

Revealing Hierarchical Structure of Leaf Venations in Plant Science via Label-Efficient Segmentation: Dataset and Method

Weizhen Liu¹, Ao Li¹, Ze Wu¹, Yue Li¹, Baobin Ge¹, Guangyu Lan¹, Shilin Chen¹, Minghe Li², Yunfei Liu¹, Xiaohui Yuan^{1,3*} and Nanqing Dong^{2*}

¹School of Computer Science and Artificial Intelligence, Wuhan University of Technology

²Shanghai Artificial Intelligence Laboratory

³Yazhouwan National Laboratory

{liuweizhen, liao2022, yuanxiaohui}@whut.edu.cn, dongnanqing@pjlab.org.cn

Abstract

Hierarchical leaf vein segmentation is a crucial but under-explored task in agricultural sciences, where analysis of the hierarchical structure of plant leaf venation can contribute to plant breeding. While current segmentation techniques rely on data-driven models, there is no publicly available dataset specifically designed for hierarchical leaf vein segmentation. To address this gap, we introduce the HierArchical Leaf Vein Segmentation (HALVS) dataset, the first public hierarchical leaf vein segmentation dataset. HALVS comprises 5,057 real-scanned high-resolution leaf images collected from three plant species: soybean, sweet cherry, and London planetree. It also includes human-annotated ground truth for three orders of leaf veins, with a total labeling effort of 83.8 person-days. Based on HALVS, we further develop a label-efficient learning paradigm that leverages partial label information, *i.e.* missing annotations for tertiary veins. Empirical studies are performed on HALVS, revealing new observations, challenges, and research directions on leaf vein segmentation. Our dataset and code are available at <https://github.com/WeizhenLiuBioinform/HALVS-Hierarchical-Vein-Segment>.

1 Introduction

Analyzing the detailed structure at different hierarchical levels of leaf venation is a fundamental step for botanists, crop breeders, and ecologists to understand the impact of these intricate structures on important physiological functions of leaves, such as photosynthesis, transpiration, respiration, and transportation. This understanding can facilitate the breeding of plant species with high yield, quality, or economic value [Sack *et al.*, 2013], thus not only targeting the United Nations' Sustainable Development Goals of No Poverty and Zero Hunger (SDG1 & SDG2) [United Nations, 2023a] but also fulfilling the Leave No One Behind Principle (LNOB) [United Nations, 2023b]. To achieve this task, a

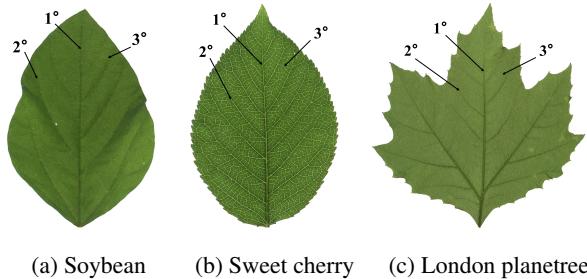


Figure 1: Illustration of three orders of veins (1° , 2° , and 3°) for three species of interest in HALVS.

non-negligible process is to segment the leaf venation. With the development of imaging technology and computer vision techniques, researchers have started to utilize image analysis methods to analyze leaf venation structures. But most studies focus on segmenting the entire leaf venation network without differentiating the order of veins [Price *et al.*, 2011; Dhondt *et al.*, 2012; Bühler *et al.*, 2015; Salima *et al.*, 2015; Grinblat *et al.*, 2016; Lasser and Katifori, 2017; Xu *et al.*, 2021; Li *et al.*, 2022; Liu *et al.*, 2022; Iwamasa and Noshita, 2023]. The task of interest, hierarchical vein segmentation from leaf images, involves not only isolating and identifying the venation network from other leaf components but also classifying each order of veins into distinct semantic categories. It is a challenging computer vision task due to the complexity of vein structures and the high similarity in color and texture among veins of different orders [Liu *et al.*, 2022]. The first three orders are the primary (1°) vein or veins emerging from the leaf base to the apex, the smaller secondary (2°) veins branching at intervals from the 1° veins towards the leaf margin, and even finer tertiary (3°) veins [Ellis *et al.*, 2009], as shown in Fig. 1. Until now, only two studies have attempted hierarchical vein segmentation [Gan *et al.*, 2019; Jin and Han, 2020], which used directional morphological filtering and region-growing-based methods, respectively. But these traditional digital image processing techniques cannot extract rich semantic information from the highly dense leaf venation networks and have not yet addressed the semantic segmentation of 3° veins.

Deep learning seems to be a promising technique for hierarchical leaf vein segmentation. However, as a data-driven

*Corresponding authors.

Species	1° veins	2° veins	3° veins	Total Time
Soybean	14.5 ± 5.2	29.9 ± 6.4	80.8 ± 15.2	125.5 ± 14.2
Sweet cherry	5.5 ± 1.1	39.1 ± 3.7	164.3 ± 10.0	208.9 ± 10.6
London planetree	21.6 ± 2.6	132.6 ± 8.4	315.7 ± 14.5	469.8 ± 19.9

Table 1: Statistics of annotation time for each order of veins in each leaf image across three species. Times are measured in minutes and are presented in the format of mean \pm standard deviation.

method, the biggest obstacle is the lack of labeled training data. The existing public datasets have limitations in leaf image acquisition methods and vein labeling. For data collection, chemical cleaning and vein staining are commonly used practices for treating leaves before imaging to increase the color contrast between leaf veins and lamina [Perez-Harguindeguy *et al.*, 2016; Xu *et al.*, 2021; Iwamasa and Noshita, 2023]. But it is quite time-consuming and can easily damage the leaf blade. X-ray imaging makes it possible without chemical treatment but requires expensive specialized facilities [Schneider *et al.*, 2018; Zhu *et al.*, 2020]. More importantly, the existing public datasets annotate all the veins in a leaf as one semantic class **without** recognizing vein orders, which cannot serve as the ground truth for deep learning-based hierarchical leaf vein segmentation.

