

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Biological Systems Engineering: Papers and Publications

Biological Systems Engineering

2023

OSC-CO²: Coattention and Cosegmentation Framework for Plant State Change with Multiple Features

Rubi Quiñones

University of Nebraska-Lincoln and Southern Illinois University, rquinon@siue.edu

Ashok Samal

University of Nebraska-Lincoln, samal@unl.edu

Sruti Das Choudhury

University of Nebraska-Lincoln, S.D.Choudhury@unl.edu

Francisco Muñoz-Arriola

University of Nebraska-Lincoln, fmunoz@unl.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/biosysengfacpub>



Part of the [Bioresource and Agricultural Engineering Commons](#), [Environmental Engineering Commons](#), and the [Other Civil and Environmental Engineering Commons](#)

Quiñones, Rubi; Samal, Ashok; Choudhury, Sruti Das; and Muñoz-Arriola, Francisco, "OSC-CO²: Coattention and Cosegmentation Framework for Plant State Change with Multiple Features" (2023). *Biological Systems Engineering: Papers and Publications*. 882.

<https://digitalcommons.unl.edu/biosysengfacpub/882>

This Article is brought to you for free and open access by the Biological Systems Engineering at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Biological Systems Engineering: Papers and Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Published in *Frontiers in Plant Science* 14 (2023), 1211409; doi: 10.3389/fpls.2023.1211409
Copyright © 2023 Rubi Quiñones, Ashok Samal, Sruti Das Choudhury, Francisco Muñoz-Arriola.
Published by Frontiers Media. Used by permission. CC BY 4.0.
Submitted April 24, 2023; accepted October 6, 2023; published October 31, 2023.

OSC-CO²: Coattention and Cosegmentation Framework for Plant State Change with Multiple Features

Rubi Quiñones,^{1,2} Ashok Samal,¹ Sruti Das Choudhury,^{1,3} and
Francisco Muñoz-Arriola^{3,4}

1. School of Computing, University of Nebraska–Lincoln, Lincoln, Nebraska, USA
2. Computer Science Department, Southern Illinois University, Edwardsville, Edwardsville, Illinois, USA
3. School of Natural Resources, University of Nebraska–Lincoln, Lincoln, Nebraska, USA
4. Department of Biological Systems Engineering, University of Nebraska–Lincoln, Lincoln, Nebraska, USA

Corresponding author – Rubi Quiñones, email rquinon@siue.edu

Abstract

Cosegmentation and coattention are extensions of traditional segmentation methods aimed at detecting a common object (or objects) in a group of images. Current cosegmentation and coattention methods are ineffective for objects, such as plants, that change their morphological state while being captured in different modalities and views. The Object State Change using Coattention-Cosegmentation (OSC-CO²) is an end-to-end unsupervised deep-learning framework that enhances traditional segmentation techniques, processing, analyzing, selecting, and combining suitable segmentation results that may contain most of our target object's pixels, and then displaying a final segmented image. The framework leverages coattention-based convolutional neural networks (CNNs) and cosegmentation-based dense Conditional Random Fields (CRFs) to address segmentation accuracy in high-dimensional plant imagery with evolving plant objects. The efficacy of OSC-CO² is demonstrated using plant growth sequences imaged with infrared, visible, and fluorescence cameras in multiple views using a remote sensing, high-throughput phenotyping platform, and is evaluated using Jacard index and precision measures. We also introduce CosegPP+, a dataset that is structured and can

provide quantitative information on the efficacy of our framework. Results show that OSC-CO² outperformed state-of-the-art segmentation and cosegmentation methods by improving segmentation accuracy by 3% to 45%.

Keywords: segmentation, cosegmentation, image analysis, high-throughput plant phenotyping, image sequences, object state change, multiple features, multiple dimensions

1. Introduction

Segmentation is a widely used technique to extract the foreground object from the background before information extraction (Langan et al., 1998; Rezaee et al., 2000; Patz and Preusser, 2012; Fan et al., 2018; Liu et al., 2021). Image segmentation has been used in many application domains, including medicine (Das and Kundu, 2013; Lian et al., 2019; Zhou et al., 2019), traffic safety (Alessandretti et al., 2007; Chang et al., 2019; Chen et al., 2020), and earth system diagnostics (Hoerser and Kuenzer, 2020). However, the success of segmentation algorithms has been limited by the complexity and diversity of the imagery. Cosegmentation is a technique developed to address the problem of segmenting an object in a set of images (Rother et al., 2006; Tao et al., 2015; Ren et al., 2018; Tao et al., 2019). Since its introduction, it has been used in many domains, including plant imagery (Quiñones et al., 2021), PET-CT images (Zhong et al., 2019), and video-based person reidentification (Subramaniam et al., 2019). Current cosegmentation methods have been developed for RGB images (Chen et al., 2014a; Lin et al., 2014; Subramaniam et al., 2019) for objects with no defined quantitative or qualitative features (e.g., environmental conditions, perspectives, temporality, among others.) (Quiñones et al., 2021). Currently, datasets lack specific labeling for cosegmentation, limiting the success and application of these visualization methods. Furthermore, some methods are dependent on training data, which make them tedious to generate and time consuming for training (Chen et al., 2014b; Meng et al., 2016; Hsu et al., 2019).

Although engineered features, such as Scale-Invariant Feature Transform (Lowe, 2004) and Histogram of Oriented Gradients (Dalal and Triggs, 2005), have been widely used in conventional cosegmentation methods, they are no longer optimal cosegmentation analytics because of their predesigned network features. Convolutional neural networks (CNNs), on the other hand, have demonstrated their effectiveness in producing feature extraction in image pairs (Krizhevsky et al., 2012). Yuan et al (Yuan et al., 2017). proposed a CNN-based supervised method for object cosegmentation that would produce the masks for an object in a pair of images. However, their method requires additional training data for the CNN model in the form of object masking.

With the explosion in variety, velocity, and volume of plant imagery datasets, traditional segmentation algorithms grapple with processing images and achieving high accuracy due to challenges, including occlusion and overlap. Cosegmentation algorithms have the potential to overcome these issues, but they achieve high accuracy only for the dataset with which they were trained.

This paper presents Object State Change using Coattention-Cosegmentation (OSC-CO²), an end-to-end trainable unsupervised coattention- and cosegmentation-based framework

for the increased segmentation of multiple feature objects that undergo state changes in high-throughput datasets. The state is referred to the object's (plant's) shape, orientation, and size at a specific point in time. OSC-CO² is designed to process datasets that contain a variety of features, such as perspective (V), species (S), temporality (T), environmental conditions (E), and modality (M) (VSTEM) (Fig. 1). The code for OSC-CO² is publicly available at [https://github.com/rubi quinones/OSC-CO²](https://github.com/rubi quinones/OSC-CO2).

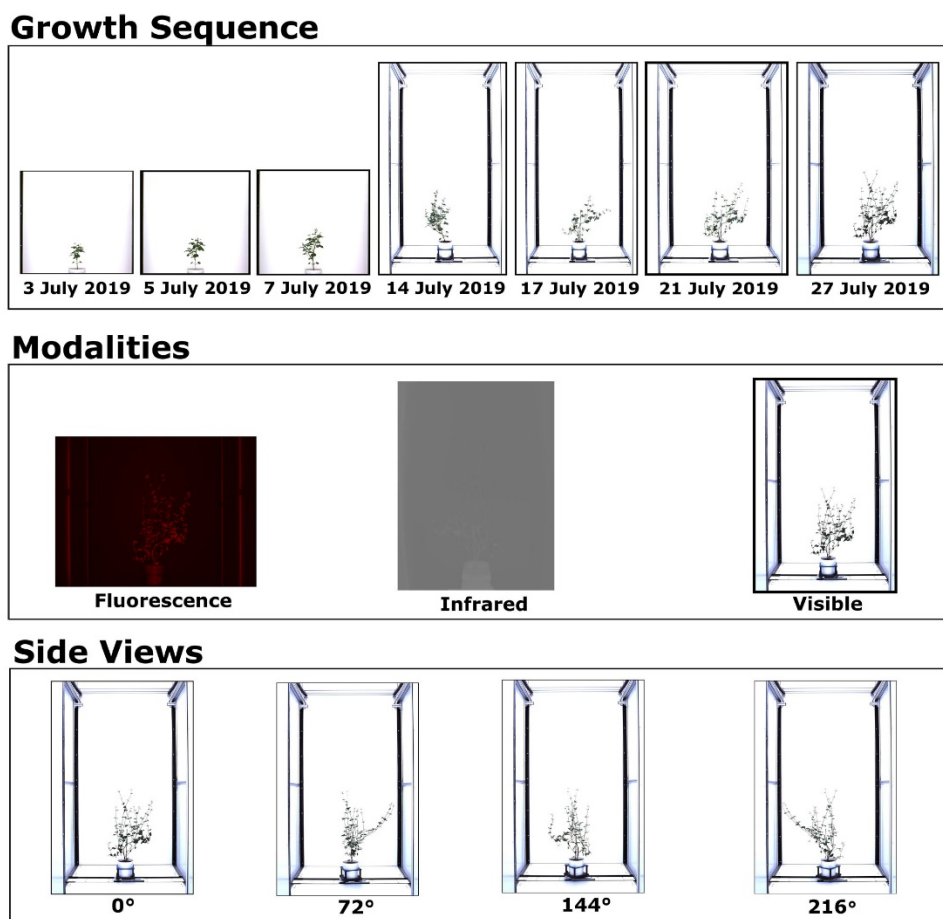


Figure 1. A preview of a VSTEM Dataset. This work will use the CosegPP dataset (Quiñones et al., 2021) and modify it as CosegPP+ and categorize it as a VSTEM dataset for our problem definition. The first row shows the growth sequence of a Buckwheat plant from 3rd July 2019 to 27th July 2019. The second row shows the three modality types of a Buckwheat plant on 27th July 2019. The third row shows the four different side views of a Buckwheat plant on 27th July 2019.

OSC-CO² is systematically evaluated using a VSTEM dataset retrieved from a high-throughput plant phenotyping facility at the University of Nebraska–Lincoln, USA. This paper will use CosegPP+ [an extension of CosegPP (Quiñones et al., 2021)], which consists

of growth sequences of multiple plants in their early growth period during which the plant grows, changes shape and appearance, develops new organs (leaves), and sometimes even rotates about a vertical axis for optimal light interception. The plants are grown in a greenhouse and are imaged daily in a high-throughput imaging platform using infrared, fluorescence, and RGB cameras with multiple views. Thus, the dataset exemplifies the VSTEM imagery and is manifestly more challenging than any dataset used in cosegmentation research because it categorizes the many features that pertain to the object while other datasets are random in their object selection with undefined features. The CosegPP+ dataset is available at <https://doi.org/10.5281/zenodo.6863013>.

