

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

Jinhui Yi*¹

jinhui.yi@uni-bonn.de

Yanan Luo*¹

ylo@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

1 The MiPlo Datasets

The Mini Plot (MiPlo) datasets are large-scale RGB image datasets consisting of high-resolution images with multiple types of nutrient treatments annotated by agronomy experts. It consists of the Mini Plot Barley (MiPlo-B) dataset from Deichmann *et al.* [2] and the newly collected Mini Plot Winter Wheat (MiPlo-WW) dataset.

Experimental Setup The MiPlot experiments enable controlled trials of nutrient deficiencies under simulated field conditions. The crops (two genotypes) were grown in mineral soil in containers (“Big Box”, L x B x H: 111 x 71 x 61 cm, vol.: 535L) and sown in rows in a density according to agricultural practice. The soil was a mixture of nutrient-poor loess from a 5 meter depth of an opencast pit mine and quartz sand 0.5 – 1 mm. To expose the plants to environmental factors, *e.g.*, wind, radiation, and precipitation, the containers were positioned in an outdoor area and equipped with a fertigation system of porous drip pipes to allow additional water supply and individual fertilization with nutrient solutions per container. To transfer soil microorganisms to the experiments, the containers were inoculated with soil slurry from the non-fertilized plot of a long-term fertilizer-field experiment. For each genotype, the containers were placed in three rows of ten containers each on a leveled concrete platform. The 30 containers were divided into seven treatments (ctrl, -N, -P, -K, -B, -S, unfertilized) with four replicates each, as well as two additional containers for invasive investigations, in a randomized block design. In this work, 24 containers with six nutrient

treatments (ctrl, -N, -P, -K, -B, -S) were selected for evaluation, because the containers with unfertilized treatment showed distinct patterns, *i.e.*, only pipes.

Image Acquisition Protocol The RGB images in the MiPlo datasets were taken from 24 containers, each of which was subjected to a type of nutrient treatment. All of the images with the size of 7296×5472 were captured by a Huawei P20 Pro smartphone with a triple camera from Leica under natural lighting conditions. Specifically, the images were taken under different conditions in terms of height, viewpoint, light, and weather to reflect realistic conditions. As a result, crops within each container have been captured over the growing season multiple times and each time from multiple views (20 views on average). Example images are shown in Figure 2. The images were annotated by the date, genotype, and six nutrient treatments (ctrl, -N, -P, -K, -S, -B), where “-” stands for the omission of the corresponding nutrient (N: nitrogen, P: phosphorous, K: potassium, B: Boron, S:Sulfur). Plants in the control group ‘ctrl’ do not suffer from nutrient deficiencies.

Statistics The statistics of the MiPlo-B dataset [1] and our proposed MiPlo-WW dataset are presented in Tables 1, 2, and 3. **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 \approx 75%:25%). **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 dates). It contains two genotypes: Julius (6253 images) and Meister (6213 images). The ID settings are the same as above. Although most annotations have a similar amount of images, there is a small imbalance of the sample distribution among different dates.

Table 1: Statistics of the MiPlo datasets.

Dataset	#Images (k)	#Class	Dates	#Views	Year
MiPlo-B (Barley) [1]	18.6	6	16	20	2022
MiPlo-WW (Winter Wheat)	12.5	6	13	20	2023

2 Experimental Details

If not specifically pointed out, the default setting of our model adopts the loss term L_s^{sa2} , L_t^{sa2} and L_t^{sa1} , five subsampled views as well as a threshold of 0.8 for hard pseudo-labels for the ablation studies. And the default hyper-parameters in our experiments are as follows: The original image was resized to 1344×1344 and normalized with a mean value of [0.485, 0.456, 0.406] and a standard deviation of [0.229, 0.224, 0.225] calculated from ImageNet. For weak augmentation, we apply random horizontal flip with a probability of 50%. Following previous work [1], we adopted RandAugment, a variant of AutoAugment that does

Table 2: The number of images in the Mini Plot Barley (MiPlo-B) dataset with two genotypes: Barke and Hanna. **06/21** denotes 21 June 2022, where 2022 is omitted for simplification. “-” stands for the omission of the corresponding nutrient (N: nitrogen, P: phosphorous, K: potassium, B: Boron, S:Sulfur).

