



LUCAS MOHIMONT, AMINE CHEMCHAM, MARINE RONDEAU, MATHIAS ROESLER, FRANÇOIS ALIN, NATHALIE GAVEAU, LUIZ ANGELO STEFFENEL
Traitement d'Images et Apprentissage Automatique pour la Viticulture de Précision

Volume 2, n° 1 (2021), p. 33-63.

http://roia.centre-mersenne.org/item?id=ROIA_2021__2_1_33_0

© Association pour la diffusion de la recherche francophone en intelligence artificielle et les auteurs, 2021, certains droits réservés.



Cet article est diffusé sous la licence
CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENSE.
<http://creativecommons.org/licenses/by/4.0/>



*La Revue Ouverte d'Intelligence Artificielle est membre du
Centre Mersenne pour l'édition scientifique ouverte*
www.centre-mersenne.org

Traitement d'Images et Apprentissage Automatique pour la Viticulture de Précision

Lucas Mohimont^a, Amine Chemchem^b, Marine Rondeau^c, Mathias Roesler^a, François Alin^a, Nathalie Gaveau^c, Luiz Angelo Steffene^a

^a Université de Reims Champagne Ardenne, Laboratoire LICIS - LRC CEA DIGIT, 51097 Reims Cedex 2, France

E-mail : lucas.mohimont@univ-reims.fr, mathias.roesler@univ-reims.fr, francois.alin@univ-reims.fr, angelo.steffene@univ-reims.fr

^b Pôle Intelligence Artificielle, ATOS. Rue du Mas de Verchant, 34000 Montpellier
E-mail : lamine.chemchem@atos.net

^c Université de Reims Champagne Ardenne, Laboratoire RIBP - USC INRAE 1488 51687 Reims Cedex 2, France

E-mail : marine.rondeau@univ-reims.fr, nathalie.gaveau@univ-reims.fr.

RÉSUMÉ. — Au cours des dernières décennies, les chercheurs ont mis au point des méthodes informatiques novatrices pour aider les viticulteurs à résoudre leurs problèmes, principalement liés à la prévision des rendements avant la récolte. L'objectif de cet article est de résumer les travaux de recherche existants liés aux techniques de vision par ordinateur qui traitent de ces questions. Il se concentre sur les approches utilisant des images RGB obtenues directement sur les parcelles, en utilisant différentes méthodes allant des algorithmes classiques de traitement d'image, en passant par les approches d'apprentissage automatique jusqu'à de nouvelles méthodes basées sur l'apprentissage profond. Nous avons l'intention de fournir un examen complet accessible également aux lecteurs non spécialisés, afin de discuter des progrès récents de l'intelligence artificielle dans la viticulture. À cette fin, nous présenterons dans les premières parties les travaux portant sur la détection des fleurs ainsi que des grappes et des raisins, puis dans la dernière partie nous présenterons les différentes méthodes d'estimation du rendement et les problématiques associées.

MOTS-CLÉS. — Viticulture, prédictions de rendement, apprentissage automatique, apprentissage profond.

1. INTRODUCTION

La viticulture de précision, l'application du concept d'agriculture de précision à la viticulture, est une méthode de gestion des parcelles qui a pour but d'optimiser les rendements et les coûts (intrants, travail requis, etc.) en prenant en compte la variabilité du milieu de culture [10, 112, 9]. Il s'agit donc d'une méthode de gestion plus fine qui intègre l'hétérogénéité des parcelles dans le processus de décision. La viticulture

de précision est rendue possible par de nombreuses technologies qui permettent d'acquérir de grandes quantités de données géolocalisées. L'objectif étant de contrôler le rendement de la récolte, éviter l'apparition et la prolifération des maladies de la vigne et obtenir des fruits de bonne qualité. Actuellement, un travail manuel est requis pour effectuer plusieurs tâches répétitives au vignoble : sélectionner les vignes et compter les grappes et les baies pour estimer le rendement, inspecter les vignes pour détecter les premiers signes de maladies, etc. Ces tâches peuvent être très chronophages car les parcelles sont très grandes, elles contiennent plusieurs milliers de pieds de vigne. De plus, les opérateurs humains peuvent faire de nombreuses erreurs (formation et compétences variables selon l'opérateur, erreur causée par le rythme de travail ou la fatigue). Les progrès scientifiques et technologiques permettent d'envisager une automatisation, au moins partielle, de ces tâches.

De très nombreuses technologies différentes ont été étudiées pour des applications pratiques. Par exemple, les réseaux de capteurs sans fil permettent d'acquérir des données à différents placements de la parcelle. Les données récoltées, telles que la température et l'humidité, sont utilisées pour prédire l'apparition des maladies [85]. Les capteurs peuvent inclure une caméra embarquée pour détecter la présence de symptômes de maladies ou de carences [66]. D'autres capteurs comme les lasers permettent d'estimer la taille de la canopée et le nombre de pieds manquants [39].

Les caméras RGB (Red, Green, Blue ou Rouge, Vert, Bleu en français), thermiques et spectrales ont également de nombreuses applications. Elles sont le plus souvent fixées à un véhicule pour couvrir une grande surface, et des drones télécommandés équipés de caméras ont ainsi été utilisés pour la détection des maladies [49], l'estimation du stress hydrique, de la vigueur et du nombre de pieds manquants [75] et pour l'estimation du rendement [28], avec l'avantage de pouvoir couvrir très vite l'ensemble d'une parcelle. Plusieurs robots dédiés à la viticulture ont également été proposés dans les années 2010. Ils sont équipés de caméras et de lampes, ce qui permet l'acquisition d'images de haute qualité dans le vignoble et à grande échelle. Ils permettent d'accomplir plusieurs actions sur le terrain comme : la détection des maladies [87], la taille automatique des vignes en hiver [17], l'estimation de la vigueur et la détection des fruits [48], la fertilisation des raisins [68], l'estimation de la taille des grappes [56], le phénotypage à grande échelle [52, 51] ou l'ensachage automatique des grappes de raisins [111]. Une veille technologique complète est disponible dans [76].

L'évolution technologique s'est accompagnée d'une évolution des techniques d'analyse d'images. L'apparition des appareils photos numériques et l'augmentation des capacités de stockage des disques durs ont permis le développement de l'imagerie numérique et de la vision par ordinateur. Les progrès de l'intelligence artificielle, et plus précisément des méthodes d'apprentissage automatique, permettent désormais d'analyser des scènes complexes et d'automatiser certaines tâches. Une tâche importante dans la viticulture est l'estimation du rendement. Elle est nécessaire pour organiser les vendanges et pour maintenir la qualité du raisin. Il s'agit alors de compter les raisins afin de faire une estimation de la récolte à venir. L'automatisation du comptage des raisins, et des fruits en général, est une des problématiques centrales de l'agriculture

connectée. De nombreuses méthodes ont été proposées ces dernières années. Certaines méthodes se basent sur une approche dite classique de l'analyse d'images qui consiste à développer des algorithmes de segmentation, de reconnaissance de formes et d'extraction de caractéristiques spécifiques à un problème donné. Cette approche a été appliquée à la détection des oranges [73], des poivrons [103] et des citrons [34]. Une autre approche se base sur l'apprentissage profond et plus particulièrement sur les réseaux de neurones convolutifs [57]. Ce type de réseau de neurones permet d'effectuer des tâches de classification, de segmentation ou de détection d'objet en apprenant automatiquement les bonnes représentations nécessaires à cette tâche. Cette approche se base donc sur les données disponibles plutôt que sur des critères subjectifs et des algorithmes spécifiques développés par un humain. L'apprentissage profond est très populaire depuis 2012 [55] et constitue désormais l'état de l'art des méthodes de classification et détection d'objets et des fruits [22, 90, 105, 65]. Les méthodes d'apprentissage profond pour l'agriculture ont été synthétisées par Kamilaris et al. [46] et Gikunda et al. [37].

Cette publication a pour objectif de synthétiser les différentes méthodes de détection et de comptage des fruits appliquées à la viticulture pour l'estimation du rendement. Cet article, relatant des méthodes basées sur de l'apprentissage profond, complète la revue de [98]. Nous présenterons tout d'abord les différentes problématiques de l'évaluation des méthodes ainsi que les métriques d'évaluation des performances. Nous présenterons ensuite les méthodes de détection des inflorescences et comptages des fleurs, puis nous présenterons les méthodes de détection des grappes et les méthodes de comptage des baies de raisins, et enfin nous présenterons les méthodes de modélisation nécessaire à l'estimation du rendement à partir de l'analyse d'images. Pour terminer, nous présenterons un bilan comparatif des méthodes classiques et l'apprentissage profond ainsi que les défis et perspectives des recherches en cours.

2. LA PROBLÉMATIQUE DE L'ÉVALUATION

2.1. L'ÉVALUATION DES MÉTHODES D'ESTIMATION DU RENDEMENT

La comparaison des travaux existants est difficile pour de nombreuses raisons. Les travaux existants sont limités à quelques cépages étudiés à un moment précis et dans des parcelles différentes partout dans le monde. Il y a donc de nombreuses variations : cépages, stades phénologiques et positionnement des vignes. Le nombre d'images prises et de pieds étudiés sont parfois très différents. Cette différence s'explique par la difficulté de l'acquisition des données. De très nombreuses images peuvent être prises mais uniquement dans une période limitée de l'année et sur un nombre de parcelles potentiellement restreint. L'étiquetage de ces images pose une problématique supplémentaire car cette étape est très chronophage. Et l'acquisition de données comme le nombre de grappes et de baies ainsi que la masse est limitée par la main-d'œuvre disponible. Il n'y a donc que peu d'études complètes portant sur plusieurs cépages, sur plusieurs centaines de pieds de vigne et sur plusieurs années. La dernière difficulté de comparaison est l'utilisation de métriques de performances différentes. De manière

générale, il n’y a pas de standards établis permettant une comparaison directe des performances des méthodes de détection des raisins et de l’estimation du rendement.

2.2. ÉVALUATION DES MÉTHODES DE DÉTECTION DES FRUITS

Le traitement d’images pour l’estimation du rendement de la vigne se base sur une des tâches suivantes : la classification d’images, la détection d’objets et la segmentation. La classification associe une classe à une image. La détection d’objet effectue une classification et une localisation, l’objet détecté étant entouré par une boîte englobante. La segmentation va plus loin en associant une classe pour chaque pixel de l’image. L’évaluation est effectuée en divisant la base de données en deux parties : une base d’apprentissage utilisée pour entraîner le modèle et une base de validation. Différentes métriques de performances sont calculées sur ces deux bases de données. La différence entre les performances d’apprentissage et de validation doit être minime (quelques pourcents) sinon cela indique un sur-ajustement (le modèle n’est pas applicable correctement sur de nouvelles images). Les métriques couramment utilisées sont : le rappel (le taux d’objets détectés), la précision (le taux d’objets détectés parmi l’ensemble des prédictions, le score F1 (ou F-mesure, la moyenne harmonique du rappel et de la précision) et l’exactitude (le taux de prédictions correctes).

Ces métriques sont calculées à partir de la matrice de confusion montrée dans la Figure 2.1. Le rappel, la précision et le F1 sont calculés pour chaque classe (la plupart des travaux mentionnés n’ont qu’une seule classe : les grappes, raisins ou fleurs). L’exactitude peut être faussée lorsque qu’il y a une classe minoritaire dans la base de données, le score F1 est alors préférable.

		Prédictions	
		Négatif	Positif
Observations	Négatif	Vrai négatif (TN)	Faux positif (FP)
	Positif	Faux négatif (FN)	Vrai positif (TP)

$$\text{Exactitude} = \frac{TP + TN}{TP + FN + TN + FP}$$

$$\text{Rappel} = \frac{TP}{TP + FN}$$

$$\text{Précision} = \frac{TP}{TP + FP}$$

$$F1 = \frac{2 \times \text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}}$$

FIGURE 2.1. Matrice de confusion et métriques de classification.

3. COMPTAGE DES FLEURS

Une première application présentée dans cette synthèse est la détection et le comptage des fleurs (un exemple est illustré dans la Figure 3.1). Il s’agit de compter les inflorescences (les futures grappes) et le nombre de fleurs (les futures baies). Le comptage des fleurs permet d’estimer très tôt le rendement potentiel. Il s’agit d’une estimation incertaine car certaines fleurs vont tomber (coulure) lors de la nouaison. Il s’agit donc d’une estimation potentielle maximale qui doit être affinée après la nouaison. La problématique de détection des fleurs est très proche de celle de la détection des

baies. Cette tâche est néanmoins plus difficile car les fleurs sont très petites, environ 3 mm de diamètre, et ont une couleur verte presque identique au feuillage. Plusieurs méthodes ont été proposées et la plupart nécessitent un contrôle partiel de l'environnement (fond artificiel, éclairage) pour rendre la tâche plus facile. Les différentes approches présentées dans cette partie sont résumées dans le Tableau 3.1.

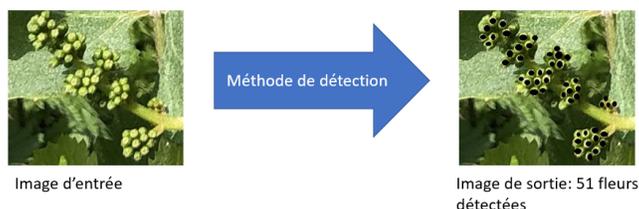


FIGURE 3.1. Méthode de détection des fleurs.

