

MV-Match: Multi-View Matching for Domain-Adaptive Identification of Plant Nutrient Deficiencies

Jinhui Yi*¹

jinhui.yi@uni-bonn.de

Yanan Luo*¹

yluo@uni-bonn.de

Marion Deichmann²

madeic@uni-bonn.de

Gabriel Schaaf²

gabriel.schaaf@uni-bonn.de

Juergen Gall^{1,3}

gall@iai.uni-bonn.de

¹ Computer Vision Group

University of Bonn

Bonn, Germany

² Plant Nutrition Group

University of Bonn

Bonn, Germany

³ Lamarr Institute for Machine Learning

and Artificial Intelligence

Germany

* indicates equal contribution

Abstract

An early, non-invasive, and on-site detection of nutrient deficiencies is critical to enable timely actions to prevent major losses of crops caused by lack of nutrients. While acquiring labeled data is very expensive, collecting images from multiple views of a crop is straightforward. Despite its relevance for practical applications, unsupervised domain adaptation where multiple views are available for the labeled source domain as well as the unlabeled target domain is an unexplored research area. In this work, we thus propose an approach that leverages multiple camera views in the source and target domain for unsupervised domain adaptation. We evaluate the proposed approach on two nutrient deficiency datasets. The proposed method achieves state-of-the-art results on both datasets compared to other unsupervised domain adaptation methods. The dataset and source code are available at <https://github.com/jh-yi/MV-Match>.

1 Introduction

In recent years, nutrient status monitoring has become a popular topic for the precision management of fertilizer in smart farming [0, 24, 27, 33, 39]. An early, non-invasive, and on-site identification of plant nutrient deficiencies is critical to enable timely actions to prevent major losses of crops caused by lack of nutrients. This can help farmers improve crop yields and prevent excess fertilization with negative environmental consequences, such as nitrous oxide emissions or groundwater pollution. For rapid and non-invasive diagnosis of nutrient deficiencies, deep learning methods applied to RGB images have been widely adopted [13, 30, 36, 40, 43, 46, 56]. These approaches, however, do not generalize to genotypes that

have not been observed during training. Furthermore, collecting labeled data is extremely expensive since it requires expert knowledge to recognize nutrient deficiency and it is even difficult for experts to distinguish plant stress caused by nutrient imbalances from other causes like droughts or diseases. Detecting nutrient deficiencies across different genotypes is thus a highly relevant application for evaluating and developing unsupervised domain adaptation methods [28, 44, 51].

Current datasets and approaches for domain adaptation consider only a single view, i.e., each object or scene in the source and target domain have been taken from a single view. Multiple views of the same scene, however, provide additional information since it allows to learn or discard variability of the appearance that is caused by viewpoint changes and not necessarily by changes of the domain. Utilizing this additional information is very practical since it is straightforward for many applications like nutrient deficiency detection to collect multiple views of the same location both for the source and target domain, in particular if the images can be simply taken by a smartphone without the need of any camera calibration or expensive camera setup. Despite its practical relevance, unsupervised domain adaptation under multi-view scenarios is a widely unexplored research area. An exception is a recent work by Lu et al. [27] that addresses domain adaptation for object detection in a surveillance setting where the camera views of the target domain overlap. This specific setting, however, does not generalize to other tasks where the assumption of overlapping views cannot be guaranteed.

In this work, we thus investigate the task of unsupervised multi-view domain adaptation. In our context, multi-view refers to multiple images of the same scene taken from different views, which is different to the multi-view learning approach investigated in Xia et al. [53] where views refer to different modalities. To study this challenging task, we collected a dataset of images of crops that suffer from different nutrient deficiencies. Each crop has been captured over the growing season multiple times and each time from multiple views as illustrated in Figure 1. While each view does not necessarily contain exactly the same plant, the multiple views contain very closely located plants with the same nutrient status. The images are annotated by the date, genotype, and nutrient deficiency. As a second contribution, we propose an approach that leverages the multiple views for domain adaptation. The proposed approach, which we term Multi-View Match (MV-Match), enforces the consistency of the predictions among multiple views. In addition, we propose a Similarity-guided View Mining (SgVM) mechanism to automatically select the most dissimilar views that contain complementary information given a query image.

We evaluate our approach on two nutrient deficiency benchmarks and show that our approach achieves state-of-the-art performance compared to other unsupervised domain adaptation approaches.

2 Related Work

Plant Nutrient Deficiency. Better matching of the timing and amount of fertilizer inputs to plant requirements will improve nutrient use efficiency and crop yields. Driven by tremendous economic potential, deep learning applied to RGB images is widely adopted for early, non-invasive, and on-site monitoring of plant nutrient deficiency in plants [40], including rice [56, 43], sugar beet [56], hydroponic basil [13], winter wheat [57], and tomato [46]. These approaches evaluated various convolutional neural networks (CNNs) on their proposed datasets [13, 46, 56, 57] or applied an ensemble of CNNs to achieve better perfor-

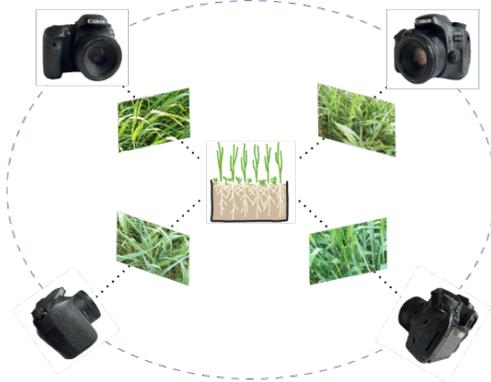


Figure 1: Illustration of our multi-view setting. Related views are expected to share the same nutrient status, both in the labeled source and the unlabeled target domain. We do not assume that each view contains exactly the same plant, but very closely located plants with the same nutrient status, which makes the data capturing in an open field very simple.

mance on publicly available datasets [56, 43]. However, they detected plant nutrient deficiency in images of the same domain as the labeled training data, whereas we aim to identify the nutrient status of plants in out-of-domain data.

