vinetur

Researchers developed an AI model that maps grapevine branches in vineyard images with 88.25% intersection over union

The system outperformed established image-segmentation tools in winter vineyard tests, a step toward more consistent robotic pruning

Saturday, November 14, 2026



A team of agricultural engineers has developed an artificial intelligence system that can separate grapevine stems and pruning branches from complex vineyard images, a step that could help future robots perform winter pruning with greater consistency. The work, published in the journal Smart Agricultural Technology and described on Friday, November 13, focuses on computer vision rather than a complete autonomous pruning machine.

Pruning is one of the most important routine jobs in vineyards. It helps control plant growth, fruit production, vine vigor and, ultimately, grape quality. In many wine-growing regions, the work still depends heavily on skilled field crews. That makes it expensive, time-consuming and hard to standardize across large vineyards, especially when labor is short. Because pruning is a major cost for grape producers, better automation could eventually matter for the wine business if the technology proves reliable in commercial fields.

The new model is called the Dual-branch Transformer-Mamba Fusion Network, or DTMF. Its purpose is to identify woody vine structures in outdoor images taken during the dormant season, when vines are bare and ready for winter pruning. That may sound like a narrow task, but it addresses a basic obstacle in vineyard robotics: a machine cannot make a cut if it cannot first tell the difference between the vine, the canes that should be pruned and the surrounding background.

That background is often difficult for software to interpret. Thin branches can blend into trellis wires, soil, weeds and other visual clutter. The researchers said standard image-segmentation systems often struggle to capture both the broad shape of a vine and the very fine detail at the edges of slender canes. To address that, they designed a dual system. One branch of the network analyzes the overall scene and long-range relationships in the image, while the other focuses on fine textures, edges and frequency patterns that can help trace narrow woody structures.

In technical terms, the model combines a vision transformer, a discrete wavelet transform and Mamba state-space modules. The transformer side is meant to capture global context across the whole image. The wavelet and Mamba side is meant to preserve small structural details and continuity along thin branches. Instead of forcing

both streams together too early, the system keeps them separate through most of the process and merges their predictions only at the final stage. The researchers also used a two-stage training method that shifts from basic pixel classification early in training to stronger boundary refinement later, including a superpixel-based loss designed to keep predictions aligned with natural object edges.

The team tested the system on 2,150 high-resolution RGB images collected in Xinjiang, China, during the winter dormant season. The dataset included 1,250 images taken in November 2024 at an experimental vineyard at Shihezi University and 900 images taken in December 2024 at a commercial production orchard in Turpan. All images were manually labeled at the pixel level to mark stems and branches. The researchers split the dataset into 1,505 training images, 215 validation images and 430 test images. After data augmentation, the effective training set expanded to 4,515 images.

According to the study, the full DTMF model reached a mean pixel accuracy of 94.24% and a mean intersection over union, a common segmentation measure, of 88.25% on the test set. In comparisons with several established models, DTMF outperformed DeepLabV3, PSPNet, FCN, Segformer, U-Net and a transformer-based U-Net. The gap was especially large in branch detection, which is the harder part of the task because branches are thinner and easier to lose against the background. DTMF achieved an 85.12% branch IoU, compared with 70.13% for the strongest baseline in that category, an improvement of 14.99 points. Stem segmentation was also strong, at 87.68%, though other modern models came closer on that class.

The researchers also ran ablation tests, which remove parts of the system to see how much each component contributes. Those tests suggested that the dual-encoder and dual-decoder structure, the wavelet-Mamba fusion module and the two-stage loss strategy all improved performance. When the model used only the transformer branch, its mean intersection over union was 71.98%. When the full design was used, that number rose to 88.25%. In separate tests of the training method, adding the later-stage supervision and superpixel constraints raised branch IoU from 74.33% to 85.12%.

Visual examples in the paper showed the model tracing canes across difficult backgrounds such as wires and uneven soil with fewer broken or missing segments than comparison systems. The authors also tried the model on outside images of other woody plants. Those demonstrations suggested the method may have broader structural usefulness, but the paper said those results were only illustrative and were not a formal test of how well the system generalizes beyond the vineyard data.

The study’s authors were careful to note that the work does not amount to a complete robotic pruning solution. Accurate image segmentation is only one part of the larger problem. A field-ready pruning robot would still need to reconstruct the vine structure in three dimensions, locate cutting points, plan movement paths and control a cutting tool in real vineyard conditions. The researchers also said the model’s parallel architecture increases computational cost, which could make deployment harder on embedded hardware mounted on agricultural robots.

The paper identified several other limits. The images came only from two regions in Xinjiang and only from the dormant winter season, so the model has not yet been validated across other growing regions, trellis systems or summer conditions with heavy foliage. The dataset split was done at the image level rather than by individual plant, and the team did not report repeated training runs to measure variation across different random starts. Those limits mean the strong benchmark results should be read as evidence of progress in perception, not as proof that the system is ready for broad commercial use.

Even so, the work points to a practical direction for vineyard automation. If future versions can be made faster and validated in more real-world settings, the approach could help reduce one of the main technical barriers to robotic pruning in grape production. For vineyards that supply wine and other grape-based beverages, that could eventually support more stable labor planning and more uniform vine management, though that outcome still depends on field integration and testing beyond image segmentation alone.