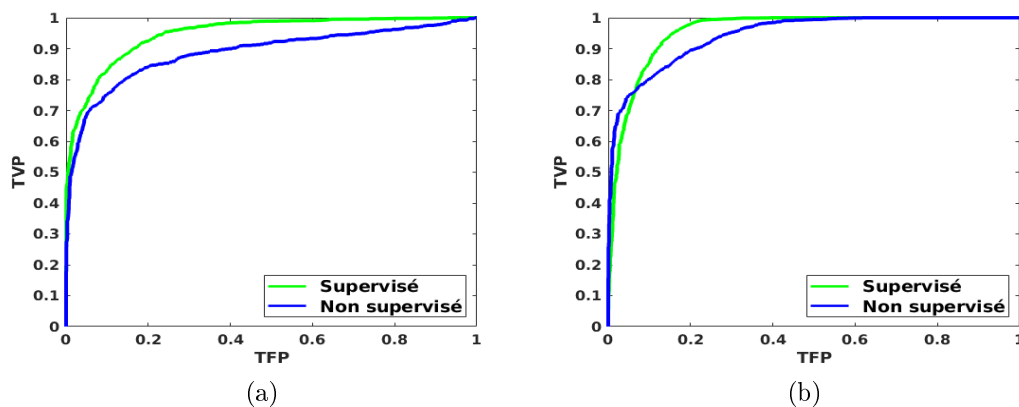


FIGURE 5.10 – Courbes ROC des données de test avec l'étiquetage non supervisé et supervisé des données. De gauche à droite, les courbes ROC calculées à partir des données de test du haricot (a) et des épinards (b). Dans le champ de haricot, les aires sous la courbe AUC sont de 88,73 % pour les données non supervisées et de 94,84 % pour les données supervisées. Dans le champ d'épinard, les AUC sont de 94,34 % pour les données non supervisées et de 95,70 % pour les données supervisées. Par données supervisées et non supervisées, nous entendons respectivement les données étiquetées par un humain et étiquetées automatiquement. TVP et TVF représentent respectivement les taux de vrais et faux positifs.

au champ d'épinard et le modèle créé avec l'étiquetage non supervisé dans le champ d'épinard est plus sensible à la présence de jeunes pousses de pomme de terre dans les échantillons d'adventices.

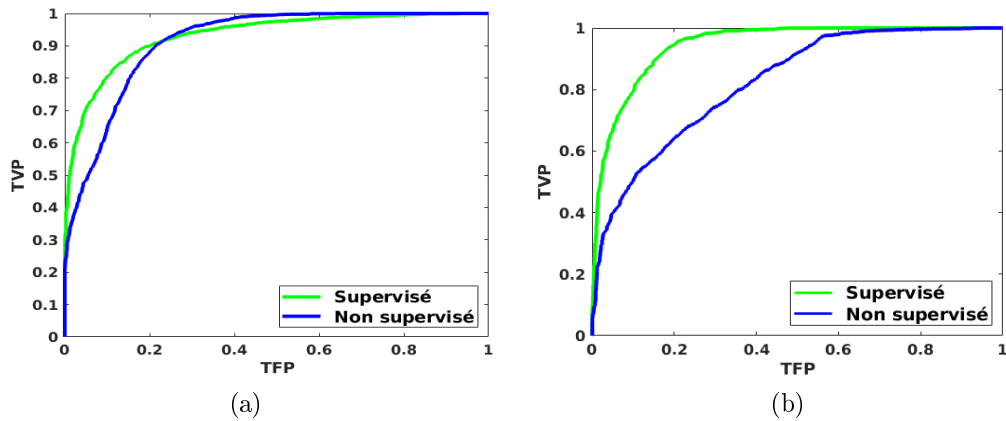


FIGURE 5.11 – Courbes ROC des données de test avec les données sur les mauvaises herbes du champ de haricot échangées avec celles du champ d'épinard. De gauche à droite, les courbes ROC obtenues à partir des données de test du haricot (a) et des épinards (b). Dans le champ de haricot, les aires sous la courbe AUC sont de 91,37 % pour les données non supervisées et de 93,25 % pour les données supervisées. Dans le champ d'épinard, l'aire sous la courbe (AUC) est de 82,70 % pour les données non supervisées et de 94,34 % pour les données supervisées. Par données supervisées et non supervisées, on entend respectivement les données étiquetées de manière supervisée et non supervisée.

### 5.4.3 Classification d'images

Dans le but de détecter les adventices sur l'ensemble d'une image une fenêtre glissante a été appliquée. Pour chaque position de la fenêtre, le modèle entraîné fournit la probabilité d'appartenir à la classe culture et adventice. Ainsi, le centre de l'imagette est marqué par un point coloré selon la classe qui a la probabilité la plus élevée. Les points bleus, rouges et blancs signifient respectivement que l'imagette est identifiée comme culture, mauvaise herbe et décision incertaine (Figures 5.12a et 5.12c). Une décision incertaine signifie que les deux probabilités sont très proches de 0,5. Par la suite, les informations des rangées de culture et les superpixels sont employés pour classer tous les pixels de l'image. Un superpixel est classé comme culture ou mauvaise herbe si les points majoritairement englobés sont respectivement en bleu ou en rouge. Pour les superpixels qui englobent majoritairement des points blancs, leur position par rapport aux rangées de culture a été étudiée. Par conséquent, les superpixels qui se trouvent dans les rangées de culture sont considérés comme des cultures et les autres sont des mauvaises herbes. Enfin l'arrière-plan a été masqué. Les Figures 5.12b et 5.12d présentent les résultats de la classification dans les parties des champs d'épinard et de haricot. Sur ces images, nous remarquons que les mauvaises herbes inter-rang et intra-rang ont été détectées avec une faible surdéttection. Les surdéttections se trouvent principalement sur les bords des rangées de culture où la fenêtre n'arrive pas à couvrir entièrement les plantes.

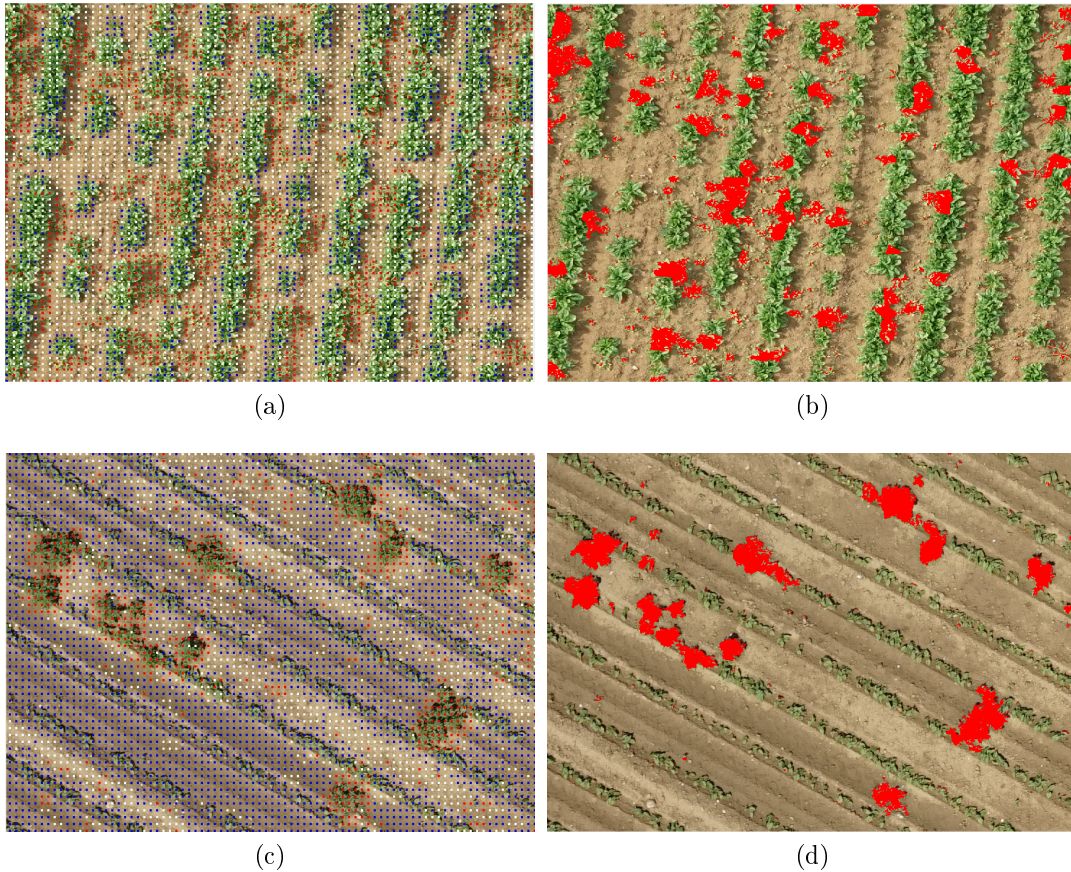


FIGURE 5.12 – Exemples de classification d’images de drones avec des modèles créés par des données non supervisées dans deux champs différents. De la première à la deuxième ligne, nous avons des échantillons de culture d’épinard et de haricot. Sur la première colonne, nous avons les échantillons obtenus après l’utilisation de la fenêtre coulissante, sans les lignes de culture et les informations de fond. Les points bleus, rouges et blancs signifient que les plantes sont identifiées respectivement comme des cultures, des mauvaises herbes et des décisions incertaines. Dans la deuxième colonne, nous avons les mauvaises herbes détectées en rouge après l’application des lignes de culture et des informations de base.

## 5.5 Conclusion

Dans ce chapitre, une nouvelle méthode de détection d'adventices avec un apprentissage entièrement automatique a été proposée. Elle utilise les réseaux neuronaux convolutionnels avec une collecte non supervisée d'ensembles de données d'entraînement. Les résultats obtenus ont montré des performances comparables de celles des données labellisées de manière supervisée. Les différences de AUC sont de 1,5 % dans le champ d'épinard et de 6% dans le champ de haricot. L'étiquetage supervisé est une tâche fastidieuse pour les experts humains ; au vu des écarts d'efficacité entre l'étiquetage supervisé et non supervisé, cette méthode peut être une bonne alternative pour la détection des mauvaises herbes, surtout lorsque les rangées de culture sont espacées. Cette méthode est aussi intéressante en termes de flexibilité et d'adaptabilité car un modèle peut être facilement généré sur un nouvel ensemble de données. Il a également été constaté qu'une architecture de CNN (réseaux de neurones convolutifs) telle que ResNet18 est capable de fournir des résultats pertinents avec la technique d'apprentissage par transfert. Toutefois, pour un grand ensemble de données, il serait intéressant d'exploiter des réseaux jugés plus performants sur la base ImageNet. Cependant, la méthode de collecte de données non supervisée dépend fortement de l'efficacité de la méthode de détection de rangées de culture ainsi que de la présence d'adventices dans l'inter-rang. Dans le chapitre suivant nous allons aborder un autre type de méthode qui permet de ne s'intéresser qu'à la culture en question en ignorant la classe des adventices.



# Chapitre 6

## Apprentissage avec une seule classe

### Sommaire

---

6.1	Procédure de collecte des données . . . . .	88
6.2	Extraction de descripteurs profonds . . . . .	89
6.3	SVM à classe unique ou SVM 1-classe . . . . .	90
6.4	Résultats et discussion . . . . .	90
6.5	Conclusion . . . . .	93

---

Dans le chapitre précédent il a été constaté que la méthode de collecte des données dépend fortement de la présence d'adventices dans l'inter-rang. Dans ce nouveau chapitre, la méthode proposée consiste maintenant à collecter uniquement les plantes situées dans les rangées de culture (la culture) pour former un classifieur à classe unique. Pour la collecte de données, la procédure est presque la même que celle employée dans le Chapitre 5. Le choix des descripteurs étant déterminant pour une classification réussie, une extraction automatique de caractéristiques est effectuée sur un réseau convolutionnel profond déjà entraîné sur la base ImageNet.

Ce chapitre est divisé en 5 sections. Dans la première section la méthode de collecte de données est présentée. La deuxième section décrit l'extraction des descripteurs profonds. La troisième section explique le principe du SVM 1-classe. Dans la quatrième section les résultats sont commentés et enfin la cinquième section conclut le chapitre.

## 6.1 Procédure de collecte des données

Avant toute opération, l'indice de végétation par différence normalisée NDVI [156] (Equation 6.1) avec le seuillage adaptatif de Otsu [127] a été utilisé pour segmenter la végétation et le fond dans les images (Figure 6.1).

$$NDVI = \frac{NIR - R}{NIR + R} \quad (6.1)$$

où NIR et R représentent respectivement l'image dans le proche infrarouge et le rouge.

