Semantic Segmentation of Sugar Beet Fields with Pseudo-Attention Mechanisms

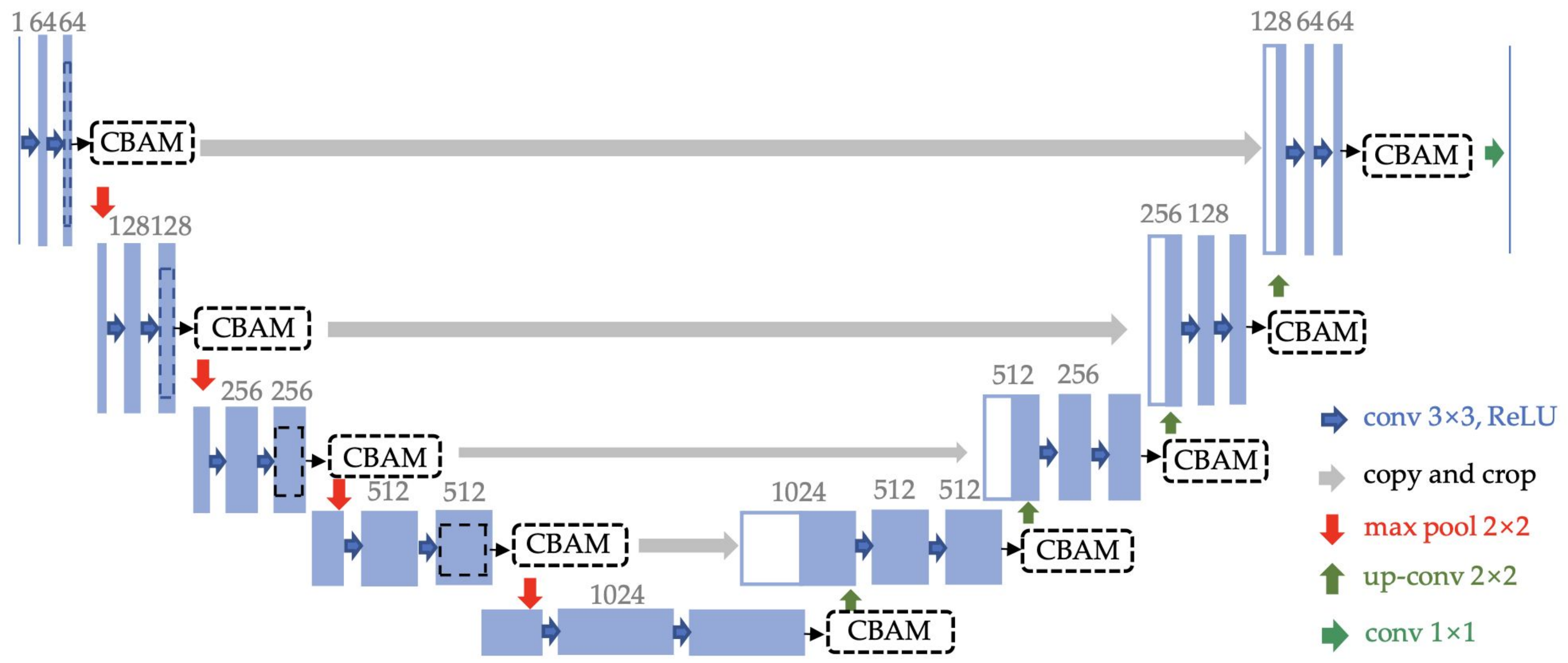
Computer Science
UNIVERSITY OF COLORADO **BOULDER**

Miles Mena & Ryan Slocum

Introduction

Agricultural robotics is a fast-growing application of deep learning in computer vision, with compelling outcomes. By being able to accurately detect the difference between weeds, crops, and other objects in an image, devices such as Verdant Robotics' Sharpshooter precisely spray or laser weeds, dramatically reducing herbicide usage. In addition, the data gathered from these computer vision applications can assist farmers in better managing their crops and maximizing yields.

Critical to this task is the proper and thorough segmentation of weeds. However, due to the relatively small size of most weeds in the sugar beet fields, these plants are easily missed by most current models. In the PhenoBench competition, in which this work enters a submission, the best models only achieve an Intersection over Union (IoU) score of less than 70% on weeds, which we believe could be significantly improved.



U-NET + CBAM architecture

Related Works

There are many existing solutions in agricultural semantic segmentation. U-NET models are commonly used for semantic segmentation, with a variety of modifications made, including Squeeze and Excite (SE) modules, Convolutional Block Attention Modules (CBAM), and ResNetX modules.

