Hellenic Robotics Forum (HRF) 2026

Crop and weed multispectral semantic segmentation with Vision-Language Foundation models

Ilias Papadeas, Georgios-Miltiadis Sandalis, Georgios Zamanakos, Ioannis Pratikakis
Visual Computing Group, Department of Electrical and Computer Engineering, Democritus University of Thrace

Motivation

Accurate crop and weed segmentation is essential for precision agriculture, enabling site-specific weed management and reducing unnecessary herbicide use.

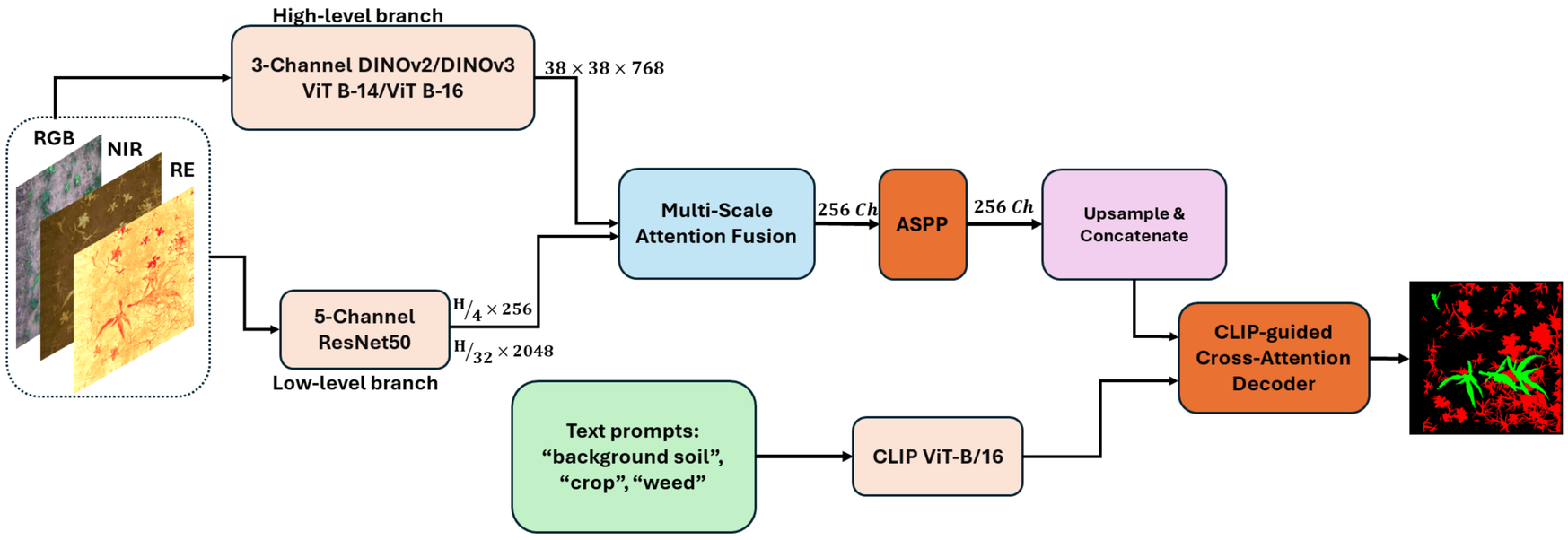
However, many existing models rely mainly on RGB imagery and struggle to generalize across different field conditions, growth stages, crop types, and weed species.

We address this limitation by combining **multispectral UAV imagery** with **vision–language foundation models** for robust crop and weed semantic segmentation.

Proposed Method

The proposed framework combines three complementary sources of information:

- **5-channel ResNet50 branch:** extracts low-level multispectral spatial features from RGB, Near-infrared (NIR), and Red-Edge (RE) bands.
- **DINOv3 branch:** provides high-level self-supervised visual representations and global semantic context.
- **CLIP-guided cross-attention decoder:** injects class-level language priors using text prompts such as “background soil”, “crop”, and “weed”.



Overview of the proposed crop & weed multispectral semantic segmentation architecture.

Novel Contributions

- **Multi-scale attention fusion mechanism** combines multispectral CNN features with DINOv3 semantic representations.
- **CLIP-guided cross-attention decoder** injects language-based class priors into the segmentation decoder.
- The proposed framework unifies multispectral imagery with vision–language foundation models for crop and weed segmentation.