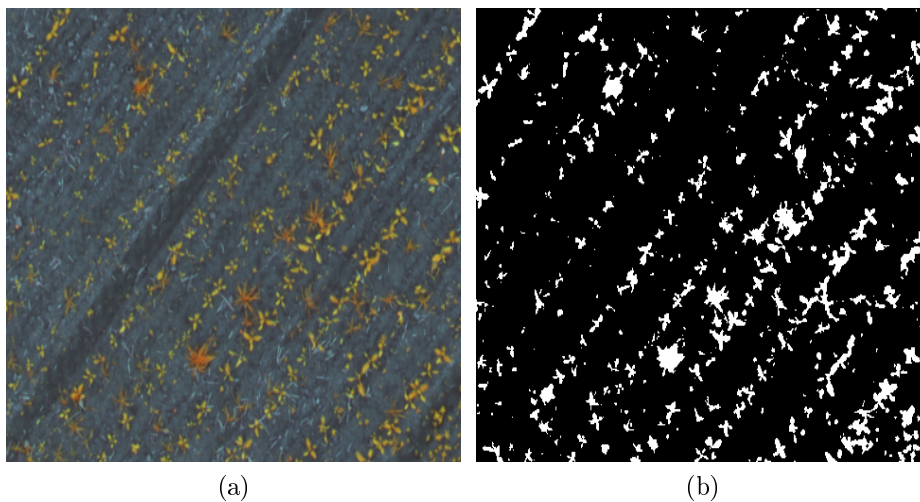


FIGURE 6.1 – (a) Exemple d'une image en fausse couleur (CIR), dont les 3 canaux RGB sont respectivement le proche infrarouge, le vert et le rouge. Sur l'image (b), la végétation est représentée en blanc.

En général, les méthodes utilisées pour collecter les données d'apprentissage appliquent soit des fenêtres glissantes, soit des boîtes englobantes sur des superpixels obtenus, par l'algorithme de simple regroupement linéaire itératif (SLIC). L'utilisation de la fenêtre glissante suppose que nous connaissions la taille de la fenêtre à sélectionner et son contenu

dépend de la position de la fenêtre par rapport à la plante. De plus, le SLIC crée un regroupement local de pixels basé sur les valeurs spectrales des plantes définies par l'espace couleur CIELAB et leur proximité spatiale (section 4.1.3). Mais les environnements agricoles extérieurs sont affectés par des conditions d'éclairage incontrôlées et variables (ombres soudaines, éclairage excessif ou faible) qui affectent la qualité de l'image. Ainsi l'éclairage non uniforme sur les feuilles et l'hétérogénéité de la couleur conduisent le SLIC à diviser une plante en plusieurs objets de végétation (Figure 6.2a).

Cependant, en nous référant à la disposition spatiale des plantes, la végétation peut être segmentée selon l'irrégularité de l'alignement des plantes si nous utilisons la segmentation par la ligne de partage des eaux (*watershed* en anglais) [19]. Watershed est une méthode de segmentation d'images qui considère l'image niveaux de gris comme une surface topographique : plus la valeur d'un pixel est grande, plus le pixel est en hauteur (montagnes, ...) et inversement, les plus petites représentent des dépressions dont la profondeur dépend de la valeur du pixel. En supposant que nous décidions d'inonder le relief en remplissant d'eau tous les creux, les barrières qui seront construites pour empêcher l'eau des creux de se rencontrer représentent les lignes de partage des eaux. Dans notre cas, les pixels blancs représentent les monticules et les noirs les creux. Ainsi les rangées de culture sont segmentées en superpixels en utilisant les creux présents sur les contours.

Sur la Figure 6.2b, nous constatons que les superpixels obtenus contiennent soit une seule plante, soit un groupe de plantes (*culture, culture – adventices ou adventices*) et leurs formes dépendent de la disposition de ces plantes. Les superpixels qui sont en contact avec les rangées de culture détectées sont appelés culture et les autres sont considérés comme étant des adventices. Enfin les boîtes englobantes de tous les superpixels sont extraites pour former les données d'apprentissage.

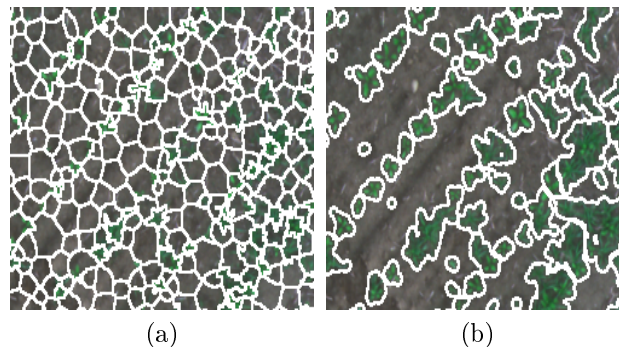


FIGURE 6.2 – (a) et (b) sont respectivement des images avec superpixels (limites blanches) générées avec la méthode SLIC (compacité = 10 et nombre de superpixels = 1000) et la méthode watershed appliquée sur la version dilatée de l'image segmentée avec un élément structurant sphérique de rayon 5.

## 6.2 Extraction de descripteurs profonds

Les descripteurs profonds (*deep features* en anglais) proviennent de l'extraction de caractéristiques de différentes couches d'un réseau neuronal convolutif (CNN) préalablement entraîné [173, 174]. L'extraction de descripteurs profonds est inspirée de l'idée que

les personnes peuvent utiliser intuitivement leur expérience acquise précédemment pour définir et résoudre de nouveaux problèmes. Dans ce travail, les descripteurs ont été extraits à différentes profondeurs du réseau ResNet18 préalablement entraîné sur la base ImageNet. Chaque descripteur a une dimension égale au nombre de filtres présents sur la couche de convolution, où chaque valeur correspond à la moyenne de la sortie de chaque filtre. Cela signifie que les dimensions des caractéristiques F1, F2, F3, F4, F5, F6, F7 et F8 sont respectivement de 64, 64, 128, 128, 256, 256, 512 et 512.

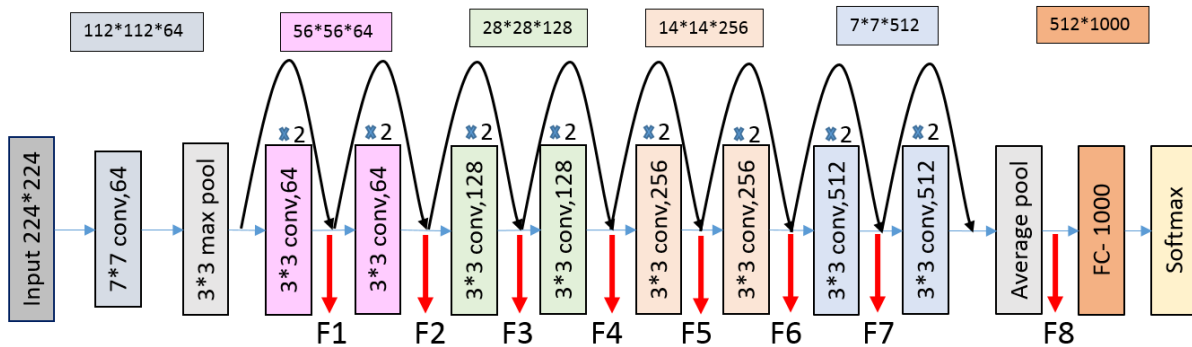


FIGURE 6.3 – Architecture de ResNet18 avec extraction des descripteurs profonds. Les flèches rouges indiquent où les descripteurs ont été extraits. Les rectangles horizontaux indiquent la taille à la sortie des couches.

## 6.3 SVM à classe unique ou SVM 1-classe

Le classifieur SVM à classe unique a été choisi comme machine d'apprentissage. C'est un algorithme qui apprend une fonction de décision pour la détection d'anomalie, classant ainsi les nouvelles données comme similaires ou différentes des données apprises [144, 113]. Conceptuellement, cet algorithme va permettre de construire un hyperplan séparant les adventices des données d'apprentissage majoritairement constituées de culture.

## 6.4 Résultats et discussion

Les expériences ont été menées sur les images de la parcelle de betterave acquises avec un capteur RedEdge-M (Chapitre 3, section 3). Pour chaque image orthomosaïquée, des données ont été collectées de façon supervisée (S000, S001, S002, S003 et S004) et non supervisée (US000, US001, US002, US003 et US004) sur les images en RGB et en fausses couleurs (CIR). Les données supervisées ont été collectées en exploitant la vérité terrain fournie avec les données. Tout comme dans le Chapitre 5, les données non supervisées et supervisées correspondent respectivement à des données étiquetées de manière non supervisée et à celles étiquetées par un expert. De plus, deux nouveaux ensembles de données ont été formés en regroupant les différents échantillons formés à partir des 5 images orthomosaïquées : S0-4 et US0-4 correspondent respectivement aux bases de test et

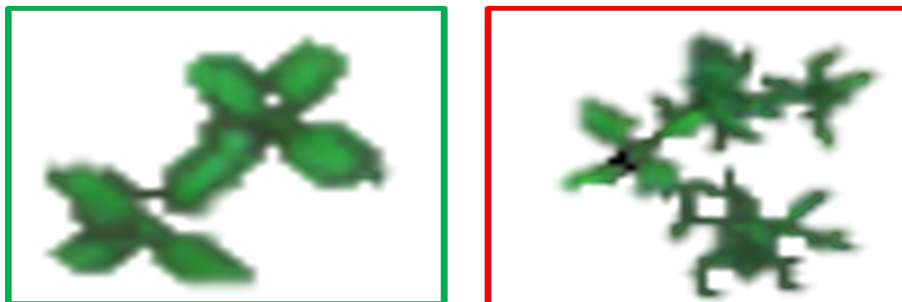


FIGURE 6.4 – Exemple d'échantillons collectés (en RGB). Les rectangles verts et rouges sont respectivement des échantillons de culture et de mauvaises herbes.

TABLE 6.1 – Échantillons prélevés pour l'apprentissage et le test.

Classe	Images		000	001	002	003	004	Total (0-4)
	Supervisées							
Culture	Non		742	1791	1501	2260	680	6974
Culture	Oui		4317	1972	7216	5620	1735	20860
Adventices	Oui		7144	5538	2888	2461	867	18898

d'apprentissage qui regroupent l'ensemble des échantillons collectés de manière supervisée et non supervisée.

Pour collecter les données non supervisées, une segmentation de l'arrière-plan a été effectuée suivie d'une dilatation sur l'image segmentée. Par la suite, des superpixels ont été générés et enfin des boîtes englobantes ont été appliquées pour extraire tous les superpixels.

Pour nous assurer que les plantes de différents superpixels n'apparaissent pas dans la même boîte englobante, un masque a été appliqué autour de chaque superpixel avant la génération de la boîte englobante. Le même processus a été appliqué pour les données supervisées. Seuls les superpixels à l'intérieur des rangs de culture ont été sélectionnés pour construire l'ensemble des données non supervisées. Les échantillons de cultures dont la taille est inférieure à  $25 \times 25$  pixels ont été supprimés.

Pour évaluer la méthode, les échantillons des deux classes ont été sélectionnés de façon supervisée. Ainsi, seuls les échantillons qui ne contiennent pas de pixels d'adventices sont sélectionnés pour la classe culture. Un échantillon est classé comme adventice si les superpixels contiennent au moins 10 % de pixels étiquetés comme adventice puisque, pour les agriculteurs, comme indiqué précédemment, il est préférable d'avoir une surdétection que de manquer des mauvaises herbes. De plus, en raison de la résolution spatiale des images de drone, il est parfois préférable de se concentrer sur une agrégation de plantes pour évaluer l'infestation par les adventices plutôt que sur une seule plante, même pour un expert. Le nombre total d'échantillons collectés est présenté dans la Table 6.1.

Une fois les données collectées, les huit descripteurs profonds ont été extraits du réseau pré-entraîné ResNet18 (section 6.2) puis combinés pour obtenir un neuvième descripteur (F1-8), vecteur de caractéristiques de 1920 éléments. La méthode d'analyse en composantes principales (ACP) [165, 125] a été utilisée pour réduire la taille de F1-8 à 200 (PCA200). Les composantes principales retenues représentent alors 98% de l'information.

En résumé, le nombre de descripteurs ou *features vector* (Nfeatures) est de 10 et le

nombre de données d'apprentissage ( $N_{\text{train}}$ ) et de test ( $N_{\text{test}}$ ) est de 6.

Pour la classification, le SVM à classe unique a été utilisé avec un noyau gaussien puisque comparé, au noyau linéaire, il avait fourni un meilleur résultat dans la section 5.4.1. Lors de l'apprentissage, le classifieur SVM à classe unique trouve les limites séparant les valeurs aberrantes du reste des données d'apprentissage selon un taux de valeurs aberrantes fixé ( $\mu$ ). Les données d'apprentissage étant collectées de façon non supervisée, une valeur de  $\mu = 10\%$  a été choisie en raison du risque d'apparition de mauvaises herbes dans les données non supervisées. Ainsi, pour chaque données d'apprentissage et chaque descripteur, un modèle a été entraîné, ce qui nous donne un nombre total de modèles égal  $N_{\text{Features}} \times N_{\text{train}}$ .

Afin d'évaluer chaque descripteur et chaque modèle, l'AUC a été calculée à partir de la courbe ROC. Les descripteurs profonds ont été comparés en construisant un modèle avec chaque descripteur sur l'ensemble de données US0-4 et en effectuant le test sur l'ensemble de données S0-4. Dans la Figure 6.5, nous pouvons remarquer que F2 présente les meilleurs résultats et inversement, F8 et PCA200 présentent les résultats les plus faibles. De plus, les descripteurs les moins profonds sont plus efficaces. En nous basant sur le fonctionnement connu des CNN, nous pouvons affirmer que la forme et la texture semblent plus appropriées pour la détection des mauvaises herbes dans ce champ de betterave. Nous notons également que les jeux de données au format CIR permettent d'obtenir de meilleurs résultats que ceux en RGB. Ainsi, dans les parties suivantes, seuls les jeux de données CIR ont été utilisés pour l'entraînement et le test.