To address the aforementioned limitations, in this work, we construct and release a new dataset for **HierArchical Leaf Vein Segmentation** (HALVS). To acquire **high-resolution** and **high-contrast** leaf venation images from the raw leaf samples, we adopt a flat-bed scanner in transmission scanning mode, which can provide backlight illumination throughout the scanning process. To the best of our knowledge, HALVS is the first public dataset that leverages this efficient and cost-effective imaging system. The HALVS comprises 5,057 leaf images from three representative species: soybean, sweet cherry, and London planetree (see Fig. 1). Each species was selected for its unique significance in diverse contexts. Soybean is widely acknowledged as a fundamental global provider of protein and oil [Vollmann, 2016] with both ecological and economic values worldwide. Sweet cherry represents a high-class fruit with significant economic value, and London planetree was selected for its relevance in urban landscaping, complementing the dataset by providing diversity in leaf venation patterns. Furthermore, we have also provided detailed pixel-level annotations for the first three orders of veins of these plants, considering that their vein traits are of significance for biological and ecological research. It is worth noting that hierarchical annotated leaf veins are extremely time-consuming, as shown in Tab. 1. The dataset contribution will be described in detail in Sec. HALVS Dataset.

This paper also makes a methodological advancement in label-efficient learning. Given the label scarcity challenge, fully supervised deep learning methods are not optimal for hierarchical leaf vein segmentation tasks. Instead, semi-supervised learning (SSL) [Chapelle *et al.*, 2009; Papandreou *et al.*, 2015], which learns a model with only a few labeled data and numerous unlabeled ones, looks like a feasible solution. We notice that annotations of 3° veins take much longer than those of the 1° and 2° veins (see Tab. 1). It is more prac-

tical to prepare partially labeled data with only annotations of 1° and 2° veins when the fully labeled data with annotations of veins of all three orders are unavailable. This has led to an increasing interest in *learning from partially labeled data*. Several attempts have been made towards *partially supervised learning* (PSL) in medical image analysis [Dong *et al.*, 2022a], where human annotation costs are also high because annotators are required to have clinical expertise. For the first time, we study PSL in the domain of agricultural science. In this paper, we propose a label-efficient learning paradigm with partial supervision to adequately benefit from learning unlabeled, partially labeled, and fully labeled data, simultaneously. This method guides the segmentation of 3° veins based on the existing annotations of 1° and 2° veins. We conduct extensive ablation studies to understand the proposed method and HALVS. The empirical results also provide insights for future research on the task of interest under label scarcity. For example, we notice that the annotation efficiency varies significantly among different species. We are the first to investigate cross-species learning in the leaf vein segmentation task, where the segmentation model is trained by leaf venation data of species that are relatively more amenable to be annotated and makes predictions on the other species whose venation data are challenging to be annotated. The experiments show that cross-species learning is a challenging task under label scarcity and the annotations in HALVS can benefit future studies.

We summarize our contributions as follows:

- To the best of our knowledge, we release the **first** hierarchical leaf venation segmentation dataset, namely, HALVS, thereby establishing a benchmark for subsequent work and benefiting future studies.
- This is the first study that leverages partially labeled data in the plant phenomics domain and presents a new practical task domain for partially supervised learning.
- We evaluate the proposed method extensively on HALVS and achieve superior performance against state-of-the-art baselines. The experimental results also present initial empirical findings on hierarchical vein segmentation, paving the way for future investigations on the task of interest in the context of data scarcity.

2 Related Work

2.1 Leaf Vein Datasets

Due to the high human cost of pixel-level annotation, there are only a few public leaf vein datasets. To our best knowledge, none of these datasets offer annotations for leaf veins with hierarchical orders. To differentiate HALVS from the existing datasets, the limitations of these datasets are summarized in Tab. 2 in the aspects of image clarity, completeness of leaf blade, whether the raw leaf is chemically treated, and hierarchy of vein annotations. For example, LVD2021 dataset [Li *et al.*, 2022] contains 4,977 low-contrast and low-resolution images captured by smartphone cameras and pixel-wise annotations do not provide detailed structural information. Additionally, only 20 high-quality images have been

Dataset	Hier. Annot.	Untreated	Complete	Clear	# Images
Leaf Vein Dataset [Blonder <i>et al.</i> , 2019]				✓	726
LVD2021 [Li <i>et al.</i> , 2022]		✓	✓		4,977
Cleared Leaf Database [Iwamasa and Noshita, 2023]			✓	✓	328
Untreated Leaf Dataset [Iwamasa and Noshita, 2023]		✓	✓	✓	479
HALVS (Ours)	✓	✓	✓	✓	5,057

Table 2: Comparison of HALVS with existing leaf vein segmentation datasets. “Hier. Annot.” means providing annotations of 1° , 2° , and 3° veins. “Untreated” indicates that the leaf is not chemically treated before imaging. “Complete” indicates providing the complete leaf blade. “Clear” refers to an image clearly displaying the detailed structural information of the 3° veins.

annotated with non-human techniques using histogram equalization and binarization algorithms in [Iwamasa and Noshita, 2023]. HALVS is the first large-scale annotated dataset for high-quality images.

2.2 Partially Supervised Learning

An important research field of PSL is medical image analysis, where partially labeled datasets of different classes of interest are collected from different data sources due to high annotation cost [Dong *et al.*, 2022b]. Many efforts have been paid into partially supervised medical image segmentation [Zhou *et al.*, 2019; Fang and Yan, 2020; Shi *et al.*, 2021; Dong *et al.*, 2022a], a similar yet different task to leaf vein segmentation. First, in contrast to medical image segmentation, leaf vein segmentation is still an under-explored task. For medical images, transfer learning has been proven as an efficient solution for cross-sites and cross-modality setups, because of human structural similarity [Dong *et al.*, 2022a]. However, this remains unclear in the leaf vein segmentation. Second, a direct application of PSL methods designed for medical images is infeasible for leaf images because of different image characteristics and task setups. For example, compared with leaf images in HALVS, medical images in previous PSL studies are “low-quality” in terms of resolution. Besides, the pixels of leaf veins are scarce, in contrast to most organs or human structures. The desired algorithms should be sensitive to pixel-level information. In this work, we propose the first PSL solution to leaf vein segmentation.