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Total
class/date	06/21	06/22	06/23	06/24	06/27	06/29	06/30	07/01	07/04	07/05	07/06	07/07	07/08	07/11	07/19	07/20	
Barke																	
-N	80	84	80	80	161	160	78	80	80	80	156	80	80	168	60	99	1606
-P	80	60	81	79	160	160	60	83	60	80	134	80	80	124	60	100	1481
-K	80	80	80	80	163	158	80	80	60	80	152	80	80	148	39	99	1539
-B	81	80	81	80	164	140	83	80	80	81	137	80	80	130	58	100	1535
-S	80	82	80	80	161	160	83	80	80	80	135	80	80	138	59	100	1558
ctrl	80	80	80	80	160	160	82	80	81	80	154	82	80	148	59	100	1586
total	481	466	482	479	969	938	466	483	441	481	868	482	480	856	335	598	9305
Hanna																	
-N	80	80	80	80	160	140	82	79	80	80	154	80	80	144	59	99	1557
-P	80	80	80	80	148	161	60	80	61	80	124	80	80	128	59	99	1481
-K	80	80	80	80	160	160	81	80	60	80	152	80	80	154	39	100	1547
-B	83	82	81	80	161	140	81	81	80	81	137	80	80	134	59	101	1539
-S	80	80	81	80	162	162	81	80	81	80	134	80	80	144	60	100	1565
ctrl	80	80	80	82	162	158	84	80	80	80	154	80	80	145	40	100	1565
total	483	482	482	482	953	921	469	480	442	481	855	480	480	849	316	599	9254

Table 3: The number of images in the Mini Plot Winter Wheat (MiPlo-WW) dataset with two genotypes: Julius and Meister. **05/12** denotes 12 May 2023, where 2023 is omitted for simplification. “-” stands for the omission of the corresponding nutrient (N: nitrogen, P: phosphorous, K: potassium, B: Boron, S:Sulfur).

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	Total
class/date	05/12	05/13	05/14	05/15	05/16	05/17	05/18	05/19	05/20	05/21	05/22	05/23	05/24	
Julius														
-N	80	20	145	80	75	80	80	80	81	80	59	78	80	1018
-P	80	61	79	80	80	82	83	81	80	80	81	81	83	1031
-K	81	20	142	81	82	84	80	80	81	80	78	78	85	1052
-B	62	40	140	79	81	87	83	81	81	80	83	82	80	1059
-S	81	40	117	81	80	70	80	80	85	80	101	82	80	1057
ctrl	81	20	143	80	83	62	84	80	80	80	81	81	81	1036
total	465	201	766	481	481	465	490	482	488	480	483	482	489	6253
Meister														
-N	80	20	140	78	80	80	82	83	86	81	77	79	73	1039
-P	80	60	78	80	83	80	81	86	80	60	81	71	81	1001
-K	80	20	144	87	86	82	83	81	81	80	80	75	80	1059
-B	78	40	150	81	81	82	83	80	80	80	81	78	74	1068
-S	81	40	120	80	80	82	83	80	81	80	78	79	75	1039
ctrl	80	22	122	82	83	63	79	83	80	80	80	78	75	1007
total	479	202	754	488	493	469	491	493	488	461	477	460	458	6213

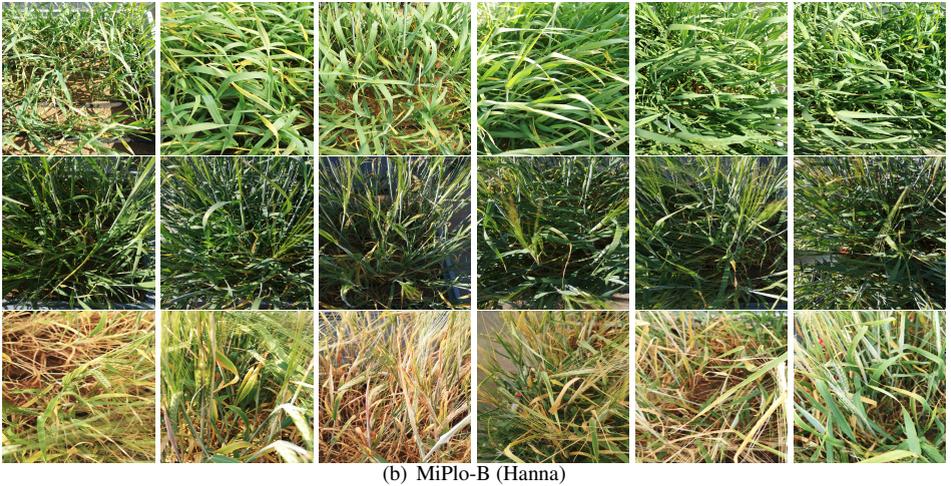
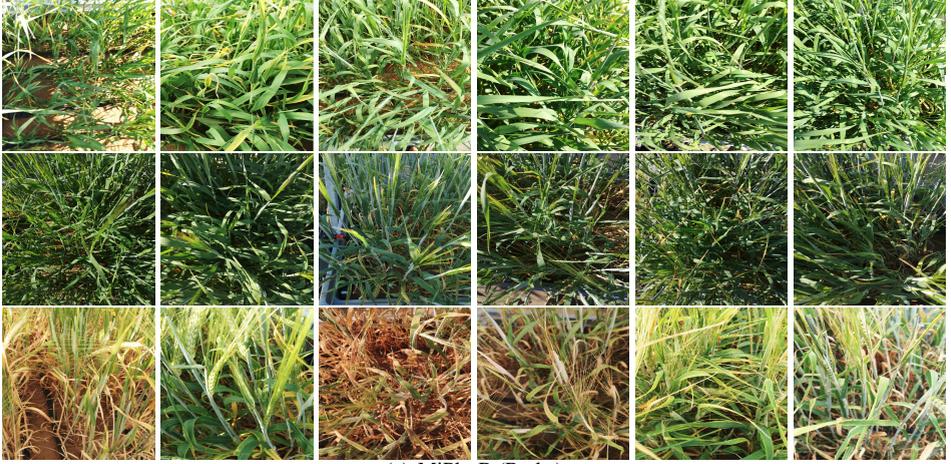