3.1. MÉTHODES CLASSIQUES - COMPTAGE PAR DÉTECTION DE POINTS-CLÉS

Une première approche consiste à détecter des fleurs candidates puis à les filtrer pour ne conserver que les vraies fleurs [7, 78, 30]. Un critère de détection courant est la réflexion de la lumière sur la surface des fleurs (ce critère est également utilisé pour les baies des grappes de raisin). Les fleurs réfléchissent la lumière et elles correspondent donc à des maxima locaux de l'image. Ces maxima sont détectables avec des algorithmes comme la « h-maxima transform » (détecte tous les maxima locaux supérieurs à un seuil h). Il y a néanmoins deux difficultés : le fond de l'image et le critère de détection qui produisent de nombreux faux positifs. Une manière simple de supprimer les faux positifs causés par le fond naturel de l'image consiste à utiliser un fond artificiel de couleur unie (noire dans le cas des fleurs). L'utilisation d'un fond noir simplifie fortement le traitement préliminaire de l'image, une méthode de segmentation simple comme la méthode d'Otsu permet alors de retirer le fond. Les autres faux positifs, causés par le critère de détection, sont retirés avec plusieurs méthodes de filtrage. Par exemple, les méthodes de [30, 78] utilisent la taille, la forme, et la distance des fleurs candidates comme critères de filtrage. Ces deux méthodes nécessitent donc un fond artificiel noir. La méthode proposée par [7] améliore la méthode de [30] en utilisant des opérateurs morphologiques spécialisés pour la suppression du fond naturel (un fond artificiel n'est donc plus nécessaire) et la détection des fleurs candidates.

Plusieurs faiblesses des travaux précédents ont été recensées par [60]. Les méthodes proposées par [30, 7, 78] dépendent toutes de paramètres choisis manuellement (la taille des opérateurs morphologiques par exemple). Ces méthodes sont donc sensibles à la couleur et à la taille apparente des fleurs. De plus, le motif de réflexion de la lumière utilisé comme critère de détection n'est pas robuste car il varie fortement selon le cépage. Une méthode a donc été proposée afin de corriger ces faiblesses [60]. Les tiges et les inflorescences sont détectées avec un algorithme de segmentation par croissance de régions à partir de leurs contours. Les fleurs potentielles sont sélectionnées avec la méthode de détection de points-clés génériques SURF. Elles sont filtrées avec

une méthode non-supervisée, la méthode des K-means, car les auteurs supposent qu'une méthode supervisée ne serait pas assez robuste aux variations causées par un éclairage naturel [59]. L'analyse de l'impact du stade phénologique montre que les performances de détection sont constantes après l'agglomération des boutons floraux (stade BBCH 55). Des travaux plus récents appliquent une méthode d'analyse de particule à la détection des fleurs [104]. Cette méthode possède les mêmes faiblesses que les travaux mentionnés par [60] car la taille apparente des fleurs doit être définie en pixels. De plus, les travaux de [60] et [104] n'adressent pas la problématique de la segmentation du fond et utilisent des fonds artificiels noirs.

3.2. APPRENTISSAGE PROFOND ET PERFORMANCES

À ce jour, deux méthodes d'apprentissages profonds ont été proposées pour le comptage des fleurs. Les premiers travaux publiés par [95] utilisent un modèle de segmentation de type Fully-Convolutional Network (FCN) pour détecter les inflorescences. Le modèle FCN a été proposé par [99] pour la segmentation sémantique. Il est une adaptation des réseaux de neurones convolutifs (CNN) classiques tel que Alex-Net [55] : les couches entièrement connectées sont remplacées par des couches de convolutions et d'up-sampling. FCN produit ainsi une sortie « dense », pixel par pixel, de manière efficace. Un des avantages de FCN est sa simplicité, l'apprentissage de bout-en-bout permet au modèle de construire automatiquement les bons extracteurs de caractéristiques nécessaires à la tâche souhaitée. Une grande quantité d'images étiquetées est nécessaire pour entraîner ces modèles. L'étiquetage est manuel ce qui est très chronophage. Il existe des outils pour automatiser partiellement l'étiquetage. Dans le cas de la segmentation sémantique, la création d'un premier masque peut être réalisé avec une méthode générant des super-pixels ou par segmentation watershed. Ces outils permettent de gagner du temps mais l'étiquetage reste malgré tout fastidieux car l'apprentissage profond nécessite des dizaines ou centaines, voire des milliers, d'images de haute résolution.

Un modèle FCN a donc été utilisé par [95] pour détecter les inflorescences dans les images prises en conditions naturelles. Cette méthode permet de supprimer le fond de l'image. Ensuite, la transformée de Hough est utilisée pour détecter les fleurs. Le modèle de segmentation proposé par [38] a également été appliqué à la détection des fleurs en conditions naturelles. Le modèle n'utilise pas d'algorithme supplémentaire et détecte directement les fleurs (représentées par des points dans les masques étiquetés). Les performances des méthodes proposées par [60] et [38] sont proches avec respectivement 84,3 % de taux de réussite et 86,7 % de score F1. La méthode de [38] a l'avantage de fonctionner directement en conditions naturelles mais l'évaluation réalisée par [60] semble plus robuste (validation croisée sur 12 bases de données). La méthode de [38] nécessite également un GPU de forte puissance pour l'apprentissage et l'application du modèle (entre 4 et 8 secondes de temps d'exécution par image, le temps de calcul pour [60] n'est pas précisé). Des performances similaires sont atteintes par les méthodes de [7], 85 % de rappel et 83,38 % de précision, de [78], 86 % de rappel et 84 % de précision, et [104], entre 12,3 % et 18,4 % d'erreur de comptage.

Un score plus faible est atteint par la méthode de [30] avec seulement 74,3 % de rappel et 92,9 % de précision. Les performances plus faibles observées par [30] sont causées par la complexité des images naturelles (le fond est artificiel mais l'éclairage est naturel) et le manque de robustesse de la méthode de détection (les autres travaux utilisent des méthodes plus robustes mais aussi plus complexes). Ces deux méthodes ont aussi les inconvénients d'utiliser des paramètres sélectionnés manuellement ce qui les rendent sensibles aux variations d'échelle (la taille des fleurs en pixels doit être connue). La méthode d'apprentissage profond proposée par [95] semble limitée par les performances de la transformée de Hough. Le comptage n'atteint que 75,2 % F1 avec le résultat de la segmentation et ne monte qu'à 80 % en utilisant les masques étiquetés. La combinaison du modèle de segmentation et d'un algorithme robuste, comme celui proposé par [60] par exemple, permettrait peut-être d'améliorer les performances et d'envisager une application en conditions réelles.

TABLE 3.1. Comparaison des différentes approches pour le comptage des fleurs

Approches	Matériel	Techniques	Résultats
Détection de maxima locaux ou de cercles [30, 7, 78, 104]	132 images réparties en 11 cépages, stade BBCH 53-57 (une inflorescence par image, avec fond artificiel) [78]	Détection de la réflexion de la lumière sur la surface des fleurs (h-maxima transform). [78]	Jusqu'à 86 % de rappel et 84 % de précision. Calibration et fond artificiel nécessaire (sensible à la couleur). [78]
Détection de points-clés génériques [60]	533 images réparties en 4 cépages, stade BBCH 18-61 (une inflorescence par image, avec fond artificiel)	Détection de fleurs candidates et filtrage non supervisé (Détecteur SURF et K-means).	84,3 % d'exactitude. Besoin d'un fond artificiel.
Apprentissage profond [95, 38]	30 images réparties en 2 cépages, stade BBCH 73 (plusieurs inflorescences par image, fond naturel) [38]	Segmentation sémantique avec un Fully-Convolutional Network (FCN) [38]	86,7 % F1. GPU nécessaire pour entraîner le modèle. [38]

4. DÉTECTION DES GRAPPES DE RAISINS

Une deuxième application présentée dans cette synthèse est la détection des grappes de raisins. L'algorithme de détection doit prendre en entrée une image contenant les grappes et nous donner en sortie une image contenant la localisation des grappes ainsi que leur comptage (voir Figure 4.1). Les différentes approches de détection des grappes sont résumées dans le Tableau 4.1.

Il s'agit d'une étape importante car elle permet de réaliser plusieurs applications pratiques comme l'estimation du rendement, la récolte automatique avec un

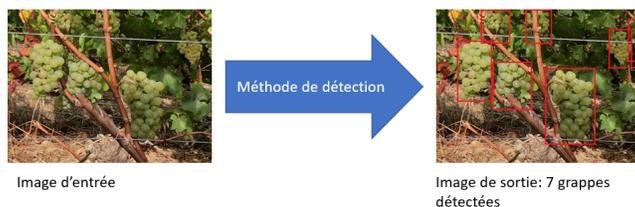


FIGURE 4.1. Méthode de détection des grappes de raisins.

robot [91, 71, 69], la pulvérisation automatique des hormones de croissances [16] et la caractérisation pour le phénotypage [94]. Cette tâche est potentiellement difficile pour plusieurs raisons : (i) il y a beaucoup de sources de variations (éclairage, distance, fond complexe) dans les images naturelles, (ii) le feuillage peut cacher les grappes et (iii) le feuillage est de couleur verte proche de celle des cépages blancs.

4.1. SEGMENTATION PAR SEUILLAGE

Une première approche simple utilise la segmentation par seuillage. Un ou plusieurs seuils sont choisis et sont appliqués sur les pixels pour ne conserver que les zones correspondantes aux fruits. La segmentation par seuillage est une méthode simple qui a été utilisée pour la première fois par [35] pour évaluer le potentiel de l'imagerie pour l'estimation du rendement. Les algorithmes de segmentation par seuillage ont l'avantage d'être très rapides en temps d'exécution et très simples à développer. Ces algorithmes ont néanmoins plusieurs limites importantes qui rendent leurs utilisations difficiles sur le terrain à moins de contrôler l'éclairage. La limite principale est l'utilisation de seuils stricts qui ne sont pas robustes aux changements de couleurs provoqués par l'éclairage naturel. Les images prises par [35] ne contenaient que des cépages noirs dans des plans fixes bien cadrés. Elle a été appliquée sur le terrain sur les cépages noirs et blancs mais uniquement la nuit avec un éclairage [91]. Elle a également été étudiée en laboratoire [41] et en conditions contrôlées avec un fond artificiel [107]. Une meilleure méthode de seuillage consiste à appliquer la distance de Mahalanobis pour comparer les pixels de l'image à des pixels de référence représentant les différentes classes. La classification d'un pixel est effectuée alors en choisissant la classe qui minimise la distance de Mahalanobis. Cette technique a été utilisée pour la détection des grappes et l'estimation de la surface foliaire [32], et pour l'estimation de la porosité, de l'exposition de la canopée, et des grappes [29]. Un contrôle de l'environnement, avec un fond artificiel ou l'utilisation d'un éclairage la nuit, est nécessaire dans les deux cas. Cette approche nécessite donc des conditions contrôlées pour ne pas être impactée par les variations d'éclairage et de fond. De plus, les seuils doivent être déterminés pour chaque situation. Cette approche est appropriée pour des études en laboratoire ou sur le terrain avec un véhicule adapté au phénotypage [51]. Elle peut également être employée en conditions naturelles sur des cas simples comme la détection des grappes de cépages noirs après défoliation [28]. Il faut cependant attendre que les grappes soient à maturité pour que la couleur soit discernable du sol et surtout du

feuillage. Pour finir, avec cette approche il est possible d'automatiser la segmentation par seuillage en utilisant un algorithme de classification non-supervisé comme Fuzzy C-Means ou K-Means [25]. Les faux positifs sont ensuite filtrés avec un classifieur SVM. Cette méthode reste sensible aux variations d'éclairage et n'a été évaluée que sur des cépages noirs avec un fond artificiel.

4.2. SEGMENTATION BASÉE SUR LES CONTOURS

Une deuxième approche utilise des méthodes de segmentation plus complexe. La segmentation par contours actifs a été proposée pour la détection des grappes de cépage blanc pour la récolte automatique [109]. Elle reste limitée à une utilisation la nuit avec un éclairage, ce qui permet de gommer le fond de l'image (le ciel, le sol et les rangs en arrière-plan). Les auteurs de [16] ont proposé une méthode hybride combinant seuillage binaire et détection de contours. Plus précisément, il s'agit de détecter les grappes de raisins en utilisant la densité des contours dans l'image comme critère de détection. Cette méthode nécessite néanmoins la détermination manuelle de plusieurs paramètres (taille des filtres de convolutions, seuils, etc.) pour fonctionner correctement. La segmentation par seuillage peut être combinée avec d'autres informations, comme la profondeur, pour obtenir de meilleurs résultats. La profondeur permet de discriminer plus facilement le premier plan et le fond de l'image. Les travaux de [54] combinent ainsi une carte de profondeur, construite avec une image stéréoscopique, une analyse des contours et une segmentation basée sur la couleur. Cette méthode a été évaluée en conditions naturelles mais uniquement sur des cépages noirs, ce qui facilite très fortement la détection des grappes.

4.3. CLASSIFICATION PAR PIXELS

Une troisième approche utilise des méthodes d'apprentissage classique pour rendre la détection des grappes plus robuste aux variations d'éclairage. Cela consiste à réaliser une segmentation pixel par pixel en utilisant les voisinages des pixels, ou blocs, comme entrée d'un modèle de classification.