3 The HALVS Dataset

3.1 Data Collection

We collect 5,057 high-quality leaf images with precise venation details from three characteristic species to represent distinct venation characters, including 2,610 images of soybean, 1,947 images of sweet cherry, and 500 images of London planetree (see Fig. 1). The leaf images are acquired using a simple but efficient transmission scanning method, proposed by [Gan *et al.*, 2019]. After being freshly picked from

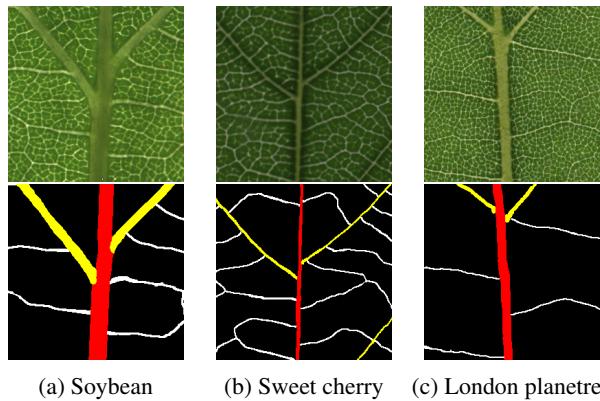


Figure 2: Visualizations of leaves and corresponding vein annotations in HALVS (red: 1° , yellow: 2° , white 3°). Leaf patches are cropped in size of 256×256 for illustration purposes.

the tree and simply wiped clean, leaves are placed in the flatbed scanner (Epson Perfection V850 pro scanner) with the transmission scan mode. Images with 4800×6000 pixels at a resolution of 600 DPI are captured and saved in the 48-bit RGB PNG format. The leaf petiole can interfere with the segmentation of the primary vein, so in each leaf image, we crop it out and only retain the leaf blade. Then, we apply the minimum bounding rectangle algorithm to eliminate redundant background pixels and generate the final leaf image data.

3.2 Classes and Annotations

Determining the categories, or *orders*, of veins is the first and the most important step during the process of vein annotations. We recognize 1° , 2° , and 3° veins starting at the widest 1° vein and progressing to the finest 3° veins, following the roles and vein characteristics described in [Hickey, 1973] and [Ellis *et al.*, 2009]. Generally, it is relatively easier to recognize the 1° and 3° veins, but sometimes the 2° veins are more complex as they may comprise several subsets with varying widths and courses. Nevertheless, all the vein subsets located between the 1° and 3° veins are considered to be 2° veins. Fig. 2 displays examples of our annotations. For each image, all three orders of veins are traced manually at full width. Despite the efforts, the annotation process for the HALVS dataset remains notably time-consuming. As detailed in Tab. 1, the average time for fully labeling the three orders of veins in a leaf blade image is about 2.1, 3.5, and 7.8 hours for soybean, sweet cherry, and London planetree, respectively. The significantly longer annotation time for the London planetree compared to the other two species is due to the larger leaf size. Labeling the 3° veins consumes considerably more time than labeling the 1° and 2° veins. As a result, 150 images (50 images per species) are densely annotated with fine-grained semantic segmentation labels, *i.e.* 7,831,792 pixels for 1° veins, 10,459,325 pixels for 2° veins, and 16,075,087 pixels for 3° veins. This is equivalent to around 13,000 non-overlapping patches in a size of 256×256 . The whole annotation process takes up to 83.8 person-days by four experienced annotators following biological instructions.

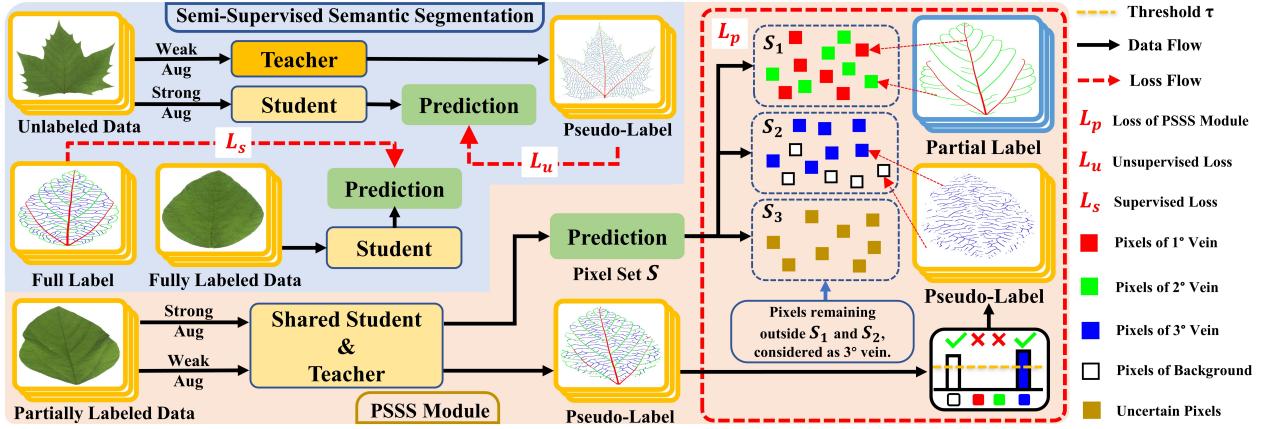


Figure 3: Illustration of the proposed label-efficient learning framework for hierarchical leaf vein segmentation. The input of this framework includes unlabeled, partially labeled (*i.e.* 1° and 2° vein), and fully labeled (*i.e.* background, 1°, 2°, and 3° veins) data which are partitioned into leaf patches with a size of 256 × 256 pixels. Intuitively, complete leaf images instead of patches are used for illustration purposes. The labeled and unlabeled data are handled by a semi-supervised semantic segmentation method. The proposed partially supervised semantic segmentation (PSSS) module can be integrated with any semi-supervised learning framework to handle the partially labeled data. Here, a teacher-student model [Wang *et al.*, 2022] is depicted as an example (blue region). In PSSS (orange region), the teacher model generates pseudo-labels from the weakly augmented data, while the student model generates set of predictions \mathcal{S} from the strongly augmented data. \mathcal{S}_1 contains the pixels that can be directly supervised by the ground truth of 1° and 2° veins from the partial labels. \mathcal{S}_2 contains the pixels that are predicted as pseudo-labels of background and 3° vein with high confidence. All remaining pixels are considered as 3° vein. Three sets of pixels are trained with three different losses.