Experiments

The proposed model was evaluated on the **WeedsGalore** UAV multispectral dataset for both 3-class and 6-class crop and weed segmentation tasks.

It achieves **85.42% mIoU** on the 3-class task and approximately **60.0% mIoU** on the more challenging 6-class task, outperforming DeepLabV3+, MaskFormer, and DINOv2-based variants.

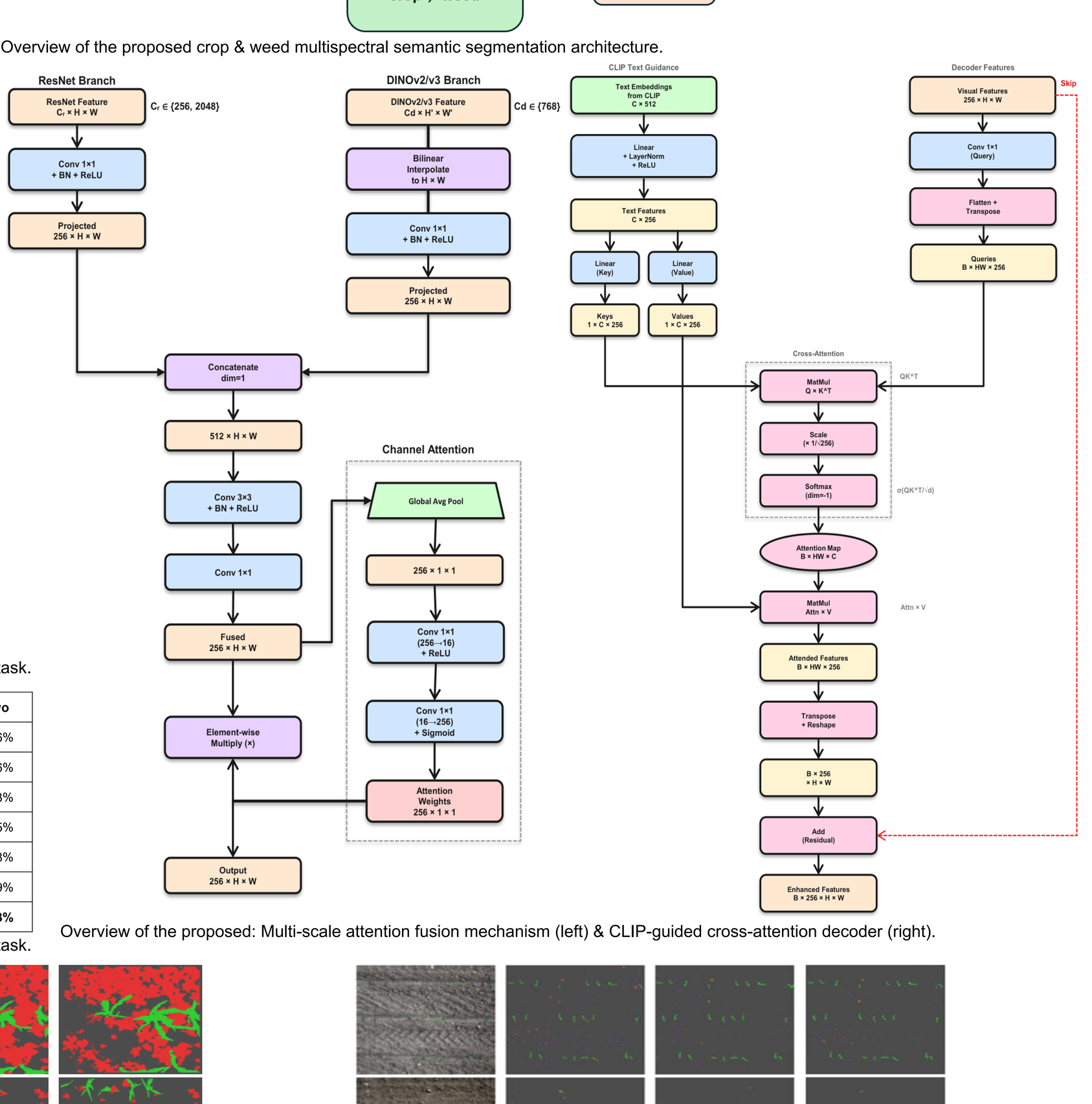
The results show that combining multispectral features with **DINOv3 visual priors** and **CLIP semantic guidance** improves robustness, especially under class imbalance and fine-grained weed species variation.