Unsupervised Domain Adaptation. Unsupervised domain adaptation approaches [28, 44, 50] aim to adapt a model trained on a labeled source domain to an unlabeled target domain. Due to ubiquitous domain shifts in real life, unsupervised domain adaptation approaches were proposed for various applications, including image classification [5, 9, 19], semantic segmentation [3, 17, 50], and smart farming [11, 10, 23, 52, 35]. Most of these approaches involve the minimization of discrepancies, adversarial training, or self-training. Methods with discrepancy minimization aim to minimize the domain gap by adopting a statistical distance function such as maximum mean discrepancy [19, 20], correlation alignment [41, 42], or entropy minimization [12, 50]. In adversarial training, encoders and domain discriminators are trained to learn domain-invariant inputs [11, 15], features [9, 21], or outputs [34, 47, 50]. Recently, unsupervised domain adaptation approaches using pseudo-labels were proposed to generate artificial supervised signals based on confidence thresholds [25, 50] or pseudo-label prototypes [29, 59] for the unlabeled target data. Although unsupervised domain adaptation for smart farming [23] is attracting more attention for species recognition [10], disease detection [1], drought stress [35], and relative water content prediction [52], the exploration of unsupervised domain adaptation for identification of plant nutrient deficiency [56] remains an open area of investigation. Moreover, leveraging multiple views of a plant or object in the source and target domain is an unexplored research area in the field of domain adaptation.

Multi-View Learning Multi-view learning is commonly adopted in 3D reconstruction [52], 3D object recognition [51], and perception of 2D images. For 2D images, multi-view consistency is a form of consistency regularization in semi-supervised learning, which is often applied to ensure consistency over different data augmentations [3, 6, 26] and different crops [16, 18]. However, the aforementioned approaches for 2D images only consider multi-views as augmented versions of the same image, which inevitably contain local patterns

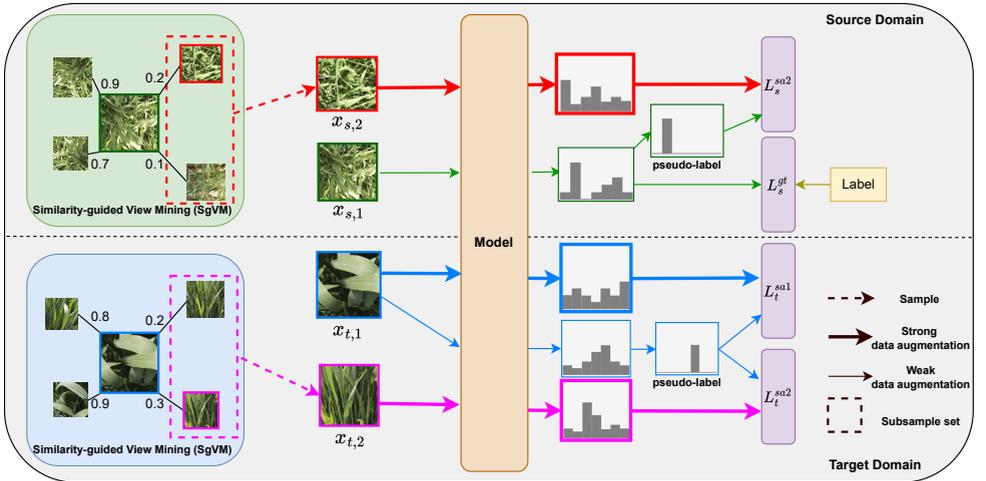


Figure 2: Proposed approach for unsupervised domain adaptation. Given multiple views of a crop in the labeled source domain (top) and the unlabeled target domain (bottom), a random query image is sampled from the source and the target domain (center of the green and blue box). The **Similarity-guided View Mining (SgVM)** module then computes the normalized mutual information between each query-view pair to select the top n dissimilar views of the same crop (red dashed rectangles). From these two sets, we randomly select a second image for each query image. We then apply weak or strong data augmentation to the four images, i.e., the two query images and their corresponding view pair images, and feed them to a shared model for predicting nutrient deficiencies. While the prediction of the query image of the source domain is supervised by the ground-truth label, the other predictions are enforced to be consistent with the corresponding view pair. For this, the query image with weak augmentation is always considered as a reference prediction, both for the source and target domain.

instead of the holistic appearance of a given object. In this work, we make use of images of plants with various camera angles to include different light conditions, viewpoints, and crop parts, to incorporate holistic details of diagnosed plants. We enforce consistency among these multi-view images instead of augmented images. Concurrent works [22, 53] evaluated similar settings for object detection, image classification, and 3D reconstruction, but none of them can be applied to the task of nutrient deficiency detection.

3 MV-Match

In this work, we propose an approach for unsupervised multi-view domain adaptation. This means that we have multiple views both in the source and the training domain, as illustrated in Fig. 1. Such a scenario is highly relevant for applications since capturing multiple views of the same crop or object is straightforward whereas annotating images by trained experts is often very expensive.