The specific contributions of this research include:

- An end-to-end unsupervised deep learning algorithm, OSC-CO², to cosegment VSTEM imagery,
- Novel cosegmentation and temporal loss functions to adapt to the challenges of a high volume and variety VSTEM dataset,
- The demonstration of the efficacy of OSC-CO² in a complex application domain—plant phenotyping. What makes plant phenotyping a complex domain is the dynamic growth and environmental interaction of an object (plant).

The literature is summarized in the next section, with a discussion on open problems, dataset biases, and how OSC-CO² addresses these research gaps. The metrics used to evaluate the efficacy of the cosegmentation algorithms are also briefly described. Section 3 presents the problem definition, introduces our overall framework, and provides details of the different parts of the framework. Section 4 discusses the quantitative and qualitative results based on our evaluation using the precision and Jaccard index measures. Section 5 provides a final discussion, our conclusions, and potential directions for future work.

2. Literature review

This section summarizes the traditional and learning-based approaches for segmentation and cosegmentation algorithms and the datasets used for their evaluation. We also identify the research gaps and how our work addresses some of them.

2.1 Traditional segmentation algorithms

Traditional segmentation algorithms use a sequence of common image-processing steps to obtain a semantic and instant region of interest in an image. One technique is called frame differencing and is most used in high-throughput plant phenotyping systems since it is capable of imaging plants in a constant fixed camera, and lighting. It has been used by researchers (Woebbecke et al., 1992; Das Choudhury et al., 2017; Choudhury et al., 2018; Das Choudhury et al., 2020) by subtracting a fixed background image that includes only an empty pot, from an image that includes the plant while ensuring the background is constant. Although the segmentation technique is quick and low in computational demands, it does require images to come from a high-throughput system and is susceptible to residual noise around the plant region.

Color-based segmentation can address the issue of residual noise in imagery while having the flexibility to be used without a high-throughput system. This technique partitions parts of the image into different color regions based on its color features while assuming the color features are homogenous. Different color-based segmentations algorithms use either RGB (red, green, and blue) or Lab (where L represents lightness, a indicates the red [positive] or green [negative], and b represents the yellow [positive] and blue [negative], or HSV [hue, saturation, and value] depending on the color space application). It was first inspired by Woebbecke et al. (1992), which used the green and red channels to derive the normalized difference index. Then, other researchers leveraged this idea by utilizing the red channel to derive the excess red index (Meyer and Neto, 2008); red, green, and blue channels to derive the color index of vegetation (Kataoka et al., 2003); and others (Hunt et al., 2005; Meyer and Neto, 2008; Zheng et al., 2009). It is a common technique to use for semantic analysis since a majority of a plant can be green compared to the background, but it does not handle multiple-colored objects (common in stress-induced plants) or when the background color is like the object.

Shape modeling-based segmentation is commonly used for leaf or stem semantic analysis (Yin et al., 2014; Agapito et al., 2015; Thorp et al., 2016; Chen et al., 2017; Li et al., 2017; Chen et al., 2019; Roggiolani et al., 2023; Williams et al., 2023), flower instant analysis (Thorp et al., 2016; Zhang et al., 2022; Mahajan et al., 2023), and fruits (Grift et al., 2017; Fu et al., 2019). Chen et al. (2017) had RGB images as an input to their framework, where it transformed the images into a polar coordinate system by using a plant's density center as the origin. Thorp et al. (2016) used RGB images as input and transformed them into an HSI (hue, saturation, intensity) color space and then segmented the flowers using a Monte Carlo approach. Pape and Klukas (2015) attempted to reduce the impacts of illumination variability by modeling 3D histograms of LAB color space to aid in the segmentation process for rosette plants. Schar et al. (2016) applied a super pixel-based unsupervised approach that can extract various regions of interests by implementing a seeded region growing algorithm. These techniques usually do not produce a high segmentation accuracy with varying accuracies of 40% to 80% (with superficial image edits, such as cropping, to improve the segmentation).

In addition to color being an obstacle for segmenting the region of interest from the background, the irregular physical characteristics of a plant and inconsistencies in lighting during data acquisition limit the effectiveness of simple, traditional segmentation methods. Therefore, approaches based on thresholding (Otsu, 1979; Sezgin and Sankur, 2004; Meng et al., 2019), frame differencing (Choudhury et al., 2018), color-based (Woebbecke et al., 1992; Zheng et al., 2009), and morphological operations (Zhou et al., 2021) have, in general, proven to be ineffective for segmentation in high-throughput plant imaging datasets. These techniques are unable to overcome image acquisition inconsistencies, including lighting variation, shadows, and plant positions, and more significantly, do not consider or leverage the dynamic nature of the plants' evolving physical characteristics (Choudhury et al., 2018; Choudhury, 2020).

2.2 Learning-based segmentation algorithms

Learning-based segmentation algorithms are the preferred method especially since traditional segmentation algorithms often yield unsatisfactory results due to the plant being a complex object (Yin et al., 2014). Traditional segmentation algorithms also have issues overcoming common computer vision problems such as occlusion and large-scale lighting variations.

Clustering-based segmentation is a classification technique that attempts to find relational information among pixels in an image and classify them based on a similarity measure. These algorithms are the prerequisite for pursuing further complex phenotypic traits. They can eliminate noisy spots (Lee et al., 2018; Guo et al., 2021) and obtain homogenous regions (Ojeda-Magaña et al., 2010; Liu et al., 2018). Some segmentation algorithms target the semantic segmentation of plants, while others target the instant segmentation of plant parts, such as leaves, flowers, and fruit. Liu et al (Liu et al., 2018). utilized a 3D point cloud and spectral clustering to semantically segment *Ixora*, *Brassica*, *Wheat*, and *Basic* plants. Then, they further segmented down to each of the plant part's leaves and stem. Their technique was unique, and their framework was capable of segmenting a variety of plant species. Valliammal et al (Valliammal and Geethalakshmi, 2012). proposed a study where they applied wavelet transformation and fuzzy clustering to segment leaves. They was able to provide good segmentation results while achieving high identification of the leaf's edges. Another study (Wang et al., 2018) proposed a framework that combined the Sobel operator and the Chan-Vese model to segment cucumber leaves with complex background and occlusion issues. A downside to these algorithms is that they are sensitive to high levels of noise and gray inhomogeneity and are difficult to determine the initial parameters (Li et al., 2020).

Researchers have begun using Convolutional Neural Networks (CNNs) in their applications since 2012 due to their promising performance in semantic and instance segmentation (Jiang and Li, 2020) by utilizing the foreground object's features. Most applications combine CNNs with deep learning libraries such as Caffe (Pound et al., 2017), TensorFlow (Koh et al., 2021), PyTorch (Zhou et al., 2021), and Keras (Gong et al., 2021) for their analysis. Researchers have attempted to utilize neural network-based segmentation algorithms to count plant organs that have replaced some traditional-based and clustering-based algorithms. Most of these neural network-based algorithms (Bolya et al., 2019; Chen et al., 2020; Kirillov et al., 2020) require plenty of images with pixel-level annotation and available training data. Neural network-based algorithms are also used for data augmentation strategies for plant organ identification, segmentation, and counting (Das Choudhury et al., 2017; Das Choudhury et al., 2020; Mazis et al., 2020). Studies that have used CNNs have shown to have achieved accuracies from 87% to 99% for stress-based application and classification (Mohanty et al., 2016; Cruz et al., 2017; DeChant et al., 2017; Fuentes et al., 2017; Lu et al., 2017; Wang et al., 2017; Barbedo, 2018; Barbedo, 2018; Ferentinos, 2018; Liu et al., 2018; Suh et al., 2018; Nazki et al., 2020) but with manual or naïve modifications of the binary masks after processing.

2.3 Cosegmentation algorithms

Merdassi et al. (Merdassi et al., 2020), categorized cosegmentation algorithms into eight categories: Markov Random Fields-based Cosegmentation (MRF-Coseg), Co-Saliency-based Cosegmentation (CoS-Coseg), Image Decomposition-based Cosegmentation (ID-Coseg), Random Walker-based Cosegmentation (RW-Coseg), Maps-based Cosegmentation (M-Coseg), Active Contours-based Cosegmentation (AC-Coseg), Clustering-based Cosegmentation (Cl-Coseg), and Deep Learning-based Cosegmentation (DL-Coseg). They quantified that almost all algorithms in these categories used only color and texture features. This presents an issue because if the algorithms are intended to only recognize color and features, then the algorithm cannot detect a heterogeneous object that may consist of multiple distinctive regions. Complex, objects, such as plants, are heterogenous and can vary in color and texture as time progresses.

Several DL-Coseg studies (Hsu et al., 2018; Li et al., 2018; Meng et al., 2019) have found that using a CNN-based framework is optimal for detecting, extracting, and map-generating an object's features for a set of images. Hsu (Hsu et al., 2019) used CNNs to detect co-peaks for an image pair and its features to determine segmentation masks. Li (Li et al., 2018) utilized a CNN-based Siamese encoder-decoder architecture to extract semantic features of the objects in a set of images. Hsu (Hsu et al., 2018) generated heat maps for each image and transformed the results for cosegmentation via dense CRFs. These algorithms require a largescale set of images to achieve results, but that is extremely time-consuming. Although some algorithms tackle this problem, they end up being semi-serviced learning-based methods (Kim et al., 2011; Wang and Liu, 2013).

Recent cosegmentation algorithms have tackled important issues such as occlusion (Jerripothula et al., 2021; Meng and Zhao, 2022) by leveraging a combination of techniques to aid in object detection. Also, the literature supports incipient cosegmentation applications in pancreas research (Liu et al., 2022). Other sources of imagery, such as UAV-based high throughput platforms (Rico et al., 2020; Rico et al., 2021) or applications to improve the predictability of phenotypes (Sarzaeim et al., 2022; Sarzaeim et al., 2023), illustrate the potential of the use of cosegmentation algorithms and datasets for a variety of object types.

2.4 Cosegmentation datasets

Several datasets to demonstrate the efficacy of cosegmentation algorithms have been proposed in the literature, including iCoseg (Batra et al., 2010), MSRC (Winn et al., 2005), Internet (Rubinstein et al., 2013), Flickr-MFC (Kim and Xing, 2012), and PASCAL-VOC (Everingham et al., 2010). However, these datasets do not reflect the complexity in many application domains where the shape of the objects change over time (temporality), the objects are imaged in different imaging sensors (modality), and under different environmental conditions. These datasets are extremely big in image count that it is difficult to parse or even provide ground truth data for them. Furthermore, most of the objects in the datasets are random and have yet to be used in domain-specific research. The datasets are also limited in their ability to be used in current problems that require temporality, object state change, and multiple modalities for a diverse set of data points. CosegPP (Quiñones et al., 2021), on the other hand, contains many features that could be leveraged to aid

current problems to advance cosegmentation research in general and in application-specific work-plant phenotyping.