Le modèle produit alors une sortie (grappe ou pas-grappe) qui sera appliquée au pixel central ou au bloc entier. Les méthodes d'apprentissage automatique « classiques » ne fonctionnent pas sur des images brutes, une étape d'extraction de caractéristiques est donc nécessaire pour le traitement de chaque bloc. Un exemple très simple serait d'utiliser la moyenne des canaux R, G et B pour produire un vecteur ne contenant que 3 composantes. En pratique, de très nombreuses méthodes sont disponibles et ont été appliquées à cette problématique. Cette approche a été proposée dès 2006 par [21] pour la récolte automatique des grappes. Les moments de Zernike, un ensemble de descripteurs très robustes, étaient alors utilisés comme caractéristiques pour entraîner un SVM. La moyenne des canaux RGB a été utilisée dans plusieurs études [19, 70, 36]. Un algorithme génétique a été proposé pour sélectionner les bons canaux de couleurs parmi plusieurs espaces possibles (RGB, HSV, CIELAB) [14]. Une autre caractéristique simple à calculer est l'histogramme de couleur, néanmoins elle reste limitée à des conditions contrôlées [61]. Des méthodes plus complexes, comme le

tenseur de structure locale associé à un classifieur bayésien, permettent de réaliser la segmentation des grappes et du feuillage en conditions semi-contrôlées (fond naturel, flash pour uniformiser l'éclairage) [1, 2]. Plus récemment, les CNN ont été utilisés pour réaliser à la fois l'extraction de caractéristiques et la classification [20, 74, 77].

Le choix des extracteurs est important car la qualité des caractéristiques va impacter le résultat final. Le choix de la méthode de classification importe moins, de bons résultats peuvent être obtenus avec un classifieur comme les SVM ou les perceptrons multicouches (MLP). Cette méthode souffre de plusieurs limites dont la sensibilité à la couleur (cépage) et le temps d'exécution (potentiellement long). Le temps d'exécution peut être raccourci mais cela demande un effort d'optimisation supplémentaire. Finalement, l'évaluation de cette approche est souvent limitée et n'est donc pas représentative des performances réelles de la méthode proposée. Cette approche est évaluée de deux façons en calculant les performances pour la classification des blocs et pour la détection des grappes. Or, certains travaux ne contiennent que les performances pour la classification des blocs [14, 101]. En pratique de très bonnes performances, au-delà de 95 %, peuvent être obtenues sur les blocs mais elles ne sont que rarement observées ensuite sur les images entières. Un écart d'environ 3 % est, par exemple, obtenu par [70] avec 96,56 % de taux de réussite sur les blocs contre 93,74 % taux de bonne détection pour les grappes individuelles.

Des optimisations sont possibles pour réduire le temps de calcul. Les travaux de [84] proposent une segmentation en deux étapes : (1) classification des pixels en sept classes avec un modèle de régression logistique et (2) un filtrage des faux positifs pour la classe « grappe » avec un modèle classique sac de mots, avec descripteurs SURF, et SVM. L'utilisation d'un classifieur simple directement sur les pixels permet de fortement réduire le temps de calcul et limite l'utilisation du modèle SVM à une zone restreinte de l'image. Cette méthode n'a été évaluée que dans un contexte très contrôlé : cépage noir et éclairage artificiel la nuit [84]. Cette approche a plusieurs limites. Elle n'exploite les voisinages des pixels que dans un contexte de petite taille (quelques dizaines de pixels au plus) et dépend fortement de la méthode d'extraction de caractéristique sélectionnée. Par exemple, certaines caractéristiques ne fonctionneront que sur les cépages noirs (plus facilement discernables par rapport au feuillage). Ces limites rendent certaines méthodes difficilement applicables en conditions réelles.

Finalement, cette approche peut être appliquée après une reconstruction 3D des vignes à partir d'images stéréoscopiques ou avec une méthode *structure-from-motion*. Il ne s'agit plus de classifier des pixels mais des points 3D. L'extraction de caractéristiques est plus complexe car il faut prendre en compte la profondeur des objets. La reconstruction 3D permettrait de caractériser plus précisément les grappes de raisins en estimant leur taille, leur volume, leur compacité, etc. Les travaux de [27] utilisent une méthode *structure-from-motion*, qui permet la reconstruction 3D à partir d'une séquence d'images et un classifieur SVM. Les travaux de [94] utilisent une caméra stéréoscopique et un classifieur Import Vector Machine, une amélioration du modèle SVM, pour détecter les grappes la nuit avec un éclairage artificiel. Malgré les conditions contrôlées, la méthode proposée n'atteint qu'au maximum 82 % de détection.

Ces performances peuvent s'expliquer par l'extracteur de caractéristiques, nécessaire pour traiter autant de variables, et le classifieur choisi. Elle nécessite surtout 10Go de stockage et 8h de temps calcul pour 25m de rang traité, ce qui la rend inutilisable sans optimisation conséquente. La reconstruction 3D peut être aussi effectuée à partir de vidéos prises par une caméra monoculaire, la reconstruction utilise alors le mouvement pour estimer la position 3D des objets (similaire à la *structure-from-motion* citée précédemment). Une méthode de cartographie et localisation simultanées (ou SLAM, Simultaneous Localization And Mapping) a été étudiée pour estimer le volume des grappes de raisins, et en déduire la taille des baies, en conditions partiellement contrôlées (fond et éclairage naturel, vignes défoliées) avec 93 % de réussite [96]. Cette méthode obtient de meilleurs résultats que [94] mais elle peut être impactée négativement par le vent (les objets doivent être immobiles dans la séquence traitée).

L'apprentissage automatique peut également être utilisé d'une autre manière. Une approche similaire à celle présentée lors de la partie précédente consiste à détecter des baies candidates puis à les filtrer avec un modèle de classification. Elle peut donc également être utilisée pour détecter les grappes. Elle a l'avantage de limiter le nombre de calculs en supprimant un maximum de pixels inutiles. Elle a été utilisée sur le terrain sur un cépage blanc [86] et nécessite un algorithme assez complexe mêlant détection de points-clés, classification et clustering. Elle a également été appliquée sur les raisins noirs [62]. Cette approche connaît les mêmes limites inhérentes à l'utilisation d'extracteurs de caractéristiques et de modèles d'apprentissage automatique classiques. Elle est aussi plus complexe car la méthode de détection des zones candidates doit être suffisamment performante pour retenir un maximum de vrais positifs avant le filtrage. De manière générale, ces approches sont complexes car elles mélangent plusieurs algorithmes qui nécessitent des réglages précis pour fonctionner.

4.4. APPRENTISSAGE PROFOND

L'apprentissage profond a été récemment appliqué à la problématique de la détection et comptage des grappes. Une approche naïve consiste à effectuer une classification bloc par bloc (ou pixel par pixel) avec un réseau de neurones convolutifs [67, 20, 77]. L'utilisation d'un CNN permet de simplifier fortement l'algorithme de détection des grappes car le modèle aura appris les bonnes caractéristiques à partir des données. Néanmoins cette approche est toujours limitée par la faible taille des blocs et le long temps de calcul requis pour segmenter une seule image. Plusieurs CNN ont été étudiés afin de prédire automatiquement la masse des grappes de raisins [100]. L'erreur de prédiction est relativement faible, 11 % en conditions contrôlées avec fond artificiel, mais la méthode proposée a plusieurs limites. La distance des grappes par rapport à la caméra varie ce qui impacte le résultat. Cette limite peut être dépassée en utilisant un capteur de profondeur, il faudrait également évaluer cette méthode sur plusieurs rangs et plusieurs cépages pour avoir une meilleure estimation de ses performances réelles. Plusieurs modèles populaires de détection d'objets, Faster R-CNN, R-FCN, et SSD, ont été appliqués à la détection des grappes et à leur comptage à partir de vidéos [43]. Le modèle Mask R-CNN [42], qui permet de détecter les objets et de les segmenter,

a également été appliqué pour la détection des grappes [97]. Le comptage à partir de vidéos est corrigé grâce à la méthode *structure-from-motion* qui permet d'estimer la position 3D à partir du mouvement de la caméra. La position 3D sert alors d'identifiant unique qui permet d'éviter les doublons lors du comptage.

4.5. COMPARAISON DES PERFORMANCES

Concernant les performances, l'approche basée sur le seuillage n'est pas applicable en situation réelle à moins d'imposer des contraintes fortes comme l'utilisation de l'éclairage la nuit [91]. Dans ces conditions 91 % des grappes blanches et 97 % des grappes noires sont détectées. Néanmoins, cette méthode reste limitée car les seuils pour chaque situation doivent être bien choisis. La méthode de [32] affiche des performances de comptage élevées avec 98 % de réussite. Ces performances sont à relativiser car l'évaluation n'a été effectuée que sur 10 vignes qui ont été défoliées progressivement pour obtenir plus d'images. Les performances obtenues par [29] sont proches de celle de [91] avec 92 % de score F1. Ces deux méthodes nécessitent également un éclairage la nuit ou un fond artificiel. La méthode de [16] obtient un taux de détection proche des méthodes précédentes avec 90 % de détection mais avec un très fort taux de faux positifs (70 %). Des performances similaires sont affichées par les méthodes de [109], avec 91,67 % de réussite, et [25] avec 95 %. De manière générale, cette approche est généralement limitée aux études effectuées en conditions contrôlées plutôt que pour une application sur le terrain : étude de la corrélation des pixels appartenant aux grappes / masses [35], comparaison de méthodes de détection 2D et 3D [41], ou estimation du nombre de grappes non visibles [107].

La deuxième approche permet des applications en conditions partiellement contrôlées. La méthode proposée par [21] atteint presque 100 % de réussite mais elle n'a été évaluée que sur 18 images de faible résolution. Un taux de détection d'environ 94 % en conditions naturelles, sur cépage noir, est présenté par [70]. Un taux similaire de 93 % est atteint par [19] mais les performances sont mauvaises sur les cépages blancs. Un rappel de 85 % et une précision de 93 % sont obtenus par [2] sur un cépage noir avant véraison et avec un flash. Il s'agit d'une amélioration de leur travaux précédents [1]. Les performances peuvent varier énormément selon les techniques sélectionnées et les conditions de capture des images. Un taux de détection de 90 % est atteint sur des images prises la nuit avec éclairage et contenant des grappes noires bien visibles. Un score très élevé de 99 % de réussite est obtenu par [14] mais il s'agit uniquement des performances de classification pour la base de données extraites pour l'apprentissage du classifieur. La méthode proposée n'a pas été évaluée sur les images complètes et le nombre de blocs utilisés pour la création et l'évaluation du modèle sont très limités (quelques dizaines de blocs). De manière générale, un taux de détection et une précision dépassant les 90 % sont envisageables avec un flash ce qui permet une application pratique. Les méthodes proposées par [86, 62] sont potentiellement moins performantes car elles dépendent de la précision de la méthode de segmentation préliminaire (ou de détection de point-clés) utilisée. Le taux de détection obtenu par la méthode de [62]

reste élevé avec 90 %, en conditions réelles, mais elle n'a été testée que sur des raisins noirs.

La méthode proposée par [86] n'obtient que 80,34 % de rappel (avec une bonne précision de 88 %) mais elle a été évaluée sur plusieurs bases de données différentes incluant des cépages noirs et blancs et dans des conditions naturelles ou contrôlées (nuit avec éclairage). Ces performances, plus faibles comparées à d'autres travaux, peuvent s'expliquer par la méthode de validation plus exhaustive et par la complexité de la méthode proposée. Un score F1, d'environ 80 %, est atteint par la méthode de [84]. Pour finir, la méthode de reconstruction 3D de [27] donne de bons résultats en conditions naturelles avec un AUC Roc de 0,98 avant la véraison et de 0,96 après la véraison. Le temps d'exécution de cette méthode n'est cependant pas précisé, un temps de calcul court et un faible besoin en mémoire faciliterait son utilisation pratique.

Concernant l'apprentissage profond, la méthode Mask R-CNN atteint 90 % de score F1 en conditions naturelles et sans éclairage artificiel [97]. Les meilleures performances sont obtenues par [43] avec 99 % de précision moyenne en conditions réelles. Un taux de réussite de 99 % est également affiché par [20] mais il s'agit des performances évaluées sur des blocs extraits des images. Les méthodes similaires proposées par [77] et [74] atteignent respectivement 91,52 % de réussite pour le comptage des grappes et 80,58 % de réussite de segmentation. Ces trois méthodes n'exploitent pas complètement le potentiel de l'apprentissage profond en effectuant la segmentation bloc par bloc. Cela peut s'expliquer par le manque de puissance de calcul et le manque de mémoire vive. En effet, un modèle de classification par bloc demande moins de puissance de calcul (et de mémoire) qu'un modèle spécialisé pour la segmentation sémantique. Malgré cette limitation, ces méthodes obtiennent de bonnes performances en conditions naturelles et sont capables de distinguer les grappes vertes du feuillage.

De manière générale, la puissance de calcul et le besoin énorme en données étiquetées sont les limites principales à l'utilisation de l'apprentissage profond. La plupart des bases de données ne sont pas suffisamment grandes et elles sont donc augmentées artificiellement en appliquant des transformations aléatoires pour créer de « nouvelles » images à partir des images réelles. De plus, les réseaux de neurones convolutifs appliqués sur des images nécessitent un accélérateur GPU puissant pour que l'apprentissage soit effectué en un temps raisonnable. Il faut donc une machine dotée à la fois de CPU et GPU puissants, ce qui rend la création de modèle d'apprentissage profond coûteuse et parfois inaccessible. Cette limite concerne également certaines méthodes de reconstruction 3D [94].

