vinetur

Investigadores desenvolvem modelo de IA que mapeia sarmentos de videira em imagens de vinhas com 88,25% de interseção sobre união

O sistema superou ferramentas estabelecidas de segmentação de imagens em testes de inverno em vinhas, um passo rumo a uma poda robótica mais consistente

sábado, 14 de novembro de 2026



Uma equipa de engenheiros agrónomos desenvolveu um sistema de inteligência artificial capaz de distinguir troncos e sarmentos de videira em imagens complexas de vinhas, um passo que poderá ajudar futuros robôs a realizar a poda de inverno de forma mais consistente. O trabalho, publicado na revista Smart Agricultural Technology e divulgado na sexta-feira, 13 de novembro, centra-se na visão por computador, e não no desenvolvimento de uma máquina de poda totalmente autónoma.

A poda é uma das tarefas regulares mais importantes nas vinhas. Ajuda a controlar o crescimento das plantas, a produção de frutos e o vigor das videiras e, em última análise, a qualidade das uvas. Em muitas regiões vitivinícolas, o trabalho continua a depender fortemente de equipas experientes no campo. Isso torna-o dispendioso, demorado e difícil de uniformizar em vinhas de grande dimensão, sobretudo quando há falta de mão de obra. Como a poda representa um custo significativo para os produtores de uvas, uma melhor automatização poderá vir a ter impacto no setor do vinho, caso a tecnologia se revele fiável em vinhas comerciais.

O novo modelo chama-se Dual-branch Transformer-Mamba Fusion Network, ou DTMF. O seu objetivo é identificar estruturas lenhosas da videira em imagens captadas ao ar livre durante o período de repouso vegetativo, quando as videiras estão despidas e prontas para a poda de inverno. Pode parecer uma tarefa muito específica, mas responde a um obstáculo básico da robótica vitícola: uma máquina não consegue fazer um corte se não conseguir primeiro distinguir a videira, os sarmentos que devem ser podados e o fundo envolvente.

Esse fundo é muitas vezes difícil de interpretar por software. Os ramos finos podem confundir-se com os arames da estrutura de suporte, o solo, as ervas daninhas e outros elementos visuais. Os investigadores afirmaram que os sistemas convencionais de segmentação de imagens têm frequentemente dificuldade em captar tanto a forma geral da videira como os pormenores muito finos nas extremidades dos sarmentos estreitos. Para resolver esse problema, conceberam um sistema com duas ramificações. Uma analisa a cena no seu conjunto e as relações de longo alcance na imagem, enquanto a outra se concentra em texturas finas, contornos e padrões de frequência que podem ajudar a delinear estruturas lenhosas estreitas.

Em termos técnicos, o modelo combina um vision transformer, uma transformada wavelet discreta e módulos Mamba de espaço de estados. A componente transformer destina-se a captar o contexto global de toda a imagem. A componente wavelet e Mamba visa preservar pequenos pormenores estruturais e a continuidade ao longo dos ramos finos. Em vez de fundir as duas vias demasiado cedo, o sistema mantém-nas separadas durante a maior parte do processo e combina as suas previsões apenas na fase final. Os investigadores utilizaram também um método de treino em duas fases, que passa da classificação básica dos píxeis no início do treino para um refinamento mais rigoroso dos contornos numa fase posterior, incluindo uma função de perda baseada em superpíxeis, concebida para manter as previsões alinhadas com os contornos naturais dos objetos.

A equipa testou o sistema em 2.150 imagens RGB de alta resolução recolhidas em Xinjiang, na China, durante o período de repouso vegetativo de inverno. O conjunto de dados incluía 1.250 imagens captadas em novembro de 2024 numa vinha experimental da Universidade de Shihezi e 900 imagens captadas em dezembro de 2024 numa vinha

comercial de produção em Turpan. Todas as imagens foram anotadas manualmente ao nível dos píxeis para assinalar os troncos e os ramos. Os investigadores dividiram o conjunto de dados em 1.505 imagens de treino, 215 imagens de validação e 430 imagens de teste. Após o aumento de dados, o conjunto de treino efetivo foi alargado para 4.515 imagens.