4 Method

4.1 Problem Definition

In the task setup of interest, the dataset $\mathcal{D} = \mathcal{D}_l \cup \mathcal{D}_u \cup \mathcal{D}_p$ consists of fully labeled, unlabeled, and partially labeled sets. $\mathcal{D}_l = \{(x_i, y_i)\}_{i=0}^{N_l}$ is the fully labeled set, where x_i represents an image and each pixel belongs to one of four classes (*i.e.* the background, 1°, 2°, and 3° veins). The corresponding y_i provides pixel-wise labels for x_i . $\mathcal{D}_u = \{x_j\}_{j=0}^{N_u}$ is the unlabeled set. $\mathcal{D}_p = \{(x_k, y_k)\}_{k=0}^{N_p}$ is the partially labeled set where y_k only contains label information regarding 1° and 2° veins, missing the background and 3° veins, due to extremely high annotation cost of 3° vein. That is to say, we are only certain whether the pixel belongs to either 1° or 2° veins.

As shown in Tab. 1, annotating veins are time-consuming and expertise-demanding, and particularly, obtaining 3° vein labels is more challenging. We simulate a practical situation: \mathcal{D}_l and \mathcal{D}_p are **small-scale** datasets and \mathcal{D}_u is a **large-scale** dataset, *i.e.* $N_l \ll N_u$ and $N_p \ll N_u$. The learning objective is to train a leaf vein segmentation model with \mathcal{D} . In contrast to standard SSL, we conjecture that the existence of \mathcal{D}_p plays an important role in enhancing the model performance across all four classes.

4.2 Partially Supervised Semantic Segmentation

We propose a pluggable module for partially supervised semantic segmentation, which we denote as PSSS in the following context. The PSSS module can be easily integrated with state-of-the-art SSL frameworks [Wang *et al.*, 2022; Yang *et al.*, 2023]. The overall framework is illustrated in Fig. 3. The training on \mathcal{D}_l and \mathcal{D}_u follows standard SSL

methods, where we use L_S and L_U to denote the supervised loss and unsupervised loss, respectively.

For simplicity, we assume that f is a neural network and $(x, y) \in \mathcal{D}_p$ is a sample-label pair where y is a *partial* label with respect to 1° and 2° veins. Following FixMatch [Sohn *et al.*, 2020], there are a strong augmentation A_s and a weak augmentation A_w . We use $p = f(A_s(x))$ as the prediction output and $\hat{y} = \arg \max(f(A_w(x)))$ as the pseudo-label. Let \mathcal{S} denote the set of pixels. \mathcal{S} can be split into three sets, *i.e.* $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$. \mathcal{S}_1 consists of the pixels corresponding to the 1° and 2° veins in the partial label y , *i.e.*

$$\mathcal{S}_1 = \{m | y^m = 1\}_{m \in \mathcal{S}} \cup \{m | y^m = 2\}_{m \in \mathcal{S}}, \quad (1)$$

where m is the pixel of interest. \mathcal{S}_2 contains the pixels corresponding to the background and 3° veins with confidence higher than a threshold τ in the pseudo-label \hat{y} , *i.e.*

$$\mathcal{S}_2 = \{m | \hat{y}^m > \tau\}_{m \in \mathcal{S}}, \quad (2)$$

where $\tau \in (0, 1)$ is the hyperparameter. \mathcal{S}_3 is the complement of $\mathcal{S}_1 \cup \mathcal{S}_2$ under \mathcal{S} .

For \mathcal{S}_1 , the ground truth labels are available. The standard supervised learning is applied.

$$L_P^s = \frac{1}{|\mathcal{S}_1|} \sum_{m \in \mathcal{S}_1} H(y^m, f(A_s(x))^m) \quad (3)$$

In Eq. (3), $|\cdot|$ denotes the cardinality of the set (*i.e.* the number of pixels) and $H(\cdot, \cdot)$ denotes the cross-entropy.

For \mathcal{S}_2 , an unsupervised loss is computed using the pseudo-label \hat{y} .

$$L_P^u = \frac{1}{|\mathcal{S}_2|} \sum_{m \in \mathcal{S}_2} H(\hat{y}^m, f(A_s(x))^m) \quad (4)$$

L_P^u utilizes consistency regularization to improve the model’s generalization on unlabeled pixels and mitigate the overfitting. Due to class imbalance, the learned pseudo-labels might be dominated by the majority classes, thus exacerbating the overfitting. Note, in Eq. (2), only those pseudo-labels with high confidence participate in L_P^u . This excludes unreliable pseudo-labels that might sabotage the training. We will further discuss the impact of τ in the experiments.

The pixels in S_3 are uncertain pixels, *i.e.* pixels with missing ground truth labels and with low confidence of the pseudo-labels. As the classes of interest are mutually exclusive, a pixel can only belong to one class. Though the remaining pixels in S_3 can be either the background or 3° vein, they are highly possible to be 3° vein. The motivation here is that if the model can reliably predict the pixels belonging to the background, 1° , and 2° veins, the remaining pixels shall be the 3° vein. We hypothesize that the background class is much easier to predict in contrast to the vein classes. With L_S and L_U , the model can efficiently learn the patterns of the background as there are more non-vein pixels than vein pixels. Utilizing this prior knowledge allows the model to focus more on predicting the most challenging 3° vein class, making the segmentation task straightforward. We design a class-specific exclusion loss [Shi *et al.*, 2021] to leverage these uncertain pixels:

$$L_P^c = \frac{1}{|S_3|} \sum_{m \in S_3} e \cdot \log(1 + f(A_s(x))^m), \quad (5)$$

where $e = [1, 1, 1, 0]$. As the labels for the 3° vein is scarce, L_P^c also regularizes the class imbalance.

To summarize, the loss for D_p is the sum of the three component losses above, *i.e.* $L_P = L_P^s + L_P^u + L_P^c$. And the total loss for the optimization process is $L = L_S + L_U + \lambda L_P$, where λ is a hyperparameter controlling the weight of L_P .

5 Experiments

The purpose of experimental design is twofold. First, we aim to evaluate the efficiency of PSSS on HALVS under label scarcity. Second, we aim to leverage HALVS to understand the challenges in cross-species transfer learning.