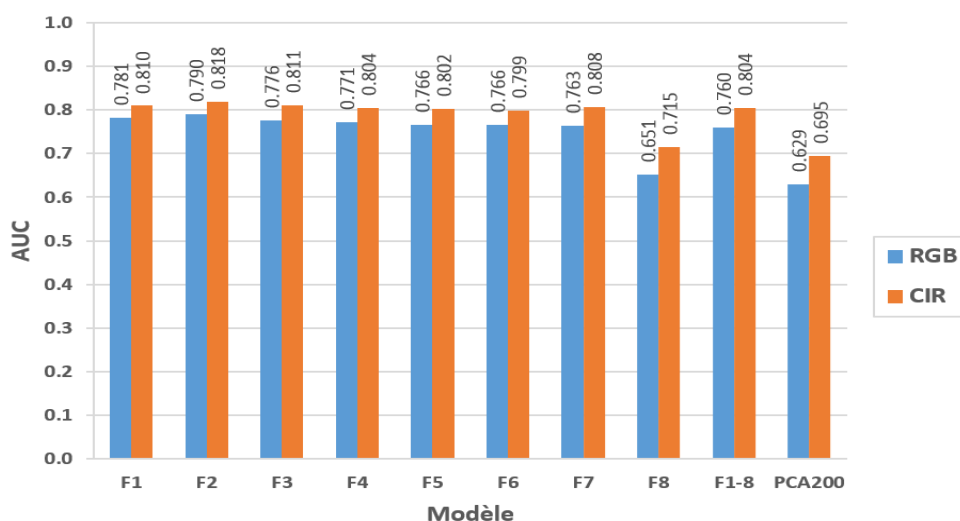


FIGURE 6.5 – AUC obtenue avec chaque descripteur en utilisant US0-4 comme base d'apprentissage et S0-4 comme base de test. RGB et CIR signifient que le modèle a été entraîné sur des données en vraies et fausses couleurs, respectivement.

F2 ayant fourni le meilleur résultat, tous les modèles entraînés avec celui-ci ont été évalués sur les six bases de test (S000, S001, S002, S003, S004 et S0-4). La Table 6.2 présente les résultats. Dans cette table, les Model0000 à Model004 sont les modèles entraînés par SVM 1-classe sur les 5 bases de données non supervisées, et le Model0-4 est le modèle entraîné sur la fusion des 5 bases de données d'apprentissage. ResNet18-000 est, quant à lui, le modèle créé avec la base de données S000. Nous pouvons observer sur la Table 6.2

TABLE 6.2 – AUC des modèles obtenus avec SVM 1-classe et ResNet18. Les Model000 à Model004 sont les modèles entraînés par SVM 1-classe sur les 5 bases de données non supervisées. Le Model0-4 est le modèle entraîné sur la fusion des 5 bases de données d'apprentissage. ResNet18-000 est le modèle créé avec la base de données S000.

Modèles	AUC sur l'ensemble de données supervisées					
	S000	S001	S002	S003	S004	S0-4
Model000	<b>0,812</b>	0,810	0,680	0,721	0,684	0,719
Model001	0,703	<b>0,757</b>	0,619	0,665	0,647	0,609
Model002	0,914	0,869	<b>0,850</b>	0,850	0,776	0,885
Model003	0,902	0,868	0,826	<b>0,848</b>	0,763	0,873
Model004	0,827	0,781	0,744	0,738	<b>0,715</b>	0,805
Model0-4	0,835	0,820	0,778	0,802	0,738	0,810
ResNet18-000	-	0,929	0,926	0,929	0,873	0,957

que les modèles entraînés et évalués dans la même parcelle donnent des résultats supérieurs à 80 %, à l'exception des Model001 et Model004. De plus, les modèles Model002 et Model003 sont les plus efficaces, avec des AUC supérieures à 0,90 sur la base de données de test S000. Sur S0-4, leurs AUC sont respectivement de 0,885 et 0,873. Le Model001 a donné le résultat le plus faible et son meilleur résultat est obtenu sur S001 (AUC = 0,757). Ces résultats s'expliquent par le fait que les données d'apprentissage US002 et US003 ont été collectées dans les deux images orthomosaiquées les moins infestées en adventices et inversement pour les données US000 et US001. Ceci dit, pour un apprentissage à classe unique, il est préférable d'entraîner le modèle sur des zones du champ considérées moins infestées pour ensuite le généraliser pour la détection d'anomalie. L'image 004 est moins infestée que US000 et US001 mais son résultat peut être justifié par la présence de flou sur sa partie inférieure.

Par la suite, les résultats sont comparés avec ceux obtenus avec le réseau ResNet18 entraîné avec des données supervisées. La base de données S000 a été utilisée pour l'entraînement du réseau (ResNet18-000). Cette base a été choisie parce qu'elle est plus équilibrée en termes de nombre d'échantillons issus des deux classes. ResNet18 a été entraîné avec 100 epochs, une taille de mini-lot de 10 et un pas d'apprentissage de 0,001. Les résultats sont présentés dans la Table 6.2. ResNet18-000 a fourni un meilleur résultat que les modèles entraînés avec SVM 1-classe. Son AUC sur S0-4 est de 95,7 %. Ce résultat est de 7 à 8 % supérieur à ceux obtenus avec Model002 et Model003.

## 6.5 Conclusion

Dans ce chapitre, nous avons proposé d'utiliser des caractéristiques extraites des réseaux profonds "deep features" et un classifieur SVM à classe unique (1-classe) sur des données non supervisées pour la détection des adventices dans des images acquises par drone. Les résultats ont montré qu'avec des jeux de données non supervisés collectés dans une image orthomosaiquée qu'il est possible de détecter des anomalies (mauvaises herbes) dans la même image avec un AUC allant de 0,715 à 0,85. L'AUC calculée sur tous les jeux de données de test se situe entre 0.609 et 0.885. Ces résultats démontrent qu'en étudiant

une partie d'un champ avec peu d'infestation en adventices il est possible d'avoir une efficacité sur l'ensemble du champ allant jusqu'à 88.5 %. Nous avons également remarqué que les caractéristiques extraites de la couche moins profonde du ResNet18 conviennent à la classification en une seule classe de la culture et des mauvaises herbes dans le champ de betterave sucrière. Les résultats obtenus se sont montrés moins efficaces qu'un modèle entraîné par le réseau de deep learning ResNet18 sur des données supervisées (ResNet18-000). Cependant, l'étiquetage supervisé est une tâche coûteuse pour les experts humains, et selon les résultats obtenus, la classification à une classe peut être un bon compromis pour la détection des mauvaises herbes, surtout dans les champs avec peu ou pas de mauvaises herbes entre les rangs.

# Conclusion et perspectives

Les mauvaises herbes sont responsables de la plupart des pertes de rendement agricole. Pour faire face à cette menace, les agriculteurs ont recours à la pulvérisation de pesticides sur l'ensemble de leurs cultures. Cette méthode nécessite non seulement d'énormes quantités d'herbicides mais elle a aussi des répercussions non négligeables sur l'environnement et la santé humaine. Une façon de réduire les coûts et l'impact sur l'environnement est alors d'allouer les bonnes doses d'herbicides au bon endroit et au bon moment ou d'intervenir par des actions mécanisées et ciblées. De nos jours, le drone est devenu un système d'acquisition intéressant pour la localisation et la gestion des mauvaises herbes de par sa capacité à acquérir des images de l'ensemble de la parcelle agricole avec une très haute résolution spatiale et à faible coût. Malgré les progrès importants réalisés dans les systèmes d'acquisition par drone, la détection automatique des mauvaises herbes dans les images demeure un problème difficile à résoudre en raison de la forte ressemblance entre les cultures et les mauvaises herbes. L'objectif de cette thèse est de proposer des méthodes de traitement d'images pour la localisation des adventices en grandes cultures. Elle a été effectuée dans le cadre d'un projet de recherche financé par la Région Centre-Val de Loire.

Au début de cette thèse tout comme au départ de la plupart des exercices scientifiques, un tour d'horizon des articles qui s'intéressent à la détection des adventices a été effectué. Il en est ressorti de cette phase bibliographique que la détection d'adventices dans des images de drone est généralement réalisée en trois étapes. La première étape consiste à segmenter l'image en deux classes, une classe végétation et une classe non végétation ; cette phase est primordiale car une erreur de segmentation importante peut perturber le reste de la chaîne de traitement. Dans l'agriculture moderne, les cultures sont semées en rang et souvent avec une distance inter-rang identique. Cette information est exploitée dans la deuxième phase pour détecter les rangées de culture. Une fois les rangées de culture détectées, toute végétation qui se situe dans l'inter-rang est considérée comme adventice dans la troisième phase. Parmi ces trois étapes, les grands défis se situent sur les deux dernières. En effet, pour séparer efficacement la végétation du sol, des études ont montré qu'un seuillage adaptatif sur des indices tels que l'ExG (pour les images couleur) et le NDVI (pour les images multispectrales) suffisait.

Ainsi, en se focalisant sur les ces deux dernières étapes, nous avons proposé une nouvelle méthode de détection des rangées de culture appelée CR-Hough-SLIC. Cette méthode applique dans un premier temps la transformée de Hough sur le squelette des l'image segmentée. Les squelettes des rangées ont été utilisés car nous avons constaté qu'ils faisaient ressortir la structure globale du champ, à savoir l'orientation, la périodicité et le nombre de rangées avec peu de pixels. En plus de l'utilisation de la transformée de Hough, une segmentation par superpixels a été employée pour identifier automatiquement les ran-

gées de culture sans connaître préalablement la distance inter-rang. Généralement, les méthodes de détection des rangées de culture se répartissent en quelques catégories selon leur principe de détection, comme la transformée de Hough, la régression linéaire, l'analyse de taches (blob), la vision stéréoscopique ou les bandes horizontales. Mais ces méthodes dépendent non seulement du prétraitement appliqué pour l'extraction de la végétation mais également des paramètres qui leur sont propres. C'est la raison pour laquelle une nouvelle méthode de détection des rangées de culture utilisant le deep learning a été proposée (CRowNet). Cette méthode s'est montrée plus robuste face à des méthodes de la littérature utilisant la transformée de Hough. De plus, contrairement à la majorité des méthodes rencontrées, l'approche proposée n'exige pas de connaître les caractéristiques des rangées de culture (distance inter-rang, largeur des rangées de culture et orientation).

Une fois les rangées détectées, toute région connexe de végétation avec au moins un pixel appartenant à ces rangées est considérée comme culture, mauvaises herbes dans le cas contraire : c'est la méthode de détection des adventices inter-rang. Cette méthode ne permet cependant pas d'identifier les adventices intra-rang.

A défaut de pouvoir détecter les adventices situées à l'intérieur des rangées de culture, nous avons alors proposé d'utiliser une méthode de détection des adventices *potentielles*, c'est-à-dire des régions de végétation qui sont en contact avec les rangées de culture détectées et se distinguent beaucoup plus dans l'inter-rang par rapport à leurs voisines. Nous sommes partis de l'hypothèse que la présence d'une adventice à une position  $p$  donnée, à l'intérieur d'une rangée de culture, va entraîner une perturbation de l'alignement des plantes à cette position. Cette méthode a ainsi démarré par la création de superpixels sur l'image segmentée puis pour l'étude des relations spatiales entre les superpixels en contact avec les rangées détectées pour enfin détecter les adventices potentielles. La méthode s'est montrée utile pour la détection des adventices très proches des rangées de culture, avec une certaine surdétection.

Pour améliorer la détection des adventices et gérer de manière simultanée les adventices intra-rang et inter-rang, nous avons exploré deux approches d'apprentissage : une approche qui s'appuie sur les méthodes d'apprentissage machine classiques (sélection manuelle de descripteurs et de classifieurs) et une approche de deep learning qui utilise le CNN (Convolutional Neural Network).

Pour ce qui est de la première approche d'apprentissage, nous savons que les indices de couleur sont utiles pour faire la distinction entre végétation et arrière-plan (sol, cailloux, ...), mais ils deviennent moins efficaces lorsqu'ils sont utilisés pour classer les espèces de plante. En effet, la couleur des mauvaises herbes et des feuilles des cultures est parfois presque la même. De plus, le résultat n'est plus fiable dans des conditions d'éclairage différentes. Pour résoudre ce problème, plusieurs caractéristiques de l'image ont été analysées : des caractéristiques statistiques, de forme et de texture. Une procédure de sélection des caractéristiques a été utilisée pour analyser les caractéristiques les plus appropriées à notre problématique. Les algorithmes de classification utilisés sont le SVM et le Random Forest. La deuxième approche d'apprentissage utilise le deep learning apparu récemment qui a donné des résultats impressionnants dans différents problèmes de classification complexes. Cette approche a été naturellement exploitée, dans notre cas, en développant de nouvelles méthodes basées sur les CNN et plus particulièrement les réseaux résiduels (ResNet).

Mais pour fonctionner correctement, ces deux approches d'apprentissage supervisé

ont besoin de beaucoup de données d'apprentissage. Et pour une bonne optimisation des modèles d'apprentissage profond, il est nécessaire d'avoir un grand nombre de données étiquetées. Mais la création de grands ensembles de données agricoles comportant des annotations au niveau des pixels est une tâche qui prend énormément de temps. Compte tenu de ce problème de collecte de données labellisées, une nouvelle méthode de création automatique de base d'apprentissage a été développée. Cette méthode utilise les adventices inter-rang et potentielles pour former la base d'apprentissage. Pour évaluer cette base d'apprentissage, des modèles ont été générés en utilisant les méthodes de machine learning classiques et le deep learning (ResNet18).