Our approach, which is illustrated in Fig. 2, thus aims to leverage the multiple views in the source and target domain. While we have labeled images for the source domain, where $x_{s,i}$ denotes an image of the source domain and $y_{s,i} \in \{1, 2, \dots, C\}$ is the corresponding label,

the images $x_{t,i}$ of the target domain are unlabeled. In our case, the images of the source and target domain contain different genotypes and the goal is to recognize different types of nutrient deficiencies in the target domain.

As illustrated in Fig. 2, we train a joint model for the source and target domain and aim to learn a model that is agnostic to the domain. During training, we select for each batch the same number of query images from the source and target domain. For each query image, we sample a second image that shows the same crop but from a different view. This step will be described in Section 3.2. For a query image of the source domain $x_{s,i}$, we can directly apply the cross-entropy loss using the ground-truth label $y_{s,i}$:

$$\mathcal{L}_s^{gt}(x_{s,i}, y_{s,i}) = H(y_{s,i}, P_{x_{s,i}}^{wa}) = - \sum_{c=1}^C \mathbf{1}(c = y_{s,i}) \log P_{x_{s,i}}^{wa}(c), \quad (1)$$

where P_x is the prediction of the network for image x and $P_x(c)$ is the predicted probability for class c . The indicator function $\mathbf{1}(c = y_{s,i})$ is 1 for the ground-truth label and 0 otherwise. x^{wa} is the weakly augmented version of a given image x , which is a random horizontal flip with a probability of 50%.

For the other images, we introduce consistency loss functions, which are denoted by \mathcal{L}_s^{sa2} , \mathcal{L}_t^{sa1} , and \mathcal{L}_t^{sa2} . These loss functions measure the prediction consistency across views for the source and the target domain, as illustrated in Fig. 2. We describe them in the next section more in detail.

3.1 Multi View Matching

For each query image of the source or target domain, we have a set of images $\mathcal{M}_{s,i}$ or $\mathcal{M}_{t,i}$ that show the same crop but from a different view. In Section 3.2, we will describe how these two sets are obtained. For each query image $x_{s,i}$ and $x_{t,i}$, we randomly select a second image from the corresponding set $x_{s,j} \in \mathcal{M}_{s,i}$ or $x_{t,j} \in \mathcal{M}_{t,i}$. As illustrated in Fig. 2, we apply strong augmentation to the sampled views, which we denote by $x_{s,j}^{sa}$ and $x_{t,j}^{sa}$. For this, we adopt AutoAugment [2] as strong augmentation, which is described more in detail in the supplementary material. For sampled views of the source domain, we compute the cross-entropy to enforce that the prediction is the same as for the query image of the source domain:

$$\mathcal{L}_s^{sa2}(x_{s,i}, x_{s,j}) = \mathbf{1}(\max_c P_{x_{s,i}}^{wa}(c) \geq \tau) H(\hat{y}_{s,i}, P_{x_{s,j}}^{sa}). \quad (2)$$

For computing the cross-entropy H , we consider in the experiments two cases: a) $\hat{y}_{s,i} = \arg\max_c P_{x_{s,i}}^{wa}(c)$ or b) $\hat{y}_{s,i} = P_{x_{s,i}}^{wa}$. While a) generates a hard pseudo-label, b) is a soft labeling based on the predicted probabilities of all classes. In the case of a hard pseudo-label, the loss is only applied if the prediction on the query image reaches a certain confidence, i.e., $\max_c P_{x_{s,i}}^{wa}(c) \geq \tau$.

For the target images, we define a similar loss function:

$$\mathcal{L}_t^{sa2}(x_{t,i}, x_{t,j}) = \mathbf{1}(\max_c P_{x_{t,i}}^{wa}(c) \geq \tau) H(\hat{y}_{t,i}, P_{x_{t,j}}^{sa}), \quad (3)$$

which enforces that the predictions of the two views of the target domain are consistent. In addition, we compute a loss that measures the consistency of the weak and strong augmentation of the target query image:

$$\mathcal{L}_t^{sa1}(x_{t,i}) = \mathbf{1}(\max_c P_{x_{t,i}}^{wa}(c) \geq \tau) H(\hat{y}_{t,i}, P_{x_{t,i}}^{sa}). \quad (4)$$

For training, we combine all four loss terms without any weighting:

$$\mathcal{L} = \mathcal{L}_s^{gl}(x_{s,i}, y_{s,i}) + \mathcal{L}_s^{sa2}(x_{s,i}, x_{s,j}) + \mathcal{L}_t^{sa2}(x_{t,i}, x_{t,j}) + \mathcal{L}_t^{sa1}(x_{t,i}) \quad (5)$$

The impact of the loss functions except \mathcal{L}_s^{gl} , which is always needed, is evaluated in the experiments.

3.2 Similarity-guided View Mining

For a given query image $x_{s,i}$ or $x_{t,i}$, we need to select the set of images $\mathcal{M}_{s,i}$ or $\mathcal{M}_{t,i}$. We first select all images that have been taken from the same location as the query image but from a different view. Since the nutrient status of the same plant might change over time, we only consider images that have been taken on the same day as the query image.

In the experiments, we show that it is better to first select a subset of images instead of randomly sampling $x_{s,j}$ from $\mathcal{M}_{s,i}$. To this end, we select a subset of images that are most dissimilar to the query image, but that show the same crop as the query image. In this way, we select views that are not very similar to the query view.