Traditional and learning-based algorithms are not advanced enough to handle a VSTEM dataset since they currently rely on naïve modifications or are unable to leverage the necessary information for deeper plant analysis. Cosegmentation algorithms tend to overcome the issues with traditional and feature-based algorithms and address complex challenges, such as occlusion, but are efficient only for a specific type of dataset. An overview of cosegmentation datasets suggests that they may not be useful for domain-specific applications and motivates the need for domain-specific datasets. Our work introduces an end-to-end unsupervised deep learning framework and a VSTEM dataset that is (1) high dimensional and (2) contains a small number of (7 to 14) images. Furthermore, it is the first cosegmentation-based algorithm proposed and tested for plant phenotyping.

3. Methods

OSC-CO² uses an information fusion approach by leveraging the outputs from multiple segmentation and cosegmentation methods to learn and refine the segmentation of VSTEM images. Specifically, for the images in a VSTEM image dataset, the object of interest will exhibit a variety of state changes over time (but captured at specific temporal points) and is captured in multiple imaging modalities at multiple views. All the available information can be leveraged to aid segmentation. Iteratively, the whole VSTEM dataset is segmented by OSC-CO² by determining the object using coattention and then cosegmenting the object with one pair of images. A novelty in OSC-CO² is that, unlike traditional CNNs, it is completely unsupervised, which suggests that no additional data annotations are needed. OSC-CO² is implemented in three stages: Object Mask Generation (OMG), Object Mask Refinement (OMR), and Final Joint Mask Generation (FJMG). Figure 2 shows the three stages and how they are processing the dataset input and sending information across stages to generate the metric output.

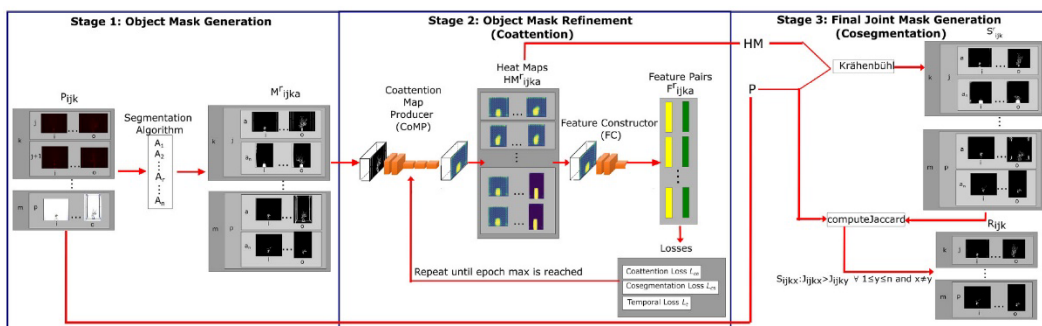


Figure 2. Block diagram of OSC-CO². This method contains three stages. The Object Mask Generation stage takes imagery, P , and produces binary imagery, M . The Object Mask Refinement stage takes the imagery, M , and processes it through the Coattention Map Producer, CoMP, generating an HM set of heat maps. Then the maps go through the Feature Constructor, FC, to produce a two-column tensor of feature information of the

background and object. The two-column tensor is used as input to the coattention loss (l_{ca}) (Hsu et al., 2018) and our novel cosegmentation (l_{cs}) and temporal (l_t) loss. This stage stops after the epoch max is reached. For the Final Joint Mask Generation stage, we adopted dense conditional random fields (CRFs) as our refinement, taking HM and P as input for Krähenbühl's algorithm, and then giving a final dataset, R, after computing Jaccard index similarity between S and P.

3.1 Problem definition

The problem including VSTEM imagery can be formally defined as follows:

Given a plant P, imaged at m time points, o modalities, and in p side views, i.e., $P = \{P_{ijk}\}$, $1 \leq i \leq m$, $1 \leq j \leq o$, $1 \leq k \leq p$, where P_{ijk} is the image of plant P at time i , view j , and modality k , determine $R = \{R_{ijk}\}$, $1 \leq i \leq o$, $1 \leq j \leq p$, $1 \leq k \leq m$, where R_{ijk} is the final segmented mask for the plant P_{ijk} .

The proposed OSC-CO² algorithm incorporates a dynamic and expandable approach using coattention and cosegmentation analytics. It consists of three stages: (1) Object Mask Generation, (2) Object Mask Refinement, and (3) Final Joint Mask Generation. This model is designed to handle high-throughput image datasets to generate accurate separation of dynamic and evolving objects using a deep learning framework. This approach generates an ensemble of binary masks for a VSTEM image set and addresses some common challenges in segmentation, including background noise and the evolution of the object's morphology. The proposed OSC-CO² framework is summarized in Algorithm 1 below.

Input: $P = \{P_{ijk}\}$, $1 \leq i \leq o$, $1 \leq j \leq p$, $1 \leq k \leq m$, where P_{ijk} is the image of plant P at time i , in view j , modality k . C_{jk} is the class at modality k , and view j . $A = \{A_1, A_2, \dots, A_r, \dots, A_n\}$, where A_r is the r^{th} segmentation algorithm that takes a plant image and generates its mask.

Output: $R = \{R\}$, $1 \leq i \leq m$, $1 \leq j \leq o$, $1 \leq k \leq p$, where R_{ijka} is the refined binary masks for the plant P_{ijka}

Stage 1: Object Mask Generation (P, A)

Begin

For $1 \leq a \leq n$ **do**

For $1 \leq i \leq o$ **do**

For $1 \leq j \leq p$ **do**

For $1 \leq k \leq m$ **do**

$M_{ijk} = \text{Segment}(A_i, P_{ijk});$

Return (M)

End

Stage 2: Object Mask Refinement (M)

Begin

For $1 \leq i \leq o$ **do**

For $1 \leq j \leq p$ **do**

For $1 \leq k \leq m$ **do**

For $1 \leq a \leq n$ **do**

Begin//process the image temporal sequence.

 Initialize (CoMP, FC)

```

For  $1 \leq k \leq m - 1$  do
  Begin
     $I_{prev} = M^{r_{i,j,k}}$ ;  $I_{current} = M^{r_{i,j,(k+1)}}$ 
     $L_{ca} = L_{co} = L_t = 0$ 
  Repeat
     $HM^{r_{i,j,ka}} = COMP(I_{prev}, I_{current})$  //heat maps
     $F^{r_{i,j,ka}} = FC(I_{prev}, I_{current})$  //features
    Compute  $L_{ca}$  using Eq. (1)
    Compute  $L_{cs}$  using Eq. (4)
    Compute  $L_t$  using Eq. (7)
     $Loss_{Total} = 1 * L_{ca} + 0.5 * L_{co} + 0.5 * L_t$ 
    Backpropagate loss and update weights
  Until epoch max is reached
  End
End
Return ( $HM$ )
End
Stage 3: Final Joint Mask Generation ( $P, HM$ )
For  $1 \leq i \leq o$  do
  For  $1 \leq j \leq p$  do
    For  $1 \leq k \leq m$  do
      For  $1 \leq a \leq n$  do
         $S^{r_{ijk}} = Kr\ddot{a}henb\ddot{u}hl(P_{ijk}, HM^{r_{ijk}})$ 
         $J^{r_{ijk}} = computeJaccard(S^{r_{ijk}}, P_{ijk})$ 
      End
       $R_{ijk} = S_{ijkx} : J_{ijkx} > J_{ijk y} \forall 1 \leq y \leq n$  and  $x \neq y$ 
    End
  End
End
Return ( $R$ )
    
```

Algorithm 1. Proposed OSC-CO² framework.

3.2 Overview

Given a set of images of a single plant P in different modalities and views at different time points, we start with a set of basic segmentation algorithms to generate initial masks. Segmented images in the temporal sequence are reconciled in order using deep neural networks with novel loss functions. The final segmentation results are derived by analyzing the refined segmentation results from different algorithms. Figure 2 shows an overview of the OSC-CO² framework. As shown in the figure, OSC-CO² consists of three key stages: initial mask generation, mask refinement, and final mask creation.

3.3 Object mask generation

The goal of the OMG stage is to generate the initial segmentations for all the images of the plant, including all modalities, views, and time points. The masks are generated for each segmentation algorithm and are used in Stage 2 to refine them. In the OMG stage, the input images, P , are processed through a set of algorithms selected by the user defines to produce a set of binary images, M , for all the algorithms. This stage has no limit to the number of algorithms and images that it can handle, but it could be limited by a computer's processing power.

3.4 Object mask refinement

The OMR stage takes the binary imagery, M , through a neural network called the Coattention Map Producer (CoMP) to create heat maps for each image. Our definition of a heat map is shown in Figure 3. The heat maps are passed through another network called the Feature Constructor (FC) that computes the features of the estimated objects and the background. The CoMP learns by optimizing multiple loss functions designed to address the challenges in cosegmenting a VSTEM dataset with evolving objects. The three functions are temporal, cosegmentation, and coattention loss. Temporal loss measures the inter-image object difference defined by the distance between the feature pairs of the current image and that of the previously computed image. The cosegmentation loss measures the foreground-background discrepancy within each image. The third is coattention loss adapted from Hsu et al. (2018), which enhances inter-image object similarity and intra-image figure-ground distinctness per image. Finally, the P and HM imagery are inputs to the FJMG stage to the dense conditional random fields (CRFs) cosegmentation algorithm to produce our framework's final joint masks, R .

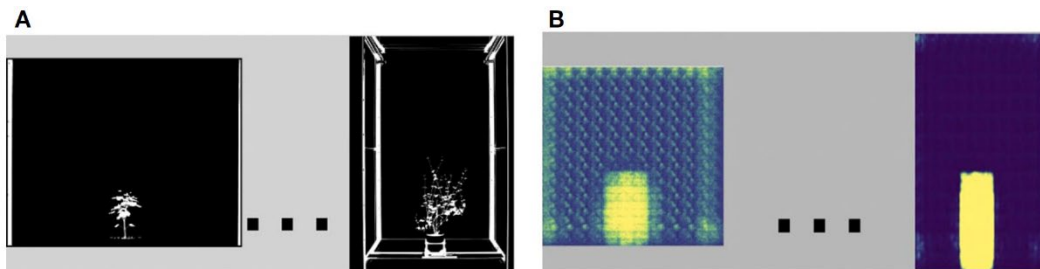


Figure 3. Our heat maps follow the standard color definition, in which purple and blue are the “coldest” (weak object prediction) and red and yellow are the “hottest” (strong object prediction). (A) shows some binary imagery with its heat maps (B).