Il en est ressorti que les résultats obtenus avec les données étiquetées de façon automatique et manuelle sont comparables. De plus, il a été constaté que même si nous parvenons à sélectionner les caractéristiques les plus appropriées pour identifier les mauvaises herbes dans un champ, il est peu probable que ces caractéristiques soient adaptées à un autre champ avec un type de culture différent. Il a été également remarqué que les caractéristiques considérées comme meilleures par un classifieur ne sont pas nécessairement les meilleures pour un autre. Un autre constat établit aussi que, dans les champs, d'une année à l'autre, il est possible de trouver de nouvelles espèces de mauvaises herbes et le niveau de croissance des plantes peut parfois entraîner une confusion entre les mauvaises herbes et les cultures, ce qui doit conduire à une nouvelle collecte de données sur les mauvaises herbes/cultures et à une nouvelle sélection de caractéristiques. Ainsi, pour une classification efficace, il serait intéressant d'utiliser un outil capable de générer automatiquement des échantillons et des caractéristiques pertinentes pour détecter les mauvaises herbes, d'où l'intérêt d'utiliser un apprentissage profond avec un étiquetage non supervisé.

Cependant si peu d'adventices inter-rang ou d'adventices potentielles sont détectées, le réseau du deep learning devient difficilement utilisable à cause d'un déséquilibre énorme dans la base de données d'apprentissage. Nous avons alors proposé une dernière méthode qui consiste à collecter de façon non supervisée uniquement des échantillons de plantes contenus dans les rangées de culture. A l'aide d'un réseau d'apprentissage profond, les caractéristiques de ces échantillons sont ensuite extraites puis placées en entrée d'un classifieur à classe unique (*one-class classifier*). Le modèle ainsi généré à partir de ce classifieur a été utilisé pour la détection des mauvaises herbes sur l'ensemble des images. Les résultats ont montré qu'avec des jeux de données non supervisés collectés dans une image orthomosaïquée, nous sommes capables, avec un SVM à classe unique, de détecter des anomalies (mauvaises herbes) dans la même image avec un AUC allant de 0.715 à 0,85 en fonction du taux d'infestation en adventice, et de la qualité de l'image. Les résultats ont été comparés avec ceux obtenus avec le ResNet18 sur les données des deux classes étiquetées de façon supervisée. La différence obtenue est de 10 % pour les images très infestées dans l'inter-rang et moins de 7% pour celles faiblement infestées.

Dans les travaux futurs, il serait intéressant de renforcer la phase de segmentation des plantes lors de la détection des rangées de culture avec la méthode CR-Hough-SLIC car elle en dépend fortement. Parfois à cause de la topologie du sol les rangées de culture ne sont pas linéaires, c'est pourquoi on souhaite dans les perspectives améliorer la robustesse des méthodes de détection des rangées de culture. De plus, même si le CRowNet a prouvé son efficacité sur les deux bases de données employées pour le test, il serait souhaitable,

dans l'avenir, de simplifier sa structure globale en proposant une architecture de réseau convolutionnel simple et plus robuste, combinant les phases d'extraction de végétation et de localisation précise des rangées de culture. Certes, la sélection automatique des échantillons d'adventices dans l'inter-rang et de la culture dans l'intra-rang ont permis d'obtenir des résultats comparables à ceux obtenus avec des données étiquetées manuellement, mais pour renforcer cette approche, nous proposons d'exploiter les récentes avancées du deep learning sur la génération de données artificielles. Cela permettra de renforcer la base d'apprentissage pour une détection plus robuste et toujours automatique des adventices. Étant données que les doses d'herbicide utilisées sont en corrélation avec le type d'adventices, il serait utile de les reconnaître et de les catégoriser après détection.

Dans les perspectives, nous pensons également à des études qui permettront d'évaluer non seulement l'impact du niveau de croissance des plantes sur les performances des méthodes développées, mais aussi l'impact des erreurs de détections de rangées de culture dans la formation de modèle à partir de données collectées de façon non supervisée.

L'ensemble des méthodes seront intégrées dans un outil de traitement rapide et doté d'une interface adaptée aux utilisateurs. Ainsi des cartes d'infestations géoréférencées seront générées en fonction des besoins : désherbage par zone ou par plante. Le désherbage par zone fait allusion à l'utilisation des tracteurs pour pulvériser par zonage en fonction du taux d'infestation en adventices tout en réduisant au mieux les fréquences de fermeture et d'ouverture des vannes de pulvérisation. Le désherbage par plante correspond à la carte obtenue en sortie des méthodes de traitement d'images, qui peut être utilisée par des robots désherbeurs utilisant des actions mécanisées ciblées.

# Publications

## Réviews

BAH, M. Dian, HAFIANE, Adel, et CANALS, Raphael. CRowNet : Deep network for Crop row detection in UAV images. *IEEE Access*, 2019, vol. 8, p. 5189-5200.

BAH, M. Dian, HAFIANE, Adel, et CANALS, Raphael. Deep learning with unsupervised data labeling for weed detection in line crops in UAV images. *Remote sensing*, 2018, vol. 10, no 11, p. 1690.

## Conférences

BAH, M. Dian, HAFIANE, Adel, CANALS, Raphael, et EMILE, Bruno. Deep features and One-class classification with unsupervised data for weed detection in UAV images. In : 2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA). IEEE, 2019. p. 1-5.

BAH, M. Dian, DERICQUEBOURG, Eric, HAFIANE, Adel, et al. Deep learning based classification system for identifying weeds using high-resolution UAV imagery. In : Science and Information Conference. Springer, Cham, 2018. p. 176-187.

BAH, M. Dian, HAFIANE, Adel, et CANALS, Raphael. Weeds detection in UAV imagery using SLIC and the hough transform. In : 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA). IEEE, 2017. p. 1-6.



# Annexe A

## Fonctionnement des réseaux de neurones artificiels

Un réseau de neurones est une association en graphe plus ou moins complexe d'objets élémentaires (les neurones formels). Le réseau se distingue par l'organisation du graphe, c'est à dire nombre de neurones, présence ou non de boucle de retro-action. Pour les réseaux à une couche, les neurones sont tous connectés aux entrées tandis que pour les réseaux à plusieurs couches (perceptron multicouches), certains neurones sont connectés aux sorties d'autres neurones.

### A.1 Neurones formels

Comme pour un neurone biologique, un neurone formel est caractérisé par un état interne  $Y_{sortie}$ , des signaux d'entrée  $X_1, X_2, \dots, X_n$  et une fonction d'activation  $\varphi$ . Pour un neurone  $J$  la sortie  $\hat{Y}$  est représentée par l'équation A.1.

$$\hat{Y} = \varphi_j(Z_j) \tag{A.1}$$

avec  $Z_j = b_j + \sum_{k=1}^n x_k w_{jk}$

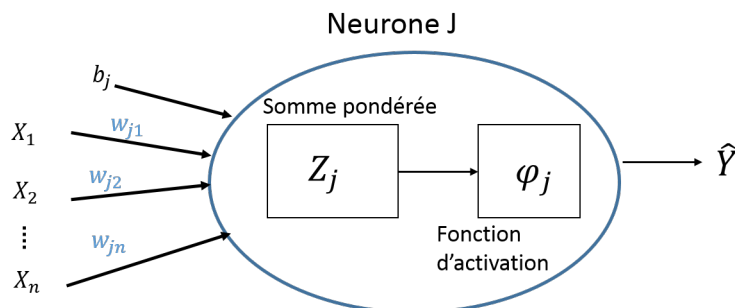


FIGURE A.1 – Exemple de neurones ou de perceptron à une couche cachée.

La fonction d'activation opère une transformation d'une combinaison affine des signaux d'entrée. Chaque neurone dispose d'un vecteur poids  $[w_{j1}, w_{j2}, \dots, w_{jn}]$  dont les valeurs sont

estimées dans la phase d'apprentissage. Ils constituent "la mémoire" ou "connaissances reparties" du réseau. Les types de neurones se distinguent par la nature  $\varphi$  de leur fonction d'activation (sigmoïde, linéaire, seuil, ...). Mais les plus utilisées restent le modèle linéaire (*ReLU* ou *rectified linear unit* en anglais) et la fonction sigmoïde. Ces deux fonctions doivent leurs succès à leurs capacités d'adaptation aux algorithmes d'apprentissage impliquant une retropropagation du gradient grâce à leurs fonctions d'activation qui sont différentiables.

## A.2 Perceptron multicouche (MLP)

Un perceptron multicouche est composé de plusieurs couches de discriminant linéaires (neurones) successives dont chaque couche permet de projeter les données dans un nouvel espace. Contrairement au réseau à une couche (perceptron), le perceptron multicouche permet de classer des données qu'elles soient linéairement ou non linéairement séparables.

Le MLP est composé de 3 types de couches de neurones (Figure A.3) : la couche d'entrée, les couches intermédiaires ou cachées et la couche de sortie.

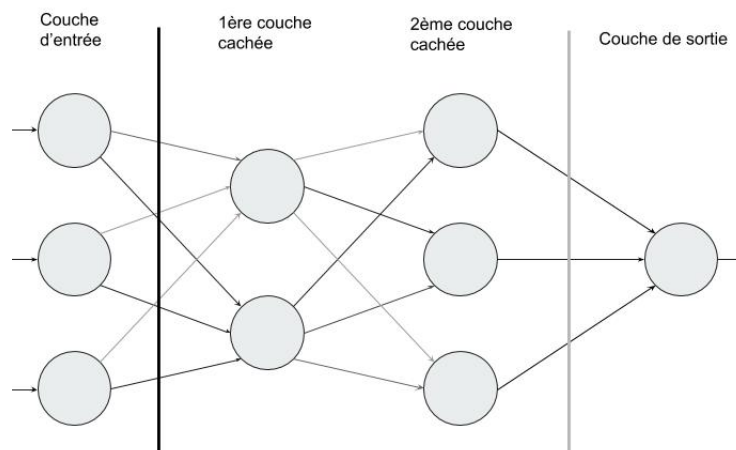


FIGURE A.2 – Perceptron multicouche à deux couches cachées et une couche de sortie.

- Une couche d'entrée qui reçoit le signal d'entrée
- Une couche de sortie fournit la réponse du réseau
- Une couche intermédiaire est composée de neurones connectés en entrée à chacune des couches précédentes et en sortie à chaque neurone de la couche suivante.

Plus les couches intermédiaires sont nombreuses plus le réseau est complexe et difficile à entraîner, cependant, il peut être capable de fournir un résultat robuste, il faut toutefois faire attention aux problèmes de sur-apprentissage (*overfitting*).

En fonction de la topologie des couches, différentes frontières de décisions peuvent être obtenues (Figure A.2), ceci lui a valu le nom d'approximateur universel [37]. Un réseau avec une seule couche cachée possède des frontières convexes et un réseau avec deux ou plusieurs couches possède des frontières concaves.

Le MLP est souvent utilisé dans des problèmes d'apprentissage supervisé. Le principe est d'apprendre les poids des classificateurs (neurones) pour rendre les problèmes linéairement séparables.

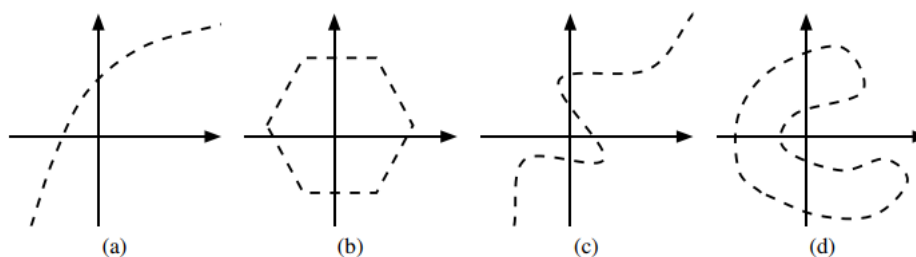


FIGURE A.3 – Exemples de frontières de décision : (a) convexe ouverte ; (b) convexe fermée ; (c) concave ouverte ; et (d) concave fermée.

Lors de l'apprentissage dans un premier temps, les entrées sont propagées vers l'avant jusqu'à obtenir une sortie calculée par le réseau. La seconde étape compare la sortie calculée à la sortie réelle connue ou cible. On modifie alors les poids de telle sorte qu'à la prochaine itération, l'erreur commise entre la sortie calculée et connue soit minimisée. On rétro-propage alors l'erreur commise vers l'arrière jusqu'à la couche d'entrée tout en ajustant les poids des différents neurones. Le processus est répété jusqu'à obtenir une erreur de sortie considérée comme négligeable (convergence). L'algorithme d'optimisation le plus simple est la descente de gradient, dont le principe est de partir d'un point aléatoire puis de se déplacer dans la direction de la plus forte pente.

### A.2.1 Descente du gradient

La descente de gradient est un algorithme d'optimisation itératif du premier ordre permettant de rechercher le minimum d'une fonction. Par définition, le gradient donne la direction (locale) de la plus grande pente. Pour trouver un minimum local par cette méthode, on se dirige vers la direction opposée au gradient.

Il existe deux façons de mettre à jour les poids :

- La première fait des mises à jour à chaque fois qu'un échantillon d'apprentissage est utilisé, cette méthode est appelée descente de gradient stochastique. Dans le cas de grosse quantité de données une mise à jour par lot ou *batch* en anglais est utilisée. Le but de cette dernière est d'appliquer la mise à jour après chaque  $N$  échantillons. Le cas stochastique est un cas particulier avec  $N=1$  ;
- La deuxième attend d'avoir évalué l'ensemble des échantillons pour faire une mise à jour à partir de la moyenne des pertes.

