vinetur

Des chercheurs conçoivent un modèle d’IA qui cartographie les sarments de vigne sur des images avec un IoU de 88,25 %

Testé en hiver dans des vignobles, le système dépasse les outils établis de segmentation d’images et ouvre la voie à une taille robotisée plus homogène.

samedi 14 novembre 2026



Une équipe d’ingénieurs agronomes a mis au point un système d’intelligence artificielle capable de distinguer les ceps et les sarments à tailler dans des images complexes de vignobles. Cette avancée pourrait aider de futurs robots à effectuer la taille hivernale de manière plus homogène. Publié dans la revue Smart Agricultural Technology et présenté vendredi 13 novembre, ces travaux portent sur la vision par ordinateur, et non sur la conception d’un robot de taille entièrement autonome.

La taille est l’une des tâches courantes les plus importantes dans les vignobles. Elle permet de maîtriser la croissance des plants, la production de fruits et la vigueur de la vigne et, en définitive, la qualité des raisins. Dans de nombreuses régions viticoles, elle repose encore largement sur des équipes de terrain qualifiées. Elle est donc coûteuse, chronophage et difficile à standardiser dans les grands vignobles, en particulier lorsque la main-d’œuvre manque. La taille représentant un poste de dépense important pour les producteurs de raisin, une meilleure automatisation pourrait à terme compter pour la filière si la technologie fait ses preuves dans les parcelles commerciales.

Le nouveau modèle, baptisé Dual-branch Transformer-Mamba Fusion Network, ou DTMF, vise à identifier les structures ligneuses de la vigne sur des images prises en extérieur pendant le repos végétatif, lorsque les vignes sont dépouillées et prêtes pour la taille hivernale. La tâche peut sembler très spécialisée, mais elle répond à un obstacle fondamental de la robotique viticole : une machine ne peut pas effectuer une coupe si elle ne sait pas d’abord distinguer la vigne, les sarments à tailler et l’arrière-plan.

Cet arrière-plan est souvent difficile à interpréter pour les logiciels. Les sarments fins peuvent se confondre avec les fils de palissage, le sol, les mauvaises herbes et d’autres éléments visuels parasites. Selon les chercheurs, les systèmes classiques de segmentation d’images peinent souvent à saisir à la fois la forme générale d’une vigne et les détails très fins des contours des sarments. Pour résoudre ce problème, ils ont conçu un système à deux branches. L’une analyse la scène dans son ensemble et les relations à longue distance dans l’image, tandis que l’autre se concentre sur les textures fines, les contours et les motifs fréquentiels susceptibles de faciliter le repérage des structures ligneuses étroites.

Sur le plan technique, le modèle associe un transformeur de vision, une transformée en ondelettes discrète et des modules d'espace d'état Mamba. Le transformeur vise à saisir le contexte global de l'image. L'association des ondelettes et de Mamba doit préserver les petits détails structurels et la continuité des sarments fins. Au lieu de fusionner les deux flux trop tôt, le système les traite séparément pendant la majeure partie du processus et ne combine leurs prédictions qu'à l'étape finale. Les chercheurs ont également utilisé une méthode d'apprentissage en deux étapes, qui passe d'une classification élémentaire des pixels au début de l'entraînement à un affinage plus poussé des contours par la suite, notamment à l'aide d'une fonction de perte fondée sur des superpixels, conçue pour aligner les prédictions sur les contours naturels des objets.

L'équipe a testé le système sur 2 150 images RVB haute résolution recueillies au Xinjiang, en Chine, pendant le repos végétatif hivernal. Le jeu de données comprenait 1 250 images prises en novembre 2024 dans un vignoble expérimental de l'université de Shihezi et 900 images prises en décembre 2024 dans un verger de production commerciale à Turpan. Toutes les images ont été annotées manuellement au niveau du pixel afin de repérer les ceps et les sarments. Les chercheurs ont réparti le jeu de données entre 1 505 images d'entraînement, 215 images de validation et 430 images de test. Après augmentation des données, l'ensemble d'entraînement effectif a atteint 4 515 images.