5.1 Experimental Setup

Baselines. The PSSS module can easily integrate into semi-supervised semantic segmentation methods without changing the training pipeline, making it convenient for a comprehensive and practical evaluation. We choose three representative approaches as our baselines, namely: FixMatch [Sohn *et al.*, 2020], U²PL [Wang *et al.*, 2022], and UniMatch [Yang *et al.*, 2023]. FixMatch is a seminal semi-supervised classification method that combines consistency regularization and pseudo-labeling. We adapt FixMatch to semantic segmentation as a strong baseline. U²PL and UniMatch are two recent state-of-the-art approaches in semi-supervised semantic segmentation. Additionally, we provide an “Oracle”, where the training data \mathcal{D} is completely labeled. “Oracle” is considered as the performance upper bound that a semi-supervised method can achieve. Similarly, we provide a performance

lower bound, only using the labeled information to train a supervised model. We denote this baseline as “Supervised”.

Implementation. All experiments are implemented by PyTorch [Paszke *et al.*, 2019] on two NVIDIA 4090 GPUs with 24G memory. The main purpose of the experiments is to evaluate the efficiency of PSSS in the default segmentation network DeepLabv3+ [Chen *et al.*, 2018] with ResNet-101 [He *et al.*, 2016] backbone, a strong semantic segmentation architecture. Following U²PL and UniMatch, we use a standard stochastic gradient descent optimizer with a batch size of 4. For a fair comparison, we adopt the poly learning strategy [Chen *et al.*, 2017] for all methods, where the initial learning rate is 10^{-3} and the weight decay is 10^{-4} . We train all methods for 80 epochs and report the best performance measured in mean Intersection-over-Union (mIoU). By default, we set $\lambda = 1$ and $\tau = 0.95$. To retain image details in our high-resolution data, we partition the original images into patches of size 256×256 . The default data ratio of fully labeled, partially labeled, and unlabeled data is $1 : 1 : 10$, where one basic unit denotes 6 images with full leaf blades (around 600 patches). For the class-balance purpose, the default class ratio among three species is $1 : 1 : 1$.

5.2 Efficiency of PSSS

The evaluation results of PSSS with three seminal SSL methods are shown in Tab. 3. While the numerical results between different SSL baselines vary, the pattern is clear: adding PSSS to the SSL training can significantly improve the 3° vein IoU by 14.37%, 3.89%, and 11% on FixMatch, U²PL, and UniMatch, respectively. It also improves the mIoU of FixMatch, U²PL, and UniMatch by 8.55%, 2.57%, and 6.82%, respectively. It is worth mentioning that with PSSS, SSL baselines can achieve competitive performance with “Oracle”, especially on the most challenging 3° vein.

In addition to quantitative evaluation, we perform qualitative comparison in Fig. 4. We use UniMatch as the baseline SSL method and visualize the segmentation results of “Supervised”, “UniMatch” (Semi-Supervised), and “UniMatch + PSSS” (Ours) from the Tab. 3. It can be clearly seen that with PSSS, the segmentation performance is significantly improved across all three categories compared to the supervised and semi-supervised baselines, especially on the 3° vein. Meanwhile, we should also realize that hierarchical leaf vein segmentation is a challenging task and there are still lots of space for improvement.

5.3 Ablation Studies

In this section, we use UniMatch as the baseline SSL method to study the robustness of PSSS under the following setups.

Effect of Component Losses in PSSS. As shown in Tab. 4, all three loss components of PSSS can improve the overall performance. It is interesting that adding our designed L_P^c not only significantly improves the performance of 3° vein, but also benefits the other two classes (2^{nd} row vs 4^{th} row, 3^{rd} row vs 5^{th} row). We conclude that L_P^c plays an important role in PSSS.

Sensitivity to τ and λ . As shown in Fig. 5, the model performance is unsatisfactory when the weight λ is too high or too low. This might be because such values disrupt the balance of

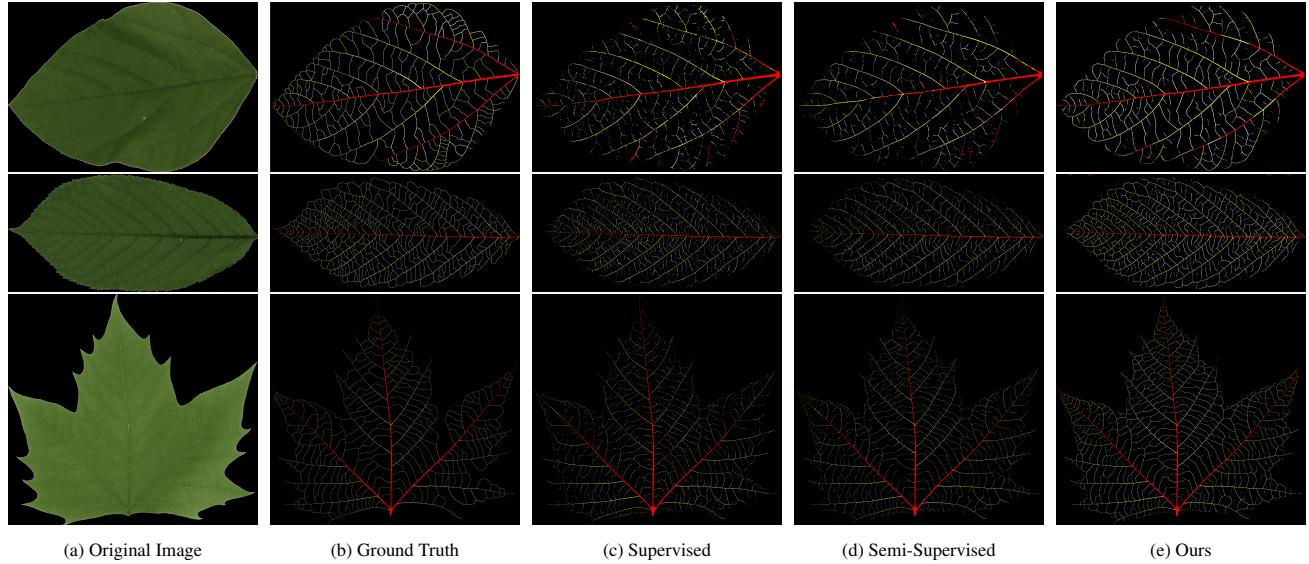


Figure 4: Visualizations of the segmentation results on the HALVS dataset. From top to bottom: soybean, sweet cherry, and London planetree. The colors of red, yellow, and white represent 1° , 2° , and 3° veins, respectively. The qualitative performance of the 3° vein in (e) is notably superior to the counterparts in (c) and (d). Best viewed with digital zoom.