La première est considérée plus efficace lorsqu'on a un grand ensemble de données d'entraînement.

En prenant comme exemple le réseau de la figure A.1, faire un apprentissage avec ce réseau consiste à apprendre les poids du neurone afin que la sortie prédite ( $\hat{Y}$ ) soit proche de celle visée ( $Y$ ) pour les entrées  $X_1, X_2, \dots, X_n$ .

Une fonction de perte  $\xi$  est calculée pour évaluer dans quelle mesure nous sommes proches ou non de l'objectif (l'erreur). En utilisant l'erreur quadratique moyenne la perte est exprimée par l'équation A.2.

$$\xi(Y, \hat{Y}) = \frac{1}{2}(Y - \hat{Y})^2 \quad (\text{A.2})$$

## A.2. PERCEPTRON MULTICOUCHE (MLP)

La propagation avant de ce réseau correspond au résultat de l'équation A.2. A l'aide de la règle de retropropagation du gradient les poids sont mis à jour. Ainsi la dérivée de la perte est calculée dans le but d'ajuster chaque poids en fonction de son influence sur la sortie. La mise à jour est effectuée par la règle présentée sur les equations A.3 et A.4.

$$w_{ji} = w_{ji} - \alpha \frac{d\xi}{dw_{ji}} \quad (\text{A.3})$$

$$b_j = b_j - \alpha \frac{d\xi}{db_j} \quad (\text{A.4})$$

$\alpha$  est une constante positive appelée taux d'apprentissage. Elle permet d'ajuster empiriquement les poids. Si  $\alpha$  est trop grande, l'algorithme n'est pas stable et oscille autour d'une solution, et si  $\alpha$  est trop petite, un très grand nombre d'itérations sera nécessaire pour converger vers la solution. En fonction du signe du gradient, chaque poids est soit augmenté ou diminué. Les processus de propagation avant et de retropropagation sont répétés jusqu'à la convergence de l'erreur d'apprentissage.

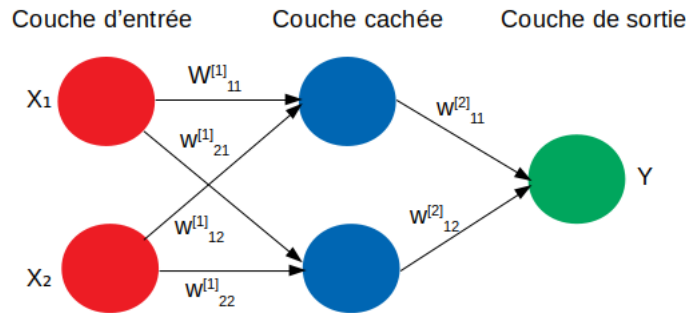


FIGURE A.4 – Exemple d'un réseau de neurones avec 2 entrées, 1 couche cachée composée de 2 neurones et 1 couche de sortie à 1 neurone.

Maintenant considérons le réseau de neurone de la figure A.4. Nous laisserons  $n_c$  indiquer le nombre de couches dans notre réseau ; ainsi  $n_c = 3$  dans notre exemple. Nous marquons la couche  $c$  comme  $C_c$ , donc la couche  $C_1$  est la couche d'entrée, et la couche  $C_{n_c}$  la couche de sortie. Notre réseau neuronal a des paramètres  $(W, b) = (W^{[1]}, b^{[1]}, W^{[2]}, b^{[2]})$ , où nous écrivons  $W_{ji}^{[c]}$  pour désigner le paramètre (ou poids) associé à la connexion entre le neurone  $i$  dans la couche  $c$  et le neurone  $j$  dans la couche  $c + 1$ . De plus,  $b_j^{[c]}$  est le biais associé au neurone  $j$  de la couche  $c + 1$ . Ainsi, dans notre exemple, nous avons  $W^{[1]} \in \mathbb{R}^{2 \times 2}$ , et  $W^{[2]} \in \mathbb{R}^{1 \times 2}$ . Notez que les biais n'ont pas d'entrées ou de connexions qui y entrent. De plus, chaque sortie de neurone  $j$  dans la couche  $c$  sera noté  $a_j^{[c]}$  et soit également  $z_j^{[c]}$  la somme pondérée totale des entrées du neurone  $j$  de la couche  $c$ , y compris le terme de biais.

### Propagation avant

Pour la propagation avant

$$Z_1^{[2]} = W_{11}^{[1]} X_1 + W_{12}^{[1]} X_2 + b_1^{[1]} \Rightarrow a_1^{[2]} = \varphi_1^{[2]}(Z_1^{[2]}) \quad (\text{A.5})$$

$$Z_2^{[2]} = W_{21}^{[1]}X_1 + W_{22}^{[1]}X_2 + b_2^{[1]} \Rightarrow a_2^{[2]} = \varphi_2^{[2]}(Z_2^{[2]}) \quad (\text{A.6})$$

$$Z_1^{[3]} = W_{11}^{[3]}a_1^{[2]} + W_{12}^{[3]}a_2^{[2]} + b_1^{[2]} \Rightarrow \widehat{Y} = a_1^{[3]} = \varphi_1^{[3]}(Z_1^{[3]}) \quad (\text{A.7})$$

De façon plus compacte les équations A.5-A.7 s'écrivent :

$$Z^{[2]} = W^{[1]}X + b^{[2]} \Rightarrow a_1^{[2]} = \varphi^{[2]}(Z^{[2]}) \quad (\text{A.8})$$

$$Z^{[3]} = W^{[2]}a^{[2]} + b^{[2]} \Rightarrow \widehat{Y} = a^{[3]} = \varphi^{[3]}(Z^{[3]}) \quad (\text{A.9})$$

Pour une formulation généralisée les entrées du réseau sont considérées comme les sorties des neurones de la couche  $c=1$  et peuvent donc être représentée par  $a^{[1]}$ . En assumant que les neurones ont la même fonction d'activation  $\varphi$  on obtient :

$$Z^{[c+1]} = W^{[c]}a^{[c]} + b^{[c]} \Rightarrow \widehat{Y} = a^{[c+1]} = \varphi(Z^{[c+1]}) \quad (\text{A.10})$$

## Retropropagation

En supposant que l'ensemble d'apprentissage utilisé est composé de  $m$  observations (échantillons d'apprentissage)  $X^{[k]}$  dont chaque échantillon est accompagné par sa vérité terrain correspondante  $Y^{[k]}$  avec  $k$  allant de 1 à  $m$ . Lors de l'apprentissage l'ajustement des poids se fait par descente de gradient. Pour une observation  $X$  et sa sortie  $Y$ , la fonction de coût est définie par l'équation A.11.

$$\xi(Y, \widehat{Y}) = \frac{1}{2}(\widehat{Y} - Y)^2 \quad (\text{A.11})$$

Une fois la propagation avant effectuée, ensuite, pour chaque neurone  $j$  de la couche  $c$ , la perte est calculée pour mesurer de combien ce neurone influence l'erreur global du réseau. Pour un neurone de la couche de sortie, on peut mesurer directement la différence entre l'activation du réseau et la valeur de la vérité terrain. En détail, voici l'algorithme de rétropropagation :

Calculons d'abord l'impact des poids de la couche 2 ( $W^{[2]}$ ) sur l'erreur  $\xi(Y, \widehat{Y})$ .

$$\frac{\delta \xi(Y, \widehat{Y})}{\delta W^{[2]}} = \frac{\delta \xi}{\delta a^{[3]}} \frac{\delta a^{[3]}}{\delta W^{[2]}} = \frac{\delta \xi}{\delta a^{[3]}} \frac{\delta a^{[3]}}{\delta Z^{[3]}} \frac{\delta Z^{[3]}}{\delta W^{[2]}} = \frac{\delta \xi}{\delta Z^{[3]}} \frac{\delta Z^{[3]}}{\delta W^{[2]}} \quad (\text{A.12})$$

L'équation A.12 est obtenue par utilisation du théorème de dérivation des fonctions composées. On note aussi dans cette équation il est important d'avoir une fonction d'activation dérivable ( $\frac{\delta a^{[3]}}{\delta Z^{[3]}}$ ).

$$\frac{\delta Z^{[3]}}{\delta W^{[2]}} = \frac{\delta(W^{[2]}a^{[2]} + b^{[2]})}{\delta W^{[2]}} = a^{[2]} \quad (\text{A.13})$$

Alors :

$$\frac{\delta\xi(Y, \widehat{Y})}{\delta W^{[2]}} = a^{[2]} \frac{\delta\xi}{\delta Z^{[3]}} \quad (\text{A.14})$$

Calculer les dérivées partielles peut être ardu si nous avons une multitude de couche dans le réseau. Ce calcul est facilité en utilisant la règle de dérivation en chaîne. La dérivation en chaîne consiste à utiliser le fait que la dérivée pour un neurone à la couche  $c$  peut être calculée à partir de la dérivée des neurones connectés à la couche  $c + 1$ .

$$\frac{\delta\xi(Y, \widehat{Y})}{\delta W^{[c]}} = a^{[c]} \frac{\delta\xi}{\delta Z^{[c+1]}} \quad (\text{A.15})$$

L'impact du poids  $i$  au neurone  $j$  à la couche  $c$  et des biais  $b$  des neurones  $j$  est calculé.

$$\frac{\delta\xi(Y, \widehat{Y})}{\delta W_{ji}^{[c]}} = a_i^{[c]} \frac{\delta\xi}{\delta Z_j^{[c+1]}} \quad (\text{A.16})$$

$$\frac{\delta\xi(Y, \widehat{Y})}{\delta b_j^{[c]}} = \frac{\delta\xi}{\delta Z_j^{[c+1]}} \quad (\text{A.17})$$

Ainsi en utilisant la règle de mise à jour on obtient :

$$W_{ji}^{[c]} = W_{ji}^{[c]} - \alpha a_i^{[c]} \frac{\delta\xi}{\delta Z_j^{[c+1]}} \quad (\text{A.18})$$

$$b_j^{[c]} = b_j^{[c]} - \alpha \frac{\delta\xi}{\delta Z_j^{[c+1]}} \quad (\text{A.19})$$

Les processus de propagation avant et de retropropagation sont répétés jusqu'à la convergence ou obtention d'un minimum. Même si cette méthode est capable de résoudre la plupart des problèmes rencontrés en apprentissage profond, quelques faiblesses sont cependant à signaler, notamment sa convergence lente et son risque de tomber sur un minimum local.

Dans le but d'accélérer la convergence de la descente du gradient un élément temporel est rajouté et qui s'accumule à chaque itération, ainsi il permet de réduire les risques de rester bloquer sur un minimum local : C'est le principe du SGDM ou descent de gradient stochastique avec momentum [105].

## Annexe B

# Résultats détaillés de l'apprentissage machine

TABLE B.1 – SVM avec des noyaux gaussien (RBF) et linéaire sur les données supervisées et non supervisées dans le champ de haricot.

	Supervisé	Non Supervisé	Supervisé	Non Supervisé
	SVM (RBF)	SVM (RBF)	SVM (Linéaire)	SVM (Linéaire)
Color	0,47237	0,57079	0,56903	0,573
Gabor	0,49098	0,49563	0,5024	0,51724
HOG	0,64976	0,58354	0,56521	0,54243
Haralick	0,63161	0,59512	0,59502	0,56887
Edensity	0,49811	0,5164	0,51644	0,54746
Geo3	0,50488	0,49044	0,4796	0,47774
HOG+color	0,68783	0,63849	0,58026	0,58963
Gabor+Color	0,63996	0,61249	0,58926	0,59126
Geo3+Color	0,5742	0,58823	0,57127	0,56203
Hara+Gabor	0,63343	0,60919	0,58716	0,56527
Hara+HOG	0,70992	0,61503	0,58771	0,55378
Hara+Color	0,71908	0,67668	0,60469	0,61458
Hara+Edensity+Geo3	0,70409	0,62767	0,61222	0,57037
Color+Edensity+Geo3	0,60019	0,60622	0,57275	0,5717
Color+HOG+Gabor	0,72293	0,64645	0,5966	0,60074
Hara+Color+HOG+Gabor	0,76587	0,67111	0,61514	0,61376
HOG+Color+Hara	0,76575	0,67021	0,61339	0,61781
Tous les descripteurs	0,78759	0,68923	0,63897	0,63041

TABLE B.2 – Résultat du random forest (RF) avec 100 et 200 arbres sur les données supervisées et non supervisées dans le champ de haricot.

	Supervisé	Non Supervisé	Supervisé	Non Supervisé
	100	100	200	200
Color	0,63904	0,61526	0,63482	0,61623
Gabor	0,51561	0,52369	0,51934	0,51743
HOG	0,63649	0,57038	0,6426	0,58086
Haralick	0,65453	0,63739	0,65167	0,63693
Edensity	0,5165	0,52892	0,51409	0,52153
Geo3	0,49911	0,44673	0,49663	0,44555
HOG+color	0,65981	0,6139	0,66267	0,61979
Gabor+Color	0,6496	0,60695	0,64806	0,61028
Geo3+Color	0,64536	0,6158	0,64054	0,61519
Hara+Gabor	0,65174	0,63698	0,65155	0,63703
Hara+HOG	0,69302	0,63103	0,69568	0,63032
Hara+Color	0,67921	0,64347	0,6811	0,64951
Hara+Edensity+Geo3	0,68637	0,63669	0,68195	0,64054
Color+Edensity+Geo3	0,66198	0,62047	0,66109	0,62445
Color+HOG+Gabor	0,66767	0,6168	0,66529	0,61745
Hara+Color+HOG+Gabor	0,69253	0,63749	0,69407	0,63975
HOG+Color+Hara	0,69639	0,63771	0,69728	0,6442
Tous les descripteurs	0,69801	0,63593	0,69767	0,64004

TABLE B.3 – SVM avec des noyaux gaussien (RBF) et linéaire sur les données supervisées et non supervisées dans le champ d'épinard.