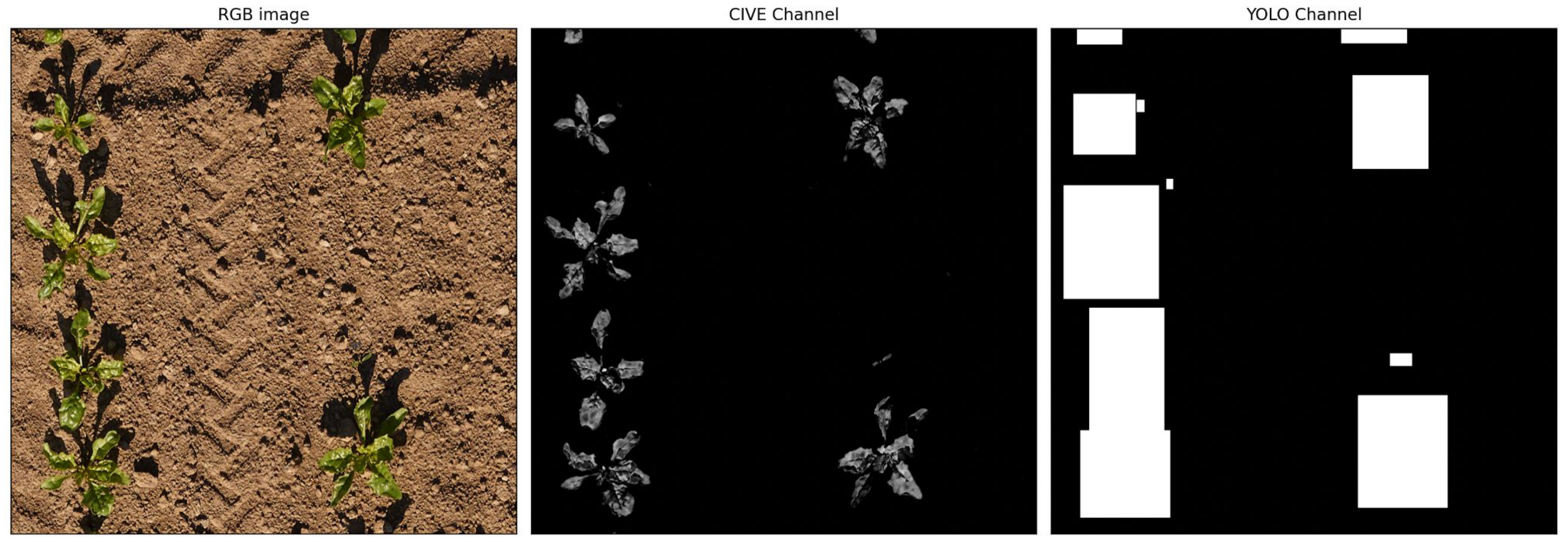
However, none of these models have achieved more than 70% mIoU on weed segmentation, meaning there is room for improvement in attention-related methods for semantic segmentation.

Methods

The methods we implemented emulate the concept of attention, but they do not instantiate the natural language method of attention as introduced in the paper *Attention Is All You Need*. We therefore call the methods we present "pseudo-attention" to differentiate them from Vision-Transformer architectures.