As shown in Figure 4, the OMR module is composed of two collaborative CNN modules to produce the heat maps (heat maps that differentiate between the object and background) and feature pairs (descriptive correlation between an image's foreground and background). They are described below.

- Coattention Map Producer (CoMP): This module produces heat maps.

- Feature Constructor (FC): Generates feature information for the object and background in each image that can be used by the loss functions for optimization.

The details of the modules and their architecture are described in detail below.

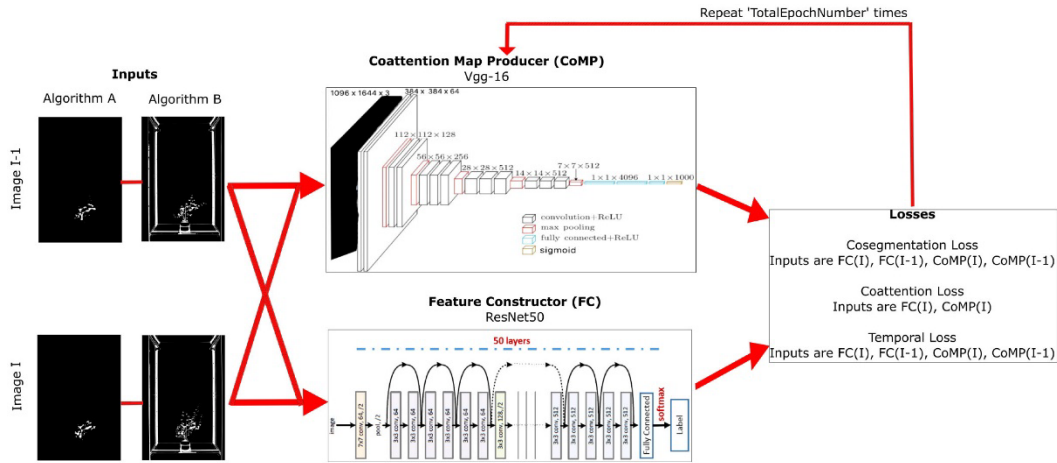


Figure 4. A detailed block diagram of Stage 2: Object Mask Refinement using Coattention. The inputs are simplified in this block diagram to show only the previous and current image with its corresponding imagery for the selected algorithms. These images are inputs to the CoMP and the FC. The generated heat maps from CoMP, and the numerical pairs from FC are used to compute the three losses. These losses are propagated back into CoMP, and this process repeats the number of epochs that is defined in the architecture.

3.4.1 Coattention map producer

The *CoMP* is a fully convolutional network (FCN) (Long, 2015) that has a modified ReLU layer to avoid modifying the data directly and avoid allocating any additional memory. An FCN was used since the architecture does not contain any dense layers, meaning the FCN can handle a wide range of image sizes because all connections are local. This is useful for VSTEM datasets that contain hundreds of images due to the temporal resolution. For each input image to *CoMP*, it estimates its heat map, i.e., $HM^{r_{ijk}} = CoMP(M^{r_{ijk}})$. We used the VGG-16 (Simonyan and Zisserman, 2015) setting of the FCN (Long et al., 2020) to create *CoMP*. Following Hsu et al. (2018), we replaced the last activation function *softmax* layer with a *sigmoid* function layer, which provides the heat maps as output. We also kept the learning rate set to 10^{-6} and fixed it during the optimization process following Hsu et al. (2018).

3.4.2 Feature constructor

FC is a Resnet50 (He et al., 2016) that takes in a segmented image, $HM^{r_{ijk}}$, and computes the semantic features object (I^n_o) and background (I^n_b) using Equations 2 and 3. A Resnet50 architecture was used since it has many layers that can be trained easily without increasing the training error while overcoming the vanishing gradient problem in the VGG-16 architecture in *CoMP*. This approach is useful for our imagery since our evolving objects contain

many parts, textures, shadows, and colors. These features are from the last fully connected layer of FC (Hsu et al., 2018) since VGG-16 (Simonyan and Zisserman, 2015) sometimes suffers from the vanishing gradient problem. Our method recognizes that early heat maps are too unstable and compensates with resilient adjustments. Furthermore, FC is an off-the-shelf model pretrained with ImageNet (Deng et al., 2009). We have set the features extracted in FC as inputs to the last fully connected layer.

3.4.3 Loss functions

A novel contribution of OSC-CO² is a loss function developed to address the unique properties of the VSTEM datasets. The overall loss function is defined as

$$L_{EMVS} = 1 * L_{ca} + 0.5 * L_{cs} + 0.5 * L_t,$$

where L_{ca} is the coattention loss, L_{cs} is the cosegmentation loss, and L_t is the temporal loss.

The coattention loss is designed to enhance both inter-image object similarity and intra-image figure-ground distinctness in each image, aiding in extracting our object type. Our novel cosegmentation loss optimizes the images by using the object and background's features for a high foreground object similarity across the output masks and a high foreground-background dissimilarity within each image. This loss will aid in extracting information across the different views and modalities. Our novel temporal loss optimizes the similarity of the foreground objects across two sequential images. This loss will help in providing information about a specific object's type for each environmental condition. All these losses target all the aspects of a VSTEM dataset.

3.4.3.1 Coattention Loss. The coattention loss is defined by Hsu et al. (2018) and is meant to guide *CoMP*'s training of optimal coattention masks by referring to the current object and background features that are computed by FC. The loss function is defined below:

$$l_c(\{I_n\}_{n=1}^N; w) = -\sum_{i=1}^N \sum_{j \neq i} \log(p_{ij}) \quad (1)$$

3.4.3.2 Cosegmentation Loss. One assumption we made about the VSTEM dataset is that the object relatively stays in the same position but grows outwards. We exploit this advantage to aid in object alignment. The proposed cosegmentation loss is designed to guide *CoMP* to generate a high foreground object similarity across the images and high foreground-background dissimilarity within each image. Given the current and previously computed image pairs (I_A, I_B) with the current and previously computed generated mask pairs from *CoMP* (S_A, S_B) , we produce the object and background features. We generate the object (I_n^o) and background (I_n^b) features by

$$I_n^o = FC(I_n \otimes S_n) \text{ and} \quad (2)$$

$$I_n^b = FC(I_n \otimes (1 - S_n)) \text{ for } n \in \{A, B\}, \quad (3)$$

where \otimes denotes the pixel-wise multiplication between the two operands. The cosegmentation loss (L_{cs}) is defined by

$$L_{cs}(I_A, I_B, F) = d_{AB}^+ + d_{AB}^-, \quad (4)$$

where d_{AB}^+ and d_{AB}^- are defined as

$$d_{AB}^+ = \frac{1}{c} \|F(I_A^0) - F(I_A^b)\|^2 \text{ and} \quad (5)$$

$$d_{AB}^- = \max\left(0, m - \left(\frac{1}{2c} \left(\|F(I_A^0) - F(I_A^b)\|^2 + \|F(I_B^0) - F(I_B^b)\|^2\right)\right)\right) \quad (6)$$

The margin m enlarges the difference between classes to enhance classification ability. If the margin is too large, the probabilities become unreliable, leading to a large loss for almost all samples (Zhang et al., 2019). For our framework, it is set to 2 as the cutoff threshold. Equation (5) aims to minimize inter-image foreground object distinctiveness, and Equation (6) maximizes the intra-image foreground-background discrepancy. Even though the cosegmentation loss L_{cs} is like the loss described in Chen et al. (2020), there is a significant difference. Our cosegmentation loss measures the mean squared error (MSE) (squared L2 norm) between each element in the input x and target y (the variable definition used by PyTorch) instead of using the dimension of the features as the constant c . In addition, since MSE penalizes prediction that is far away from the previously computed by applying a squared operator, we used that as our criterion to stop the computation of our losses when near convergence. To the best of our knowledge, computing a loss with results from previous iterations to model the temporal effect has not been explored.

3.4.3.3 Temporal Loss. This loss assumes that there exists an object that changes in shape because of environmental conditions, thus insinuating a discrepancy between the foreground and background as time progresses. The temporal loss is designed to measure the inter-image object distance between the current and previously computed image feature pairs of (I_A, I_B) . We generate the object (I_n^o) and background (I_n^b) features based on Equations (2) and (3). The temporal loss (L_t) is defined by

$$L_t(\{I_n\}_{n=1}^N) = -\sum_{i=1}^N \sum_{j \neq i} \log(p_{ij}), \quad (7)$$

where p_{ij} is defined as

$$P_{ij} = \frac{\exp(-ob_{ij}^+)}{\exp(-ob_{ij}^+) + \exp(-ob_{ij}^-)} \quad (8)$$

$$ob_{ij}^+ = \frac{1}{c} \|F(I_{B,i}^o) - (F(I_{B,j}^o))\|^2, \text{ and} \quad (9)$$

$$ob_{ij}^- = \frac{1}{c} (\|F(I_{A,i}^0) - F(I_{A,j}^0)\|)^2 \quad (10)$$

CoMP generates heat maps to optimize low inter-image object distances for both current and previously computed feature pairs using Equations (9) and (10). Our temporal loss is motivated by Hsu et al.'s (2018) coattention loss, but the difference is that ours ignores the intra-image figure-ground dissimilarity and computes the inter-image object distance with the current and previously computed image feature pairs.

3.5 Final joint mask generation

This stage uses the dense CRF approach proposed in Krähenbühl and Koltun (2011), where the unary and the pairwise terms are set to the two heat maps generated from the results of two segmentation algorithms and bilateral filtering, respectively. For each pixel in the heat maps, we define a probability that the pixel belongs to undefined classes. The hyperparameters for the network are summarized in Table 1. This stage outputs the final binary masks, *R*, by computing and selecting the mask with the highest Jaccard index similarity between the plant imagery dataset, *P*, and the segmented masks dataset, (*S*), using the dense CRF approach (Krähenbühl and Koltun, 2011).

Table 1. Hyperparameter values of the dense CR network approach (Gong et al., 2021)

Hyperparameter	Value
Weight of the bilateral term	10
Spatial standard deviation	80
RGB standard deviation	13
Weight of the spatial term	3.0
Spatial standard deviation	3
Number of iterations	5.0

3.6 Optimization process

OSC-CO² uses the ADAM optimizer to derive its hyperparameters because of its widespread use and rapid convergence (Hsu et al., 2018; Mehta et al., 2019; Melinte and Vladareanu, 2020) properties. The final parameters determined by ADAM include a 0.01 learning rate and a 0.0005 weight decay for *CoMP*'s parameters. At the start of processing each one-pair of images, the optimizer sets all the gradients to zero.