Descripteurs	Supervisé	Non Supervisé	Supervisé	Non Supervisé
	SVM (rbf)	SVM (rbf)	SVM (Linéaire)	SVM (Linéaire)
Color	0,91346	0,87528	0,94454	0,82844
Gabor	0,612	0,80838	0,67141	0,62513
HOG	0,78069	0,71955	0,87158	0,83144
Haralick	0,81764	0,89099	0,84862	0,80681
Edensity	0,47344	0,58584	0,48588	0,71795
Geo3	0,65489	0,75894	0,75206	0,3886
HOG+color	0,93949	0,90958	0,94982	0,87157
Gabor+Color	0,94277	0,90438	0,95404	0,82706
Geo3+Color	0,92657	0,90079	0,94751	0,84222
Hara+Gabor	0,91862	0,83887	0,88194	0,833
Hara+HOG	0,88109	0,85643	0,85656	0,81148
Hara+Color	0,95904	0,91202	0,93638	0,90686
Hara+Edensity+Geo3	0,91588	0,83169	0,84398	0,79524
Color+Edensity+Geo3	0,93866	0,91014	0,9498	0,89175
Color+HOG+Gabor	0,94867	0,88843	0,96427	0,87713
Hara+Color+HOG+Gabor	0,96252	0,89054	0,9537	0,89915
HOG+Color+Hara	0,95997	0,90346	0,95104	0,88296
Tous les descripteurs	0,96781	0,88562	0,95441	0,89997

TABLE B.4 – Résultat du random forest (RF) avec 100 et 200 arbres sur les données supervisées et non supervisées dans le champ d'épinard.

Descripteurs	Supervisé	Non Supervisé	Supervisé	Non Supervisé
	100	100	200	200
Color	0,95146	0,81356	0,9554	0,8239
Gabor	0,64615	0,79046	0,6542	0,7923
HOG	0,81727	0,80295	0,8209	0,7941
Haralick	0,87846	0,90099	0,8776	0,9038
Edensity	0,50406	0,62688	0,4928	0,6204
Geo3	0,68025	0,78811	0,6722	0,773
HOG+color	0,93627	0,85608	0,94286	0,8602
Gabor+Color	0,94782	0,91452	0,94601	0,91205
Geo3+Color	0,95044	0,88034	0,95229	0,89318
Hara+Gabor	0,89185	0,90871	0,88547	0,91633
Hara+HOG	0,91511	0,93529	0,91546	0,93894
Hara+Color	0,96165	0,91687	0,959	0,92697
Hara+Edensity+Geo3	0,85631	0,94857	0,8479	0,94997
Color+Edensity+Geo3	0,95018	0,91607	0,94513	0,91075
Color+HOG+Gabor	0,95217	0,93982	0,95207	0,93784
Hara+Color+HOG+Gabor	0,94586	0,9619	0,95053	0,96333
HOG+Color+Hara	0,95461	0,93906	0,96275	0,93325
Tous les descripteurs	0,96869	0,95147	0,96866	0,95468

---

# Bibliographie

- [1] Introduction to deep learning : What is deep learning? video - matlab. <https://fr.mathworks.com/videos/introduction-to-deep-learning-what-is-deep-learning--1489502328819.html>. (Accessed on 01/20/2020).
- [2] Le plan Écophyto, qu'est-ce que c'est ? | alim'agri. <https://agriculture.gouv.fr/le-plan-ecophyto-quest-ce-que-cest>. (Accessed on 01/19/2020).
- [3] New vision technologies for real-world applications. <https://semiengineering.com/new-vision-technologies-for-real-world-applications/>. (Accessed on 01/20/2020).
- [4] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(11) :2274–2282, 2011.
- [5] Kamal N Agrawal, Karan Singh, Ganesh C Bora, and Dongqing Lin. Weed Recognition Using Image-Processing Technique Based on Leaf Parameters. *Journal of Agricultural Science and Technology. B*, 2(8B) :1939–1250, 2012.
- [6] Irshad Ahmad, Muhammad Hameed Siddiqi, Iram Fatima, Sungyoung Lee, and Young-Koo Lee. Weed classification based on Haar wavelet transform via k-Nearest Neighbor (k-NN) for real-time automatic sprayer control system. In *Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication (ICUIMC)*, pages 1–6, New York, NY, USA, 2011.
- [7] Jamil Ahmad, Khan Muhammad, Imran Ahmad, Wakeel Ahmad, Melvyn L Smith, Lyndon N Smith, Deepak Kumar Jain, Haoxiang Wang, and Irfan Mehmood. Visual features based boosted classification of weeds for real-time selective herbicide sprayer systems. *Computers in Industry*, 98 :23–33, 2018.
- [8] Faisal Ahmed, Hawlader Abdullah Al-Mamun, A. S M Hossain Bari, Emam Hossain, and Paul Kwan. Classification of crops and weeds from digital images : A support vector machine approach. *Crop Protection*, 40 :98–104, 2012.
- [9] Gerald L Anderson, James H Everitt, Arthur J Richardson, and David E Escobar. Using satellite data to map false broomweed (*ericameria austrotexana*) infestations on south texas rangelands. *Weed Technology*, 7(4) :865–871, 1993.
- [10] Dionisio Andújar, A Ribeiro, R Carmona, César FERNÁNDEZ-QUINTANILLA, and José Dorado. An assessment of the accuracy and consistency of human perception of weed cover. *Weed Research*, 50(6) :638–647, 2010.

- [11] Ron Appel, Thomas Fuchs, Piotr Dollár, and Pietro Perona. Quickly boosting decision trees—pruning underachieving features early. In *International conference on machine learning*, pages 594–602, 2013.
- [12] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet : A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12) :2481–2495, 2017.
- [13] M. Dian Bah, Adel Hafiane, and Raphael Canals. Weeds detection in UAV imagery using SLIC and the hough transform. In *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6. IEEE, 2017.
- [14] Adel Bakhshipour and Abdolabbas Jafari. Evaluation of support vector machine and artificial neural networks in weed detection using shape features. *Computers and Electronics in Agriculture*, 145 :153–160, 2018.
- [15] Adel Bakhshipour, Abdolabbas Jafari, Seyed Mehdi Nassiri, and Dariush Zare. Weed segmentation using texture features extracted from wavelet sub-images. *Biosystems Engineering*, 157 :1–12, 2017.
- [16] Tijmen Bakker, Hendrik Wouters, Kees van Asselt, Jan Bontsema, Lie Tang, Joachim Müller, and Gerrit van Straten. A vision based row detection system for sugar beet. *Computers and Electronics in Agriculture*, 60(1) :87–95, 2008.
- [17] Bilal Bataineh. An iterative thinning algorithm for binary images based on sequential and parallel approaches. *Pattern Recognition and Image Analysis*, 28(1) :34–43, 2018.
- [18] Michel Bertrand and Thierry Doré. Comment intégrer la maîtrise de la flore adventice dans le cadre général d’un système de production intégrée. *Innovations Agronomiques*, 3 :1–13, 2008.
- [19] Serge Beucher et al. The watershed transformation applied to image segmentation. *SCANNING MICROSCOPY-SUPPLEMENT*, 6 :299–314, 1992.
- [20] Thomas Blaschke. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1) :2–16, jan 2010.
- [21] Alan Conrad Bovik, Marianna Clark, and Wilson S. Geisler. Multichannel Texture Analysis Using Localized Spatial Filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1) :55–73, 1990.
- [22] Rebecca Brown, Jeffrey G Paine, S Saylam, Thomas A Tremblay, John R Andrews, and Aaron Averett. Mangrove monitoring using airborne vnir in the espiritu santo bay area, central texas coast. *Report of the Bureau of Economic Geology, The University of Texas at Austin. Report prepared for the General Land Office under contract, (14-078) :000*, 2016.
- [23] A.H.C. Van Bruggen, M.M. He, K. Shin, V. Mai, K.C. Jeong, M.R. Finckh, and J.G. Morris. Environmental and health effects of the herbicide glyphosate. *Science of the Total Environment*, 616 :255–268, 2018.
- [24] Xavier P. Burgos-Artizzu, Angela Ribeiro, Maria Guijarro, and Gonzalo Pajares. Real-time image processing for crop / weed discrimination in maize fields. *Computers and Electronics in Agriculture*, 75(2) :337–346, 2011.

- [25] Jean-Pierre Butault, Nathalie Delame, Florence Jacquet, and Guillaume Zardet. L'utilisation des pesticides en france : Etat des lieux et perspectives de réduction. (32) :7–26, 2011.
- [26] Cristina Campi, Annalisa Perasso, Mauro C Beltrametti, Anna Maria Massone, Gianmario Sambuceti, and Michele Piana. Pattern recognition in medical imaging by means of the hough transform of curves. In *2013 8th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pages 280–283. IEEE, 2013.
- [27] Jean-Pierre Caussanel. Nuisibilité et seuils de nuisibilité des mauvaises herbes dans une culture annuelle : situation de concurrence bispécifique. *Agronomie*, 9(3) :219–240, 2007.
- [28] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab : Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4) :834–848, 2017.
- [29] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv :1406.1078*, 2014.
- [30] Lorenzo Comba, Paolo Gay, Jacopo Primicerio, and Davide Ricauda Aimonino. Vineyard detection from unmanned aerial systems images. *Computers and Electronics in Agriculture*, 114 :78–87, 2015.
- [31] Claude Compagnone, Florence Hellec, K Macé, Pierre Morlon, Nicolas Munier-Jolain, and L Quéré. Raisonement des pratiques et des changements de pratiques en matière de désherbage : regards agronomique et sociologique à partir d'enquêtes chez des agriculteurs. *Innovations agronomiques*, 3 :89–105, 2008.
- [32] Michael Confino. *Systèmes agraires et progrès agricole : L'assolement triennal en Russie aux XVIII–XIX siècles. Étude d'économie et de sociologie rurales*, volume 14. Walter de Gruyter GmbH & Co KG, 2019.
- [33] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3) :273–297, 1995.
- [34] Sadie Costello, Myles Cockburn, Jeff Bronstein, Xinbo Zhang, and Beate Ritz. Parkinson's disease and residential exposure to maneb and paraquat from agricultural applications in the central valley of california. *American journal of epidemiology*, 169(8) :919–926, 2009.
- [35] Antonio Criminisi and J. Shotton. *Decision Forests for Computer Vision and Medical Image Analysis*. Springer Publishing Company, Incorporated, 2013.
- [36] Nello Cristianini and John Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press, 2000.
- [37] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4) :303–314, 1989.
- [38] Navneet Dalal, William Triggs, and Bill Triggs. Histograms of Oriented Gradients for Human Detection, 2004.

- [39] Alain Derville. L'assolement triennal dans la France du nord au moyen âge. *Revue historique*, 280(Fasc. 2 (568)) :337–376, 1988.
- [40] Maurilio Di Cicco, Ciro Potena, Giorgio Grisetti, and Alberto Pretto. Automatic model based dataset generation for fast and accurate crop and weeds detection. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5188–5195. IEEE, sep 2017.
- [41] Bureau Dominique and Chalmin Philippe. *Perspectives agricoles en France et en Europe*. Documentation Française (La), 2007.
- [42] William W Donald. Between-observer differences in relative corn yield vs. rated weed control. *Weed technology*, 20(1) :41–51, 2006.
- [43] Alessandro Dos Santos Ferreira, Daniel Matte Freitas, Gercina Gonçalves da Silva, Hemerson Pistori, and Marcelo Theophilo Folhes. Weed detection in soybean crops using ConvNets. *Computers and Electronics in Agriculture*, 143 :314–324, 2017.
- [44] Daniel Downey, D. Ken Giles, David C Slaughter, et al. Weeds accurately mapped using dgps and ground-based vision identification. *California agriculture*, 58(4) :218–221, 2004.
- [45] Qian Du, Ni-Bin Chang, Chenghai Yang, and Kanth R Srilakshmi. Combination of multispectral remote sensing, variable rate technology and environmental modeling for citrus pest management. *Journal of Environmental Management*, 86(1) :14–26, 2008.
- [46] Richard O. Duda and Peter E. Hart. Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1) :11–15, jan 1972.
- [47] Jean-Louis Dumas. Liebig et son empreinte sur l'agronomie moderne : Biographie sommaire de liebig. *Revue d'histoire des sciences et de leurs applications*, 18(1) :73–108, 1965.
- [48] European Crop Protection. With or without pesticides ? | ECPA, 2017.
- [49] James H Everitt, David E Escobar, Mario A Alaniz, Ricardo Villarreal, and Michael R Davis. Distinguishing brush and weeds on rangelands using video remote sensing. *Weed technology*, 6(4) :913–921, 1992.
- [50] James H Everitt, David E Escobar, Ricardo Villarreal, Mario A Alaniz, and Michael R Davis. Canopy light reflectance and remote sensing of shin oak (*quercus havardii*) and associated vegetation. *Weed Science*, 41(2) :291–297, 1993.
- [51] FAO. Adventices : la menace qui plane sur la tête des agriculteurs. <http://www.fao.org/news/story/fr/item/29424/icode/>. (Accessed on 01/13/2020).
- [52] Filip Feyaerts and Luc J. Van Gool. Multi-spectral vision system for weed detection. *Pattern Recognition Letters*, 22(6-7) :667–674, 2001.
- [53] Veronique Fontaine and Trever G Crowe. Development of line-detection algorithms for local positioning in densely seeded crops. *Canadian biosystems engineering / Le genie des biosystemes au Canada : La revue de la societe canadienne de genie agroalimentaire et biologique*, 2006.
- [54] Md Ferdouse Ahmed Foysal, Mohammad Shakirul Islam, Asif Karim, and Nafis Neehal. Shot-net : A convolutional neural network for classifying different cricket shots. In *International Conference on Recent Trends in Image Processing and Pattern Recognition*, pages 111–120. Springer, 2018.

