Abnormality Detection in Maize Fields Using Selective Domain Adaptation–Driven Data Augmentation

Aminul Huq[1], Dimitris Zermas[2], and George Bebis[1]

[1] University of Nevada Reno, Reno NV 89512, USA
[2] Sentera
ahuq@unr.edu,dimitris.zermas@sentera.com,bebis@unr.edu

Abstract. Accurate and timely identification of plant abnormalities is vital in agriculture, since delays can severely reduce both crop health and yield. Automating this task has the potential to generate significant environmental and economic benefits. In this work, we investigate the detection and quantification of abnormalities in maize fields using high-resolution RGB imagery gathered by Unmanned Aerial Vehicles (UAVs). A limitation of many existing methods is that they depend on relatively small datasets that are also tuned to the characteristics of specific fields. As a result, models trained on one field do not perform well on data from another field due to *domain shift*. While data augmentation and synthetic image generation have been used to increase dataset size and diversity, these techniques often fail to provide the data variability and fidelity required for robust agricultural applications. An alternative strategy is to merge data from multiple fields; however, this approach is ineffective without first normalizing the data to account for irrelevant variations such as lighting conditions, sensor characteristics, and soil types. To address this challenge, we propose a framework for standardizing multi-field data by mapping them into a *shared domain* using unsupervised domain adaptation (UDA). To further improve quality, we introduce a *selective* UDA strategy that filters out poorly adapted images prone to artifacts. During training, images captured under diverse conditions are aggregated in this shared domain, enriching the training set with realistically domain-adapted data rather than relying solely on synthetic augmentations. For detection and quantification, we leverage Vision Transformers (ViTs), while an ensemble of CycleGANs is used for domain adaptation. We validate our framework on a publicly available dataset of UAV-based, high-resolution RGB images of both healthy and abnormal maize plants across multiple growth stages, collected from two fields with diverse environmental conditions.

Keywords: Deep Learning · Maize Plants · Abnormality Detection.

1 Introduction

Plant abnormalities may arise from nutrient deficiencies, drought stress, diseases, or pest infestations. Detecting these issues early enables timely, targeted

interventions that protect yield quality and overall productivity. Traditionally, detection has relied on manual field inspections, where farmers examine rows and leaves. This approach is labor-intensive, prone to error, and impractical for large-scale farming, as inspections must be repeated throughout the growth cycle. Such limitations often lead to misapplication or overuse of fertilizers and pesticides, creating environmental risks. In this study, we focus on detecting and quantifying abnormalities in maize plants, one of the most critical staple crops in the US, with an estimated 15.2 billion bushels harvested in 2023 [1]. UAVs and robotic platforms integrated with DL provide a scalable solution for crop monitoring [2] [3]. While UAVs enable efficient large-scale data acquisition, several challenges remain. First, DL models require large annotated datasets, which are often limited. Transfer learning with fine-tuning [4] and data augmentation [5] partially mitigate this issue. Generating synthetic data which can represent various realistic field conditions along with various growth stages of the crops is also challenging and computationally expensive. Second, DL models trained in one field often suffer from *domain shift*, leading to degraded performance when applied to new fields or even to the same field under changing conditions. Factors such as soil composition, crop stage, illumination, and sensor variation contribute to this challenge, limiting broad adoption.

The main contribution of this work is the development of an effective framework to integrate data from different fields to increase the diversity of the training while removing irrelevant information in the data due to variations in lighting conditions, sensor characteristics, and soil types. To achieve this, we employ CycleGANs [6] to transform images from diverse fields into a shared domain, producing consistent representations. Unlike standard augmentation and synthetic data generation approaches, our approach does not rely solely on synthetic data, though synthetic samples can be easily integrated into the framework. To avoid performance degradation from low-quality domain-adapted images, we introduce a *selective* UDA strategy that filters poorly adapted samples. For quantification, we apply a Vision Transformer (ViT) regressor [7], which can also perform abnormality detection by thresholding regressor outputs. The proposed method provides farmers with actionable insights by highlighting regions needing attention and assists annotators by narrowing focus to specific field areas, improving both efficiency and accuracy [8].

2 Background

Detecting nutrient deficiencies and abnormalities caused by drought, disease, or pests remains a major challenge in agriculture. Traditional methods are invasive, requiring repeated sampling of leaves, foliage, or soil for chemical analysis. By contrast, high-resolution RGB cameras provide a cost-effective alternative capable of capturing detailed visual information.

Recent methods focus on data augmentation through Real-to-Real Unsupervised Domain Adaptation (RR-UDA) to address domain shifts between source and target datasets. Early work using contrast stretching [9], [10] achieved lim-

ited success. Magistri et al. [11] proposed an RR-UDA strategy for crop-weed-soil segmentation under varied crops and conditions, using labeled source data and unlabeled target images. Similarly to Fei et al. [12], the goal was to produce domain-adapted images resembling the target while preserving source semantics however unlike [12], their model jointly learned image translation and target-domain segmentation from source labels. A modified CUT framework generated adapted images and masks, while a new IoU-based loss enforced semantic alignment. The adapted images with source labels then trained a CNN segmentation model. Validation with UAV and ground imagery across multiple crops showed strong results on large datasets. Notably, Fréchet Inception Distance (FID), often used for translation quality, did not correlate with segmentation gains.