U-NET + CIVE + YOLOv7

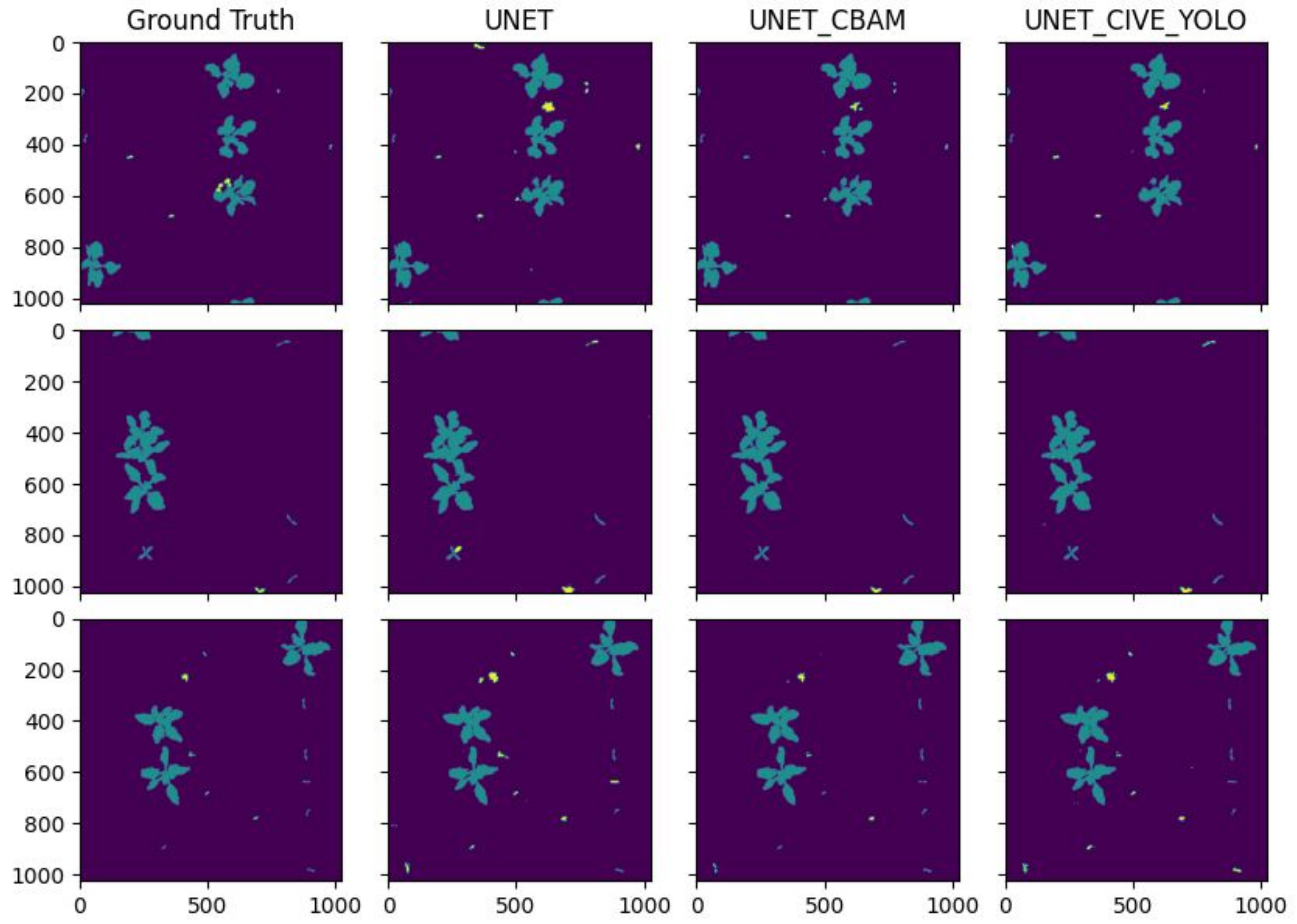
Our second pseudo-attention method involved feature engineering to transform the 3-channel image into a 5-channel image. The additional channels were created using a vegetation index, CIVE, and YOLOv7 to create bounding boxes around probable weeds and plants. The resulting channels are visible above



The five channels in the U-NET + CIVE + YOLO model

U-NET CBAM

The first pseudo-attention method we implemented was integrating Convolutional Block Attention Modules (CBAM) into our U-NET architecture. We append CBAM to the architecture before a max pooling layer and before an up-convolution layer. CBAM consists of a channel attention and spatial attention mechanism. A feature map is passed into the channel attention mechanism to inform the model of the importance of the map's channels. Then the channel refined feature map is passed into the spatial attention mechanism to similarly extract salient areas in a single channel.