La détection des grappes de raisins est une première application pratique réalisable avec un bon niveau de performances, 90 % avec une méthode d'apprentissage automatique et un flash, allant jusqu'à 99 % avec l'apprentissage profond. La détection des grappes n'est qu'une étape préliminaire pour réaliser des tâches précises. Pour la récolte automatique par exemple, il est également nécessaire de détecter le point de coupe. Cette étape supplémentaire pose une difficulté car il faut pouvoir distinguer les grappes individuelles alors que certaines d'entre elles se superposent dans l'image. Une méthode a été proposée pour résoudre ce problème avec deux grappes [69]. La

segmentation des grappes et du feuillage a été appliquée à l'optimisation de l'usage des pesticides et des hormones de croissances [16]. Ces travaux montrent qu'une pulvérisation ciblée permettrait d'économiser 30 % des pesticides. La détection des grappes et la reconstruction 3D des vignes est actuellement utilisée pour la caractérisation pour le phénotypage car elles permettent de faire des mesures précises, comme le diamètre, le volume, ou la taille, des baies et grappes, et à grande échelle. En pratique, la reconstruction 3D stéréoscopique n'est pas utilisable à échelle industrielle car elle nécessite des ressources informatiques très importantes [94, 51]. La détection des grappes est également une première étape vers l'estimation du rendement. Il faut être capable de convertir le comptage des grappes en une estimation de la récolte. Des études préliminaires ont évalué des méthodes d'estimation visuelle de la masse des grappes, en laboratoire avec 5 % d'erreur [61], sur le terrain avec une corrélation de $0,87R^2$ [41], et avec de l'apprentissage profond avec 11 % d'erreur [100]. Cette méthode d'estimation visuelle n'est pas satisfaisante car elle dépend fortement de la distance de la caméra par rapport aux fruits. Les méthodes d'estimations du rendement se basent le plus souvent sur le comptage des raisins.

5. COMPTAGE DES BAIES

Le comptage des raisins est une des méthodes d'estimation du rendement actuellement utilisée. Les grappes sont prélevées parmi des échantillons aléatoires. Cette sélection permet d'estimer le nombre de grappes par vigne, le nombre de baies par grappe et la masse des baies. Ces composantes correspondent respectivement à 60 %, 30 % et 10 % de la variance du rendement [24], et permettent ensuite d'alimenter une équation du rendement qui sera appliquée à l'ensemble de la parcelle pour obtenir une estimation en kg/ha (ou kg/pied, hL/ha). Cette méthode est destructrice, les grappes doivent être récoltées ce qui limite le nombre d'échantillons. Cette limite rend les performances d'estimation très variables selon la parcelle et l'année. De nombreux travaux proposent des méthodes de comptage automatique des raisins grâce à la vision par ordinateur. Les travaux présentés dans cette partie sont résumés dans le Tableau 5.1. Un exemple de comptage des raisins est illustré dans la Figure 5.1. Un comptage rapide peut être obtenu en laboratoire avec un scanner [13], ou avec une caméra et un fond artificiel [53], mais toujours avec un échantillonnage destructeur. Une première méthode applicable sur le terrain a été proposée en 2010 et se base sur la détection d'ellipse avec un capteur doté d'une caméra et d'un fond artificiel [88]. Cette méthode était limitée à une grappe par image et n'obtenait pas de bons résultats pour le comptage, 30 % de réussite, car l'extraction des contours des baies provoquaient de nombreux faux négatifs (seuil de qualité trop élevé).

5.1. COMPTAGE PAR DÉTECTION DE POINTS CLÉS

De nombreuses autres méthodes de comptage ont été proposées. Elles consistent à détecter d'abord les baies puis à en déduire le nombre de grappes si nécessaire (ce qui correspond donc à 90 % de la variance du rendement). Il s'agit d'une problématique de détection de forme, plus précisément d'une détection de cercles. La réflexion spéculaire

TABLE 4.1. Comparaison des différentes approches pour la détection des grappes

Approches	Matériel	Techniques	Résultats
Segmentation par seuillage [35, 91, 41, 32, 29, 51, 107, 61, 28].	190 images, cépages et stades phénologiques non précisés. [91].	Seuillage basé sur la couleur [91].	97 % et 91 % d'exactitude sur cépages noirs et blancs. Éclairage la nuit requis(ou fond artificiel). [91].
Segmentation basée sur les contours [109, 16, 54, 25].	324 images prises le jour et 637 images prises la nuit avec éclairage. Cépage blanc, prise d'image depuis un véhicule avant la récolte. [109]	Segmentation Chan-Vese (segmentation par contours actifs) [109]	91,67 % d'exactitude. La nuit avec éclairage pour la récolte automatique. [109].
Classification par pixels avec une fenêtre glissante [21, 14, 62, 19, 70, 36, 1, 2, 74, 20, 102, 84].	200 images prises en conditions naturelles avec un éclairage variable (ciel couvert ou dégagé). Cépage noir, avant la récolte. [70].	Extraction de caractéristiques pour chaque voisinage de pixels et classification : Moyenne de couleur et classifieur Adaboost. [70].	93,74 % d'exactitude Calibration nécessaire pour les cépages blancs. Temps d'exécution potentiellement long. Performances limitées par la petite taille de l'entrée. [70].
Détection de régions d'intérêts et filtrage [62, 86].	163 images prises en conditions naturelles, provenant de base de données différentes. Cépages noirs et blancs, après la véraison. [86]	Algorithme en plusieurs étapes : détection de points clés, filtrage avec classification et clustering (SVM et DBSCAN). [86]	Rappel de 80 % et précision de 88 %. Performances limitées par le détecteur utilisé. Reproduction potentiellement difficile (calibration nécessaire). [86].
Reconstruction 3D et classification [27, 94, 96]	Acquisition sur 2 rangs de 25 vignes avec un véhicule, la nuit avec éclairage. Caméra stéréoscopique. Cépage blanc (Riesling). [94]	Reconstruction 3D et classification des points 3D avec Import Vector Machines [94]	82 % de rappel et 79 % de précision. Première étape vers un phénotypage à grande échelle. Temps d'exécution et besoin mémoire trop important. [94].
Apprentissage profond [101, 20, 77, 100, 97, 43].	300 images réparties en 5 cépages noirs et blancs. Lumière naturelle. [97]	Segmentation par instance(détection d'objets + segmentation) avec Mask R-CNN. [97]	91 % F1. Grande puissance de calcul requise pour entrainer le modèle. [97]

causée par la lumière produit un motif suivant une distribution gaussienne sur la surface des baies. Les baies ont alors l'apparence de petites sphères brillantes ce qui les rend plus facilement distinguables du fond de l'image si les conditions d'éclairage sont bonnes. Détecter les raisins consiste alors à détecter des cercles ou détecter des maximums locaux produits par la réflexion spéculaire. Une des premières méthodes utilisait un noyau gaussien pour traiter des images prises par un téléphone avec un

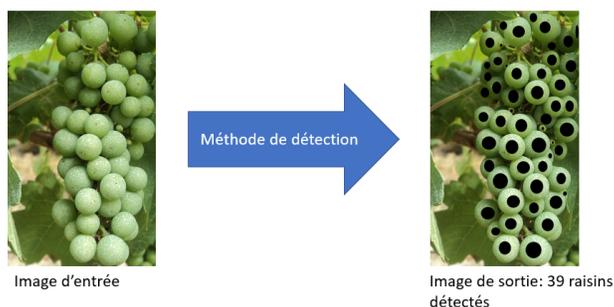


FIGURE 5.1. Méthode de comptage des baies de raisin.

flash [40]. Elle était limitée aux images centrées sur une ou plusieurs grappes afin d'éviter les faux positifs causées par le fond.

De nombreux algorithmes ont été utilisés pour détecter les raisins : la h-maxima transform [4, 5], la fast radial symmetry transform [83] et la transformée de Hough [80, 31, 63, 64, 93, 89, 47]. La fast radial symmetry transform permet de détecter les baies candidates rapidement alors que la transformée de Hough est potentiellement coûteuse en mémoire et temps de calcul et est sensible au bruit. Une alternative à la h-maxima transform, nommée invariant maxima detector, est proposé par [83] pour la détection des baies candidates avec un éclairage artificiel. Ces travaux montrent que les détecteurs basés sur les maxima locaux sont trop sensibles aux variations de l'éclairage naturel, l'utilisation d'un flash ou d'une lampe est donc recommandée. L'éclairage artificiel crée la réflexion spéculaire sur la surface des baies de manière uniforme ce qui facilite leur détection.

Cette approche nécessite une étape de filtrage supplémentaire similaire à celle abordée lors de la partie précédente. Cette approche classique est donc complexe car elle utilise plusieurs algorithmes qu'il faut paramétrer correctement pour (1) détecter les raisins candidats, (2) extraire les caractéristiques de chaque candidat et (3) filtrer les faux positifs avec un classifieur. Il faut donc sélectionner ces 3 algorithmes parmi les nombreuses méthodes existantes. Le choix de la méthode de détection des candidats est crucial car cette étape risque de limiter fortement le taux de détection final. Il faut donc choisir l'algorithme qui aura le plus fort taux de détection sans se soucier des faux positifs. L'éclairage est ici très important et un flash ou une lampe est nécessaire pour que le motif de réflexion spéculaire soit bien présent sur les raisins. La sélection des caractéristiques est également importante car elles permettent de discriminer les raisins du fond. Elles ne doivent pas se baser uniquement sur la couleur pour éviter la confusion entre les baies et le feuillage. Des caractéristiques de contour et de forme sont le plus souvent utilisées. De cette manière, il est possible de détecter les raisins sans utiliser de fond artificiel. Le choix du classifieur importe moins car des performances similaires peuvent être obtenues avec des modèles classiques, comme les SVM ou les MLP, si les caractéristiques extraites sont suffisamment discriminantes. De nombreux classifieurs courants ont été étudiés : les SVM [86, 102], les K-plus-proche-voisins [82],

les perceptrons multicouches [4, 5, 33], les random KD-Forest [83] et les réseaux de neurones convolutifs [101, 47]. Un taux de réussite très élevé, proche des 100 %, peut être obtenu avec un CNN. Et dans ce cas précis, l'apprentissage est rapide car le modèle est de taille réduite (entrée de 10x10 pixels). Pour finir, les travaux de [79] utilisent une approche inversée. Les grappes sont d'abord détectées avec la distance de Mahalanobis, la segmentation est affinée avec des filtres spécialisés puis un modèle booléen permet de détecter les baies. Le modèle booléen est plus robuste aux obstructions car capable de détecter les baies partiellement cachées. La méthode proposée nécessite néanmoins un contrôle de l'environnement (éclairage la nuit).

L'approche de détection de cercle peut également être transposée à la reconstruction 3D des vignes. Il s'agit de détecter des sphères parmi les points 3D. La transformée de Hough peut être appliquée en trois dimensions à cet effet [45]. Cette méthode a été évaluée sur une centaine de grappes dans des conditions très contrôlées (laboratoire avec fond artificiel et 4 lampes). Malgré ces bonnes conditions, la méthode proposée ne permet que de reconstruire 20 % des baies sans intervention humaine supplémentaire. Une calibration manuelle est donc nécessaire pour obtenir une reconstruction 3D correcte. La reconstruction 3D a également été appliquée sur le terrain [44, 94]. La méthode de [44] compare l'estimation du volume, de la masse et du nombre de baies obtenues avec une reconstruction automatique aux estimations obtenues de manière semi-automatique (une intervention humaine est nécessaire). La méthode d'estimation automatique est compétitive avec la méthode semi-automatique mais une calibration précise est nécessaire dans les deux cas pour que la reconstruction 3D soit correcte. De plus, cette méthode est potentiellement coûteuse car elle utilise cinq caméras. Cette approche permet donc de détecter les raisins sur le terrain à conditions de contrôler l'éclairage avec une source artificielle. L'utilisation d'un fond ou de l'éclairage la nuit n'est pas forcément indispensable dans ce cas précis.

5.2. APPRENTISSAGE PROFOND

L'apprentissage profond permet de simplifier le comptage en traitant directement les images brutes. Les CNN peuvent être utilisés comme méthode d'extraction de caractéristiques et comme classifieurs unifiés pour filtrer les baies candidates. Cette approche a été implémentée sur un Raspberry Pi [47] pour une utilisation en temps réel. Les résultats obtenus avec les CNNs ont été comparés à ceux obtenus avec les SVM et montrent l'avantage des CNNs [101]. Les CNNs ont été adaptés pour produire une carte de densité en sortie. Les zones de forte densité correspondent alors aux baies [26] et permettent le comptage avec une erreur d'environ 10 %. Les CNNs ont également été adaptés pour effectuer une segmentation à partir d'une image brute. Ces CNNs ont montré des résultats très prometteurs. Un modèle FCN a été proposé pour le comptage des baies et utilise un étiquetage en deux parties, une pour l'intérieur des raisins et une pour leur contour [110]. Le modèle proposé est plus rapide et plus performant que les modèles Mask R-CNN [42] et U-Net [92]. Cette différence s'explique par la structure des modèles, Mask R-CNN utilise une structure complexe pour détecter et segmenter les objets, et par la petite taille de la base de données.

5.3. COMPARAISON DES PERFORMANCES

Le résultat le plus encourageant est obtenu par [38]. Un seul modèle de segmentation a été évalué sur plusieurs tâches, comptage des baies, des fleurs et des rameaux, et obtient de bonnes performances avec un éclairage naturel et avec ou sans fond artificiel. De plus, de meilleures performances sont atteignables en optimisant la structure du modèle selon la tâche considérée et en utilisant plus d'images (seules 60 images de cépage noir ont été utilisées). Ce modèle utilise également une segmentation binaire plus simple (chaque baie est représentée par un point dans le masque étiqueté). L'apprentissage profond, Mask-RCNN plus précisément, a été combiné avec une technique de reconstruction 3D SLAM pour effectuer le comptage des baies en temps réel [81]. Une souffleuse est utilisée pendant la capture des images pour dégager le feuillage. Les performances affichées sont impressionnantes avec une précision de 96,62 % et un rappel de 98,81 % sur un rang de 187m. L'approche de reconstruction 3D utilisée par [94] avait également une bonne précision, environ 98 %, mais avec un rappel en dessous de 80 % et évalué sur seulement 20 grappes.