- [55] Yoav Freund and Robert Schapire. A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, 14(5) :771–780, 1999.
- [56] Nir Friedman, Dan Geiger, and Moises Goldszmidt. Bayesian network classifiers. *Machine learning*, 29(2-3) :131–163, 1997.
- [57] Junfeng Gao, Wenzhi Liao, David Nuyttens, Peter Lootens, Jürgen Vangeyte, Aleksandra Pižurica, Yong He, and Jan G. Pieters. Fusion of pixel and object-based features for weed mapping using unmanned aerial vehicle imagery. *International Journal of Applied Earth Observation and Geoinformation*, 67 :43–53, may 2018.
- [58] Junfeng Gao, David Nuyttens, Peter Lootens, Yong He, and Jan G Pieters. Recognising weeds in a maize crop using a random forest machine-learning algorithm and near-infrared snapshot mosaic hyperspectral imagery. *Biosystems Engineering*, 170 :39–50, 2018.
- [59] Christelle Gée, Jérémie Bossu, Gawain. Jones, and Frédéric Truchetet. Crop/weed discrimination in perspective agronomic images. *Computers and Electronics in Agriculture*, 60(1) :49–59, 2008.
- [60] Robin Genuer and Jean-Michel Poggi. Arbres CART et Forêts aléatoires, Importance et sélection de variables. working paper or preprint, January 2017.
- [61] Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely randomized trees. *Machine learning*, 63(1) :3–42, 2006.
- [62] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.
- [63] Michel Griffon. Vers une septième révolution agricole. *Revue Projet*, 332(1) :11–19, 2013.
- [64] M. Guijarro, G. Pajares, I. Riomoros, P. J. Herrera, X. P. Burgos-Artizzu, and A. Ribeiro. Automatic segmentation of relevant textures in agricultural images. *Computers and Electronics in Agriculture*, 75(1) :75–83, 2011.
- [65] T. Hague, N. D. Tillett, and H. Wheeler. Automated Crop and Weed Monitoring in Widely Spaced Cereals. *Precision Agriculture*, 7(1) :21–32, mar 2006.
- [66] Ryuhei Hamaguchi, Aito Fujita, Keisuke Nemoto, Tomoyuki Imaizumi, and Shuhei Hikosaka. Effective use of dilated convolutions for segmenting small object instances in remote sensing imagery. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1442–1450. IEEE, 2018.
- [67] Esmael Hamuda, Martin Glavin, and Edward Jones. A survey of image processing techniques for plant extraction and segmentation in the field. *Computers and Electronics in Agriculture*, 125 :184–199, 2016.
- [68] Robert M Haralick and Kumarasamy Shanmugam. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6) :610–621, 1973.
- [69] Shayan Hati and G Sajeevan. Plant recognition from leaf image through artificial neural network. *International Journal of Computer Applications*, 62(17), 2013.

- [70] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.
- [71] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778. IEEE, jun 2016.
- [72] David W Hosmer Jr, Stanley Lemeshow, and Rodney X Sturdivant. *Applied logistic regression*, volume 398. John Wiley & Sons, 2013.
- [73] Paul V. C. Hough. Method and means for recognizing complex patterns. *US Patent 3,069,654*, 21 :225–231, dec 1962.
- [74] Zhongling Huang, Zongxu Pan, and Bin Lei. Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data. *Remote Sensing*, 9(9) :907, aug 2017.
- [75] Alfredo Huete, Kamel Didan, Tomoaki Miura, E Patricia Rodriguez, Xiang Gao, and Laerte G Ferreira. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83(1-2) :195–213, 2002.
- [76] Calvin Hung, Zhe Xu, and Salah Sukkarieh. Feature Learning Based Approach for Weed Classification Using High Resolution Aerial Images from a Digital Camera Mounted on a UAV. *Remote Sensing*, 6(12) :12037–12054, dec 2014.
- [77] Zaidah Ibrahim, Nurbaity Sabri, and Nur Nabilah Abu Mangshor. Leaf Recognition using Texture Features for Herbal Plant Identification. *Indonesian Journal of Electrical Engineering and Computer Science*, 9(1) :152–156, 2018.
- [78] John Illingworth and Josef Kittler. A survey of the hough transform. *Computer vision, graphics, and image processing*, 44(1) :87–116, 1988.
- [79] Agricultural Sustainability Institute. What is sustainable agriculture | agricultural sustainability institute. <https://asi.ucdavis.edu/programs/ucsarep/about/what-is-sustainable-agriculture>. (Accessed on 10/20/2019).
- [80] Sergey Ioffe and Christian Szegedy. Batch normalization : Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv :1502.03167*, 2015.
- [81] Ray D Jackson. Remote sensing of vegetation characteristics for farm management. In *Remote Sensing : Critical Review of Technology*, volume 475, pages 81–97. International Society for Optics and Photonics, 1984.
- [82] Wes Jackson. *New roots for agriculture*. University of Nebraska Press, 1980.
- [83] Hong Y. Jeon, Lei F. Tian, and Heping Zhu. Robust crop and weed segmentation under uncontrolled outdoor illumination. *Sensors*, 11(6) :6270–6283, 2011.
- [84] Ronghua Ji and Lijun Qi. Crop-row detection algorithm based on Random Hough Transformation. *Mathematical and Computer Modelling*, 54(3-4) :1016–1020, aug 2011.
- [85] Gwain Jones, Christelle Gée, and Frédéric Truchetet. Modelling agronomic images for weed detection and comparison of crop/weed discrimination algorithm performance. *Precision Agriculture*, 10(1) :1–15, 2009.

- [86] Carl F. Jordan. Derivation of leaf-area index from quality of light on the forest floor. *Ecology*, 50(4) :663–666, jul 1969.
- [87] Antony Joseph. Chapter 5 - imaging of seawater motion signatures using remote sensors. In Antony Joseph, editor, *Measuring Ocean Currents*, pages 139 – 175. Elsevier, Boston, 2014.
- [88] Simon J. Julier and Jeffrey K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3) :401–422, March 2004.
- [89] Heikki Kälviäinen and Petri Hirvonen. An extension to the randomized hough transform exploiting connectivity. *Pattern Recognition Letters*, 18(1) :77–85, 1997.
- [90] Andreas Kamilaris and Francesc X. Prenafeta-Boldú. Deep learning in agriculture : A survey. *Computers and Electronics in Agriculture*, 147 :70–90, apr 2018.
- [91] Takashi Kataoka, Toshihiro Kaneko, Hiroshi Okamoto, and Shogo Hata. Crop growth estimation system using machine vision. *Proceedings 2003 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM 2003)*, 2(Aim) :1079–1083, 2003.
- [92] K. A. Kennedy and P. A. Addison. Some considerations for the use of visual estimates of plant cover in biomonitoring. *The Journal of Ecology*, pages 151–157, 1987.
- [93] Michio Kise, Qin Zhang, and Francisco Rovira Más. A stereovision-based crop row detection method for tractor-automated guidance. *Biosystems Engineering*, 90(4) :357–367, 2005.
- [94] Nicholas E Korres. Agronomic weed control : a trustworthy approach for sustainable weed management. In *Non-Chemical Weed Control*, pages 97–114. Elsevier, 2018.
- [95] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, pages 1–9, 2012.
- [96] Avinash Kumar, Sobhangi Sarkar, and Chittaranjan Pradhan. Malaria disease detection using cnn technique with sgd, rmsprop and adam optimizers. In *Deep Learning Techniques for Biomedical and Health Informatics*, pages 211–230. Springer, 2020.
- [97] Andrea S Laliberte, Craig Winters, and Albert Rango. Acquisition, orthorectification, and classification of hyperspatial uav imagery. In *Fourth Annual Symposium : Research Insights in Semiarid Scosystems, RISE, University of Arizona, Tucson*, 2007.
- [98] Louisa Lam, Seong-Whan Lee, and Ching Y Suen. Thinning methodologies-a comprehensive survey. *IEEE Transactions on pattern analysis and machine intelligence*, 14(9) :869–885, 1992.
- [99] David W. Lamb and R.B. Brown. Pa—precision agriculture : Remote-sensing and mapping of weeds in crops. *Journal of Agricultural Engineering Research*, 78(2) :117 – 125, 2001.
- [100] Ross D Lamm, David C Slaughter, and D Ken Giles. Precision weed control system for cotton. *Transactions of the ASAE*, 45(1) :231, 2002.
- [101] Mrs Latha, A Poojith, Amarnath Reddy, and Vittal Kumar. Image Processing in Agriculture. *International Journal Of Innovative Research In Electrical, Electronics, Instrumentation And Control Engineering*, 2(6) :2321–2004, 2014.

- [102] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11) :2278–2323, 1998.
- [103] W. S. Lee, D. C. Slaughter, and D. K. Giles. Robotic weed control system for tomatoes. *Precision Agriculture*, 1(1) :95–113, Jan 1999.
- [104] Breiman Leo. Random forests. *Machine learning*, 45(1) :5–32, 2001.
- [105] Nicolas Loizou and Peter Richtárik. Momentum and stochastic momentum for stochastic gradient, newton, proximal point and subspace descent methods. *arXiv preprint arXiv :1712.09677*, 2017.
- [106] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [107] Philipp Lottes, Raghav Khanna, Johannes Pfeifer, Roland Siegwart, and Cyrill Stachniss. Uav-based crop and weed classification for smart farming. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3024–3031. IEEE, 2017.
- [108] Philipp Lottes, Raghav Khanna, Johannes Pfeifer, Roland Siegwart, and Cyrill Stachniss. Uav-based crop and weed classification for smart farming. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3024–3031. IEEE, may 2017.
- [109] Marine Louargant. *Proximité des adventices par imagerie aérienne : vers un service de gestion par drone*. PhD thesis, 2016.
- [110] Marine Louargant, Gawain Jones, Romain Faroux, Jean-Noël Paoli, Thibault Maillot, Christelle Gée, and Sylvain Villette. Unsupervised classification algorithm for early weed detection in row-crops by combining spatial and spectral information. *Remote Sensing*, 10(5) :761, 2018.
- [111] JE Lutman, PJW, Rew, LJ, Cussans, GW, Miller, PCH, Paice, MER, Stafford. Development of a patch spraying system to control weeds in winter wheat, 1998.
- [112] David M Magerman. Statistical decision-tree models for parsing. In *Proceedings of the 33rd annual meeting on Association for Computational Linguistics*, pages 276–283. Association for Computational Linguistics, 1995.
- [113] Larry M Manevitz and Malik Yousef. One-class svms for document classification. *Journal of machine Learning research*, 2 :139–154, 2001.
- [114] Pascal Marnotte. La lutte contre cyperus rotundus. *Agriculture et développement*, (1) :57–58, 1994.
- [115] E. J. P. MARSHALL. Field-scale estimates of grass weed populations in arable land. *Weed Research*, 28(3) :191–198, jun 1988.
- [116] Andrew McCallum and Kamal Nigam. A comparison of event models for naive Bayes text classification. In *Learning for Text Categorization : Papers from the 1998 AAAI Workshop*, pages 41–48, 1998.
- [117] Francois Menant. Du moyen age à aujourd’hui mille ans de famines. Technical Report 383, Jan 2013.