Segundo o estudo, o modelo DTMF completo alcançou uma precisão média ao nível dos píxeis de 94,24% e uma média de interseção sobre união, uma métrica comum de segmentação, de 88,25% no conjunto de teste. Nas comparações com vários modelos estabelecidos, o DTMF superou o DeepLabV3, o PSPNet, o FCN, o Segformer, o U-Net e uma versão do U-Net baseada em transformer. A diferença foi especialmente acentuada na deteção de ramos, a parte mais difícil da tarefa, uma vez que os ramos são mais finos e mais fáceis de confundir com o fundo. O DTMF alcançou uma IoU de 85,12% na deteção de ramos, contra 70,13% do modelo de referência mais forte nessa categoria, uma melhoria de 14,99 pontos. A segmentação dos troncos também apresentou bons resultados, com 87,68%, embora outros modelos modernos tenham ficado mais próximos nesta categoria.

Os investigadores realizaram também testes de ablação, retirando componentes do sistema para avaliar o contributo de cada um. Os resultados sugeriram que a estrutura com codificador e descodificador duplos, o módulo de fusão wavelet-Mamba e a estratégia de perda em duas fases melhoraram todos o desempenho. Quando o modelo utilizou apenas a ramificação transformer, a média de interseção sobre união foi de 71,98%. Com a arquitetura completa, subiu para 88,25%. Em testes separados do método de treino, a adição da supervisão na fase final e das restrições de superpíxeis elevou a IoU dos ramos de 74,33% para 85,12%.

Os exemplos visuais apresentados no artigo mostraram o modelo a delinear sarmentos em fundos difíceis, como arames e solo irregular, com menos segmentos interrompidos ou em falta do que os sistemas de comparação. Os autores testaram também o modelo em imagens exteriores de outras plantas lenhosas. Essas demonstrações sugeriram que o método poderá ter uma utilidade estrutural mais ampla, mas o artigo esclareceu que os resultados eram apenas ilustrativos e não constituíam um teste formal da capacidade do sistema para generalizar para além dos dados das vinhas.

Os autores do estudo tiveram o cuidado de sublinhar que o trabalho não corresponde a uma solução completa de poda robótica. A segmentação precisa de imagens é apenas uma parte de um problema mais vasto. Um robô de poda pronto para utilização no campo teria ainda de reconstruir a estrutura da videira em três dimensões, localizar os pontos de corte, planear trajetórias de movimento e controlar uma ferramenta de corte em condições reais de vinha. Os investigadores referiram ainda que a arquitetura paralela do modelo aumenta os custos computacionais, o que poderá dificultar a sua instalação em hardware integrado nos robôs agrícolas.

O artigo identificou vários outros limites. As imagens provinham apenas de duas regiões de Xinjiang e apenas do período de repouso vegetativo de inverno, pelo que o modelo ainda não foi validado noutras regiões vitícolas, noutros sistemas de condução ou em condições de verão, com folhagem densa. A divisão do conjunto de dados foi feita ao nível das imagens, e não por planta, e a equipa não apresentou resultados de várias rondas de treino para medir a variação entre diferentes inicializações aleatórias. Estas limitações significam que os bons resultados de referência devem ser entendidos como prova de progresso na perceção, e não como prova de que o sistema está pronto para uma utilização comercial alargada.

Ainda assim, o trabalho aponta para uma direção prática para a automatização das vinhas. Se as versões futuras puderem ser mais rápidas e validadas em mais contextos reais, a abordagem poderá ajudar a reduzir um dos principais obstáculos técnicos à poda robótica na produção de uvas. Nas vinhas que abastecem a produção de vinho e de outras bebidas à base de uva, isso poderá, a prazo, contribuir para um planeamento mais estável da mão de obra e uma gestão mais uniforme das videiras, embora esse resultado continue a depender da integração no campo e de testes que vão além da segmentação de imagens.