En termes de performances de détection, les résultats obtenus par [81] sont les plus prometteurs. Néanmoins, l'approche proposée reste complexe car elle combine plusieurs modèles différents et les cépages utilisés pour l'évaluation ne sont pas précisés (les images suggèrent l'utilisation sur un cépage noir). Une approche plus classique comme celle proposée par [5] permet d'atteindre 87,6 % de rappel et une bonne précision de 95,8 % mais dans des conditions contrôlées (nuit avec éclairage) pour limiter les faux positifs provoqués par le fond de l'image et la variabilité de l'éclairage.

Des résultats similaires sont obtenus par [110] dans des conditions naturelles. Les performances dépendent également du positionnement des vignes. Un écart de 5 % est observé entre le positionnement Vertical Shoot Positioning (VSP) et le positionnement Semi Minimal Pruned Hedges (SMPH) (94 % contre 89 % en rappel). Le modèle de segmentation étudié par [38] atteint 89,7 % et 86,6 % en conditions naturelles et sans optimisation des hyper-paramètres du modèle spécifique à cette tâche (le même modèle a été appliqué sur des tâches différentes). Ces résultats sont confortés par les travaux de [26] affichant entre 1 % et 10 % d'erreur de comptage en conditions naturelles. De manière générale, l'apprentissage profond semble plus robuste et permet de traiter les images en conditions naturelles. Cela peut s'expliquer par la capacité de l'apprentissage profond à traiter les images entières sans utiliser de méthode de détection de baies candidates. Les méthodes plus classiques sont limitées par le taux de détection de la méthode de sélection des candidats et ces performances peuvent varier énormément selon le cépage et l'année. Par exemple, la fast radial symmetry transform étudiée par [83] a un rappel qui varie entre 66 % et 89 %. Pour finir, la méthode proposée par [79] atteint seulement 78 % de réussite en moyenne sur quatre cépages noirs et en conditions contrôlées (fond artificiel ou éclairage la nuit). Les performances allant de 47 % à 88 % selon le cépage. Pour finir, la méthode proposée par [86] atteint une précision de 99 % et un rappel allant jusqu'à 92,4 % sur un cépage blanc en conditions naturelles. Ces performances ont été obtenus sur les bases de données publiées par [102] et qui contiennent des petits blocs de 40 x 40 pixels centrés

sur des baies individuelles ou le fond de l'image. Le comptage des baies individuelles n'a pas été évalué sur d'autres images (contrairement à la détection des grappes qui a été abordée dans la partie précédente). Les performances présentées par [86] ne sont donc pas représentatives des performances attendues sur des images complètes prises en conditions naturelles.

Nous n'avons abordé jusqu'à présent que le comptage des baies visibles. Or, l'estimation du rendement doit prendre en compte l'ensemble des baies. Les travaux de [107] montrent que le feuillage cache une majorité des grappes (jusqu'à 72 % pour du Syrah). Une étape de modélisation est donc nécessaire pour prendre en compte ces grappes cachées. Les différentes méthodes d'estimation du rendement sont décrites et comparées dans la partie suivante.

6. ESTIMATION DU RENDEMENT

6.1. LE POTENTIEL DE L'IMAGERIE NUMÉRIQUE POUR L'ESTIMATION DU RENDEMENT

La détection des raisins est une première étape pour l'estimation automatique du rendement. Il s'agit de convertir le résultat du comptage en une estimation en kg/pieds, kg/ha ou hL/ha. L'intérêt des méthodes d'analyse d'images est de pouvoir traiter rapidement de grandes quantités de données ce qui permet d'éviter la sélection aléatoire et destructrice d'échantillons sur le terrain. Les différentes approches de prédiction du rendement ont été synthétisées dans le Tableau 6.1. Le potentiel de ces techniques a d'abord été évalué en laboratoire [61, 31, 45, 41]. La corrélation entre le résultat de la méthode de détection, le nombre de baies ou le comptage/ratio des pixels des fruits, et la masse est alors établie dans des conditions idéales. Une forte corrélation, variant entre $0.69R^2$ et $0.95R^2$ selon le cépage ($0.69R^2$ pour du Grenache, $0.95R^2$ pour du Bobal, cépages noirs), est ainsi obtenue par [31] et confirme l'intérêt du comptage des baies pour la prédiction du rendement. Plusieurs travaux ont également montré une forte corrélation entre le nombre (ou ratio) de pixels appartenant aux fruits et la masse : $0.85R^2$ [35], $0.93R^2$ [61], $0.73R^2$ [32], $0.95R^2$ [41]. Une étude similaire a été réalisée à partir du comptage des fleurs et a obtenu une forte corrélation, $0.79R^2$, entre le nombre de baies estimé et la masse des grappes à la vendange [78].

L'évaluation de ces méthodes a été réalisée dans des conditions contrôlées, laboratoire ou terrain avec fond artificiel, qui ne sont donc pas représentatives des conditions naturelles. Une application terrain peut être partiellement contrôlée en utilisant une lampe ou un flash pour obtenir un éclairage uniforme. L'éclairage la nuit permet également de traiter la problématique de la suppression du fond de l'image car seules les vignes au premier plan sont éclairées. Cela nécessite néanmoins du matériel supplémentaire et l'utilisation exclusive la nuit n'est pas forcément souhaitable si un travail de nuit est nécessaire (à moins de disposer d'un robot autonome). De plus, l'utilisation d'un fond artificiel est à exclure pour un traitement à grande échelle.

TABLE 5.1. Comparaison des approches pour le comptage des baies de raisin.

Approches	Matériel	Techniques	Résultats
Scanner [53, 13, 88].	250 baies extraites de 8 grappes de Pinot Noir. Stade BBCH 75. Les baies sont placées sur un scanner à plat bon marché. [13]	Segmentation watershed et comptage de particules. [13]	100 % d'exactitude Procédure potentiellement longue. [13]
Détection des baies candidates et filtrage [40, 80, 31, 63, 47, 83, 82, 101, 102, 4, 5, 47, 86].	Prises d'images sur 30 vignes pour 5 cépages noirs et blancs (150 pieds en tout). Acquisition depuis un véhicule, la nuit avec éclairage. Stades BBCH 71-85 (de l'apparition des baies jusqu'à la pré-récolte). [5]	Détection de la réflexion de la lumière sur la surface des baies et filtrage des candidats par classification. [5]	87,6 % de rappel et 95,8 % de précision. La prise d'image doit se faire la nuit avec un éclairage artificiel. Calibration nécessaire pour chaque algorithme.[5].
Détection des grappes puis comptage des baies [79].	Prise d'images sur 3 cépages noirs avant la récolte ; 64 images de vignes prises la nuit avec un éclairage, depuis un véhicule.	Segmentation des grappes puis comptage des baies individuelles (distance de Mahalanobis et modèle booléen).	78 % d'exactitude La prise d'image doit se faire la nuit avec un éclairage artificiel. N'a été testé qu'au moment de la récolte.
Apprentissage profond [47, 101, 26, 110, 38].	Acquisition d'images avec un véhicule Phenoliner. Trois cépages blancs et noirs, 3 stades différents (non précisés), 38 images. [110].	Segmentation sémantique avec un modèle Fully Convolutional Network(FCN). [110].	Jusqu'à 92 % F1. Grande puissance de calcul requise pour l'entraînement du modèle. [110]
Reconstruction 3D [81] [94] [44]	750 images de cépage noir prises dans 20 rangs. Stade phénologique et cépage non précisé. [81]	Reconstruction 3D SLAM à partir de vidéo. Comptage des baies avec modèle Mask R-CNN, SVDD et SVM. [81]	98,81 % de rappel avec 96,62 % de précision. Traitement en temps réel potentiel. Grande puissance de calcul requise, sensible au vent, méthode complexe à calibrer. [81]

6.2. LES PROBLÉMATIQUES DU COMPTAGE DES RAISINS

Une problématique est posée par l'application des algorithmes de détection des raisins dans les parcelles. Ils doivent pouvoir être appliqués sur des séquences d'images permettant ainsi de couvrir des rangs entiers de vigne. Or, il y a un risque de redondance lors du comptage. Plusieurs solutions plus ou moins complexes ont été appliquées pour limiter ce problème. Une solution simple consiste à reconstruire l'image complète du rang à partir d'une vidéo puis à conserver un maximum de comptage dans les images qui se superposent [83]. La méthode SIFT permet ce genre de reconstruction en détectant les points-clés communs entre les images d'une séquence [25]. Il est aussi possible de

suivre les grappes détectées d'une image à l'autre en se basant sur la distance entre les grappes dans les images qui se suivent [43]. La reconstruction 3D peut également permettre d'éviter les doublons en utilisant les coordonnées 3D des fruits comme identifiant unique. Une approche photogrammétrique de type structure-from-motion a été utilisée à cet effet par [97]. Une reconstruction 3D temps-réel de type SLAM a été proposée par [81], ce qui permet d'éviter un traitement coûteux en temps de calcul. Finalement, une dernière solution mécanique consiste à capturer les images selon la distance parcourue par le véhicule (une image tous les 2 mètres par exemple) [5].

La problématique principale de l'estimation du rendement concerne l'étape de modélisation qui permet de convertir le comptage en récolte estimée. Plusieurs méthodes de régression ont été proposées afin de déterminer automatiquement la masse des grappes détectées dans une image. Par exemple, le nombre de pixels de chaque grappe peut être utilisé pour former un modèle linéaire de prédiction de la masse. L'apprentissage profond a été utilisé pour réaliser cette prédiction directement à partir de l'image [100]. En pratique, ces méthodes permettent surtout d'évaluer le potentiel de prédiction [35, 41] de l'analyse d'images mais ne sont pas toujours applicables à grande échelle en conditions réelles. Il est possible d'obtenir un taux d'erreur d'environ 16 % mais dans un contexte simplifié : cépage noir après véraison et vignes partiellement défoliées pour rendre les grappes bien visibles [28].

L'utilisation du comptage des pixels, ou d'images RGB sans profondeur, rend ces méthodes très sensibles aux variations de distances. Ces méthodes ne permettraient pas de faire la différence entre une grappe lointaine ou une grappe trop petite. Cette limite peut être dépassée en utilisant un capteur de profondeur [67] ce qui permet de mieux évaluer la taille des grappes. Une autre limite de ces méthodes est la problématique principale de l'estimation du rendement à partir du comptage : il n'est effectué que sur la partie visible. Dans ce cas, un modèle qui prédit directement la masse sans effectuer de comptage ne permet pas de résoudre cette problématique et est plus difficile à interpréter. Les deux méthodes proposées par [67] et [100] sont prometteuses avec respectivement 11,8 % et 15,2 % d'erreur de prédiction du rendement. Elles ont néanmoins été évaluées que sur un nombre très limité de vignes. Un taux d'erreur de prédiction de 16 % a également été obtenu avec la méthode de [36] mais elle ne fonctionne que la nuit avec un éclairage et sur des cépages noirs.

L'approche de modélisation classique consiste à estimer le nombre total de baies (ou de fleurs) avec un modèle de régression, généralement un modèle linéaire. Le nombre total estimé est alors converti en masse en utilisant la masse moyenne historique des raisins ou en faisant des prélèvements sur le terrain. Cette approche est utilisée par [5] avec un modèle linéaire qui prédit la masse totale à partir de la masse « visible ». Une erreur moyenne de 12,83 % a été obtenue sur 5 cépages répartis dans 30 segments de 3 vignes. Plusieurs étapes de calibration ont été proposées par [83] afin de corriger les erreurs de comptage. Deux méthodes alternatives ont été testées pour estimer le nombre total de baies à partir de la partie visible de la grappe : (1) en utilisant l'enveloppe convexe de la grappe visible et en fixant une taille moyenne des baies et de l'épaisseur de la grappe ou (2) en estimant le nombre de baies contenues dans

un modèle 3D d'ellipsoïde calculé à partir de la partie visible. La première méthode alternative a obtenu les meilleures performances de prédiction de masse des fruits, 13,7 % contre 15,4 % pour la régression linéaire et 17 % pour le modèle ellipsoïde, en laboratoire. Ces deux méthodes n'ont pas pu fonctionner correctement en conditions réelles car elles nécessitent que les grappes soient bien séparées les unes des autres. La méthode proposée par [83] utilise une modélisation de l'erreur des années précédentes et une estimation de la visibilité des fruits pour corriger le comptage final. De plus, la méthode est optimisée en sélectionnant les paramètres permettant d'atteindre l'erreur globale et l'erreur spatiale la plus faible. L'erreur d'estimation est de 6,48 %, 9,07 % ou 11,65 % sur le cépage Flame Seedless (cépage noir) selon l'année de calibration utilisée pour corriger l'erreur de comptage. Une erreur très faible de -2,47 % est également atteinte sur du Chardonnay (cépage blanc). Le taux d'erreur sur le premier cépage est à relativiser car les images ont été prises 7 jours avant la récolte (et 75 jours pour le Chardonnay, 100 jours pour [5]). En pratique, les vignerons souhaiteraient obtenir un taux d'erreur le plus faible possible plusieurs semaines avant la récolte. Pour corriger le comptage, [79] propose d'utiliser un modèle booléen robuste aux occlusions partielles (certains raisins ne sont que partiellement visibles et ne sont pas forcément détectés). L'erreur d'estimation par pied de vigne est alors beaucoup plus faible par rapport à une méthode naïve basée sur le comptage des pixels. Elle reste néanmoins un peu plus élevée, 200 g/pieds, comparée à la méthode de [5] qui affichait une erreur de 160 g/pieds.