- [118] R. M. Menges, P. R. Nixon, and A. J. Richardson. Light reflectance and remote sensing of weeds in agronomic and horticultural crops. *Weed Science*, 33(4) :569–581, 1985.
- [119] George E. Meyer and João Camargo Neto. Verification of color vegetation indices for automated crop imaging applications. *Computers and Electronics in Agriculture*, 63 :282–293, 2008.
- [120] Georges E. Meyer, Timothy Hindman, and Koppolu Laksmi. Machine vision detection parameters for plant species identification. *Proceedings of SPIE*, 3543 :327–335, 1998.
- [121] Andres Milioto, Philipp Lottes, and Cyrill Stachniss. Real-time blob-wise sugar beets vs weeds classification for monitoring fields using convolutional neural networks. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W3 :41–48, aug 2017.
- [122] M. Montalvo, G. Pajares, J. M. Guerrero, J. Romeo, M. Guijarro, A. Ribeiro, J. J. Ruz, and J. M. Cruz. Automatic detection of crop rows in maize fields with high weeds pressure. *Expert Systems With Applications*, 39(15) :11889–11897, 2012.
- [123] Anders Krogh Mortensen, Mads Dyrmann, Henrik Karstoft, Rasmus Nyholm Jørgensen, and René Gislum. Semantic Segmentation of Mixed Crops using Deep Convolutional Neural Network. In *CIGR-AgEng conference*, 2016.
- [124] M Mosango. Influence des plantes adventices sur les plantes de culture : quelques résultats. *Journal d’agriculture traditionnelle et de botanique appliquée*, 30(1) :35–48, 1983.
- [125] M Mudrova and Aleš Procházka. Principal component analysis in image processing. In *Proceedings of the MATLAB Technical Computing Conference, Prague*, 2005.
- [126] Golshah A. Naghdy, Jian Wang, and Philip O. Ogunbona. Texture analysis using Gabor wavelets. 2657 :74 – 85, 1996.
- [127] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1) :62–66, jan 1979.
- [128] José Manuel Peña Barragán, Maggi Kelly, Ana Isabel de Castro, and Francisca López Granados. Object-based approach for crop row characterization in UAV images for site-specific weed management. In *Queiroz-Feitosa et al., editors. 4th International; Conference on Geographic Object-Based Image Analysis (GEOBIA 2012), Rio de Janeiro, Brazil*, pages 426–430, 2012.
- [129] Céline Pélosi, Lucile Toutous, François Chiron, Florence Dubs, Mickaël Hedde, Audrey Muratet, J-F Ponge, Sandrine Salmon, and David Makowski. Reduction of pesticide use can increase earthworm populations in wheat crops in a european temperate region. *Agriculture, ecosystems & environment*, 181 :223–230, 2013.
- [130] José Manuel Peña, Jorge Torres-Sánchez, Ana Isabel De Castro, Maggi Kelly, and Francisca López-Granados. Weed Mapping in Early-Season Maize Fields Using Object-Based Analysis of Unmanned Aerial Vehicle (UAV) Images. *PLoS ONE*, 8(10), 2013.
- [131] Alberto J. Pérez, Fernando López, José V. Benlloch, and Svend Christensen. Colour and shape analysis techniques for weed detection in cereal fields. *Computers and Electronics in Agriculture*, 25(3) :197–212, feb 2000.

- [132] María Pérez-Ortiz, José Manuel Peña, Pedro Antonio Gutiérrez, Jorge Torres-Sánchez, César Hervás-Martínez, and Francisca López-Granados. Selecting patterns and features for between- and within- crop-row weed mapping using UAV-imagery. *Expert Systems With Applications*, 47 :85–94, 2015.
- [133] Albert J Peters, Bradley C Reed, Marlen D Eve, and Kirk C McDaniel. Remote sensing of broom snakeweed (*gutierrezia sarothrae*) with noaa-10 spectral image processing. *Weed Technology*, 6(4) :1015–1020, 1992.
- [134] Ciro Potena, Daniele Nardi, and Alberto Pretto. Fast and accurate crop and weed identification with summarized train sets for precision agriculture. In *International Conference on Intelligent Autonomous Systems*, pages 105–121. Springer, 2016.
- [135] Agreste Primeur. Enquête sur la structure des exploitations agricoles. (350), 2018.
- [136] María Pérez-Ortiz, José Manuel Peña, Pedro Antonio Gutiérrez, Jorge Torres-Sánchez, César Hervás-Martínez, and Francisca López-Granados. Selecting patterns and features for between- and within- crop-row weed mapping using uav-imagery. *Expert Systems with Applications*, 47 :85 – 94, 2016.
- [137] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once : Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.
- [138] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn : Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.
- [139] Lisa J. Rew, G. W. Cussans, Moira A. Mugglestone, and P. C. H. Miller. A technique for mapping the spatial distribution of *Elymus repens*, with estimates of the potential reduction in herbicide usage from patch spraying. *Weed Research*, 36(4) :283–292, aug 1996.
- [140] A. J. Richardson, R. M. Menges, and P. R. Nixon. Distinguishing weed from crop plants using video remote sensing. *Photogrammetric Engineering and Remote Sensing*, 51(11) :1785–1790, 1985.
- [141] Inkyu Sa, Zetao Chen, Marija Popović, Raghav Khanna, Frank Liebisch, Juan Nieto, and Roland Siegwart. weednet : Dense semantic weed classification using multispectral images and mav for smart farming. *IEEE Robotics and Automation Letters*, 3(1) :588–595, 2017.
- [142] Inkyu Sa, Marija Popović, Raghav Khanna, Zetao Chen, Philipp Lottes, Frank Liebisch, Juan Nieto, Cyrill Stachniss, Achim Walter, and Roland Siegwart. Weedmap : A large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming. *MDPI Remote Sensing*, 10(9), Aug 2018.
- [143] D Koc San and M Turker. Building extraction from high resolution satellite images using hough transform. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science*, 38(8), 2010.
- [144] Bernhard Schölkopf, Robert Williamson, Alex Smola, John Shawe-Taylor, and John Platt. Support vector method for novelty detection. In *Proceedings of the 12th International Conference on Neural Information Processing Systems*, NIPS’99, page 582–588, Cambridge, MA, USA, 1999. MIT Press.

- [145] Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, and Andrew Blake. Real-time human pose recognition in parts from single depth images. In *CVPR 2011*, pages 1297–1304. Ieee, 2011.
- [146] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint arXiv :1409.1556*, pages 1–14, 2015.
- [147] Henning Tangen Sogaard and Hans Jørgen Olsen. Determination of crop rows by image analysis without segmentation. *Computers and electronics in agriculture*, 38(2) :141–158, 2003.
- [148] J V Stafford, J. M. Bars, and B Ambler. Hand-held data logger with integral GPS for producing weed maps by field walking. *Computers and Electronics in Agriculture*, 1699(95) :235–247, 1996.
- [149] Kishore C Swain, Hemantha PW Jayasuriya, and Vilas M Salokhe. Suitability of low-altitude remote sensing images for estimating nitrogen treatment variations in rice cropping for precision agriculture adoption. *Journal of Applied Remote Sensing*, 1(1), 2007.
- [150] Jiliang Tang, Salem Alelyani, and Huan Liu. *Feature selection for classification : A review*, pages 37–64. CRC Press, January 2014.
- [151] Lie Tang, LeiF Tian, and Brian L Steward. Classification of Broadleaf and Grass Weeds Using Gabor Wavelets and an Artificial Neural Network. *Transactions of the American Society of Agricultural Engineers*, 46(4) :1247–1254, 2003.
- [152] JF Thompson, JV Stafford, and PCH Miller. Potential for automatic weed detection and selective herbicide application. *Crop protection*, 10(4) :254–259, 1991.
- [153] J. Torres-Sánchez, F. López-Granados, and J. M. Peña. An automatic object-based method for optimal thresholding in UAV images : Application for vegetation detection in herbaceous crops. *Computers and Electronics in Agriculture*, 114 :43–52, 2015.
- [154] L Trémel. Lutte contre les mauvaises herbes. pages 1–12, 1990.
- [155] Compton J Tucker. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2) :127–150, 1979.
- [156] Compton J Tucker. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2) :127–150, 1979.
- [157] E. Ullah, J.A. Peterson, R.C.H. Shepherd, and J.T. Baxter. Mapping flowering Paterson’s Curse (*Echium plantagineum*) around Lake Hume, north eastern Victoria, using Landsat TM data. *Plant Protection Quarterly (Australia)*, 4 :155–157, 1985.
- [158] E Lynn Usery, Stuart Pocknee, and Broughton Boydell. Precision farming data management using geographic information systems. *Photogrammetric engineering and remote sensing*, 61(11) :1383–1392, 1995.
- [159] Trygve Utstumo, Frode Urdal, Anders Brevik, Jarle Dørum, Jan Netland, Øyvind Overskeid, Therese W Berge, and Jan Tommy Gravdahl. Robotic in-row weed control in vegetables. *Computers and electronics in agriculture*, 154 :36–45, 2018.
- [160] Michel Vanderpooten. Éléments techniques d’une révolution agricole au début de l’époque contemporaine. thèse de doctorat en histoire. *Revue d’histoire du XIX<sup>ème</sup> siècle*, (2) :41–41, 2001.

- [161] Junping Wang, Quanshi Chen, and Yong Chen. Rbf kernel based support vector machine with universal approximation and its application. In *International symposium on neural networks*, pages 512–517. Springer, 2004.
- [162] Martin Weis and Roland Gerhards. Feature extraction for the identification of weed species in digital images for the purpose of site-specific weed control. *Precision agriculture '07*, pages 537–543, 2007.
- [163] B. J. Wilson and P. Brain. Weed monitoring on a whole farm-patchiness and the stability of distribution of alopecurus myosuroides over a ten year period. pages 45–52. European Weed Research Society, 1990.
- [164] D.M. Woebbecke, G.E. Meyer, K. Von Bargen, and D.A. Mortensen. Color indices for weed identification under various soil, residue, and lighting conditions. *Transactions of the ASAE*, 38(1) :259–269, 1995.
- [165] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3) :37–52, 1987.
- [166] Haitao Xiang and Lei Tian. Development of a low-cost agricultural remote sensing system based on an autonomous unmanned aerial vehicle (UAV). *Biosystems Engineering*, 108(2) :174–190, 2011.
- [167] Xue-Yang Xiao, Rongxiang Hu, Shan-Wen Zhang, and Xiao-Feng Wang. Hog-based approach for leaf classification. In *Proceedings of the Advanced Intelligent Computing Theories and Applications, and 6th International Conference on Intelligent Computing*, ICIC'10, page 149–155, Berlin, Heidelberg, 2010. Springer-Verlag.
- [168] Lei Xu and Erkki Oja. Randomized hough transform (rht) : basic mechanisms, algorithms, and computational complexities. *CVGIP : Image understanding*, 57(2) :131–154, 1993.
- [169] Mark CK Yang, Jong-Sen Lee, Cheng-Chang Lien, and Chung-Lin Huang. Hough transform modified by line connectivity and line thickness. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(8) :905–910, 1997.
- [170] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv :1511.07122*, 2015.
- [171] Chunhua Zhang and John M. Kovacs. The application of small unmanned aerial systems for precision agriculture : a review. *Precision Agriculture*, 13(6) :693–712, dec 2012.
- [172] Fang Zheng, Sheng Luo, Song Kang, Yan Chang-Wei, and Wang Mu-Chou. Improved Lane Line Detection Algorithm Based on Hough Transform. *Pattern Recognition and Image Analysis*, 28(2) :254–260, apr 2018.
- [173] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [174] Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. Learning deep features for scene recognition using places database. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 487–495. Curran Associates, Inc., 2014.

- [175] P. Zwaenepoel and J.M. Le Bars. L'agriculture de précision. *Ingénieries eau-agriculture-territoires*, (12) :p. 67 – p. 79, 1997.
- [176] Björn Åstrand and Albert-Jan Baerveldt. A vision based row-following system for agricultural field machinery. *Mechatronics*, 15(2) :251 – 269, 2005.

**Mamadou Dian BAH**

## **Détection des Adventices par imagerie aérienne**

### Résumé :

Dans le contexte agricole actuel, il est nécessaire de réduire l'utilisation des produits phytosanitaires contre les mauvaises herbes. Le désherbage localisé présente une alternative prometteuse pour limiter les coûts et l'impact environnemental. Cependant, la localisation automatique des adventices n'est pas une tâche facile car elle présente plusieurs défis scientifiques et technologiques. L'objectif de cette thèse est de proposer des méthodes de traitement d'images et d'intelligence artificielle pour la localisation des adventices en grandes cultures. Dans ce cadre, nous avons abordé deux problématiques, la détection des rangées de culture et la détection des adventices. Deux méthodes ont été proposées pour la détection des rangées de culture. La première méthode combine la transformée de Hough et l'algorithme de regroupement linéaire itératif SLIC. La deuxième, quant à elle, utilise une approche totalement nouvelle basée sur l'apprentissage profond. Ces deux méthodes ont été utilisées pour détecter les adventices inter-rang et celles qui sont en contact avec les rangées de culture. Pour tendre vers une meilleure efficacité, deux nouvelles méthodes de détection d'adventices par apprentissage machine, entièrement automatiques ont été développées. L'originalité de ces méthodes est que l'apprentissage est effectué sur des données annotées automatiquement. La première méthode est basée sur l'apprentissage profond tandis que la seconde génère des modèles à partir de descripteurs profonds et un classifieur à classe unique. Les résultats obtenus sur des données réelles montrent l'intérêt des approches proposées.

Mots clés : détection d'adventices, détection des rangées de cultures, apprentissage profond

## **Weed Detection by Aerial Imagery**

### Summary :

In the current agricultural context, there is a need to reduce the use of pesticide for weed control. Localized weed control presents a promising option to limit costs and environmental impact. However, automatic weed detection is not an easy task and presents several scientific and technological challenges. The objective of this thesis is to propose image processing and artificial intelligence methods for weed detection in field crops. Within this framework, we addressed two issues, crop row detection and weed detection. Two methods were proposed for crop row detection. The first method combines the Hough transform and the simple linear iterative clustering SLIC. The second one uses a completely new approach using deep learning. Both methods were used to detect inter-row weeds and weeds in contact with crop rows. To achieve greater efficiency, two new fully automatic machine learning weed detection methods have been developed. The originality of these methods is that learning is carried out on automatically annotated data. The first method is based on deep learning while the second method generates models from deep features and one-class classifier. The results obtained on real data show the interest of the proposed approaches.

Keywords : weed detection, crop row detection, deep learning