Selon l'étude, le modèle DTMF complet a atteint une précision moyenne au niveau des pixels de 94,24 % et un IoU moyen, mesure courante de la segmentation, de 88,25 % sur le jeu de test. Lors de comparaisons avec plusieurs modèles établis, DTMF a surpassé DeepLabV3, PSPNet, FCN, Segformer, U-Net et un modèle U-Net fondé sur un transformeur. L'écart était particulièrement marqué pour la détection des sarments, la partie la plus difficile de la tâche, car ils sont plus fins et se confondent plus facilement avec l'arrière-plan. DTMF a obtenu un IoU de 85,12 % pour les sarments, contre 70,13 % pour le meilleur modèle de référence dans cette catégorie, soit une hausse de 14,99 points. La segmentation des ceps était également performante, à 87,68 %, même si d'autres modèles récents s'en rapprochaient davantage pour cette catégorie.

Les chercheurs ont également réalisé des tests d'ablation, qui consistent à retirer certaines parties du système pour mesurer la contribution de chacune. Ces tests indiquent que la structure à double encodeur et double décodeur, le module de fusion ondelettes-Mamba et la stratégie de fonction de perte en deux étapes ont tous amélioré les résultats. Avec la seule branche transformeur, l'IoU moyen du modèle était de 71,98 %. Avec l'architecture complète, il est passé à 88,25 %. Lors de tests distincts portant sur la méthode d'entraînement, l'ajout de la supervision à l'étape ultérieure et des contraintes fondées sur les superpixels a fait passer l'IoU des sarments de 74,33 % à 85,12 %.

Les exemples visuels présentés dans l'article montrent que le modèle suit les sarments dans des arrière-plans difficiles, comme les fils et les sols irréguliers, avec moins de segments interrompus ou manquants que les systèmes comparés. Les auteurs ont également testé le modèle sur des images extérieures d'autres plantes ligneuses. Ces démonstrations laissent entrevoir une possible utilité structurelle plus large, mais l'article précise qu'elles sont uniquement illustratives et ne constituent pas un test formel de la capacité du système à généraliser au-delà des données viticoles.

Les auteurs de l'étude soulignent que ces travaux ne constituent pas une solution complète de taille robotisée. La segmentation précise des images n'est qu'une partie du problème. Un robot de taille prêt à travailler dans les vignes devrait encore reconstituer la structure de la vigne en trois dimensions, localiser les points de coupe, planifier ses trajectoires et piloter un outil de coupe dans les conditions réelles du vignoble. Les chercheurs indiquent également que l'architecture parallèle du modèle augmente les besoins de calcul, ce qui pourrait compliquer son déploiement sur les systèmes embarqués installés sur des robots agricoles.

L'article relève plusieurs autres limites. Les images proviennent de deux régions seulement du Xinjiang et ont toutes été prises pendant le repos végétatif hivernal. Le modèle n'a donc pas encore été validé dans d'autres régions viticoles, avec d'autres systèmes de palissage ni en été, lorsque le feuillage est dense. La répartition des données a été effectuée au niveau des images, et non des plants, et l'équipe n'a pas fait état de plusieurs entraînements répétés permettant de mesurer les variations liées aux différentes initialisations aléatoires. Ces limites invitent à considérer les solides

résultats obtenus sur le jeu de référence comme un signe de progrès en matière de perception, et non comme la preuve que le système est prêt pour un déploiement commercial à grande échelle.

Cette étude ouvre néanmoins une piste concrète pour l’automatisation des vignobles. Si les prochaines versions gagnent en rapidité et sont validées dans davantage de situations réelles, cette approche pourrait contribuer à lever l’un des principaux obstacles techniques à la taille robotisée des vignes. Pour les vignobles qui fournissent du raisin destiné au vin et à d’autres boissons à base de raisin, elle pourrait à terme favoriser une planification plus stable de la main-d’œuvre et une conduite plus homogène des vignes, même si ces bénéfices dépendront encore de l’intégration sur le terrain et de tests allant au-delà de la seule segmentation d’images.