6.3. AVANCÉES RÉCENTES ET PROBLÉMATIQUES À RÉSOUDRE

Une méthode récente utilise une nouvelle technique de comptage des raisins pour estimer le nombre total de baies dans une grappe [64]. Il s'agit d'une méthode de reconstruction 3D fonctionnant à partir d'une seule image 2D. Le modèle 3D est construit en positionnant les baies de manière à remplir une enveloppe estimée à partir des contours de la grappe. Un paramètre est utilisé pour modéliser l'espacement entre les baies. Cette méthode permet donc d'estimer directement le comptage total des baies dans une grappe. Elle reste limitée à une grappe par image photographiée avec un fond artificiel. Une estimation du rendement a été effectuée en combinant cette méthode à une autre méthode proposée par les mêmes auteurs [59] qui se basait sur le comptage des rameaux. L'erreur d'estimation sur les trois parcelles étaient de 3 %, 6 % et 16 %, ce qui est comparable aux performances affichées par les autres méthodes.

L'estimation du nombre total de baies dans une grappe et des baies cachées par le feuillage sont les deux problématiques principales qui limitent actuellement les performances de prédiction du rendement. L'estimation du nombre total de baies nécessite une étape de modélisation, comme celle proposée par [64] ou [83], évaluée sur différents cépages, parcelles, et saisons. L'estimation du nombre de baies/grappes cachées par le feuillage est une étape très importante car il n'est pas envisageable pour le vigneron d'effeuiller les vignes. Une solution possible est l'utilisation d'un souffleur pour dégager le feuillage mais elle n'a pas été évaluée quantitativement [81]. Une étape de modélisation supplémentaire est donc nécessaire et pourrait utiliser des

variables supplémentaires comme la porosité de la canopée (corrélation de $0,82R^2$ entre la porosité de la canopée et le pourcentage de grappes visibles) [107].

La plupart des travaux mentionnés effectuent la prise d'images à un seul moment. Or, le stade phénologique peut impacter les résultats. Les baies peuvent être plus difficiles à compter lorsqu'ils sont plus petits. Leurs couleurs sont également très proches de celles du feuillage avant la véraison. En pratique de bons résultats ont été reportés par les auteurs de Nuske et al 2014, jusqu'à 75 jours avant la récolte [83]. Le comptage peut être plus difficile également après la fermeture de la grappe (les baies se touchent). Les travaux de Liu et al. 2020 montrent des performances d'estimation du rendement similaires avant et après la véraison [64].

TABLE 6.1. Comparaison des différentes approches pour l'estimation du rendement.

Approches	Matériel	Techniques	Résultats
Prédiction directe de la masse des grappes [67, 100, 28].	40 vignes sont photographiées de trois côtés différents, 120 images en tout. Pinot noir après la véraison et après défoliation. Un fond artificiel est placé derrière chaque pied.[67]	La masse des grappes visibles est prédite directement à partir de l'image. Par exemple avec un modèle CNN. [67]	11,8 % d'erreur d'estimation du rendement. Amélioration possible avec un capteur de profondeur. Cette méthode ne prend pas en compte les raisins cachés par le feuillage. [67].
Comptage des baies et modélisation [5, 83, 36, 79].	Prises d'images de 24 vignes de Chardonnay (cépage blanc) en 2013, 75 jours avant la récolte, la nuit avec éclairage. Caméra montée sur un véhicule. Protocole similaire en 2011 avec 636 vignes. [83]	Comptage des baies visibles, estimation du nombre total avec un modèle (le plus souvent linéaire) et calcul du rendement. [83]	-2,47 % d'erreur de prédiction du rendement en 2013(calibration avec les données de 2011). Algorithme applicable la nuit avec éclairage seulement. La correction du comptage pour prendre en compte l'occlusion et les baies cachées est une problématique non résolue. [83]
Reconstruction automatique de la structure des grappes et comptage des rameaux [59, 64].	Prises d'images de grappes de cépage noirs et blancs à différents stades phénologiques (avant et après véraison). Une grappe par image, un fond artificiel est placé derrière chaque grappe.	Reconstruction 3D de grappes de raisins et comptage des rameaux. [59].	Erreur d'estimation du rendement de 3 %, 6 % et 16 % sur trois parcelles (contre 3 %, 10 % et 24 % avec une méthode manuelle). Méthode possiblement difficile à appliquer à grande échelle. Ne prends pas en compte les grappes cachées par le feuillage.

7. CONCLUSION

La première grande approche utilise des méthodes classiques de traitement d'images d'apprentissage automatique. Elle a été appliquée dans de nombreux travaux portant sur la viticulture. Sa limite principale est la difficulté à choisir et calibrer les bons algorithmes d'extraction de caractéristiques, de détection de formes et de classification. Elle nécessite la plupart du temps un environnement partiellement contrôlé avec un fond artificiel [64], un traitement la nuit avec éclairage [5] ou l'utilisation d'un flash [83]. Certaines méthodes sont disponibles librement sous la forme d'applications mobiles [6, 58, 3].

La seconde approche utilise l'apprentissage profond pour automatiser la détection des fruits. Cette approche remplace donc la sélection manuelle des seuils de segmentation et des méthodes d'extraction de caractéristiques car les réseaux de neurones convolutifs sont capables d'apprendre seuls à partir d'images étiquetées. De plus, la communauté travaillant sur l'apprentissage profond est très active et de nombreux outils permettent d'entraîner un modèle facilement et parfois sans aucune programmation. L'apprentissage profond a néanmoins deux inconvénients : il nécessite (1) une grande quantité de données étiquetées et (2) une grande puissance de calcul. L'étiquetage manuel, ou semi-automatique, est une étape nécessaire à l'apprentissage des modèles supervisés et est très chronophage. De plus, l'apprentissage requiert au minimum un GPU afin d'accélérer fortement les calculs. La plupart des méthodes classiques sont bien plus abordables en terme de puissance de calcul et de quantité de données requises.

De nombreuses améliorations sont possibles. Une première amélioration serait une plus grande transparence des travaux existants via la publication des données ainsi que la normalisation des méthodes d'évaluation. Actuellement le benchmark le plus complet est le White Grape Instance Segmentation Dataset publié par Santos et al. [97]. Il contient 300 images réparties en 6 cépages noirs et blancs avec un étiquetage binaire (localisation des grappes ou des baies). Un benchmark plus complet devrait être établi sur plusieurs pieds de vigne, de différents cépages, pendant plusieurs années et à des stades phénologiques différents. Il devrait également inclure les données permettant la modélisation du rendement (nombre de baies, nombre de grappes, masses). La construction d'un tel benchmark demanderait beaucoup de main-d'œuvre et un grand effort d'étiquetage. Il permettrait néanmoins d'étudier plus précisément l'impact du stade phénologique sur la prédiction du rendement.

Le comptage des fruits en général est une tâche difficile car l'aspect de la plante peut varier énormément selon la prise de vue. Une prise de vue multiple, de chaque côté de la plante, permet d'affiner le comptage des mangues [11] et des pommes [108]. Les auteurs de Bargoti et Underwood 2016 font l'hypothèse que le taux d'occlusion des fruits est constant, un modèle linéaire permet alors de corriger le comptage visible des pommes [12]. Il y a également une problématique similaire d'étiquetage des images. Une méthode faiblement supervisée, l'étiquetage est limitée à une classification binaire indiquant la présence de fruits dans l'image, a été appliqué aux comptages des amandes et des pommes [15]. Le modèle est néanmoins limité à une tâche de régression (comptage uniquement). Une piste d'amélioration à explorer est l'ajout de nouveaux prédicteurs dans la phase de modélisation du rendement. Actuellement, les modèles n'utilisent que des prédicteurs liés au comptage des raisins et ignorent le feuillage. Le feuillage peut être détecté avec un modèle de segmentation par exemple [23]. L'extraction de nouveaux prédicteurs à partir des images nécessite un étiquetage plus complexe et plus chronophage. Des modèles d'apprentissage profond basés sur l'apprentissage auto-supervisé et utilisant la technique de l'attention [106], permettrait de réduire le nombre d'images à étiqueter. Les travaux de Caron et al 2021 montrent que ce type de modèle apprend à délimiter les objets dans les images de manière non supervisée [18]. L'ajout de nouveaux prédicteurs, comme le volume des grappes, est une piste à explorer du côté des méthodes de reconstruction 3D. Le coût

élevé du matériel et du traitement des données est actuellement un frein pour ce type de méthodes. Finalement, l'ajout de nouveaux indicateurs pourraient être effectués en croisant différentes sources. Par exemple, les images satellites ont été récemment utilisées pour prédire le rendement de la vigne [8]. De nouveaux modèles seront alors nécessaires pour prendre en compte des données de différentes natures (images RGB, images satellites, séries temporelles, etc.).

REMERCIEMENTS

Ce travail a été développé dans le cadre du projet H2020 AI4DI - Artificial Intelligence for Digitizing Industry (<https://ai4di.eu/>) sous contrat no. 826060.

BIBLIOGRAPHIE

- [1] F. ABDELGHAFOUR, B. KERESZTES, C. GERMAIN & J. P. DA COSTA, « Potential of on-board colour imaging for in-field detection and counting of grape bunches at early fruiting stages », *Advances in Animal Biosciences* **8** (2017), n° 22, p. 505-509.
- [2] F. ABDELGHAFOUR, R. ROSU, B. KERESZTES, C. GERMAIN & J. P. DA COSTA, « A Bayesian framework for joint structure and colour based pixel-wise classification of grapevine proximal images », *Computers and Electronics in Agriculture* **158** (2019), p. 345-357.
- [3] A. AQUINO, I. BARRIO, M.-P. DIAGO, B. MILLAN & J. TARDAGUILA, « vitisBerry : An Android-smartphone application to early evaluate the number of grapevine berries by means of image analysis », *Computers and Electronics in Agriculture* **148** (2018), p. 19-28.
- [4] A. AQUINO, M. P. DIAGO, B. MILLÁN & J. TARDÁGUILA, « A new methodology for estimating the grapevine-berry number per cluster using image analysis », *Biosystems Engineering* **156** (2017), p. 80-95.
- [5] A. AQUINO, B. MILLAN, M.-P. DIAGO & J. TARDAGUILA, « Automated early yield prediction in vineyards from on-the-go image acquisition », *Computers and Electronics in Agriculture* **144** (2018), p. 26-36.
- [6] A. AQUINO, B. MILLAN, D. GASTON, M.-P. DIAGO & J. TARDAGUILA, « vitisFlower : Development and Testing of a Novel Android-Smartphone Application for Assessing the Number of Grapevine Flowers per Inflorescence Using Artificial Vision Techniques », *Sensors* **15** (2015), n° 9, p. 21204-21218.
- [7] A. AQUINO, B. MILLAN, S. GUTIÉRREZ & J. TARDÁGUILA, « Grapevine flower estimation by applying artificial vision techniques on images with uncontrolled scene and multi-model analysis », *Computers and Electronics in Agriculture* **119** (2015), p. 92-104.
- [8] S. T. ARAB, R. NOGUCHI, S. MATSUSHITA & T. AHAMED, « Prediction of grape yields from time-series vegetation indices using satellite remote sensing and a machine-learning approach », *Remote Sensing Applications : Society and Environment* **22** (2021), article no. 100485.
- [9] J. ARNÓ, M. CASASNOVAS, M. RIBES-DASI & J. R. ROSELL, « Review. Precision Viticulture. Research topics, challenges and opportunities in site-specific vineyard management », *Spanish Journal of Agricultural Research* **7** (2009), p. 779-790.
- [10] D. ARROUAYS, J. C. BEGON, B. B. NICOUILLAUD & C. LE BAS, « La variabilité des milieux, une réalité : de la région à la plante », *Perspectives Agricoles* (1997), n° 222, p. 8-12.
- [11] P. A. B., K. B. WALSH, P. P. SUBEDI & D. JARVIS, « Estimation of mango crop yield using image analysis – Segmentation method », *Computers and Electronics in Agriculture* **91** (2013), p. 57-64.
- [12] S. BARGOTI & J. UNDERWOOD, « Image Segmentation for Fruit Detection and Yield Estimation in Apple Orchards », *Journal of Field Robotics* **34** (2016), p. 1039-1060.
- [13] M. BATTANY, « A Practical Method for Counting Berries based on Image Analysis », *Proceedings of the 2nd Annual National Viticulture Research Conference* (2008), p. 4-5.
- [14] N. BEHROOZI-KHAZAEI & M. R. MALEKI, « A robust algorithm based on color features for grape cluster segmentation », *Computers and Electronics in Agriculture* **142** (2017), p. 41-49.