For the selection, we compute the Normalized Mutual Information (NMI) [49] between the query image $x_{s,i}$ and all view images $x_{s,k}$ of the same crop as the query image:

$$NMI(x_{s,i}, x_{s,k}) = 2 \frac{H(x_{s,i}) - H(x_{s,i}|x_{s,k})}{H(x_{s,i}) + H(x_{s,k})}. \quad (6)$$

where $H(x_{s,i})$ is the entropy of image $x_{s,i}$ and $H(x_{s,i}|x_{s,k})$ is the conditional entropy. Larger $NMI(x_{s,i}, x_{s,k})$ indicates higher similarity. For building $\mathcal{M}_{s,i}$, we select the 5 images $x_{s,k}$ with the lowest NMI, i.e., we select the 5 most dissimilar images to the query image. We denote the selection as Similarity-guided View Mining (SgVM). We apply SgVM to the target images in the same way to obtain $\mathcal{M}_{t,i}$.

4 Experiments

4.1 Datasets and Metrics

We evaluate our proposed MV-Match for image classification on two nutrient deficiency benchmarks, MiPlo-B [8] and MiPlo-WW. We report the top-1 accuracy on the test set as our metric. We also report per-nutrient accuracy as it is critical for nutrient monitoring in smart farming. As a network, we use a ResNet50 [44].

MiPlo-B. The Mini Plot Barley (MiPlo-B) dataset consists of 18559 images with 6 nutrient treatments (-N, -P, -K, -B, -S, ctrl) annotated, ranging from 21.06.2022 - 20.07.2022 (16 dates). It contains two genotypes: Barke (9305 images) and Hanna (9254 images). For each genotype, each treatment was repeated 4 times, resulting in 24 containers, each of which has a unique ID. Six unique containers with six different nutrient treatments were selected as the test set while the other containers as the training set (#train:#test≈75%:25%), trying to avoid information leaks due to commonly adopted random sampling, i.e., multi-views of the same crop being separated into both training and test set.

MiPlo-WW. The Mini Plot Winter Wheat (MiPlo-WW) dataset has 12466 images with 6 treatments (-N, -P, -K, -B, -S, ctrl) annotated, ranging from 12.05.2023 - 24.05.2023 (13

Table 1: Top-1 Classification Accuracy (%) for adaptation across genotypes: **Barley: Barke** → **Hanna**. *Oracle* indicates the model was trained with full supervision on the **Hanna** training set. *Source-Only* denotes the results without adaptation. The highest accuracy is shown in bold, while the second best is underlined. *TPS* refers to throughput per second.

Model	Barley: Barke → Hanna						AVG	Train Time (min)	TPS (test)
	-N	-P	-K	-B	-S	control			
Oracle	97.3	78.5	82.1	99.1	9.1	87.1	75.4	-	-
Source-Only	95.1	26.0	80.4	84.3	14.8	24.2	54.0	-	-
DANN [9]	93.0	36.6	59.0	68.6	24.2	46.0	54.7	215.1	35.0
ADDA [13]	83.7	47.1	50.4	79.5	17.4	33.0	51.8	125.4	35.0
JAN [20]	87.2	26.1	63.7	70.8	22.9	39.6	51.9	86.6	35.0
CDAN [21]	95.2	35.0	41.5	69.8	20.4	44.8	51.3	100.3	35.0
BSP [5]	94.5	41.3	68.2	76.0	14.5	57.2	58.8	153.0	35.0
AFN [6]	94.7	25.8	73.9	64.1	24.9	68.7	59.0	233.9	35.0
Mean Teacher [15]	94.2	32.1	81.8	89.4	11.7	52.5	60.6	266.7	35.0
FixMatch [37]	98.2	38.9	83.8	86.4	25.4	53.2	64.6	257.6	35.0
FlexMatch [38]	97.2	50.0	89.1	89.4	11.0	58.0	65.9	276.4	35.0
Ours+hard	92.7	50.0	86.8	95.3	22.7	68.9	69.6	666.7	35.0
Ours+soft	97.0	53.9	75.4	92.8	20.0	64.7	<u>67.4</u>	714.3	35.0

dates). It contains two genotypes: Julius (6253 images) and Meister (6213 images). The ID settings are the same as above.

We refer to the supplementary materials for more details about the datasets and experimental details.

4.2 Comparison to State of the Art

We present experimental results of our approach compared to different baselines. *Source-only* denotes training only on the source domain without adaptation. DANN [9], ADDA [13], JAN [20], and CDAN [21] learn domain-invariant features with adversarial learning: DANN jointly trains encoder and discriminator with a gradient reversal layer; ADDA adopts separated encoders for source and target domains and alternates training of the encoder and discriminator; JAN adopts adversarial training to maximize a joint maximum mean discrepancy (JMMD); CDAN conditions the adversarial adaptation models on both the features and the classifier predictions. BSP [5] avoids deterioration of the feature discriminability after adaptation by penalizing the largest singular values. AFN [6] enlarges feature norms to enhance the transferability of features. Mean Teacher [15] constructs a teacher model by averaging model weights of a student model with an exponential moving average and forcing the consistency between the predictions of two models. FixMatch [37] takes the predictions of weakly-augmented unlabeled images as pseudo-labels and their strongly-augmented version as predictions to train the model in a supervised manner. FlexMatch [38] extends FixMatch by considering adaptive thresholds for different classes at each time step. All approaches were trained with the same amount of images, i.e., single-view approaches were trained using all views as images but they do not utilize multi-view information. In what follows, we show the performance of our model in two plant nutrient deficiency benchmarks under different settings, i.e., adaptation across genotypes (smaller domain gap) and across cultivars (larger domain gap).