Figure 1: Example images. Columns 1-6: -N, -P, -K, -B, -S, ctrl; row 1-3: 21 June 2022, 04 July 2022, and 20 July 2022.

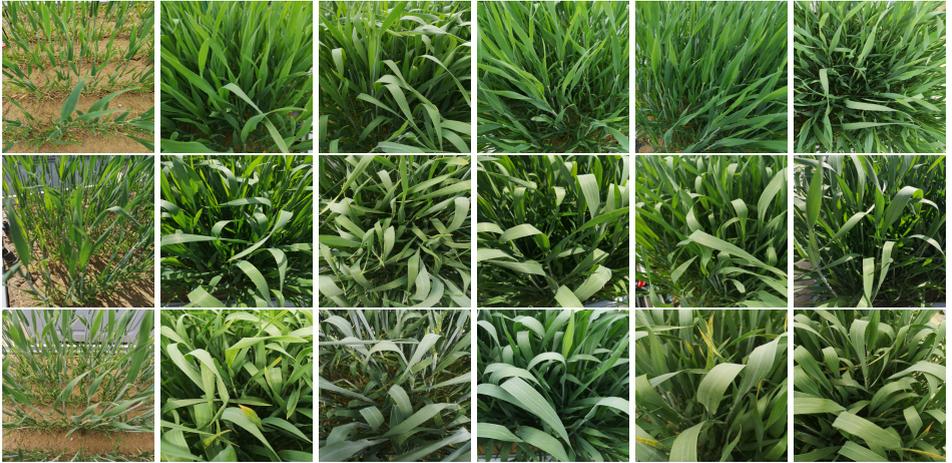


Figure 2: Example images. Column 1-6: -N, -P, -K, -B, -S, ctrl; row 1-3: 12 May 2023, 18 May 2023, and 24 May 2023.

not need to pre-train the augmentation strategy with labeled data, for strong augmentation. We used ResNet-50 as backbone that was pre-trained on ImageNet. We then trained each model for 20 epochs with a batch of four samples from the source domain and four samples from the target domain at each iteration. We used stochastic gradient descent (SGD) with an initial learning rate of 3×10^{-3} , where the momentum and weight decay were set as 0.9 and 10^{-3} , respectively. The learning rate was reduced with schedule $lr_p = \frac{lr_0}{(1+\alpha p)^\beta}$ where p is the training progress linearly changing from 0 to 1, lr_0 is the initial learning rate, $\alpha = 8$ and $\beta = 0.75$ is the decay factor. All of the experiments were conducted with a single NVIDIA RTX A6000 with 48GB VRAM. For evaluation, we report the top-1 accuracy metric on the test set of the target domain, which denotes whether the predicted category with the highest confidence matches the ground truth category.

3 Additional Ablation Study

3.1 Number of Views

To explore how many views are necessary for adaptation in the detection of plant nutrient deficiency, we sample a subset of related views by computing the similarity of each query-view pair given a query image and select the top n related views that are most dissimilar to the query image. To validate the effectiveness of our proposed SgVM mechanism, we also report the results by randomly sampling the subset. The results in Table 4 indicate that the model with a subset of five related views performs the best, and including more views does not improve the performance but might provide noisy signals due to redundant information. Notably, model sampling with our proposed SgVM mechanism consistently outperforms its counterpart with random sampling.