3.7 The data repository creation

The VSTEM imagery used to evaluate the performance of OSC-CO² is based on the CosegPP data repository (Quiñones et al., 2021). The data repository has plant images with a large inter-class variation and background noise. The images were captured using the LemnaTec Scanalyzer at the University of Nebraska–Lincoln, USA. It contains two buckwheat plants, one of which underwent drought stress and the other remained the control, and two sunflower plants, one of which underwent drought stress and the other remained the control. Each plant represents a dataset that has four side views (0°, 72°, 144°, 216°) and three modalities (fluorescence, infrared, and visible) with 7 to 14 time points.

We created an extension of CosegPP's datasets, which we will refer to as CosegPP+, by processing all four plant datasets through segmentation using Otsu's method (Otsu, 1979) and cosegmentation using Subdiscover (Meng et al., 2016). These two methods were chosen because Quiñones et al. (2021) defined these as the top methods for being able to segment some of the challenging features of computer vision. CosegPP+ is publicly available at <https://doi.org/10.5281/zenodo.6863013>.

We replaced the original images with the outputs generated by Otsu's method and Subdiscover, meaning that each time point i will have at most a binary images, where a is the number of algorithms (i.e., Otsu's method and Subdiscover) used. Some groups do not contain Subdiscover binary masks because of the method's limitation in not being able to segment the original images.

3.8 Implementation

OSC-CO² allows for the dynamic input of epoch runs, but we used 10 epoch runs for CosegPP+. It is worth noting that CosegPP began overfitting after 7 epochs. OSC-CO² also requires a minimum 2 epochs to allow for the heat maps to generate stable proposals and generate the coattention loss based on pixel-wise averaging of the masks. The cosegmentation loss and temporal loss activates at the final epoch. The batch size is set to a (the number of algorithms used as input with their binary masks). Also, all input images are resized to 384 × 384 pixel resolution prior to subsequent processing because FC can be applied only to images of the same size while using 3 × 3 kernels with a stride of 3 and 3 for height and width, respectively, and with an initial learning rate of 0.001. After the cosegmentation, we resized the images back to their original sizes for performance evaluation.

4. Results and discussion

In this section, we describe the dataset, performance metrics, experimental design, and evaluation results for OSC-CO². The results are compared with several existing methods to demonstrate the efficacy of OSC-CO².

4.1 Evaluation metrics

Our experiment uses two widely used metrics, Precision (P) and Jaccard index's (J) similarity (also known as IoU) to evaluate the final estimated masks for each time point. Precision is the measurement that identifies the percentage of correctly segmented pixels. Jaccard index is the measurement for the intersection area ratio between the detected object and ground truth. Both metrics will range from 0 to 1 where 1 is the ideal value. We chose these two metrics because of their continuous use in coattention and cosegmentation analytics (Dong et al., 2015; Meng et al., 2016; Wang and Shen, 2016; Li et al., 2018; Merdassi et al., 2020).

4.2 Quantitative results

Table 2 summarizes the performance of OSC-CO² on the CosegPP+ data repository. The precision and Jaccard index similarity scores for the four plants are presented for each of the three modalities (visible, fluorescence, and infrared). For each modality, the scores for

each of the four views are also reported. Finally, the average scores for each plant over all the views and for all the modalities are presented. OSC-CO² produces the highest precision and Jaccard index similarity scores for the fluorescence and visible modality for the buckwheat species and high scores for the infrared modality for the sunflower species.

Table 2. The performance of our OSC-CO² on the CosegPP+ data repository

		Fluo 0°	Fluo 72°	Fluo 144°	Fluo 216°	IR 0°	IR 72°	IR 144°	IR 216°	Vis 0°	Vis 72°	Vis 144°	Vis 216°	Fluo Avg	IR Avg	Vis Avg	All Avg
<i>BC</i>	P	0.93	0.93	0.93	0.92	0.90	0.90	0.90	0.90	0.99	0.99	0.99	0.99	0.93	0.90	0.99	0.94
	J	0.39	0.39	0.39	0.38	0.32	0.31	0.29	0.29	0.37	0.42	0.35	0.33	0.39	0.31	0.37	0.36
<i>BD</i>	P	0.92	0.92	0.93	0.93	0.89	0.89	0.89	0.89	0.98	0.95	0.98	0.97	0.93	0.89	0.97	0.93
	J	0.43	0.44	0.47	0.46	0.35	0.35	0.34	0.34	0.33	0.27	0.30	0.38	0.45	0.35	0.32	0.37
<i>SC</i>	P	0.93	0.86	0.86	0.87	0.92	0.95	0.95	0.95	0.91	0.90	0.92	0.91	0.88	0.95	0.91	0.91
	J	0.36	0.31	0.32	0.34	0.41	0.37	0.37	0.36	0.43	0.34	0.44	0.43	0.33	0.38	0.41	0.37
<i>SD</i>	P	0.95	0.88	0.88	0.91	0.98	0.98	0.99	0.98	0.89	0.92	0.92	0.92	0.91	0.99	0.92	0.94
	J	0.23	0.27	0.28	0.27	0.34	0.31	0.31	0.29	0.27	0.26	0.26	0.35	0.27	0.32	0.29	0.29

BC, Buckwheat (Control); BD, Buckwheat (Drought); SC, Sunflower (Control); SD, Sunflower (Drought); P, Precision; J, Jaccard index similarity; Vis, Visible imagery; Fluo, Fluorescence imagery; IR, Infrared imagery. The best average Precision and Jaccard index similarity per dataset are shown in bold.

Table 3 compares the performance of OSC-CO² against Otsu’s method, a widely used algorithm in the plant phenotyping domain, and Subdiscover, a leading cosegmentation method. The performance results for Otsu’s algorithm and Subdiscover are derived from our previous research (Quiñones et al., 2021). Table 3A summarizes the effectiveness of our method, OSC-CO², under normal growth conditions and under drought for buckwheat images. Table 3A shows that the performance of OSC-CO² is comparable to other algorithms based on the precision score; it is slightly lower in fluorescence and infrared modality but the same or higher for visible imagery. However, the Jaccard index similarity measures for OSC-CO² are significantly superior to other algorithms. This implies that OSC-CO² can properly detect the object’s pixels, but at the expense of a very slight reduction in precision.

Table 3. The comparative evaluation of OSC-CO² with Otsu’s algorithm and Subdiscover

A. Performance comparison on buckwheat species												
	Buckwheat (Control)						Buckwheat (Drought)					
	Fluo		IR		Vis		Fluo		IR		Vis	
	P	J	P	J	P	J	P	J	P	J	P	J
Otsu	0.97	0.38	0.90	0.05	0.93	0.14	0.97	0.37	0.90	0.05	0.93	0.14
Subdiscover	—	—	—	—	0.99	0.34	—	—	—	—	0.97	0.23
OSC-CO ² (ours)	0.93	0.39	0.91	0.31	0.99	0.37	0.93	0.45	0.89	0.35	0.97	0.32

B. Performance comparison on sunflower species												
	Sunflower (Control)						Sunflower (Drought)					
	Fluo		IR		Vis		Fluo		IR		Vis	
	P	J	P	J	P	J	P	J	P	J	P	J
Otsu	0.87	0.26	0.91	0.31	0.92	0.30	0.87	0.12	0.93	0.20	0.93	0.14
Subdiscover	0.84	0.10	0.94	0.19	0.98	0.47	0.84	0.00	0.98	0.09	0.98	0.16
OSC-CO ² (ours)	0.92	0.39	0.92	0.35	0.96	0.37	0.92	0.36	0.94	0.34	0.95	0.35

BC, Buckwheat (Control); BD, Buckwheat (Drought); SC, Sunflower (Control); SD, Sunflower (Drought); P, Precision; J, Jaccard index similarity; Vis, Visible imagery; Fluo, Fluorescence imagery; IR, Infrared imagery. The best average Precision and Jaccard index similarity per dataset are shown in bold. “—” means the algorithm did not generate an output.

Table 3B compares the performance of OSC-CO² for the sunflower species. The results are like those for the buckwheat species. The performance for the fluorescence and infrared modality is slightly lower for precision but significantly better overall for Jaccard index similarity. However, Subdiscover clearly outperformed OSC-CO² in the visible modality. This may be due to the inconsistent appearance of sunflower images due to the presence of flowers (yellow) with green stems and leaves.

4.3 Qualitative results

Figure 5 shows some sample images from the CosegPP+ data repository and their corresponding segmentations generated by OSC-CO². For the buckwheat images, OSC-CO² removes most of the background imaging chamber (background noise in plant phenotyping) while leaving the object (plant) intact. However, the infrared modality was not as accurate in this case. Sunflower images show similar patterns. However, it is noticeable with the Visible light modality that it came at a cost by removing most of the plant itself from the object since those temporal points began to include the flowers. Furthermore, the sunflower displays an empty result for the drought environmental conditions on the last temporal point leading to the assumption that it overfitted too soon. Thus, suggesting the future work of a dynamic epoch cutoff.