Adaptation Across Genotypes: We report in Tables 1 and 2 the performance of adaptation across genotypes in terms of per-nutrient accuracy and average accuracy. The results indicate that for both settings of adaptation across genotypes, MV-Match outperforms the

Table 2: Top-1 Classification Accuracy (%) for adaptation across genotypes: **Winter Wheat: Julius** \rightarrow **Meister**. *Oracle* indicates the model was trained with full supervision on the **Meister** training set. *Source-Only* denotes the results without adaptation. The highest accuracy is shown in bold, while the second best is underlined. *TPS* refers to throughput per second.

Model	Winter Wheat: Julius \rightarrow Meister							Train Time (min)	TPS (test)
	-N	-P	-K	-B	-S	control	AVG		
Oracle	98.9	85.4	58.9	81.9	50.2	51.7	71.0	-	-
Source-Only	100.0	88.8	5.6	46.6	1.1	61.6	50.5	-	-
DANN [D]	99.8	54.9	22.6	52.9	20.5	58.9	51.4	196.7	35.0
ADDA [Dx]	99.0	57.5	50.7	51.3	30.4	44.1	55.7	130.3	35.0
JAN [Dx]	98.9	50.0	26.7	49.6	39.9	58.2	53.9	83.3	35.0
CDAN [Dx]	99.2	44.0	33.7	55.9	29.7	63.9	54.3	95.2	35.0
BSP [D]	100.0	59.7	56.3	55.9	49.8	49.4	61.9	144.4	35.0
AFN [Dx]	99.6	67.9	24.8	68.1	19.8	61.6	56.6	199.0	35.0
Mean Teacher [Dx]	99.2	67.9	62.6	72.7	18.6	37.6	59.6	266.7	35.0
FixMatch [Dx]	99.6	75.0	66.7	62.2	25.5	64.6	<u>65.6</u>	279.1	35.0
FlexMatch [Dx]	99.2	72.8	68.1	87.4	3.0	64.6	65.5	239.3	35.0
Ours+hard	99.2	59.0	71.1	93.7	4.6	76.4	66.9	588.2	35.0
Ours+soft	100.0	60.8	59.3	76.5	29.7	50.6	62.5	555.6	35.0

baseline methods by a large margin. Comparing *Oracle* and *Source-Only*, we see that different genotypes drastically deteriorate the performance without adaptation, especially for *P* deficiency and *control* in **Barley: Hanna** \rightarrow **Barke** and *K* deficiency, *B* deficiency and *S* deficiency in **Winter Wheat: Meister** \rightarrow **Julius**. On adapting the network, most baselines based on adversarial learning or discrepancy minimization do not attain large improvements over the un-adapted model for both scenarios. Notably, Mean Teacher and FixMatch achieve a relatively large improvement. Both approaches adopt consistency regularization, which we assume is critical for plant nutrient deficiency. With the proposed multi-view consistency, our approach yields consistent and significant performance gains in both adaptation settings. MV-Match with hard pseudo-labels outperforms its counterpart with soft pseudo-labels in both settings. We furthermore report the training time and throughput per second (TPS) during inference in Tables 1 and 2. While the inference time is the same for all methods since they all use the same backbone suitable for real-time processing, our approach requires more time for training due to selecting dissimilar views.

Adaptation Across Cultivars: We also evaluate the methods for adaptation across different crop species. This setting is extremely challenging because both crops were cultivated in different years, where changing environmental conditions as well as abiotic and biotic stresses might have affected the growing status of crops. This is reflected in the results in Table 3. For many adaptation settings, baseline methods decrease the performance of the un-adapted model. Our method consistently boosts the performance across crop species. We attribute this robustness against the changing environment to our proposed multi-view matching, which incorporates holistic visual symptoms for effective identification of plant nutrient deficiency. Our model with soft pseudo-labels works better than with hard pseudo-labels. This is different to the across genotypes experiments where the domain gaps are smaller. This is not unexpected since soft pseudo-labels work better for out-of-domain unlabeled data [54]. Note that FixMatch and FlexMatch also achieve promising results in adaptation across genotypes, but their performance deteriorates for adaptation across crop species.

Table 3: Mean Top-1 Classification Accuracy (%) for adaptation across crop species. *Oracle* indicates the model was trained with full supervision on the target training set. *Source-Only* denotes the results without adaptation. The highest accuracy is shown in bold, while the second best is underlined.