- [15] E. BELLOCCHIO, T. A. CIARFUGLIA, G. COSTANTE & P. VALIGI, « Weakly Supervised Fruit Counting for Yield Estimation Using Spatial Consistency », *IEEE Robotics and Automation Letters* **4** (2019), n° 3, p. 2348-2355.
- [16] R. BERENSTEIN, O. B. SHAHAR, A. SHAPIRO & Y. EDAN, « Grape clusters and foliage detection algorithms for autonomous selective vineyard sprayer », *Intelligent Service Robotics* **3** (2010), n° 44, p. 233-243.
- [17] T. BOTTERILL, S. PAULIN, R. GREEN, S. WILLIAMS, J. LIN, V. SAXTON, S. MILLS, X. CHEN & S. CORBETT-DAVIES, « A Robot System for Pruning Grape Vines », *Journal of Field Robotics* **34** (2017), n° 6, p. 1100-1122.
- [18] M. CARON, H. TOUVRON, I. MISRA, H. JÉGOU, J. MAIRAL, P. BOJANOWSKI & A. JOULIN, « Emerging Properties in Self-Supervised Vision Transformers », <https://arxiv.org/abs/2104.14294>, 2021.
- [19] V. CASSER, « Using Feedforward Neural Networks for Color Based Grape Detection in Field Images », in *CSCUBS 2016 - Computer Science Conference for University of Bonn Students*, 2016, p. 23-33.
- [20] H. CECOTTI, A. RIVERA, M. FARHADLOO & M. A. PEDROZA, « Grape detection with convolutional neural networks », *Expert Systems with Applications* **159** (2020), article no. 113588.
- [21] R. CHAMELAT, E. ROSSO, A. CHOKSURIWONG, C. ROSENBERGER, H. LAURENT & P. BRO, « Grape Detection By Image Processing », in *IECON 2006 - 32nd Annual Conference on IEEE Industrial Electronics*, Nov 2006, p. 3697-3702.
- [22] S. W. CHEN, S. S. SHIVAKUMAR, S. DCUNHA, J. DAS, E. OKON, C. QU, C. J. TAYLOR & V. KUMAR, « Counting Apples and Oranges With Deep Learning : A Data-Driven Approach », *IEEE Robotics and Automation Letters* **2** (2017), n° 2, p. 781-788.
- [23] H. CHENG, L. DAMEROW, Y. SUN & M. BLANKE, « Early Yield Prediction Using Image Analysis of Apple Fruit and Tree Canopy Features with Neural Networks », *Journal of Imaging* **3** (2017), n° 1, article no. 6 (13 pages).
- [24] P. R. CLINGELEFFER, S. R. MARTIN, G. M. DUNN & M. P. KRSTIC, *Crop development, crop estimation and crop control to secure quality and production of major wine grape varieties : a national approach*, Adelaide, Grape and Wine Research and Development Corporation, 2001.
- [25] C. CORREA, C. VALERO & P. BARREIRO, « Characterization of Vineyard's Canopy through Fuzzy Clustering and SVM over Color Images », in *3rd CIGR International Conference of Agricultural Engineering (CIGR-AgEng2012)*, vol. 1, 2012.
- [26] L. COVIELLO, M. CRISTOFORETTI, G. JURMAN & C. FURLANELLO, « GBCNet : In-Field Grape Berries Counting for Yield Estimation by Dilated CNNs », *Applied Sciences* **10** (2020), n° 1414, article no. 4870 (15 pages).
- [27] D. DEY, L. MUMMERT & R. SUKTHANKAR, « Classification of plant structures from uncalibrated image sequences », in *2012 IEEE Workshop on the Applications of Computer Vision (WACV)*, Jan 2012, p. 329-336.
- [28] S. F. DI GENNARO, P. TOSCANO, P. CINAT, A. BERTON & A. MATESE, « A Low-Cost and Unsupervised Image Recognition Methodology for Yield Estimation in a Vineyard », *Frontiers in Plant Science* **10** (2019), article no. 559.
- [29] M. P. DIAGO, A. AQUINO, B. MILLAN, F. PALACIOS & J. TARDAGUILA, « On-the-go assessment of vineyard canopy porosity, bunch and leaf exposure by image analysis », *Australian Journal of Grape and Wine Research* **25** (2019), n° 3, p. 363-374.
- [30] M. P. DIAGO, A. SANZ-GARCIA, B. MILLAN, J. BLASCO & J. TARDAGUILA, « Assessment of flower number per inflorescence in grapevine by image analysis under field conditions », *Journal of the Science of Food and Agriculture* **94** (2014), n° 10, p. 1981-1987.
- [31] M. P. DIAGO, J. TARDAGUILA, N. ALEIXOS, B. MILLAN, J. M. PRATS-MONTALBAN, S. CUBERO & J. BLASCO, « Assessment of cluster yield components by image analysis », *Journal of the Science of Food and Agriculture* **95** (2015), n° 66, p. 1274-1282.
- [32] M.-P. DIAGO, C. CORREA, B. MILLÁN, P. BARREIRO, C. VALERO & J. TARDAGUILA, « Grapevine Yield and Leaf Area Estimation Using Supervised Classification Methodology on RGB Images Taken under Field Conditions », *Sensors* **12** (2012), n° 12, p. 16988-17006.
- [33] P. DOLEZEL, P. SKRABANEK & L. GAGO, « Detection of grapes in natural environment using feedforward neural network as a classifier », in *2016 SAI Computing Conference (SAI)*, Jul 2016, p. 1330-1334.

- [34] U.-O. DORJ, M. LEE & S. SEOK YUN, « An yield estimation in citrus orchards via fruit detection and counting using image processing », *Computers and Electronics in Agriculture* **140** (2017), p. 103-112.
- [35] G. M. DUNN & S. R. MARTIN, « Yield prediction from digital image analysis : A technique with potential for vineyard assessments prior to harvest », *Australian Journal of Grape and Wine Research* **10** (2004), n° 33, p. 196-198.
- [36] D. FONT, M. TRESANCHEZ, D. MARTÁNEZ, J. MORENO, E. CLOTET & J. PALACÁN, « Vineyard Yield Estimation Based on the Analysis of High Resolution Images Obtained with Artificial Illumination at Night », *Sensors* **15** (2015), n° 4, p. 8284-8301.
- [37] P. K. GIKUNDA & N. JOUANDEAU, « State-of-the-Art Convolutional Neural Networks for Smart Farms : A Review », in *Intelligent Computing* (Cham) (K. Arai, R. Bhatia & S. Kapoor, éd.), Springer International Publishing, 2019, p. 763-775.
- [38] J. GRIMM, K. HERZOG, F. RIST, A. KICHERER, R. TÖPFER & V. STEINHAGE, « An adaptable approach to automated visual detection of plant organs with applications in grapevine breeding », *Biosystems Engineering* **183** (2019), p. 170-183.
- [39] B. GROCHOLSKY, S. NUSKE, M. AASTED, S. ACHAR & T. BATES, « A Camera and Laser System for Automatic Vine Balance Assessment », in *American Society of Agricultural and Biological Engineers Annual International Meeting 2011, ASABE 2011*, vol. 7, Jul 2011.
- [40] M. GROSSETETE, Y. BERTHOUMIEU, J.-P. DA COSTA, C. GERMAIN, O. LAVIALLE & G. GRENIER, « Early Estimation of Vineyard Yield : site specific counting of berries by using a smartphone », in *International Conference on Agriculture Engineering (AgEng)*, Jul 2012.
- [41] C. HACKING, N. POONA, N. MANZAN & C. POBLETE-ECHEVERRÍA, « Investigating 2-D and 3-D Proximal Remote Sensing Techniques for Vineyard Yield Estimation », *Sensors* **19** (2019), n° 17, article no. 3652 (20 pages).
- [42] K. HE, G. GKIOXARI, P. DOLLÁR & R. GIRSHICK, « Mask R-CNN », <https://arxiv.org/abs/1703.06870>, 2018.
- [43] K. HEINRICH, A. ROTH, L. BREITHAUPT, B. MÖLLER & J. MARESCH, « Yield Prognosis for the Agrarian Management of Vineyards using Deep Learning for Object Counting », *Wirtschaftsinformatik 2019 Proceedings* (2019), p. 407-421.
- [44] M. HERRERO-HUERTA, D. GONZÁLEZ-AGUILERA, P. RODRIGUEZ-GONZALVEZ & D. HERNÁNDEZ-LÓPEZ, « Vineyard yield estimation by automatic 3D bunch modelling in field conditions », *Computers and Electronics in Agriculture* **110** (2015), p. 17-26.
- [45] E. IVORRA, A. J. SÁNCHEZ, J. G. CAMARASA, M. P. DIAGO & J. TARDAGUILA, « Assessment of grape cluster yield components based on 3D descriptors using stereo vision », *Food Control* **50** (2015), p. 273-282.
- [46] A. KAMILARIS & F. X. PRENAFETA-BOLDÁ, « Deep learning in agriculture : A survey », *Computers and Electronics in Agriculture* **147** (2018), p. 70-90.
- [47] B. KERESZTES, F. ABDELGHAFOUR, D. RANDRIAMANGA, J.-P. DA COSTA & C. GERMAIN, « Real-time Fruit Detection Using Deep Neural Networks », in *14th International Conference on Precision Agriculture*, 2018.
- [48] B. KERESZTES, C. GERMAIN, J.-P. DA COSTA, G. GRENIER, X. DAVID-BEAULIEU & A. DE LA FOUCHARDIÈRE, « Vineyard Vigilant and INNovative Ecological Rover (VVINNER) : an autonomous robot for automated scoring of vineyards », in *International Conference of Agricultural Engineering* (Zurich, Switzerland), 2014.
- [49] M. KERKECH, A. HAFIANE & R. CANALS, « Vine disease detection in UAV multispectral images using optimized image registration and deep learning segmentation approach », *Computers and Electronics in Agriculture* **174** (2020), p. 105446.
- [50] A. KICHERER, K. HERZOG, N. BENDEL, H.-C. KLÜCK, A. BACKHAUS, M. WIELAND, J. C. ROSE, L. KLINGBEIL, T. LÄBE, C. HOHL & ET AL., « Phenoliner : A New Field Phenotyping Platform for Grapevine Research », *Sensors* **17** (2017), n° 77, article no. 1625 (18 pages).
- [51] A. KICHERER, K. HERZOG, N. BENDEL, H.-C. KLÜCK, A. BACKHAUS, M. WIELAND, J. C. ROSE, L. KLINGBEIL, T. LÄBE, C. HOHL, W. PETRY, H. KUHLMANN, U. SEIFFERT & R. TÖPFER, « Phenoliner : A New Field Phenotyping Platform for Grapevine Research », *Sensors* **17** (2017), n° 7, article no. 1625 (18 pages).

- [52] A. KICHERER, K. HERZOG, M. PFLANZ, M. WIELAND, P. RÜGER, S. KECKE, H. KUHLMANN & R. TÖPFER, « An Automated Field Phenotyping Pipeline for Application in Grapevine Research », *Sensors* **15** (2015), n° 3, p. 4823-4836.
- [53] A. KICHERER, R. ROSCHER, K. HERZOG, S. Á IMON, W. FÖRSTNER & R. TOEPPER, « BAT (Berry Analysis Tool) : A high-throughput image interpretation tool to acquire the number, diameter, and volume of grapevine berries », *Vitis -Geilweilerhof-* **52** (2013), p. 129-135.
- [54] M. KLODT, K. HERZOG, R. TÖPFER & D. CREMERS, « Field phenotyping of grapevine growth using dense stereo reconstruction », *BMC Bioinformatics* **16** (2015), n° 1, article no. 143.
- [55] A. KRIZHEVSKY, I. SUTSKEVER & G. E. HINTON, « ImageNet Classification with Deep Convolutional Neural Networks », in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1* (Red Hook, NY, USA), NIPS'12, Curran Associates Inc., 2012, p. 1097-1105.
- [56] P. KURTSEER, O. RINGDAHL, N. ROTSTEIN, R. BERENSTEIN & Y. EDAN, « In-Field Grape Cluster Size Assessment for Vine Yield Estimation Using a Mobile Robot and a Consumer Level RGB-D Camera », *IEEE Robotics and Automation Letters* **5** (2020), n° 2, p. 2031-2038.
- [57] Y. LECUN, B. E. BOSER, J. S. DENKER, D. HENDERSON, R. E. HOWARD, W. E. HUBBARD & L. D. JACKEL, « Handwritten Digit Recognition with a Back-Propagation Network », p. 396-404, Morgan-Kaufmann, 1990.
- [58] S. LIU, X. ZENG & M. WHITTY, « 3DBunch : A Novel iOS-Smartphone Application to Evaluate the Number of Grape Berries per Bunch Using Image Analysis Techniques », *IEEE Access* **8** (2020), p. 114663-114674.
- [59] S. LIU, S. COSSELL, J. TANG, G. DUNN & M. WHITTY, « A computer vision system for early stage grape yield estimation based on shoot detection », *Computers and Electronics in Agriculture* **137** (2017), p. 88-101.
- [60] S. LIU, X. LI, H. WU, B. XIN, J. TANG, P. R. PETRIE & M. WHITTY, « A robust automated flower estimation system for grape vines », *Biosystems Engineering* **172** (2018), p. 110-123.
- [61] S. LIU, S. MARDEN & M. WHITTY, « Towards Automated Yield Estimation in Viticulture », in *Proceedings of the Australasian Conference on Robotics and Automation*, 2013, p. 9.
- [62] S. LIU & M. WHITTY, « Automatic grape bunch detection in vineyards with an SVM classifier », *Journal of Applied Logic* **13** (2015), n° 4, Part 34, Part 3, p. 643-653.
- [63] S. LIU, M. WHITTY & S. COSSELL, « A Lightweight Method for Grape Berry Counting based on Automated 3 D Bunch Reconstruction from a Single Image », in *ICRA, International Conference on Robotics and Automation (IEEE), Workshop on Robotics in Agriculture*, 2015.
- [64] S. LIU, X. ZENG & M. WHITTY, « A vision-based robust grape berry counting algorithm for fast calibration-free bunch weight estimation in the field », *Computers and Electronics in Agriculture* **173** (2020), article no. 105360 (11 pages).
- [65] X. LIU, S. W. CHEN, C. LIU, S. S. SHIVAKUMAR, J. DAS, C. J. TAYLOR, J. UNDERWOOD & V. KUMAR, « Monocular Camera Based Fruit Counting and Mapping With Semantic Data Association », *IEEE Robotics and Automation Letters* **4** (2019), n° 3, p. 2296-2303.
- [66] J. LLORET, I. BOSCH, S. SENDRA & A. SERRANO, « A Wireless Sensor Network for Vineyard Monitoring That Uses Image Processing », *Sensors* **11** (2011), n° 6, p. 6165-6196.
- [67] C. LOPES, A. TORRES, R. GUZMAN, J. GRAÇA, A. MONTEIRO, R. BRAGA, A. BARRIGUINHA, G. VICTORINO & M. REYS, « Using an Unmanned Ground Vehicle to Scout Vineyards for Non-intrusive Estimation of Canopy Features and Grape Yield », in *20th GiESCO International Meeting*, Nov 2017.
- [68] A. LOPEZ-CASTRO, A. MARROQUIN-JACOBO, A. SOTO-AMADOR, E. PADILLA-DAVILA, J. A. LOPEZ-LEYVA & M. O. CASTAÑEDA-RAMOS, « Design of a Vineyard Terrestrial Robot for Multiple Applications as Part of the Innovation of Process and Product : Preliminary Results », in *2020 IEEE International Conference on Engineering Veracruz (ICEV)*, 2020, p. 1-4.
- [69] L. LUO, Y. TANG, Q. LU, X. CHEN, P. ZHANG & X. ZOU, « A vision methodology for harvesting robot to detect cutting points on peduncles of double overlapping grape clusters in a vineyard », *Computers in Industry* **99** (2018), p. 130-139.
- [70] L. LUO, Y. TANG, X. ZOU, C. WANG, P. ZHANG & W. FENG, « Robust Grape Cluster Detection in a Vineyard by Combining the AdaBoost Framework and Multiple Color Components », *Sensors* **16** (2016), n° 12, article no. 2098 (20 pages).