Method	mIoU	IoUbg	IoUcrop	IoUweed
MaskFormer	80.27%	97.99%	69.49%	73.33%
DeepLabV3+	82.90%	98.45%	72.93%	77.31%
+DINOv2 (frozen)	83.88%	98.45%	76.38%	76.82%
+CLIP (frozen)	82.27%	98.48%	72.82%	75.51%
+DINOv2 + CLIP (frozen)	83.99%	98.50%	76.03%	77.44%
+DINOv2 + CLIP (unfreeze 2)	85.35%	98.51%	79.02%	78.53%
+DINOv3 (frozen)	84.80%	98.33%	78.12%	77.95%
+DINOv3 + CLIP poly	85.30%	98.53%	79.00%	78.37%
+DINOv3 + CLIP cosine	**85.42%**	**98.55%**	**79.09%**	**78.63%**

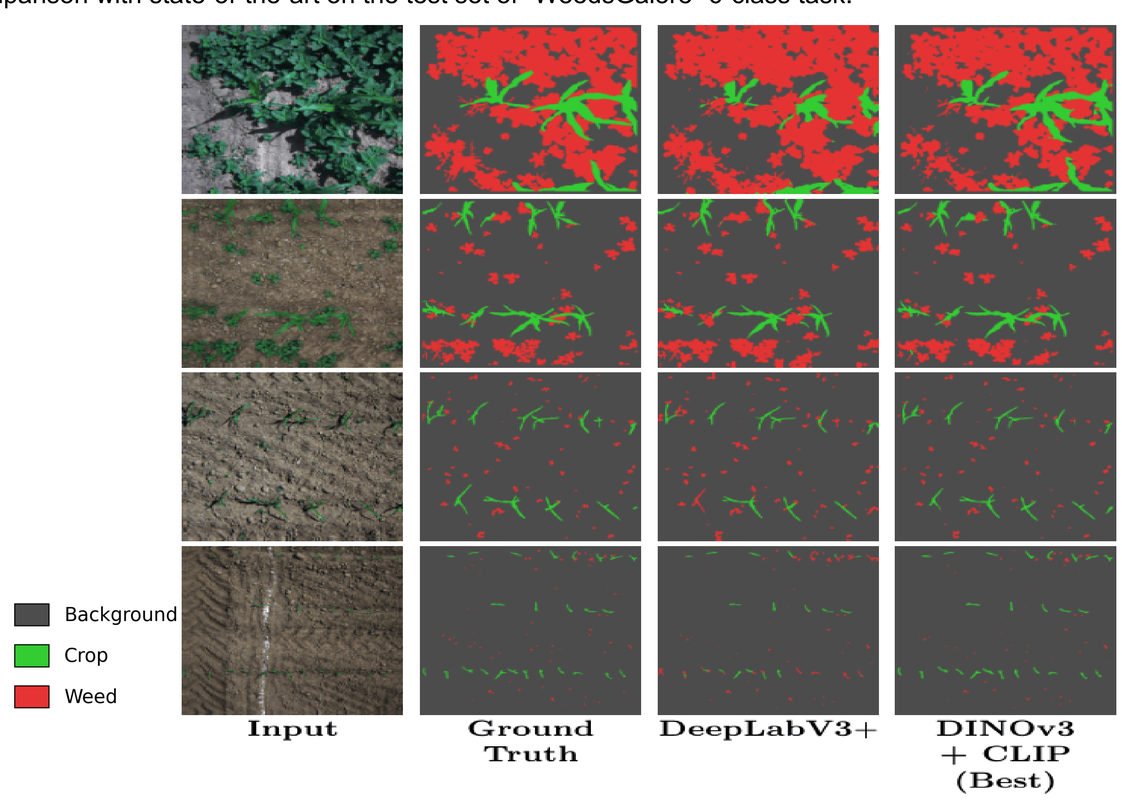
Comparison with state-of-the-art on the test set of ”WeedsGalore” 3-class task.