Model	Barley \rightarrow Winter Wheat				Winter Wheat \rightarrow Barley			
	B \rightarrow J	B \rightarrow M	H \rightarrow J	H \rightarrow M	J \rightarrow B	J \rightarrow H	M \rightarrow B	M \rightarrow H
Oracle	73.5	71.0	73.5	71.0	67.9	75.4	67.9	75.4
Source-Only	31.9	40.7	27.0	31.7	18.8	13.3	16.3	13.7
DANN [9]	<u>40.5</u>	39.4	22.2	19.3	19.0	18.7	23.8	17.6
ADDA [13]	30.6	34.4	21.3	24.4	<u>25.2</u>	16.0	21.0	19.8
JAN [24]	35.5	40.0	14.4	27.4	19.5	14.6	17.9	16.1
CDAN [24]	36.0	34.3	25.7	17.2	21.9	14.0	24.4	18.0
BSP [8]	39.9	<u>42.5</u>	30.5	29.6	22.1	16.3	21.0	18.3
AFN [5]	38.2	39.6	23.3	20.0	20.9	18.0	22.2	18.5
Mean Teacher [15]	30.6	37.6	37.1	<u>40.9</u>	23.3	<u>19.9</u>	26.0	<u>20.4</u>
FixMatch [7]	27.9	33.6	29.4	27.8	24.3	19.3	19.6	19.5
FlexMatch [13]	26.8	35.4	30.1	34.3	25.1	16.5	21.7	18.9
Ours+hard	33.7	33.7	28.9	31.8	20.9	16.8	21.6	16.5
Ours+soft	42.2	46.6	<u>34.4</u>	41.2	27.1	23.2	26.0	21.3

4.3 Ablation Study

The results in Table 4 show that forcing consistency between an unlabeled target image with its related view significantly boosts the performance, and incorporating such consistency for the source image brings extra gain in performance. This validates the effectiveness of our proposed multi-view matching mechanism. Note that we only use weak augmentation instead of strong augmentation for these results. To further explore the effects of strong augmentation in our multi-view setting, we run a series of experiments with various combinations of loss components, as shown in Table 5. The models that use L_s^{sa2} and L_t^{sa2} have better results as compared to Table 4, indicating the effectiveness of strong augmentation. By including L_t^{sa1} , we observe a decent boost in performance. The combination of L_s^{sa2} , L_t^{sa2} and L_t^{sa1} achieves the highest accuracy of 69.6% on the Barley: B \rightarrow H benchmark. Such a result indicates that complementary information is learned by multi-view matching. More ablation studies are in the supplementary material.

\mathcal{L}_s^{wa2}	\mathcal{L}_t^{wa2}	B \rightarrow H
		54.0
	✓	65.5
✓	✓	66.0

Table 4: Ablation study on effects of multi-view matching without any strong augmentation. \mathcal{L}_s^{wa2} denotes the self-training loss between weakly-augmented source image (pseudo-labels) and its weakly-augmented related view (predictions). The same applies to \mathcal{L}_t^{wa2} .

\mathcal{L}_s^{sa2}	\mathcal{L}_t^{sa2}	\mathcal{L}_t^{sa1}	B \rightarrow H
			54.0
	✓		67.2
✓	✓		68.7
		✓	64.6
	✓	✓	67.3
✓	✓	✓	69.6

Table 5: Ablation study on the impact of loss functions in the Barley: B \rightarrow H benchmark.

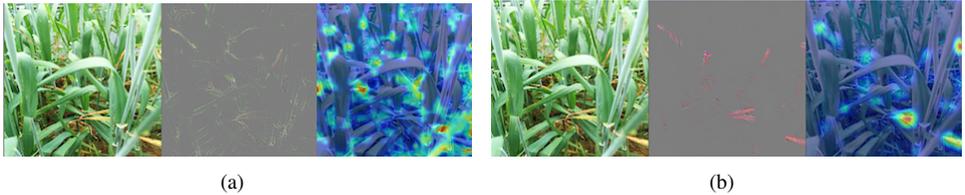


Figure 3: Saliency visualization (a) without adaptation and (b) with adaptation. For each case, we show the original image, saliency map obtained by guided backpropagation [58], and by GradCAM++ [9].

4.4 Visualization

We present an example of saliency visualization using our model before and after adaptation on the target dataset in Figure 3. The results show that the model fails to localize pathological symptoms caused by potassium deficiency in crops without adapting the model to the target domain (Figure 3a). After adaptation, the model successfully classifies the nutrient status by focusing on the discriminative pathological symptoms (Figure 3b).

5 Conclusion

In this work, we addressed unsupervised domain adaptation for plant nutrient deficiency detection. We proposed an approach that utilizes multiple views of crops both in the source and target domain. To this end, we enforce consistency between a query image and its strongly-augmented related view to incorporate complementary visual symptoms for the identification of plant nutrient deficiency. We also proposed a similarity-guided view mining mechanism, which ensures that the most dissimilar views are selected. We evaluated our approach on two datasets and provided additional ablation studies in the supplementary material. Our approach exhibited significantly improved performance compared to other methods in all the settings we studied. Despite the achieved major improvements compared to other unsupervised domain adaptation approaches, domain adaptation for nutrient deficiency detection across genotypes or cultivars remains challenging. Extending the approach to a multi-modal approach that utilizes data from other sensors, e.g., multi-spectral data, humidity, temperature, or chlorophyll fluorescence, is an interesting research direction to enhance the robustness and generalization abilities.

Acknowledgement This work has been supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy - EXC 2070–390732324 (PhenoRob), GA 1927/9-1 (KI-FOR 5351), and the Chinese Scholarship Council (202108440041).