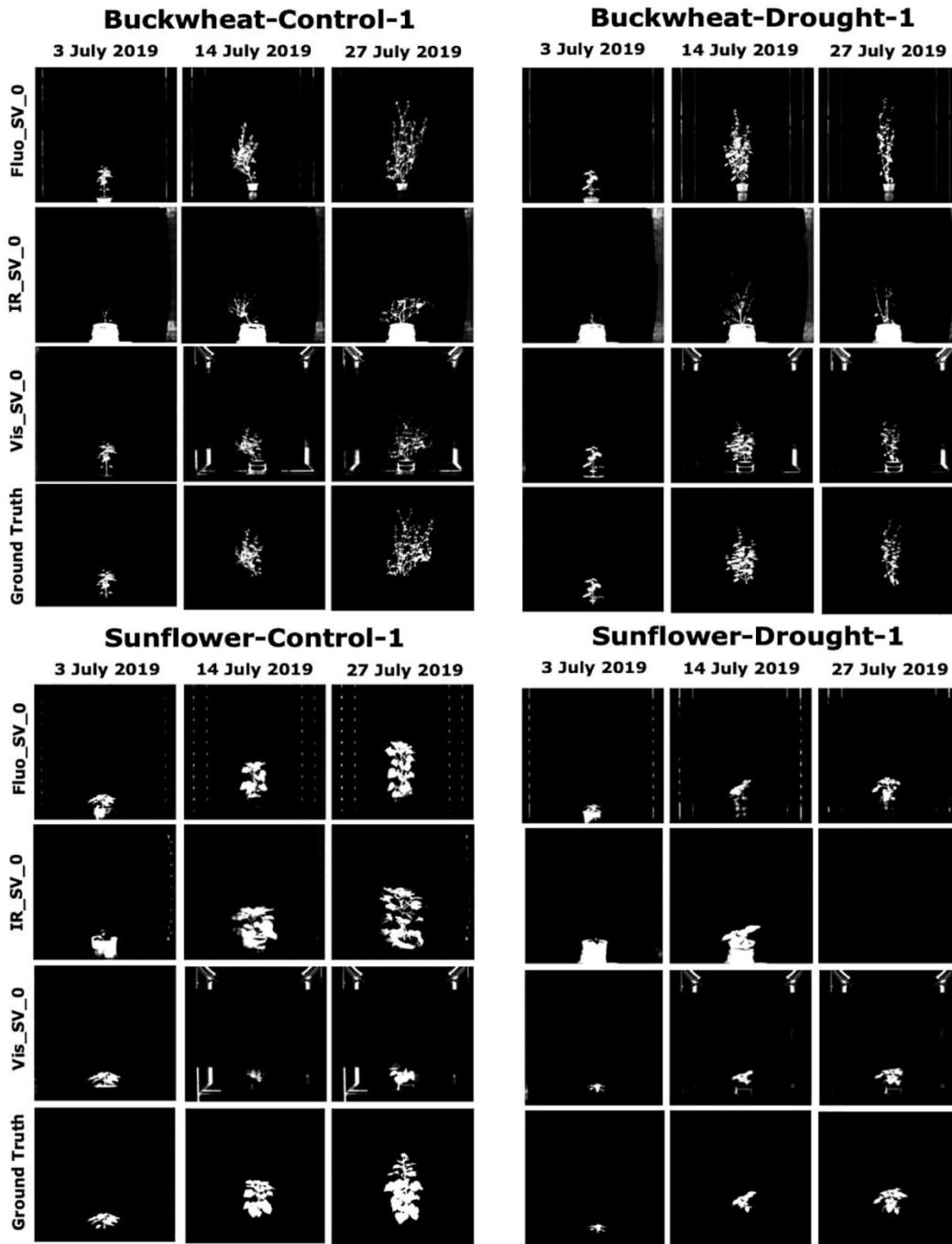


Figure 5. Illustration of qualitative performance of OSC-CO² on the CosegPP data repository. This preview shows only three temporal points (start, middle, end), and one side view for all modalities.

As evident from Figure 5, the sunflower plants tend to have thicker stems and hence generate more discernable infrared imagery, resulting in better segmentation accuracy than buckwheat. Furthermore, since buckwheat plants were imaged during the vegetative state only, they were green throughout the imaging period. In contrast, sunflower plants have yellow flowers, sometimes several, during the later stages of growth. Therefore, visible images for buckwheat were more consistent, leading to the highest accuracy. Similarly, the green organs in buckwheat in fluorescence imagery, which serves as a proxy for chlorophyll level, have higher segmentation accuracy in this modality.

5. Conclusion and future work

In this paper, we achieve our first contribution of designing an unsupervised method for cosegmenting binary plant imagery by using CNNs that outperformed previous works (Otsu, 1979; Meng et al., 2016) by improving segmentation accuracy by 3% to 45%. The model has three stages. The first stage is the Object Mask Generation which produces the necessary binary imagery from a set of user-defined algorithms. The second stage is the Object Mask Refinement which uses FCN32, VGGNet, and ResNet50. We also achieved our second contribution by designing two novel unsupervised cosegmentation and temporal loss for stage two with one unsupervised coattention loss from literature. The third stage is Final Joint Mask Generation which refines the binary image output by using the heat maps. The experimental results demonstrate a promising new technique that can learn and enhance binary masks, without training data, to refine the masks leading to higher segmentation accuracy for further object analysis.

Using CNNs for evolving objects at different temporal stages shows promising development in increasing accuracy that it may replace some traditional methods for plant phenotyping. This paper creates an unsupervised coattention and cosegmentation method for high-throughput datasets with defined quantitative and qualitative features that leverage the information from multiple algorithms' binary output. Within this framework, we have proposed two novel loss functions: cosegmentation and temporal loss that aids the coattention loss by helping the discovery of the foreground object while removing background noise.

For our third contribution, experimental evaluations of OSC-CO² on CosegPP+ demonstrate the method's great capabilities of being able to recognize the evolving, moving object. This also introduces a base analysis for different types of modalities that are being used more in plant phenotyping analytics. Our method was able to leverage these object features to produce and demonstrate its optimal performance among different modalities and environmental conditions.

This paper is a critical contribution to image segmentation to high-throughput multimodal image segmentation because it eliminates the need for researchers to perform naïve image preprocessing, such as image cropping that may skew an algorithm's performance by eliminating the complex aspect of an image that may challenge, and push future algorithmic development.

Future work includes implementing a dynamic epoch cutoff algorithm tailored to dataset varieties in terms of environmental conditions and species. An adjustment of the

coattention framework can be made to include the selection of flowers and merge it with the object. This could significantly improve segmentation accuracy. Finally, hyperparameter weights can be implemented for the different dimensions in the dataset so that the algorithm can leverage more of the appropriate dimension for higher segmentation accuracy.

Author contributions – RQ contributed as the first author in constructing and analyzing the framework, led the experimental analysis, and writing the manuscript. AS, FMA, and SDC contributed in supervising, writing, and reviewing the original draft. All authors contributed to the article and approved the submitted version.

Funding – This material is based upon work supported by the National Science Foundation under Grant No. DGE-1735362. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Also, the authors acknowledge the support provided by the Agriculture and Food Research Initiative Grant number NEB-21-176 and NEB-21-166 from the USDA National Institute of Food and Agriculture, Plant Health and Production and Plant Products: Plant Breeding for Agricultural Production.

Acknowledgments – This work was completed utilizing the Holland Computing Center of the University of Nebraska, which receives support from the Nebraska Research Initiative. We acknowledge the Agricultural Research Division, Institute of Agriculture and Natural Resources, University of Nebraska–Lincoln for providing the images from the LemnaTec.

Conflict of interest – The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Agapito, L., Bronstein, M. M., and Rother, C. (2015). “Computer vision—ECCV 2014 workshops,” in *2014 Proceedings, Part IV. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, (Zurich, Switzerland: Springer), September 6–7 and 12, Vol. 8928. 61–74. doi: 10.1007/978-3-319-16220-1
- Alessandretti, G., Broggi, A., and Cerri, P. (2007). Vehicle and guard rail detection using radar and vision data fusion. *IEEE Trans. Intell. transportation Syst.* 8 (1), 95–105. doi: 10.1109/TITS.2006.888597
- Barbedo, J. G. A. (2018). Factors influencing the use of deep learning for plant disease recognition. *Biosyst. engineering* 172, 84–91. doi: 10.1016/j.biosystemseng.2018.05.013
- Barbedo, J. G. A. (2018). Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Comput. electronics Agric.* 153, 46–53. doi: 10.1016/j.compag.2018.08.013
- Batra, D., Kowdle, A., Parikh, D., Luo, J., and Chen, T. (2010). “Icoseg: Interactive cosegmentation with intelligent scribble guidance,” in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 3169–3176, (San Francisco, California: IEEE).
- Bolya, D., Zhou, C., Xiao, F., and Lee, Y. J. (2019). “YOLACT: Real-time instance segmentation,” in *Proc. IEEE Int. Conf. Comput. Vis* (Seoul, South Korea: IEEE), 2019 Octob. 9156–9165. doi: 10.1109/ICCV.2019.00925

- Chang, D., Chirakkal, V., Goswami, S., Hasan, M., Jung, T., Kang, J., et al. (2019). "Multi-lane detection using instance segmentation and attentive voting," in *2019 19th International Conference on Control, Automation and Systems (ICCAS)* (Jeju, South Korea: IEEE), 1538–1542.
- Chen, Y., Baireddy, S., Cai, E., Yang, C., and Delp, E. J. (2019). "Leaf segmentation by functional modeling," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (Long Beach, California: IEEE), 0–0.
- Chen, Y.-C., Lin, Y.-Y., Yang, M.-H., and Huang, J.-B. (2020). "Show, match and segment: joint weakly supervised learning of semantic matching and object cosegmentation," in *IEEE transactions on pattern analysis and machine intelligence* (IEEE).
- Chen, Y., Ribera, J., Boomsma, C., and Delp, E. J. (2017). "Plant leaf segmentation for estimating phenotypic traits," in *2017 IEEE International Conference on Image Processing (ICIP)* (Beijing, China: IEEE), 3884–3888.
- Chen, X., Shrivastava, A., and Gupta, A. (2014a). "Enriching visual knowledge bases via object discovery and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Columbus, Ohio: IEEE), 2027–2034.
- Chen, X., Shrivastava, A., and Gupta, A. (2014b). "Enriching visual knowledge bases via object discovery and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*.
- Chen, H., Sun, K., Tian, Z., Shen, C., Huang, Y., and Yan, Y. (2020). "BlendMask: Top-down meets bottom-up for instance segmentation," in *Proceedings of the IEEE/ CVF Conference on Computer Vision and Pattern Recognition* (IEEE). 8573–8581.
- Chen, W., Wang, W., Wang, K., Li, Z., Li, H., and Liu, S. (2020). Lane departure warning systems and lane line detection methods based on image processing and semantic segmentation: A review. *J. Traffic Transportation Eng.* (English Edition) 7 (6), 748–774. doi: 10.1016/j.jtte.2020.10.002
- Choudhury, S. D. (2020). "Segmentation techniques and challenges in plant phenotyping," in *Intelligent Image Analysis for Plant Phenotyping* (Frontiers Media SA: CRC Press), 69–92.
- Choudhury, S. D., Bashyam, S., Qiu, Y., Samal, A., and Awada, T. (2018). Holistic and component plant phenotyping using temporal image sequence. *Plant Methods* 14 (1), 35. doi: 10.1186/s13007-018-0303-x
- Cruz, A. C., Luvisi, A., De Bellis, L., and Ampatzidis, Y. (2017). X-fido: An effective application for detecting olive quick decline syndrome with deep learning and data fusion. *Front. Plant Sci.* 8, 1741. doi: 10.3389/fpls.2017.01741
- Dalal, N., and Triggs, B. (2005). Histograms of oriented gradients for human detection. *CVPR* 1, 886–893. doi: 10.1109/CVPR.2005.177
- Das, S., and Kundu, M. K. (2013). A neuro-fuzzy approach for medical image fusion. *IEEE Transaction Biomed. Eng.* 60 (12), 3347–3353. doi: 10.1109/TBME.2013.2282461
- Das Choudhury, S., Goswami, S., Bashyam, S., Samal, A., and Awada, T. (2017). "Automated stem angle determination for temporal plant phenotyping analysis," in *Proceedings of the IEEE International Conference on Computer Vision Workshops* (Venice, Italy: IEEE). 2022–2029.
- Das Choudhury, S., Maturu, S., Samal, A., Stoerger, V., and Awada, T. (2020). Leveraging image analysis to compute 3D plant phenotypes based on voxel-grid plant reconstruction. *Front. Plant Sci.* 11, 521431. doi: 10.3389/fpls.2020.521431
- DeChant, C., Wiesner-Hanks, T., Chen, S., Stewart, E. L., Yosinski, J., Gore, M. A., et al. (2017). Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning. *Phytopathology* 107 (11), 1426–1432. doi: 10.1094/PHYTO-11-16-0417-R

- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). "ImageNet: A preview of a large-scale hierarchical database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition* (Miami, Florida: IEEE).
- Dong, X., Shen, J., Shao, L., and Yang, M.-H. (2015). Interactive cosegmentation using global and local energy optimization. *IEEE Trans. Image Process.* 24.11, 3966–3977. doi: 10.1109/TIP.2015.2456636
- Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *Int. J. Comput. Vision* 88 (2), 303–338. doi: 10.1007/s11263-009-0275-4
- Fan, Z., Lu, J., Wei, C., Huang, H., Cai, X., and Chen, X. (2018). A hierarchical image matting model for blood vessel segmentation in fundus images. *IEEE Trans. Image Process.* 28 (5), 2367–2377. doi: 10.1109/TIP.2018.2885495
- Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* 145, 311–318. doi: 10.1016/j.compag.2018.01.009
- Fu, L., Tola, E., Al-Mallahi, A., Li, R., and Cui, Y. (2019). A novel image processing algorithm to separate linearly clustered kiwifruits. *Biosyst. Eng.* 183, 184–195. doi: 10.1016/j.biosystemseng.2019.04.024
- Fuentes, A., Yoon, S., Kim, S. C., and Park, D. S. (2017). A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors* 17 (9), 2022. doi: 10.3390/s17092022
- Gong, L., Du, X., Zhu, K., Lin, C., Lin, K., Wang, T., et al. (2021). Pixel level segmentation of early-stage in bag rice root for its architecture analysis. *Comput. Electron. Agric.* 186, 106197. doi: 10.1016/j.compag.2021.106197
- Grift, T. E., Zhao, W., Momin, M. A., Zhang, Y., and Bohn, M. O. (2017). Semiautomated, machine vision based maize kernel counting on the ear. *Biosyst. Eng.* 164, 171–180. doi: 10.1016/j.biosystemseng.2017.10.010
- Guo, X., Qiu, Y., Nettleton, D., Yeh, C. T., Zheng, Z., Hey, S., et al. (2021). KAT4IA: K-means assisted training for image analysis of field-grown plant phenotypes. *Plant Phenomics* 2021. doi: 10.34133/2021/9805489
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, Nevada: IEEE).
- Hoerser, T., and Kuenzer, C. (2020). Object detection and image segmentation with deep learning on earth observation data: A review part i: Evolution and recent trends. *Remote Sens.* 12 (10), 1667. doi: 10.3390/rs12101667
- Hsu, K.-J., Lin, Y.-Y., and Chuang, Y.-Y. (2018). Co-attention CNNs for unsupervised object co-segmentation. *IJCAI* 1, 2. doi: 10.24963/ijcai.2018/104
- Hsu, K.-J., Lin, Y.-Y., and Chuang, Y.-Y. (2019). "Deepco3: Deep instance cosegmentation by co-peak search and co-saliency detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Long Beach, California: IEEE).
- Hunt, E. R., Cavigelli, M., Daughtry, C. S., McMurtrey, J. E., and Walthall, C. L. (2005). Evaluation of digital photography from model aircraft for remote sensing of crop biomass and nitrogen status. *Precis. Agric.* 6, 359–378. doi: 10.1007/s11119-005-2324-5
- Jerripothula, K. R., Cai, J., Lu, J., and Yuan, J. (2021). Image co-skeletonization via cosegmentation. *IEEE Trans. Image Process.* 30, 2784–2797. doi: 10.1109/TIP.2021.3054464
- Jiang, Y., and Li, C. (2020). Convolutional neural networks for image-based high-throughput plant phenotyping: a review. *Plant Phenomics*. doi: 10.34133/2020/4152816

- Kataoka, T., Kaneko, T., Okamoto, H., and Hata, S. (2003). "Crop growth estimation system using machine vision," in *Proceedings 2003 IEEE/ASME international conference on advanced intelligent mechatronics (AIM 2003)* (Port Island, Japan: IEEE), Vol. 2. b1079–b1083
- Kim, G., and Xing, E. P. (2012). "On multiple foreground cosegmentation," in *2012 IEEE Conference on Computer Vision and Pattern Recognition* (Providence, Rhode Island: IEEE). 837–844.
- Kim, G., Xing, E. P., Li, F.-F., and Kanade, T. (2011). "Distributed cosegmentation via submodular optimization on anisotropic diffusion," in *2011 International Conference on Computer Vision* (Barcelona, Spain: IEEE).
- Kirillov, A., Wu, Y., He, K., and Girshick, R. (2020). "Pointrend: Image segmentation as rendering," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE). 9799–9808.
- Koh, J. C. O., Spangenberg, G., and Kant, S. (2021). Automated machine learning for high-throughput image-based plant phenotyping. *Remote Sens.* 13 (5), 858. doi: 10.3390/rs13050858
- Krähenbühl, P., and Koltun, V. (2011). "Efficient inference in fully connected crfs with gaussian edge potentials," in *Advances in neural information processing systems* (Curran Associates).
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* (Curran Associates), 25.
- Langan, D. A., Modestino, J. W., and Zhang, J. (1998). Cluster validation for unsupervised stochastic model-based image segmentation. *IEEE Trans. Image Process.* 7 (2), 180–195. doi: 10.1109/83.660995
- Lee, U., Chang, S., Putra, G. A., Kim, H., and Kim, D. H. (2018). An automated, high-throughput plant phenotyping system using machine learning-based plant segmentation and image analysis. *PLoS One* 13 (4), e0196615. doi: 10.1371/journal.pone.0196615
- Li, S., Dai, L., Wang, H., Wang, Y., He, Z., and Lin, S. (2017). Estimating leaf area density of individual trees using the point cloud segmentation of terrestrial LiDAR data and a voxel-based model. *Remote Sens.* 9 (11), 1202. doi: 10.3390/rs9111202
- Li, Z., Guo, R., Li, M., Chen, Y., and Li, G. (2020). A review of computer vision technologies for plant phenotyping. *Comput. Electron. Agric.* 176, 105672. doi: 10.1016/j.compag.2020.105672
- Li, W., Jafari, O. H., and Rother, C. (2018). "Deep object co-segmentation," in *Computer Vision—ACCV 2018: 14th Asian Conference on Computer Vision*, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part III 14 (Cham: Springer) 638–653.
- Lian, C., Ruan, S., Denoix, T., Li, H., and Vera, P. (2019). Joint tumor segmentation in pet-ct images using co-clustering and fusion based on belief functions. *IEEE Trans. Image Process.* 28 (2), 755–766. doi: 10.1109/TIP.2018.2872908
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014). Microsoft coco: Common objects in context. *Eur. Conf. Comput. Vision* 13, 740–755. Springer. doi: 10.1007/978-3-319-10602-1_48
- Liu, Z., Su, J., Wang, R., Jiang, R., Song, Y.-Q., Zhang, D., et al. (2022). Pancreas Cosegmentation based on dynamic ROI extraction and VGGU-Net. *Expert Syst. Appl.* 192, 116444. doi: 10.1016/j.eswa.2021.116444
- Liu, J., Liu, Y., and Doonan, J. (2018). "Point cloud based iterative segmentation technique for 3d plant phenotyping," in *Symmetry* (MDPI), 1072–1077.
- Liu, B., Zhang, Y., He, D., and Li, Y. (2018). Identification of apple leaf diseases based on deep convolutional neural networks. *Symmetry* 10 (1), 11. doi: 10.1109/ICInfA.2018.8812589
- Liu, D., Zhang, D., Song, Y., Huang, H., and Cai, W. (2021). Panoptic feature fusion net: a novel instance segmentation paradigm for biomedical and biological images. *IEEE Trans. Image Process.* 30, 2045–2059. doi: 10.1109/TIP.2021.3050668

- Long, E. (2015). "Shelhamer and T. Darrell. Fully convolutional models for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Boston, Massachusetts: IEEE).
- Long, J., Shelhamer, E., and Darrell, T. (2015). "Fully convolutional networks for semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition 2015*.
- Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *IJCV* 60, 91-110. doi: 10.1023/B:VISI.0000029664.99615.94
- Lu, Y., Yi, S., Zeng, N., Liu, Y., and Zhang, Y. (2017). Identification of rice diseases using deep convolutional neural networks. *Neurocomputing* 267, 378–384. doi: 10.1016/j.neucom.2017.06.023
- Mahajan, S., Mittal, N., and Pandit, A. K. (2023). Image segmentation approach based on adaptive flower pollination algorithm and type II fuzzy entropy. *Multimedia Tools Appl.* 82 (6), 8537–8559. doi: 10.1007/s11042-022-13551-2
- Mazis, A., Choudhury, S. D., Morgan, P. B., Stoerger, V., Hiller, J., Ge, Y., et al. (2020). Application of high-throughput plant phenotyping for assessing biophysical traits and drought response in two oak species under controlled environment. *For. Ecol. Manage.* 465, 118101. doi: 10.1016/j.foreco.2020.118101
- Mehta, S., Paunwala, C., and Vaidya, B. (2019). "CNN based traffic sign classification using adam optimizer," in *2019 International Conference on Intelligent Computing and Control Systems (ICCS)* (Madurai, India: IEEE).
- Melinte, D. O., and Vladareanu, L. (2020). Facial expressions recognition for human-robot interaction using deep convolutional neural networks with rectified adam optimizer. *Sensors* 20.8, 2393. doi: 10.3390/s20082393
- Meng, F., Cai, J., and Li, H. (2016). Cosegmentation of multiple image groups. *Comput. Vision Image Understanding* 146, 67–76. doi: 10.1016/j.cviu.2016.02.004
- Meng, F., Luo, K., Li, H., Wu, Q., and Xu, X. (2019). "Weakly supervised semantic segmentation by a class-level multiple group cosegmentation and foreground fusion strategy," in *IEEE Transactions on Circuits and Systems for Video Technology* (IEEE), Vol. 30. 4823–4836.
- Meng, H., and Zhao, Q. (2022). "A lightweight model based on co-segmentation attention for occluded person re-identification," in *Proceedings of 2021 Chinese Intelligent Automation Conference* (Singapore: Springer).
- Merdassi, H., Barhoumi, W., and Zagrouba, E. (2020). A comprehensive overview of relevant methods of image cosegmentation. *Expert Syst. Appl.* 140, 112901. doi: 10.1016/j.eswa.2019.112901
- Meyer, G. E., and Neto, J. C. (2008). Verification of color vegetation indices for automated crop imaging applications. *Comput. Electron. Agric.* 63 (2), 282–293. doi: 10.1016/j.compag.2008.03.009
- Mohanty, S. P., Hughes, D. P., and Salathe, M. (2016). Using deep learning for image-based plant disease detection. *Front. Plant Sci.* 7, 1419. doi: 10.3389/fpls.2016.01419
- Nazki, H., Yoon, S., Fuentes, A., and Park, D. S. (2020). Unsupervised image translation using adversarial networks for improved plant disease recognition. *Comput. Electron. Agric.* 168, 105117. doi: 10.1016/j.compag.2019.105117
- Ojeda-Magaña, B., Ruelas, R., Quintanilla-Dominguez, J., and Andina, D. (2010). "Color image segmentation by partitional clustering algorithms," in *IECON 2010-36th Annual Conference on IEEE Industrial Electronics Society* (Glendale, Arizona: IEEE), 2828–2833.
- Otsu, N. (1979). "A threshold selection method from gray-level histograms," in *IEEE transactions on systems, man, and cybernetics* (IEEE), Vol. 9. 62–66.