Method	1° vein	2° vein	3° vein	mIoU
Supervised	64.93	39.71	23.56	42.73
FixMatch [Sohn <i>et al.</i> , 2020]	68.07	45.57	22.32	45.32
FixMatch + PSSS (Ours)	72.71	52.20	36.69	53.87
U ² PL [Wang <i>et al.</i> , 2022]	68.85	43.28	28.85	46.99
U ² PL + PSSS (Ours)	69.48	46.46	32.74	49.56
UniMatch [Yang <i>et al.</i> , 2023]	67.94	45.73	25.73	46.47
UniMatch + PSSS (Ours)	71.79	51.34	36.73	53.29
Oracle	73.37	51.12	33.80	52.76

Table 3: Performance comparison for SSL methods with and without the proposed PSSS module. IoU of 1° vein, 2° vein, 3° vein, and mIoU are reported. Integrating PSSS can efficiently improve the segmentation performance, especially on the 3° vein.

the overall loss. As expected, higher confidence of τ is preferable, aiding in filtering out unreliable pseudo-labels. The optimal performance is observed with $\lambda = 1$ and $\tau = 0.95$, adopted as the default setting for all experiments.

Sensitivity to Labeled Data Ratio. We study the impact of the relative ratio of partially labeled data and unlabeled data to fully labeled data. As shown in Tab. 5, the model’s performance can be improved when only raising the ratio of partially labeled or unlabeled data. Also, only a slight increase in the ratio of partially labeled data can result in a non-trivial performance gain.

Effect of Dataset Size. We increase the volume of training data three times to study the scaling effect of dataset. As shown in Fig. 6, increasing data volume can achieve better performance under all three setups. However, there are two important findings. First, this scaling effect is not linear. There seems to be a diminishing return when the size of

L_P^s	L_P^u	L_P^c	1° vein	2° vein	3° vein	mIoU
-	-	-	70.11	47.75	16.73	44.86
✓			67.53	51.26	18.28	45.69
✓	✓		70.16	57.05	22.08	49.76
✓		✓	70.37	54.52	30.50	51.80
✓	✓	✓	73.27	57.19	32.67	54.38

Table 4: Ablation study on the effectiveness of loss components in PSSS. In contrast to SSL baseline (1st row), all loss components have positive impacts on the 3° vein.

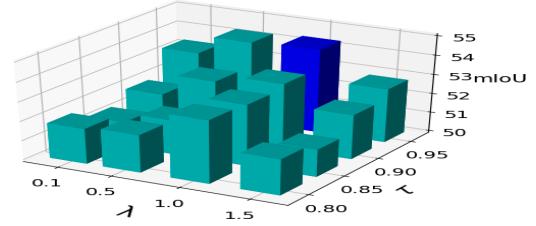


Figure 5: Comparison of performance under various combinations of score threshold τ and loss weight λ . The blue bar represents the optimal situation, with $\lambda = 1$ and $\tau = 0.95$.

dataset becomes larger. Second, while increasing data volume results in some enhancement, the performance gain on segmenting the 3° vein remains limited. Importantly, though segmenting the 3° vein is difficult, this suggests that PSSS can use only $\frac{1}{3}$ of training data to reach the performance upper bound while significantly outperforming supervised and semi-supervised baselines.

Effect of Network Backbones. We employ DeepLabV3+ as the architecture with ResNet101 and Xception [Chollet, 2017] serving as the backbones. Additionally, we select the

Method	Ratio	1° vein	2° vein	3° vein	mIoU
FixMatch	1:1:10	72.71	52.20	36.69	53.87
	1:1:20	73.50	52.89	37.50	54.63
	1:2:10	75.32	54.77	38.70	56.26
	1:2:20	76.21	55.80	37.94	56.65
	1:4:10	77.47	56.40	38.61	57.49
	1:4:20	77.08	57.30	39.54	57.97
U ² PL	1:1:10	69.48	46.46	32.74	49.56
	1:1:20	71.95	49.67	36.11	52.58
	1:2:10	73.48	51.02	36.49	53.66
	1:2:20	73.55	51.30	37.11	53.99
	1:4:10	73.75	52.50	37.29	54.51
	1:4:20	74.15	52.43	37.92	54.83
UniMatch	1:1:10	71.79	51.34	36.73	53.29
	1:1:20	74.28	54.48	38.41	55.72
	1:2:10	75.29	54.58	38.21	56.03
	1:2:20	76.41	55.79	38.79	57.00
	1:4:10	76.10	55.13	38.44	56.56
	1:4:20	77.72	56.56	38.68	57.65

Table 5: Ablation study on different ratios of fully labeled, partially labeled, and unlabeled data under three SSL baselines with PSSS. Adding partially labeled data and/or unlabeled data can increase the performance. Each ratio unit contains 6 images with complete leaf blades.

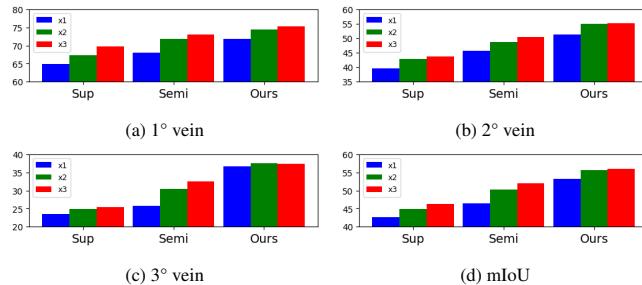


Figure 6: Comparison of performance across different data volumes. “Sup”, “Semi”, and “Ours” refer to “Supervised”, “Unimatch”, and “Unimatch + PSSS”, respectively. PSSS can efficiently utilize the labeled information, especially for the 3° vein.

Swin-Unet [Cao *et al.*, 2022] architecture that employs Swin Transformer [Liu *et al.*, 2021] as its backbone. As shown in Tab. 6, the PSSS module can improve the model’s performance on all three representative backbones, proving its superior generalization capabilities.

5.4 Analysis of Cross-species Learning

We conduct the first study of cross-species learning for leaf vein segmentation. Based on Tab. 1, we find that soybean leaves are relatively labor-sparse and easy to be annotated in contrast to the other two species. Thus, soybean is considered the *source* species, and the other two species are considered the *target* species. In the first scenario, we assume that only soybean data are available and we perform direct transfer learning to two other species unseen in the training. We use the default ratio 1:1:10 where the basic unit contains 20 images with full blades. The results are shown in Tab. 7.