References

- [1] Alwaseela Abdalla, Terry A Wheeler, Jane Dever, Zhe Lin, Joel Arce, and Wenxuan Guo. Assessing fusarium oxysporum disease severity in cotton using unmanned aerial

- system images and a hybrid domain adaptation deep learning time series model. *Biosystems Engineering*, 237:220–231, 2024.
- [2] MM Ali, Ahmed Al-Ani, Derek Eamus, and Daniel KY Tan. Leaf nitrogen determination using non-destructive techniques—a review. *Journal of Plant Nutrition*, 40(7): 928–953, 2017.
- [3] Nikita Araslanov and Stefan Roth. Self-supervised augmentation consistency for adapting semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 15384–15394, 2021.
- [4] Aditya Chattopadhyay, Anirban Sarkar, Prantik Howlader, and Vineeth N Balasubramanian. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In *WACV*, pages 839–847, 2018.
- [5] Xinyang Chen, Sinan Wang, Mingsheng Long, and Jianmin Wang. Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation. In *International conference on machine learning*, pages 1081–1090. PMLR, 2019.
- [6] Jaehoon Choi, Taekyung Kim, and Changick Kim. Self-ensembling with gan-based data augmentation for domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6830–6840, 2019.
- [7] Ekin D Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V Le. Autoaugment: Learning augmentation strategies from data. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 113–123, 2019.
- [8] Marion Deichmann, Jinhui Yi, Yeshambel E. Mihiret, Samer Zahra, Corinna Sauer, Hubert Hueging, Jens Léon, Matthias Wissuwa, Juergen Gall, and Gabriel Schaaf. Rgb image-based detection of nutrient deficiencies in barley by deep learning methods. *agriRXiv preprint*, 2024.
- [9] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario March, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of machine learning research*, 17(59):1–35, 2016.
- [10] Dario Gogoll, Philipp Lottes, Jan Weyler, Nik Petrinic, and Cyrill Stachniss. Unsupervised domain adaptation for transferring plant classification systems to new field environments, crops, and robots. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2636–2642. IEEE, 2020.
- [11] Rui Gong, Wen Li, Yuhua Chen, and Luc Van Gool. Dlow: Domain flow for adaptation and generalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2477–2486, 2019.
- [12] Yves Grandvalet and Yoshua Bengio. Semi-supervised learning by entropy minimization. *Advances in neural information processing systems*, 17, 2004.
- [13] Zeki Gul and Sebnem Bora. Exploiting pre-trained convolutional neural networks for the detection of nutrient deficiencies in hydroponic basil. *Sensors*, 23(12):5407, 2023.

- [14] 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.
- [15] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei Efros, and Trevor Darrell. Cycada: Cycle-consistent adversarial domain adaptation. In *International conference on machine learning*, pages 1989–1998. Pmlr, 2018.
- [16] Lukas Hoyer, Dengxin Dai, and Luc Van Gool. Hrda: Context-aware high-resolution domain-adaptive semantic segmentation. In *European conference on computer vision*, pages 372–391. Springer, 2022.
- [17] Lukas Hoyer, Dengxin Dai, Haoran Wang, and Luc Van Gool. Mic: Masked image consistency for context-enhanced domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11721–11732, 2023.
- [18] Xin Lai, Zhuotao Tian, Li Jiang, Shu Liu, Hengshuang Zhao, Liwei Wang, and Jiaya Jia. Semi-supervised semantic segmentation with directional context-aware consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1205–1214, 2021.
- [19] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation networks. In *International conference on machine learning*, pages 97–105. PMLR, 2015.
- [20] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. Deep transfer learning with joint adaptation networks. In *International conference on machine learning*, pages 2208–2217. PMLR, 2017.
- [21] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. *Advances in neural information processing systems*, 31, 2018.
- [22] Yan Lu, Zhun Zhong, and Yuanchao Shu. Multi-view domain adaptive object detection on camera networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 8966–8974, 2023.
- [23] Yuchi Ma, Shuo Chen, Stefano Ermon, and David B Lobell. Transfer learning in environmental remote sensing. *Remote Sensing of Environment*, 301:113924, 2024.
- [24] ChongYen Mee, Kumar Balasundram Siva, Husni Mohd Hanif Ahmad, et al. Detecting and monitoring plant nutrient stress using remote sensing approaches: a review. *Asian Journal of Plant Sciences*, 16(1):1–8, 2017.
- [25] Ke Mei, Chuang Zhu, Jiaqi Zou, and Shanghang Zhang. Instance adaptive self-training for unsupervised domain adaptation. In *European Conference on Computer Vision*, pages 415–430, 2020.
- [26] Luke Melas-Kyriazi and Arjun K Manrai. Pixmatch: Unsupervised domain adaptation via pixelwise consistency training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12435–12445, 2021.

- [27] Francisco M Padilla, Marisa Gallardo, M Teresa Peña-Fleitas, Romina De Souza, and Rodney B Thompson. Proximal optical sensors for nitrogen management of vegetable crops: A review. *Sensors*, 18(7):2083, 2018.
- [28] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.
- [29] Yingwei Pan, Ting Yao, Yehao Li, Yu Wang, Chong-Wah Ngo, and Tao Mei. Transferable prototypical networks for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2239–2247, 2019.
- [30] Ashish Patel, Priya Swaminarayan, and Mukta Patel. Identification of nutrition’s deficiency in plant and prediction of nutrition requirement using image processing. In *Proceedings of the Second International Conference on Information Management and Machine Intelligence*, pages 463–469. Springer, 2021.
- [31] Shaohua Qi, Xin Ning, Guowei Yang, Liping Zhang, Peng Long, Weiwei Cai, and Weijun Li. Review of multi-view 3d object recognition methods based on deep learning. *Displays*, 69:102053, 2021.
- [32] Tanzeel U Rehman and Jian Jin. Deep adversarial domain adaptation for hyperspectral calibration model transfer among plant phenotyping systems. *Biosystems Engineering*, 224:246–258, 2022.
- [33] Victor Rueda-Ayala, Hella Ellen Ahrends, Stefan Siebert, Thomas Gaiser, Hubert Hüging, and Frank Ewert. Impact of nutrient supply on the expression of genetic improvements of cereals and row crops—a case study using data from a long-term fertilization experiment in germany. *European journal of agronomy*, 96:34–46, 2018.
- [34] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3723–3732, 2018.
- [35] P Schmitter, J Steinrücken, C Römer, A Ballvora, J Léon, U Rascher, and L Plümer. Unsupervised domain adaptation for early detection of drought stress in hyperspectral images. *ISPRS journal of photogrammetry and remote sensing*, 131:65–76, 2017.
- [36] Mayuri Sharma, Keshab Nath, Rupam Kumar Sharma, Chandan Jyoti Kumar, and Ankit Chaudhary. Ensemble averaging of transfer learning models for identification of nutritional deficiency in rice plant. *Electronics*, 11(1):148, 2022.
- [37] 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.
- [38] J Springenberg, Alexey Dosovitskiy, Thomas Brox, and M Riedmiller. Striving for simplicity: The all convolutional net. In *ICLR Workshops*, 2015.