- [71] L. LUO, Y. TANG, X. ZOU, M. YE, W. FENG & G. LI, « Vision-based extraction of spatial information in grape clusters for harvesting robots », *Biosystems Engineering* **151** (2016), p. 90-104.
- [72] J. MACK, C. LENZ, J. TEUTRINE & V. STEINHAGE, « High-precision 3D detection and reconstruction of grapes from laser range data for efficient phenotyping based on supervised learning », *Computers and Electronics in Agriculture* **135** (2017), p. 300-311.
- [73] W. MALDONADO & J. C. BARBOSA, « Automatic green fruit counting in orange trees using digital images », *Computers and Electronics in Agriculture* **127** (2016), p. 572-581.
- [74] R. MARANI, A. MILELLA, A. PETITTI & G. REINA, « Deep neural networks for grape bunch segmentation in natural images from a consumer-grade camera », *Precision Agriculture* (2021), p. 387-413.
- [75] A. MATESE & S. F. DI GENNARO, « Practical Applications of a Multisensor UAV Platform Based on Multispectral, Thermal and RGB High Resolution Images in Precision Viticulture », *Agriculture* **8** (2018), n° 7, article no. 116 (13 pages).
- [76] A. MATESE & S. F. D. GENNARO, « Technology in precision viticulture : a state of the art review », *International Journal of Wine Research* **7** (2015), p. 69-81.
- [77] A. MILELLA, R. MARANI, A. PETITTI & G. REINA, « In-field high throughput grapevine phenotyping with a consumer-grade depth camera », *Computers and Electronics in Agriculture* **156** (2019), p. 293-306.
- [78] B. MILLAN, A. AQUINO, M. P. DIAGO & J. TARDAGUILA, « Image analysis-based modelling for flower number estimation in grapevine », *Journal of the Science of Food and Agriculture* **97** (2017), n° 3, p. 784-792.
- [79] B. MILLAN, S. VELASCO-FORERO, A. AQUINO & J. TARDAGUILA, « On-the-Go Grapevine Yield Estimation Using Image Analysis and Boolean Model », *Journal of Sensors* **2018** (2018), article no. 9634752 (15 pages).
- [80] E. A. MURILLO-BRACAMONTES, M. E. MARTINEZ-ROSAS, M. M. MIRANDA-VELASCO, H. L. MARTINEZ-REYES, J. R. MARTINEZ-SANDOVAL & H. CERVANTES-DE AVILA, « Implementation of Hough transform for fruit image segmentation », *Procedia Engineering* **35** (2012), p. 230-239.
- [81] A. K. NELLITHIMARU & G. A. KANTOR, « ROLS : Robust Object-Level SLAM for Grape Counting », in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019, p. 2648-2656.
- [82] S. NUSKE, S. ACHAR, T. BATES, S. NARASIMHAN & S. SINGH, « Yield estimation in vineyards by visual grape detection », in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sep 2011, p. 2352-2358.
- [83] S. NUSKE, K. WILSHUSEN, S. ACHAR, L. YODER & S. SINGH, « Automated Visual Yield Estimation in Vineyards », *J. Field Robot.* **31** (2014), n° 55, p. 837-860.
- [84] F. PALACIOS, M. P. DIAGO & J. TARDAGUILA, « A Non-Invasive Method Based on Computer Vision for Grapevine Cluster Compactness Assessment Using a Mobile Sensing Platform under Field Conditions », *Sensors* **19** (2019), n° 17, article no. 3799.
- [85] J. P. PÉREZ-EXPÓSITO, T. M. FERNÁNDEZ-CARAMÉS, P. FRAGA-LAMAS & L. CASTEDO, « VineSens : An Eco-Smart Decision-Support Viticulture System », *Sensors* **17** (2017), n° 3, article no. 465 (26 pages).
- [86] R. PÉREZ-ZAVALA, M. TORRES-TORRITI, F. A. CHEEIN & G. TRONI, « A pattern recognition strategy for visual grape bunch detection in vineyards », *Computers and Electronics in Agriculture* **151** (2018), p. 136-149.
- [87] S. K. PILLI, B. NALLATHAMBI, S. J. GEORGE & V. DIWANJI, « eAGROBOT - A robot for early crop disease detection using image processing », in *2014 International Conference on Electronics and Communication Systems (ICECS)*, 2014, p. 1-6.
- [88] G. RABATEL & C. GUIZARD, « Grape berry calibration by computer vision using elliptical model fitting », in *ECPA 2007, 6th European Conference on Precision Agriculture*, Jun 2007, p. 581-587.
- [89] A. RAHMAN & A. HELLICAR, « Identification of mature grape bunches using image processing and computational intelligence methods », in *2014 IEEE Symposium on Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP)*, Dec 2014, p. 1-6.
- [90] M. RAHNEMOONFAR & C. SHEPPARD, « Deep Count : Fruit Counting Based on Deep Simulated Learning », *Sensors* **17** (2017), n° 4, article no. 905 (12 pages).

- [91] M. J. C. S. REIS, R. MORAIS, E. PERES, C. PEREIRA, O. CONTENTE, S. SOARES, A. VALENTE, J. BAPTISTA, P. J. S. G. FERREIRA & J. BULAS CRUZ, « Automatic detection of bunches of grapes in natural environment from color images », *Journal of Applied Logic* **10** (2012), n° 44, p. 285-290.
- [92] O. RONNEBERGER, P. FISCHER & T. BROX, « U-Net : Convolutional Networks for Biomedical Image Segmentation », in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (N. Navab, J. Hornegger, W. M. Wells & A. F. Frangi, éd.), Lecture Notes in Computer Science, Springer International Publishing, 2015, p. 234-241.
- [93] R. ROSCHER, K. HERZOG, A. KUNDEL, A. KICHERER, R. TÖPFER & W. FÖRSTNER, « Automated image analysis framework for high-throughput determination of grapevine berry sizes using conditional random fields », *Computers and Electronics in Agriculture* **100** (2014), p. 148-158.
- [94] J. C. ROSE, A. KICHERER, M. WIELAND, L. KLINGBEIL, R. TÖPFER & H. KUHLMANN, « Towards Automated Large-Scale 3D Phenotyping of Vineyards under Field Conditions », *Sensors* **16** (2016), n° 12, article no. 2136 (25 pages).
- [95] R. RUDOLPH, K. HERZOG, R. TÖPFER & V. STEINHAGE, « Efficient identification, localization and quantification of grapevine inflorescences in unprepared field images using Fully Convolutional Networks », *Journal of Grapevine Research* **58** (2019), n° 3, p. 95-104.
- [96] T. SANTOS, L. BASSOI, H. OLDONI & R. MARTINS, *Automatic grape bunch detection in vineyards based on affordable 3D phenotyping using a consumer webcam*, XI Congresso Brasileiro de Agroinformática (SBIAgro 2017), Oct 2017.
- [97] T. T. SANTOS, L. L. DE SOUZA, A. A. DOS SANTOS & S. AVILA, « Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association », *Computers and Electronics in Agriculture* **170** (2020), article no. 105247.
- [98] K. P. SENG, L. ANG, L. M. SCHMIDTKE & S. Y. ROGIERS, « Computer Vision and Machine Learning for Viticulture Technology », *IEEE Access* **6** (2018), p. 67494-67510.
- [99] E. SHELHAMER, J. LONG & T. DARRELL, « Fully Convolutional Networks for Semantic Segmentation », *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39** (2017), n° 4, p. 640-651.
- [100] D. L. SILVER & T. MONGA, « In Vino Veritas : Estimating Vineyard Grape Yield from Images Using Deep Learning », in *Advances in Artificial Intelligence* (M.-J. Meurs & F. Rudzicz, éd.), Lecture Notes in Computer Science, Springer International Publishing, 2019, p. 212-224.
- [101] P. ŠKRABÁNEK, « DeepGrapes : Precise Detection of Grapes in Low-resolution Images », *IFAC-PapersOnLine* **51** (2018), n° 66, p. 185-189.
- [102] P. ŠKRABÁNEK & P. DOLEŽEL, « Robust Grape Detector Based on SVMs and HOG Features », *Computational Intelligence and Neuroscience* **2017** (2017), article no. 3478602.
- [103] Y. SONG, C. A. GLASBEY, G. W. HORGAN, G. POLDER, J. DIELEMAN & G. W. A. M. VAN DER HEIJDEN, « Automatic fruit recognition and counting from multiple images », *Biosystems Engineering* **118** (2014), p. 203-215.
- [104] J. TELLO, K. HERZOG, F. RIST, P. THIS & A. DOLIGEZ, « Automatic Flower Number Evaluation in Grapevine Inflorescences Using RGB Images », *American Journal of Enology and Viticulture* **71** (2019), p. 10-16.
- [105] J. P. VASCONEZ, J. DELPIANO, S. VOUGIOUKAS & F. AUAT CHEEIN, « Comparison of convolutional neural networks in fruit detection and counting : A comprehensive evaluation », *Computers and Electronics in Agriculture* **173** (2020), article no. 105348.
- [106] A. VASWANI, N. SHAZEER, N. PARMAR, J. USZKOREIT, L. JONES, A. N. GOMEZ, L. KAISER & I. POLOSUKHIN, « Attention is All you Need », in *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan & R. Garnett, éd.), vol. 30, Curran Associates, Inc., 2017.
- [107] G. VICTORINO, G. MAIA, J. QUEIROZ, R. BRAGA, J. MARQUES & C. LOPES, « Grapevine yield prediction using image analysis – improving the estimation of non-visible bunches », *European Federation for Information Technology in Agriculture, Food and the Environment (EFITA)* (2019), p. 6.
- [108] Q. WANG, S. NUSKE, M. BERGERMAN & S. SINGH, « Automated Crop Yield Estimation for Apple Orchards », p. 745-758, Springer International Publishing, Heidelberg, 2013.
- [109] J. XIONG, Z. LIU, R. LIN, R. BU, Z. HE, Z. YANG & C. LIANG, « Green Grape Detection and Picking-Point Calculation in a Night-Time Natural Environment Using a Charge-Coupled Device (CCD) Vision Sensor with Artificial Illumination », *Sensors* **18** (2018), n° 4, article no. 969 (17 pages).

- [110] L. ZABAWA, A. KICHERER, L. KLINGBEIL, R. TÖPFER, H. KUHLMANN & R. ROSCHER, « Counting of grapevine berries in images via semantic segmentation using convolutional neural networks », *ISPRS Journal of Photogrammetry and Remote Sensing* **164** (2020), p. 73-83.
- [111] K. ZHANG, L. ZHAO, S. ZHE, C. GENG & W. LI, « Design and Experiment of Intelligent Grape Bagging Robot », *Applied Mechanics and Materials* **389** (2013), p. 706-711.
- [112] P. ZWAENEPOEL & J. LE BARS, « L'agriculture de précision », *Ingénieries eau-agriculture-territoires* (1997), n° 12, p. 67-79.

ABSTRACT. — An early yield prediction is an essential tool for vinegrowers to prepare for the harvest. In the last decades, many methods have been proposed by researchers to solve this problem practically. In particular, recent computer vision advances allow for fruit counting directly in the field. The objective of this publication is to make an exhaustive review of existing works related to grape detection and yield prediction with computer vision algorithms. This review is focused on approaches using RGB proximal images, including classical image processing algorithms, feature extraction, machine learning models, and deep learning. By recalling works on flower counting, grape detection, berry counting, and yield modelization, we highlight challenges on computer vision research for viticulture applications. In addition, we believe that non-specialists can find this review helpful to understand recent advances in artificial intelligence for precision viticulture.

KEYWORDS. — Viticulture, yield prediction, machine learning, deep learning.

Manuscrit reçu le 30 mars 2021, révisé le 15 juillet 2021, accepté le 1^{er} septembre 2021.