Backbone	1° vein	2° vein	3° vein	mIoU
ResNet101	68.07	45.57	22.32	45.32
ResNet101 + PSSS	72.71	52.20	36.69	53.87
Xception	69.71	45.83	19.51	45.02
Xception + PSSS	71.08	50.73	34.49	52.10
Swin Transformer	36.88	27.22	22.73	28.94
Swin Transformer + PSSS	40.10	33.54	25.04	32.89

Table 6: Evaluation of PSSS under different network backbones. In this experiment, we use FixMatch as the SSL baseline. PSSS is robust under different network architectures.

Method	Sweet cherry			London planetree				
	1° vein	2° vein	3° vein	mIoU	1° vein	2° vein	3° vein	mIoU
Sup	29.30	23.51	9.62	20.81	26.43	12.46	7.83	15.57
Semi	35.31	24.40	8.14	22.62	31.59	13.71	8.87	18.06
Ours	59.40	19.20	13.16	30.59	30.34	26.07	26.94	27.78

Table 7: Transfer learning performance from soybean to other two species. The baseline SSL method is UniMatch.

Method	Sweet cherry			London planetree				
	1° vein	2° vein	3° vein	mIoU	1° vein	2° vein	3° vein	mIoU
Semi	25.24	3.43	0.00	9.56	35.45	1.32	0.00	12.25
Ours	35.02	33.87	19.61	29.50	48.96	26.37	1.46	25.60
Oracle	68.36	46.82	13.25	42.81	67.67	27.69	9.86	35.07

Table 8: Cross-species learning performance from soybean to other two species. The baseline SSL method is UniMatch.

Without seeing any samples from the target species, simply adding more partially labeled and unlabeled data from the source species can improve the overall model generalization on the target species, especially on the 3° vein.

In the second scenario, we consider an extreme label scarcity case: in addition to small-scale fully labeled source data, the target species can have access to unlabeled data and small-scale partially labeled data in the training. With the default ratio, there are 6 fully labeled source leaves, 6 partially labeled, and 60 unlabeled target leaves. The results are shown in Tab. 8. We also include an “Oracle” for supervised training on fully labeled target species. We notice that though our method significantly outperforms the semi-supervised counterpart, the supervised Oracle can easily achieve better overall performance. We also conclude that, due to *dataset shift* [Quiñonero-Candela *et al.*, 2008] between species, human annotations are still the key to successful applications. This further supports that HALVS is important for future research on the task of interest.

6 Conclusion

In this work, we explore the novel task of leaf hierarchical vein segmentation. We provide a finely annotated HALVS dataset for the first time and propose a label-efficient learning paradigm by considering the practical difficulties in annotating leaf veins. The empirical studies not only reveal new observations and challenges but also pose an insight into future research directions.

Acknowledgements

This study was partially supported by the National Science Foundation of China (No.32090061, No.32200331), the Major Science and Technology Research Project of Hubei Province (No.2021AFB002), the Start-Up Grant from Wuhan University of Technology of China (No.104-40120526), and Shanghai Artificial Intelligence Laboratory.

References

- [Blonder *et al.*, 2019] Benjamin Blonder, Sabine Both, Miguel Jodra, Noreen Majalap, David Burslem, Yit Arn Teh, and Yadavinder Malhi. Leaf venation networks of bornean trees: images and hand-traced segmentations. *Ecology*, 100(11), 2019.
- [Böhler *et al.*, 2015] Jonas Böhler, Louai Rishmawi, Daniel Pflugfelder, Gregor Huber, Hanno Scharr, Martin Hüskamp, Maarten Koornneef, Ulrich Schurr, and Siegfried Jahnke. phenovein—a tool for leaf vein segmentation and analysis. *Plant Physiology*, 169(4):2359–2370, 2015.
- [Cao *et al.*, 2022] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European Conference on Computer Vision*, pages 205–218. Springer, 2022.
- [Chapelle *et al.*, 2009] Olivier Chapelle, Bernhard Scholkopf, and Alexander Zien. Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. *IEEE Transactions on Neural Networks*, 20(3):542–542, 2009.
- [Chen *et al.*, 2017] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, 2017.
- [Chen *et al.*, 2018] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 801–818, 2018.
- [Chollet, 2017] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1251–1258, 2017.
- [Dhondt *et al.*, 2012] Stijn Dhondt, Dirk Van Haerenborgh, Caroline Van Cauwenbergh, Roeland MH Merks, Wilfried Philips, Gerrit TS Beemster, and Dirk Inzé. Quantitative analysis of venation patterns of arabidopsis leaves by supervised image analysis. *The Plant Journal*, 69(3):553–563, 2012.
- [Dong *et al.*, 2022a] Nanqing Dong, Michael Kampffmeyer, Xiaodan Liang, Min Xu, Irina Voiculescu, and Eric Xing. Towards robust partially supervised multi-structure medical image segmentation on small-scale data. *Applied Soft Computing*, 114:108074, 2022.
- [Dong *et al.*, 2022b] Nanqing Dong, Jiayi Wang, and Irina Voiculescu. Revisiting vicinal risk minimization for partially supervised multi-label classification under data scarcity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4212–4220, 2022.
- [Ellis *et al.*, 2009] Beth Ellis, Douglas C Daly, Leo J Hickey, Kirk R Johnson, John D Mitchell, Peter Wilf, and Scott L Wing. *Manual of Leaf Architecture*. 2009.
- [Fang and Yan, 2020] Xi Fang and Pingkun Yan. Multi-organ segmentation over partially labeled datasets with multi-scale feature abstraction. *IEEE TMI*, 2020.
- [Gan *et al.*, 2019] Yangjing Gan, Yi Rong, Fei Huang, Lun Hu, Xiaohan Yu, Pengfei Duan, Shengwu Xiong, Haiping Liu, Jing Peng, and Xiaohui Yuan. Automatic hierarchy classification in venation networks using directional morphological filtering for hierarchical structure traits extraction. *Computational Biology and Chemistry*, 80:187–194, 2019.
- [Grinblat *et al.*, 2016] Guillermo L Grinblat, Lucas C Uzal, Mónica G Larese, and Pablo M Granitto. Deep learning for plant identification using vein morphological patterns. *Computers and Electronics in Agriculture*, 127:418–424, 2016.
- [He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.
- [Hickey, 1973] Leo J Hickey. Classification of the architecture of dicotyledonous leaves. *American Journal of Botany*, 60(1):17–33, 1973.
- [Iwamasa and Noshita, 2023] Kohei Iwamasa and Koji Noshita. Network feature-based phenotyping of leaf venation robustly reconstructs the latent space. *PLOS Computational Biology*, 19(7):e1010581, 2023.
- [Jin and Han, 2020] Zhou Jin and Feixiang Han. An automatic hierarchical leaf venation segmentation based on region growing. In *2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP)*, pages 124–128. IEEE, 2020.
- [Lasser and Katifori, 2017] Jana Lasser and Eleni Katifori. Net: a new framework for the vectorization and examination of network data. *Source Code for Biology and Medicine*, 12:1–11, 2017.
- [Li *et al.*, 2022] Lei Li, Wenzheng Hu, Jiang Lu, and Changshui Zhang. Leaf vein segmentation with self-supervision. *Computers and Electronics in Agriculture*, 203:107352, 2022.
- [Liu *et al.*, 2021] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using

- shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10012–10022, 2021.
- [Liu *et al.*, 2022] Xiaobao Liu, Biao Xu, Wenjuan Gu, Yanchao Yin, and Hongcheng Wang. Plant leaf veins coupling feature representation and measurement method based on deeplabv3+. *Frontiers in Plant Science*, 13:1043884, 2022.
- [Papandreou *et al.*, 2015] George Papandreou, Liang-Chieh Chen, Kevin P Murphy, and Alan L Yuille. Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1742–1750, 2015.
- [Paszke *et al.*, 2019] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, 32, 2019.
- [Perez-Harguindeguy *et al.*, 2016] Natalia Perez-Harguindeguy, Sandra Diaz, Eric Garnier, Sandra Lavorel, Hendrik Poorter, Pedro Jaureguiberry, M Syndonia Bret-Harte, William K Cornwell, Joseph M Craine, Diego E Gurvich, et al. Corrigendum to: New handbook for standardised measurement of plant functional traits worldwide. *Australian Journal of Botany*, 64(8):715–716, 2016.
- [Price *et al.*, 2011] Charles A Price, Olga Symonova, Yuriy Mileyko, Troy Hilley, and Joshua S Weitz. Leaf extraction and analysis framework graphical user interface: segmenting and analyzing the structure of leaf veins and areoles. *Plant Physiology*, 155(1):236–245, 2011.
- [Quiñonero-Candela *et al.*, 2008] Joaquin Quiñonero-Candela, Masashi Sugiyama, Anton Schwaighofer, and N Lawrence. Covariate shift and local learning by distribution matching. *Dataset Shift in Machine Learning*, pages 131–160, 2008.
- [Sack *et al.*, 2013] Lawren Sack, Christine Scoffoni, Grace P John, Hendrik Poorter, Chase M Mason, Rodrigo Mendez-Alonso, and Lisa A Donovan. How do leaf veins influence the worldwide leaf economic spectrum? review and synthesis. *Journal of experimental botany*, 64(13):4053–4080, 2013.
- [Salima *et al.*, 2015] Adzkia Salima, Yeni Herdiyeni, and Stéphane Douady. Leaf vein segmentation of medicinal plant using hessian matrix. In *2015 International Conference on Advanced Computer Science and Information Systems (ICACIS)*, pages 275–279. IEEE, 2015.
- [Schneider *et al.*, 2018] Julio V Schneider, Renate Rabenstein, Jens Wessenberg, Karsten Wesche, Georg Zizka, and Jörg Habersetzer. Improved non-destructive 2d and 3d x-ray imaging of leaf venation. *Plant Methods*, 14(1):1–15, 2018.
- [Shi *et al.*, 2021] Gonglei Shi, Li Xiao, Yang Chen, and S Kevin Zhou. Marginal loss and exclusion loss for partially supervised multi-organ segmentation. *Medical Image Analysis*, 70:101979, 2021.
- [Sohn *et al.*, 2020] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in Neural Information Processing Systems*, 33:596–608, 2020.
- [United Nations, 2023a] United Nations. 17 sustainable development goals. <https://sdgs.un.org/goals>, 2023.
- [United Nations, 2023b] United Nations. Leave no one behind. <https://unsdg.un.org/2030-agenda/universal-values/leave-no-one-behind>, 2023.
- [Vollmann, 2016] Johann Vollmann. Soybean versus other food grain legumes: A critical appraisal of the united nations international year of pulses 2016. *Die Bodenkultur: Journal of Land Management, Food and Environment*, 67(1):17–24, 2016.
- [Wang *et al.*, 2022] Yuchao Wang, Haochen Wang, Yujun Shen, Jingjing Fei, Wei Li, Guoqiang Jin, Liwei Wu, Rui Zhao, and Xinyi Le. Semi-supervised semantic segmentation using unreliable pseudo-labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4248–4257, 2022.
- [Xu *et al.*, 2021] Hao Xu, Benjamin Blonder, Miguel Jodra, Yadvinder Malhi, and Mark Fricker. Automated and accurate segmentation of leaf venation networks via deep learning. *New Phytologist*, 229(1):631–648, 2021.
- [Yang *et al.*, 2023] Lihe Yang, Lei Qi, Litong Feng, Wayne Zhang, and Yinghuan Shi. Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7236–7246, 2023.
- [Zhou *et al.*, 2019] Yuyin Zhou, Zhe Li, Song Bai, Chong Wang, Xinlei Chen, Mei Han, Elliot Fishman, and Alan L Yuille. Prior-aware neural network for partially-supervised multi-organ segmentation. In *ICCV*, pages 10672–10681, 2019.
- [Zhu *et al.*, 2020] Jiyou Zhu, Jiangming Yao, Qiang Yu, Weijun He, Chengyang Xu, Guoming Qin, Qiuyu Zhu, Dayong Fan, and Hua Zhu. A fast and automatic method for leaf vein network extraction and vein density measurement based on object-oriented classification. *Frontiers in Plant Science*, 11:499, 2020.