- [39] Hugo Storm, Sabine Julia Seidel, Lasse Klingbeil, Frank Ewert, Harry Vereecken, Wulf Amelung, Sven Behnke, Maren Bennewitz, Jan Börner, Thomas Döring, et al. Research priorities to leverage smart digital technologies for sustainable crop production. *European Journal of Agronomy*, 156:127178, 2024.
- [40] M Sudhakar and RM Priya. Computer vision based machine learning and deep learning approaches for identification of nutrient deficiency in crops: A survey. *Nature Environment and Pollution Technology*, 22(3), 2023.
- [41] Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *ECCV Workshops*, pages 443–450. Springer, 2016.
- [42] Baochen Sun, Jiashi Feng, and Kate Saenko. Return of frustratingly easy domain adaptation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30, 2016.
- [43] Md Simul Hasan Talukder and Ajay Krishno Sarkar. Nutrients deficiency diagnosis of rice crop by weighted average ensemble learning. *Smart Agricultural Technology*, 4: 100155, 2023.
- [44] Muhammad Hassan Tanveer, Zainab Fatima, Shehnila Zardari, and David Guerra-Zubiaga. An in-depth analysis of domain adaptation in computer and robotic vision. *Applied Sciences*, 13(23):12823, 2023.
- [45] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30, 2017.
- [46] Trung-Tin Tran, Jae-Won Choi, Thien-Tu Huynh Le, and Jong-Wook Kim. A comparative study of deep cnn in forecasting and classifying the macronutrient deficiencies on development of tomato plant. *Applied Sciences*, 9(8):1601, 2019.
- [47] Yi-Hsuan Tsai, Wei-Chih Hung, Samuel Schuler, Kihyuk Sohn, Ming-Hsuan Yang, and Manmohan Chandraker. Learning to adapt structured output space for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7472–7481, 2018.
- [48] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7167–7176, 2017.
- [49] Nguyen Xuan Vinh, Julien Epps, and James Bailey. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research*, 11:2837–2854, 2010.
- [50] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Pérez. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2517–2526, 2019.
- [51] Mei Wang and Weihong Deng. Deep visual domain adaptation: A survey. *Neurocomputing*, 312:135–153, 2018.

- [52] Juhao Wu, Omar Wyman, Yadong Tang, Damiano Pasini, and Wenlong Wang. Multi-view 3d reconstruction based on deep learning: A survey and comparison of methods. *Neurocomputing*, 582:127553, 2024.
- [53] Haifeng Xia, Pu Wang, and Zhengming Ding. Incomplete multi-view domain adaptation via channel enhancement and knowledge transfer. In *European Conference on Computer Vision*, pages 200–217. Springer, 2022.
- [54] Qizhe Xie, Minh-Thang Luong, Eduard Hovy, and Quoc V Le. Self-training with noisy student improves imagenet classification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10687–10698, 2020.
- [55] Ruijia Xu, Guanbin Li, Jihan Yang, and Liang Lin. Larger norm more transferable: An adaptive feature norm approach for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1426–1435, 2019.
- [56] Jinhui Yi, Lukas Krusenbaum, Paula Unger, Hubert Hüging, Sabine J Seidel, Gabriel Schaaf, and Juergen Gall. Deep learning for non-invasive diagnosis of nutrient deficiencies in sugar beet using rgb images. *Sensors*, 20(20):5893, 2020.
- [57] Jinhui Yi, Gina Lopez, Sofia Hadir, Jan Weyler, Lasse Klingbeil, Marion Deichmann, Juergen Gall, and Sabine J Seidel. Non-invasive diagnosis of nutrient deficiencies in winter wheat and winter rye using uav-based rgb images. *Available at SSRN 4549653*, 2024.
- [58] Bowen Zhang, Yidong Wang, Wenxin Hou, Hao Wu, Jindong Wang, Manabu Okumura, and Takahiro Shinozaki. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. *Advances in Neural Information Processing Systems*, 34: 18408–18419, 2021.
- [59] Pan Zhang, Bo Zhang, Ting Zhang, Dong Chen, Yong Wang, and Fang Wen. Prototypical pseudo label denoising and target structure learning for domain adaptive semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12414–12424, 2021.
- [60] Yang Zou, Zhiding Yu, BVK Kumar, and Jinsong Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *Proceedings of the European conference on computer vision*, pages 289–305, 2018.