vinetur

Forschende entwickelten KI-Modell, das Rebzweige auf Weinbergsbildern mit 88,25 % IoU erfasst

Bei Tests im Winterweinberg übertraf das System etablierte Bildsegmentierungsverfahren – ein Schritt hin zu gleichmäßigerem robotergestütztem Rebschnitt

Samstag, 14. November 2026



Ein Team von Agraringenieuren hat ein System mit künstlicher Intelligenz entwickelt, das Rebstämme und Ruten auf komplexen Weinbergsbildern voneinander abgrenzen kann. Das könnte künftigen Robotern helfen, den Winterschnitt gleichmäßiger auszuführen. Die in der Fachzeitschrift Smart Agricultural Technology veröffentlichte und am Freitag, dem 13. November, vorgestellte Arbeit befasst sich mit Computer Vision, nicht mit einer vollständigen autonomen Schnittmaschine.

Der Rebschnitt zählt zu den wichtigsten regelmäßigen Arbeiten im Weinberg. Er trägt dazu bei, das Pflanzenwachstum, den Fruchtertrag und die Wuchskraft der Reben zu steuern und letztlich die Traubenqualität zu beeinflussen. In vielen Weinbauregionen ist die Arbeit nach wie vor stark von erfahrenen Feldteams abhängig. Dadurch ist sie teuer, zeitaufwendig und in großen Weinbergen nur schwer einheitlich auszuführen – insbesondere bei knappem Arbeitskräfteangebot. Da der Rebschnitt für Traubenerzeuger einen erheblichen Kostenfaktor darstellt, könnte eine bessere Automatisierung für die Weinbranche langfristig relevant werden, sofern sich die Technologie in der Praxis bewährt.

Das neue Modell heißt Dual-branch Transformer-Mamba Fusion Network, kurz DTMF. Es soll holzige Rebstrukturen auf Außenaufnahmen aus der Vegetationsruhe erkennen, wenn die Reben kahl und für den Winterschnitt bereit sind. Das mag nach einer eng umrissenen Aufgabe klingen, betrifft aber ein grundlegendes Hindernis für die Robotik im Weinberg: Eine Maschine kann keinen Schnitt setzen, wenn sie nicht zunächst zwischen der Rebe, den zu schneidenden Ruten und dem umgebenden Hintergrund unterscheiden kann.

Dieser Hintergrund ist für Software oft schwer zu interpretieren. Dünne Ruten können mit Drahtrahmen, Boden, Unkraut und anderen visuellen Störelementen verschmelzen. Nach Angaben der Forschenden haben gängige Bildsegmentierungssysteme oft Schwierigkeiten, sowohl die Gesamtform einer Rebe als auch feinste Details an den Rändern dünner Ruten zu erfassen. Um dem entgegenzuwirken, entwickelten sie ein duales System. Ein Zweig des Netzwerks analysiert die Gesamtszene und weitreichende Bildzusammenhänge, während der andere sich auf feine Texturen, Kanten und Frequenzmuster konzentriert, die dabei helfen können, schmale Holzstrukturen nachzuzeichnen.

Technisch kombiniert das Modell einen Vision Transformer, eine diskrete Wavelet-Transformation und Mamba-Zustandsraummodule. Der Transformer-Zweig soll den globalen Kontext im gesamten Bild erfassen. Der Wavelet-Mamba-Zweig soll kleine strukturelle Details und den Verlauf dünner Ruten bewahren. Statt beide Verarbeitungspfade früh zusammenzuführen, hält das System sie während des größten Teils des Prozesses getrennt und führt ihre Vorhersagen erst in der letzten Phase zusammen. Die Forschenden nutzten außerdem ein zweistufiges Trainingsverfahren: Zu Beginn des Trainings steht die grundlegende Pixelklassifizierung im Vordergrund, später eine stärkere Verfeinerung der Grenzen. Dazu gehört eine auf Superpixeln basierende Verlustfunktion, die Vorhersagen an natürlichen Objektkanten ausrichten soll.

Das Team testete das System anhand von 2.150 hochauflösenden RGB-Bildern, die während der winterlichen Vegetationsruhe in Xinjiang, China, aufgenommen worden waren. Der Datensatz umfasste 1.250 Bilder aus dem November 2024 von einer Versuchsanlage der Shihezi University sowie 900 Bilder aus dem Dezember 2024 von einer kommerziellen Produktionsanlage in Turpan. Auf allen Bildern waren Stämme und Ruten manuell auf Pixelebene markiert. Die Forschenden teilten den Datensatz in 1.505 Trainingsbilder, 215 Validierungsbilder und 430 Testbilder auf. Durch Datenerweiterung wuchs der effektive Trainingsdatensatz auf 4.515 Bilder.