#Views	SgVM	Random
1	65.1	62.5
5	69.6	65.3
10	67.4	64.7
20	66.8	63.3
40	69.1	64.8

Table 4: Ablation study on the number of views and sampling methods for the Barley: B \rightarrow H benchmark. *#Views* indicates the number of related views given a query image, *SgVM* denotes our proposed Similarity-guided View Mining mechanism, and *Random* refers to random sampling while constructing the set of views for each query image.

τ	B \rightarrow H	B \rightarrow M
0.3	66.0	34.0
0.5	63.0	31.2
0.8	69.6	33.7
0.9	64.4	31.2
soft	67.4	45.6

Table 5: Ablation study on threshold τ and types of pseudo-labels. B \rightarrow H indicates smaller domain gap (cross genotypes), while B \rightarrow M indicates larger domain gap (cross cultivars). *soft* refers to soft pseudo-labels instead of hard pseudo-labels with a pre-defined threshold τ .

3.2 Hard Pseudo-Label vs. Soft Pseudo-Label

To evaluate the impact of the threshold τ for hard pseudo-labels, as well as to compare hard pseudo-labels with soft pseudo-labels, we report the results in Table 5. Comparing B \rightarrow H and B \rightarrow M adaptation, we see soft pseudo-labels work better than hard pseudo-labels

when the domain gap is large ($B \rightarrow M$), *i.e.*, the adaptation across crop species. In this case, increasing the threshold will mask out most of the pseudo-labels since the confidences are in general low, while decreasing the threshold will force the model to learn from noisy pseudo-labels. If the domain gap is smaller ($B \rightarrow H$), the initial confidences are higher and the hard pseudo-labels perform better. The choice of hard and soft pseudo-labels thus depends on the domain gap, but soft pseudo-labels can always be applied.

3.3 Supervision of \mathcal{L}_s^{sa2}

While we compute \mathcal{L}_s^{gt} based on the ground-truth labels of the source images, \mathcal{L}_s^{sa2} is computed based on the pseudo-labels. In Table 6, we compare the results of computing the consistency loss \mathcal{L}_s^{sa2} for source domain images with ground-truth labels or pseudo-labels. The results show that it is better to use the pseudo-labels instead of the ground-truth labels for each view pair of the query source images, which shows that the gain of \mathcal{L}_s^{sa2} is due to measuring the prediction consistency between two views and not simply due to data augmentation.

Supervision	$B \rightarrow H$	$B \rightarrow M$
label	66.8	35.9
pseudo-label	69.6 (hard)	45.6 (soft)

Table 6: Ablation study on the supervision signals of \mathcal{L}_s^{sa2} .

3.4 Different Backbones

To explore the performance of smaller backbones with fewer parameters, as they are commonly used for applications with very limited computational resources, we also report the results for MobileNetV3 (large and small versions) [14] as backbone. The results in Table 7 show that our approach outperforms other baselines consistently with ResNet50 [15] as well as various efficient MobileNetV3 architectures as backbones.

3.5 Confusion Matrices

We finally show confusion matrices before and after adaptation in Figure 3.

References

- [1] 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.
- [2] 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.

Table 7: Top-1 Classification Accuracy (%) for adaptation across genotypes: **Barley: Barke** → **Hanna** with different backbones. *FLOPs* stands for floating point operations. *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.

Model	Barley: Barke → Hanna		
	ResNet50	MobileNetV3Large	MobileNetV3Small
Parameters (M)	25.6	5.5	2.5
FLOPs (G)	4.09	0.22	0.06
Oracle	75.4	70.2	67.7
Source-Only	54.0	53.0	48.0
DANN [8]	54.7	47.0	41.0
ADDA [14]	51.8	39.5	27.8
JAN [8]	51.9	49.4	43.2
CDAN [8]	51.3	46.3	37.1
BSP [14]	58.8	49.3	44.6
AFN [14]	59.0	51.3	45.1
Mean Teacher [8]	60.6	56.2	48.7
FixMatch [8]	64.6	<u>57.5</u>	49.6
FlexMatch [14]	65.9	57.1	<u>50.1</u>
Ours+hard	69.6	59.6	53.1
Ours+soft	<u>67.4</u>	55.4	45.6

- [3] 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.
- [4] 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.
- [5] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1314–1324, 2019.
- [6] 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.
- [7] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. *Advances in neural information processing systems*, 31, 2018.
- [8] 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.



Figure 3: Confusion matrices (x-axis: predicted treatment, y-axis: real treatment) without adaptation (a) and our approach (b) for **Barley: Barke** \rightarrow **Hanna**. The confusion matrices (c,d) are without adaptation and our approach for **Barley** \rightarrow **Winter Wheat (B** \rightarrow **J)**. AVG denotes average accuracy.

- [9] 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.
- [10] 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.
- [11] 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.
- [12] 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.