Example masks generated for each model, contrasted with the ground truth mask

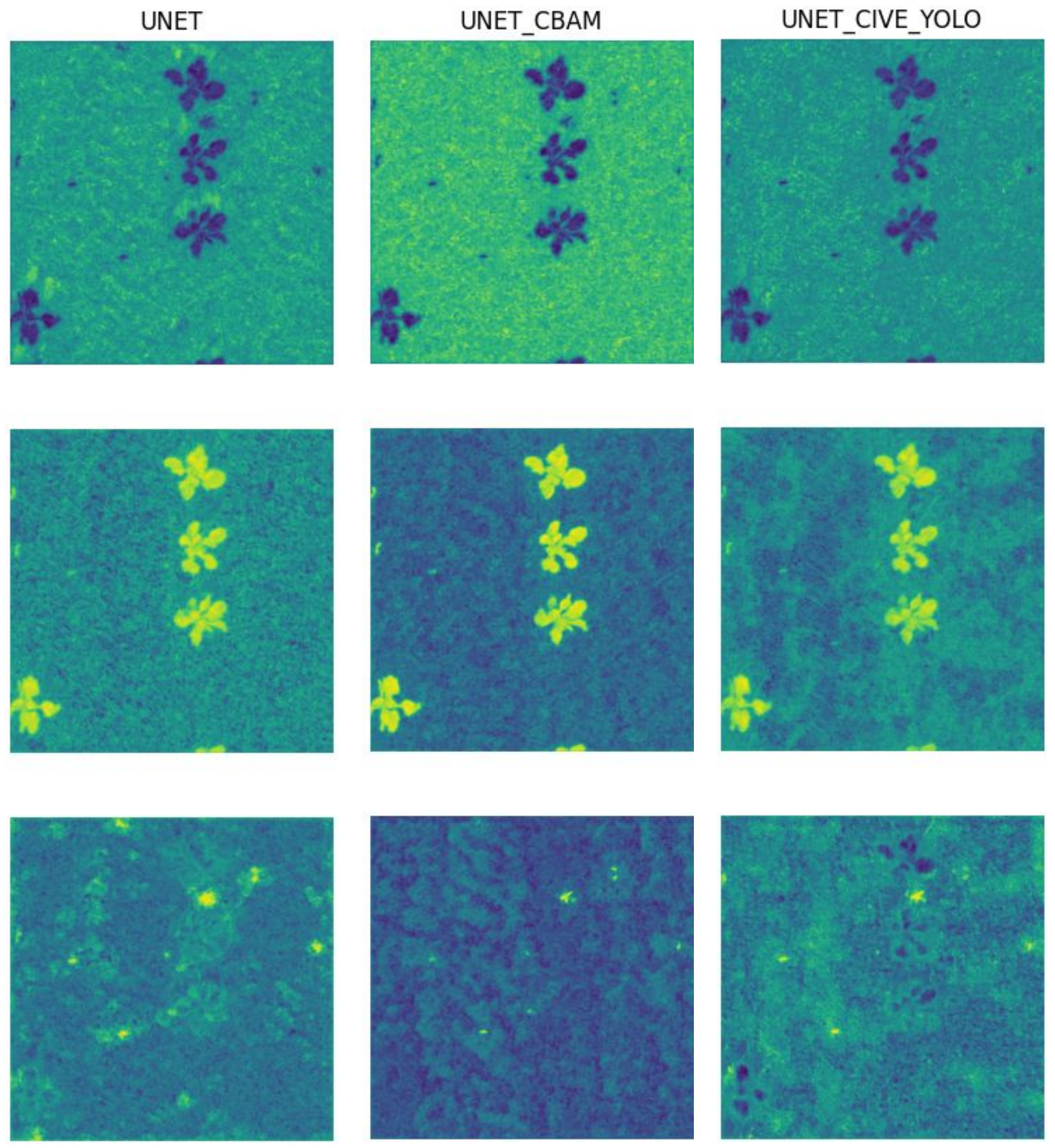
Results

To evaluate the performance improvement of the two pseudo-attention mechanisms we've implemented, we use a control "vanilla" U-NET model with the same training parameters as the two test models. Below we report the mIOU for the test partition, as computed by the PhenoBench competition website.

Model	mIOU Soil	mIOU Crops	mIOU Weeds	avg mIOU
Vanilla U-NET	98.99	**93.01**	48.96	80.32
U-NET CIVE+YOLOv7	99.01	92.81	51.18	81.0
U-NET CBAM	**99.11**	92.99	**58.49**	**83.53**

Results from the test dataset for each of our three models

It's clear that both of the implemented pseudo-attention mechanisms made marginal gains over the control model. In the soil and crop categories, all three of the models performed at roughly the same level, with no noticeable differences in segmentation efficacy. However, there was a substantial increase in performance for the mIOU score for weeds. Given the nearly doubled labeling time of these UNET CBAM, though, this is not necessarily a sufficient performance gain to consider switching.



The final feature map of each model, which occurs directly before the softmax. This shows the activation level of each class before a formal prediction is made. Row 1 represents the soil class, row 2 represents the crop class, and row 3 represents the weed class. Yellow is higher activation and blue is lower activation.

References

Weyler, J., Magistri, F., Marks, E., Chong, Y.L., Sodano, M., Roggiolani, G., Chebrolu, N., Stachniss, C., Behley, J.: Phenobench–a large dataset and benchmarks for semantic image interpretation in the agricultural domain. arXiv preprint arXiv:2306.04557 (2023)

Zhao, J., Wang, J., Qian, H., Zhan, Y., Lei, Y.: Extraction of winter-wheat planting areas using a combination of u-net and cbam. Agronomy 12 (12) (2022) 2965