Laut der Studie erreichte das vollständige DTMF-Modell im Testsatz eine mittlere Pixelgenauigkeit von 94,24 % und einen mittleren Intersection-over-Union-Wert (IoU), ein gängiges Segmentierungsmaß, von 88,25 %. Im Vergleich mit mehreren etablierten Modellen schnitt DTMF besser ab als DeepLabV3, PSPNet, FCN, Segformer, U-Net und ein Transformer-basiertes U-Net. Besonders groß war der Abstand bei der Erkennung von Ruten. Dieser Teil der Aufgabe ist schwieriger, weil Ruten dünner sind und sich leichter vor dem Hintergrund verlieren. DTMF erreichte bei Ruten einen IoU-Wert von 85,12 %, gegenüber 70,13 % beim stärksten Vergleichsmodell in dieser Kategorie – eine Verbesserung um 14,99 Punkte. Auch die Segmentierung der Stämme war mit 87,68 % gut, wenngleich andere moderne Modelle bei dieser Klasse näher herankamen.

Die Forschenden führten außerdem Ablationstests durch, bei denen Teile des Systems entfernt werden, um den Beitrag der einzelnen Komponenten zu ermitteln. Die Tests deuteten darauf hin, dass die Struktur mit dualem Encoder und Decoder, das Wavelet-Mamba-Fusionsmodul und die zweistufige Verluststrategie die Leistung jeweils verbesserten. Bei alleiniger Verwendung des Transformer-Zweigs lag der mittlere Intersection-over-Union-Wert des Modells bei 71,98 %. Mit dem vollständigen Aufbau stieg er auf 88,25 %. In separaten Tests des Trainingsverfahrens erhöhte die zusätzliche Überwachung in der späteren Phase zusammen mit Superpixel-Beschränkungen den IoU-Wert für Ruten von 74,33 % auf 85,12 %.

Visuelle Beispiele in der Arbeit zeigten, dass das Modell Ruten auch vor schwierigen Hintergründen wie Drähten und unebenem Boden nachverfolgen konnte. Dabei wies es weniger Unterbrechungen und fehlende Abschnitte auf als die Vergleichssysteme. Die Autoren testeten das Modell außerdem mit Außenaufnahmen anderer holziger Pflanzen. Diese Demonstrationen deuteten auf einen möglichen breiteren Nutzen für die Erfassung von Strukturen hin. Die Arbeit betonte jedoch, dass es sich dabei lediglich um Anschauungsbeispiele und nicht um einen formalen Test der Übertragbarkeit auf andere Bereiche als die Weinbergsdaten handelte.

Die Autoren der Studie wiesen ausdrücklich darauf hin, dass es sich bei der Arbeit nicht um eine vollständige Lösung für den robotergestützten Rebschnitt handelt. Eine präzise Bildsegmentierung ist nur ein Teil der größeren Herausforderung. Ein praxistauglicher Schnittroboter müsste die Rebenstruktur weiterhin dreidimensional rekonstruieren, Schnittpunkte lokalisieren, Bewegungsabläufe planen und ein Schneidwerkzeug unter realen Bedingungen im Weinberg steuern. Die Forschenden erklärten außerdem, dass die parallele Architektur des Modells den Rechenaufwand erhöht. Das könnte den Einsatz auf eingebetteter Hardware erschweren, die an Landmaschinenrobotern montiert ist.

Die Arbeit benannte mehrere weitere Einschränkungen. Die Bilder stammten ausschließlich aus zwei Regionen in Xinjiang und aus der winterlichen Vegetationsruhe. Das Modell wurde daher noch nicht in anderen Anbaugebieten, bei anderen Erziehungssystemen oder unter sommerlichen Bedingungen mit dichtem Laub validiert. Die Aufteilung des Datensatzes erfolgte auf Bildebene und nicht nach einzelnen

Pflanzen. Zudem berichtete das Team nicht über wiederholte Trainingsläufe, mit denen sich die Abweichungen bei unterschiedlichen zufälligen Startwerten messen ließen. Die starken Benchmark-Ergebnisse sind daher als Beleg für Fortschritte bei der visuellen Erfassung zu verstehen, nicht als Nachweis dafür, dass das System bereits für den breiten kommerziellen Einsatz bereit ist.

Dennoch weist die Arbeit einen praktischen Weg für die Automatisierung im Weinberg auf. Wenn künftige Versionen schneller werden und sich in weiteren realen Einsatzumgebungen bewähren, könnte der Ansatz dazu beitragen, eines der größten technischen Hindernisse für den robotergestützten Rebschnitt zu verringern. Für Weinberge, die Trauben für Wein und andere Getränke auf Traubenbasis liefern, könnte das langfristig eine verlässlichere Personalplanung und eine einheitlichere Bewirtschaftung der Reben unterstützen. Das hängt jedoch weiterhin von der Integration in den Feldbetrieb und von Tests ab, die über die reine Bildsegmentierung hinausgehen.