- Pape, J.-M., and Klukas, C. (2015). Utilizing machine learning approaches to improve the prediction of leaf counts and individual leaf segmentation of rosette plant images. *Proc. Comput. Vis. Probl. Plant Phenotyping* 3, 1–12. doi: 10.5244/C.29.CVPPP.3
- Patz, T., and Preusser, T. (2012). Segmentation of stochastic images with a stochastic random walker method. *IEEE Trans. Image Process.* 21 (5), 2424–2433. doi: 10.1109/TIP.2012.2187531
- Pound, M. P., Atkinson, J. A., Townsend, A. J., Wilson, M. H., Griffiths, M., Jackson, A. S., et al. (2017). Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. *Gigascience* 6 (10), gix083. doi: 10.1093/gigascience/gix083
- Quiñones, R., Muñoz-Arriola, F., Choudhury, S. D., and Samal, A. (2021). Multifeature data repository development and analytics for image cosegmentation in high-throughput plant phenotyping. *PLoS One* 16 (9), 1–21. doi: 10.1371/journal.pone.0257001
- Ren, Y., Jiao, L., Yang, S., and Wang, S. (2018). Mutual learning between saliency and similarity: Image cosegmentation via tree structured sparsity and tree graph matching. *IEEE Trans. Image Process.* 27 (9), 4690–4704. doi: 10.1109/TIP.2018.2842207
- Rezaee, M. R., van der Zwet, P. M., Lelieveldt, B. P. E., van der Geest, R. J., and Reiber, J. H. (2000). A multiresolution image segmentation technique based on pyramidal segmentation and fuzzy clustering. *IEEE Trans. Image Process.* 9 (7), 1238–1248. doi: 10.1109/83.847836
- Rico, D. A., Detweiler, C., and Muñoz-Arriola, F. (2020). “Power-over-tether UAS leveraged for nearly-indefinite meteorological data acquisition,” in *2020 ASABE Annual International Meeting, Paper No. 1345* (American Society of Agricultural and Biological Engineers). doi: 10.13031/aim.202001345
- Rico, D. A., Muñoz-Arriola, F., and Detweiler, C. (2021). Trajectory selection for power-over-tether atmospheric sensing UAS. *2021 IEEE/RSJ Int. Conf. Intelligent Robots Syst. (IROS), 2021*. 2321–2328. doi: 10.1109/IROS51168.2021.9636364
- Roggiolani, G., Sodano, M., Guadagnino, T., Magistri, F., Behley, J., and Stachniss, C. (2023). “Hierarchical approach for joint semantic, plant instance, and leaf instance segmentation in the agricultural domain,” in *2023 IEEE International Conference on Robotics and Automation (ICRA)* (London, United Kingdom: IEEE), 9601–9607.
- Rother, C., Kolmogorov, V., Minka, T., and Blake, A. (2006). “Cosegmentation of image pairs by histogram matching-incorporating a global constraint into mrfs,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (New York, New York: IEEE), 993–1000.
- Rubinstein, M., Joulin, A., Kopf, J., and Liu, C. (2013). “Unsupervised joint object discovery and segmentation in internet images,” in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Portland, Oregon: IEEE), 1939–1946.
- Sarzaeim, P., Muñoz-Arriola, F., and Jarquin, D. (2022). Climate and genetic data enhancement using deep learning analytics to improve maize yield predictability. *J. Exp. Bot.* vol. 73 (15), 5336–5354. doi: 10.1093/jxb/erac146
- Sarzaeim, P., Muñoz-Arriola, F., Jarquin, D., Aslam, H., and De Leon Gatti, N. (2023). CLIM4OMICS: a geospatially comprehensive climate and multi-OMICS database for Maize phenotype predictability in the U.S. and Canada. *Earth Syst. Sci. Data* 15 (9), 3963–3990. doi: 10.5194/essd-2023-11
- Scharr, H., Minervini, M., French, A. P., Klukas, C., Kramer, D. M., Liu, X., et al. (2016). Leaf segmentation in plant phenotyping: a collation study. *Mach. Vis. Appl.* 27, 585–606. doi: 10.1007/s00138-015-0737-3
- Sezgin, M., and Sankur, B. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *J. Electronic Imaging* 13 (1), 146–166. doi: 10.1117/1.1631315

- Simonyan, K., and Zisserman, A. (2015). "Very deep convolutional networks for large-scale image recognition," in arXiv preprint arXiv:1409.1556 (Cornell University).
- Subramaniam, A., Nambiar, A., and Mittal, A. (2019). "Co-segmentation inspired attention networks for video-based person re-identification," in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (Seoul, South Korea: IEEE), 562–572.
- Suh, H. K., Ijsselmuiden, J., Hofstee, J. W., and van Henten, E. J. (2018). Transfer learning for the classification of sugar beet and volunteer potato under field conditions. *Biosyst. Eng.* 174, 50–65. doi: 10.1016/j.biosystemseng.2018.06.017
- Tao, W., Li, K., and Sun, K. (2015). SaCoseg: Object cosegmentation by shape conformability. *IEEE Trans. Image Process.* 24 (3), 943–955. doi: 10.1109/TIP.2014.2387384
- Tao, Z., Liu, H., Fu, H., and Fu, Y. (2019). Multi-view saliency-guided clustering for image cosegmentation. *IEEE Trans. Image Process.* 28 (9), 4634–4645. doi: 10.1109/TIP.2019.2913555
- Thorp, K. R., Wang, G., Badaruddin, M., and Bronson, K. F. (2016). Lesquerella seed yield estimation using color image segmentation to track flowering dynamics in response to variable water and nitrogen management. *Ind. Crops Products* 86, 186–195. doi: 10.1016/j.indcrop.2016.03.035
- Valliammal, N., and Geethalakshmi, S. N. (2012). A novel approach for plant leaf image segmentation using fuzzy clustering. *Int. J. Comput. Appl.* 44 (3), 10–20. doi: 10.5120/6322-8669
- Wang, Z., and Liu, R. (2013). "Semi-supervised learning for large scale image cosegmentation," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (Sydney, Australia: IEEE).
- Wang, G., Sun, Y., and Wang, J. (2017). Automatic image-based plant disease severity estimation using deep learning. *Comput. Intell. Neurosci.* doi: 10.1155/2017/2917536
- Wang, Z., Wang, K., Yang, F., Pan, S., and Han, Y. (2018). Image segmentation of overlapping leaves based on Chan–Vese model and Sobel operator. *Inf. Process. Agric.* 5 (1), 1–10. doi: 10.1016/j.inpa.2017.09.005
- Wang, W., and Shen, J. (2016). Higher-order image co-segmentation. *IEEE Trans. Multimedia* 18.6, 1011–1021. doi: 10.1109/TMM.2016.2545409
- Williams, D., MacFarlane, F., and Britten, A. (2023). Leaf only SAM: A segment anything pipeline for zero-shot automated leaf segmentation. arXiv preprint arXiv 2305.09418. doi: 10.48550/arXiv.2305.09418
- Winn, J., Criminisi, A., and Minka, T. (2005). "Object categorization by learned universal visual dictionary," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1* (Beijing, China: IEEE), Vol. 2. 1800–1807.
- Woebbecke, D. M., Meyer, G. E., Von Bargen, K., and Mortensen, D. A. (1992). "Plant species identification, size, and enumeration using machine vision techniques on near-binary images [1836–20]," in *Optics in Agriculture and Forestry* (SPIE). 208–208.
- Yin, X., Liu, X., Chen, J., and Kramer, D. M. (2014). "Multi-leaf alignment from fluorescence plant images," in *IEEE Winter Conference on Applications of Computer Vision* (Steamboat Springs, Colorado: IEEE), Vol. 2014. 437–444. doi: 10.1109/WACV.2014.6836067
- Yuan, Z., Lu, T., and Wu, Y. (2017). "Deep-Dense Conditional Random Fields for Object Co-segmentation," in *IJCAI*.
- Zhang, X., Zhao, R., Qiao, Y., Wang, X., and Li, H. (2019). "Adacos: Adaptively scaling cosine logits for effectively learning deep face representations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Long Beach, California: IEEE).
- Zhang, C., Mouton, C., Valente, J., Kooistra, L., van Ooteghem, R., de Hoog, D., et al. (2022). Automatic flower cluster estimation in apple orchards using aerial and ground based point clouds. *Biosyst. Eng.* 221, 164–180. doi: 10.1016/j.biosystemseng.2022.05.004

- Zheng, L., Zhang, J., and Wang, Q. (2009). Mean-shift-based color segmentation of images containing green vegetation. *Comput. Electron. Agric.* 65 (1), 93–98. doi: 10.1016/j.compag.2008.08.002
- Zhong, Z., Kim, Y., Plichta, K., Allen, B. G., Zhou, L., Buatti, J., et al. (2019). Simultaneous cosegmentation of tumors in pet-ct images using deep fully convolutional networks. *Med. Phys.* 46 (2), 619–633. doi: 10.1002/mp.13331
- Zhou, S., Chai, X., Yang, Z., Wang, H., Yang, C., and Sun, T. (2021). Maize-ias: a maize image analysis software using deep learning for high-throughput plant phenotyping. *Plant Methods* 17 (1), 1–17. doi: 10.1186/s13007-021-00747-0
- Zhou, T., Ruan, S., and Canu, S. (2019). A review: Deep learning for medical image segmentation using multimodality fusion. *Array* 3, 100004. doi: 10.1016/j.array.2019.100004