Method	mIoU	IoUbg	IoUcrop	IoUam	IoUgr	IoUqw	IoUwo
MaskFormer	50.65%	97.90%	68.04%	73.11%	45.34%	8.66%	10.86%
DeepLabV3+	55.52%	98.37%	73.03%	76.17%	53.55%	21.11%	10.86%
+DINOv2 (no CLIP)	59.79%	98.42%	76.54%	79.37%	55.16%	27.12%	22.13%
+DINOv3 (no CLIP)	59.23%	98.50%	78.06%	79.91%	55.63%	22.55%	20.75%
+DINOv2 + CLIP	58.90%	98.50%	77.55%	80.96%	56.30%	21.71%	18.38%
+DINOv3 + CLIP (poly)	58.16%	98.39%	78.59%	79.32%	54.84%	22.34%	15.49%
+DINOv3 + CLIP (cosine)	**59.97%**	**98.46%**	**78.66%**	**80.64%**	**56.09%**	**22.33%**	**23.63%**

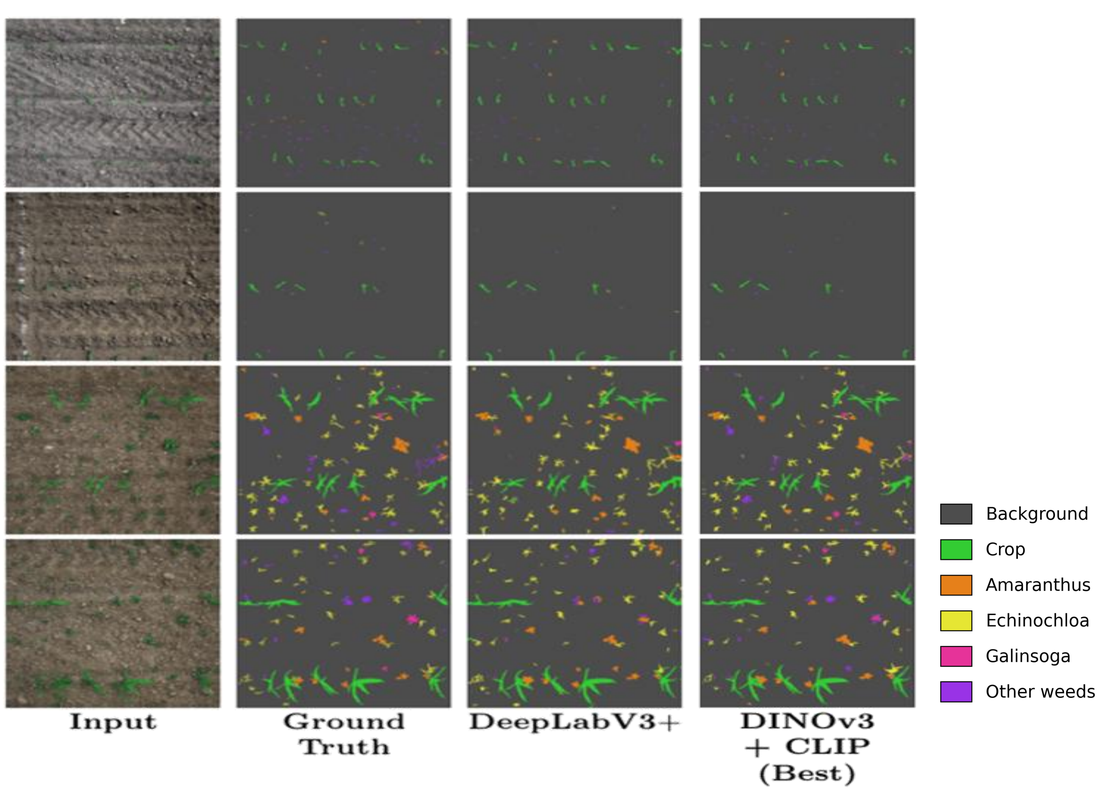
Comparison with state-of-the-art on the test set of ”WeedsGalore” 6-class task.



Overview of the proposed: Multi-scale attention fusion mechanism (left) & CLIP-guided cross-attention decoder (right).



Qualitative results on the 3-class "WeedsGalore" segmentation task.



Qualitative results on the 6-class "WeedsGalore" segmentation task.

Conclusion

By combining **multispectral imagery**, **DINOv3 representations**, and **CLIP-guided semantic priors**, the proposed model achieves state-of-the-art crop and weed segmentation on WeedsGalore.
The proposed approach supports **scalable precision agriculture** by enabling more reliable **site-specific weed management** and reducing unnecessary herbicide use.

Acknowledgment: This research was funded by the European Union – NextGenerationEU and national resources through the Recovery and Resilience Facility under the “Clusters of Research Excellence – CREs” Action, Greece 2.0, project code: ΥΠ3ΤΑ-0559722.