Other studies have maintained semantic consistency through different strategies. Gogoll et al. [13] trained a CycleGAN for adaptation alongside a supervised segmentation model, while a second, frozen segmentation network enforced cross-domain consistency using IoU-based loss terms, as proposed by Magistri et al. [11] and inspired by Chen et al. [14]. Building on this, Bertoglio et al. [15] introduced a Fourier phase loss, further improving weed segmentation. Beyond segmentation, RR-UDA has also been applied to leaf counting [16], fruit counting [17], and potato defect detection [18]. In this work, we adopt RR-UDA to mitigate performance degradation caused by domain shifts. Our approach leverages domain adaptation as a form of data augmentation during training. Unlike synthetic-only augmentation, it incorporates real datasets capturing diverse growth stages, abnormalities, and environmental conditions. All data are first mapped into a shared domain via RR-UDA, producing a unified representation. As new data become available, they can be mapped into this shared domain and seamlessly integrated into training, thereby enhancing both dataset size and diversity.

3 Methodology

Using UAV-captured high-resolution RGB imagery of maize fields, our objective is to detect abnormal plants and quantify overall abnormality. We employ a ViT regressor [7], denoted ViT_{rg}, applied to non-overlapping patches via a sliding window. A $W \times W$ window moves across the image with stride W, and ViT_{rg} predicts the fraction of abnormal pixels per patch (see Section 4). The image-level Abnormality Probability (AP) is obtained by summing patch scores and dividing by the total pixel count. An AP above zero (or a set threshold) indicates abnormalities, allowing ViT_{rg} to act as both detector and quantifier. We also evaluated ResNet, EfficientNet, and DenseNet backbones, but ViT consistently outperformed them. To accelerate sliding-window inference, we replaced the ViT's final fully connected layer with a convolutional layer [19].

The reliability of AP depends on patch-level accuracy of ViT_{rg}. Training uses randomly sampled $W \times W$ patches with augmentation via domain adaptation. Training on a single field often fails to generalize due to dataset bias, whereas including multiple fields improves performance on both familiar and unseen data.

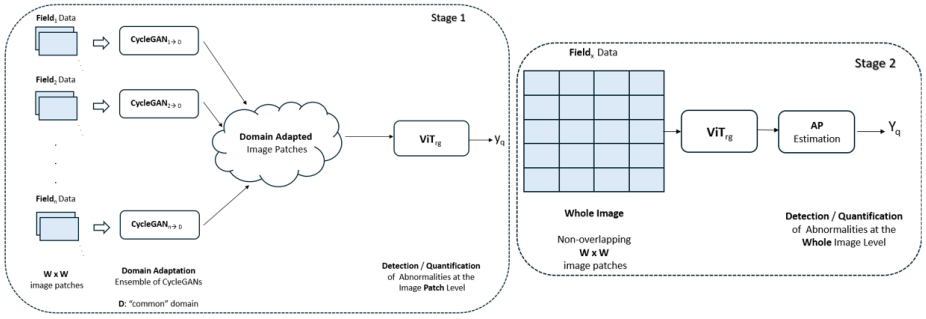


Fig. 1: Overview of the proposed methodology. In Stage 1, image patches from different fields are transformed into a shared domain D and used to train the ViT_{rg} model for patch-level abnormality detection and quantification (y_q). In Stage 2, non-overlapping patches from test images are passed through the ViT_{rg} model and the AP value is estimated at the whole-image level (Y_q).

Simple merging of datasets is ineffective, as irrelevant variations in illumination, soil, or background introduce spurious signals. To address this, we transform images from different field conditions into a shared domain D before aggregation. Data from n fields are mapped into D with an ensemble of CycleGANs [6], one trained per field. Using a single CycleGAN model for all fields is also possible as discussed in Section 6. Combining all adapted datasets produces a unified training set in D, enabling ViT_{rg} to learn realistic field-specific variations beyond synthetic augmentation. Figure 1 (Stage 1) illustrates this pipeline, where y_q denotes patch-level predictions. New datasets can be adapted using an existing or newly trained CycleGAN. The ensemble helps mitigate imbalance, though a single CycleGAN with imbalance-aware training (e.g., [20]) could also be used. At test time (Figure 1, Stage 2), all non-overlapping patched images are passed through the ViT_{rg} model for AP estimation, where Y_q represents whole-image predictions. Since poor adaptations degrade performance, we apply selective domain adaptation. Low-quality images are detected using Peak Signal-to-Noise Ratio (PSNR) between original and adapted versions. By comparing PSNR values to the CycleGAN's loss distribution, outliers are removed from the training stage. Further examples and details appear in Section 5. In the following section, we also define reconstruction loss as the combination of cycle-consistency and identity losses in the CycleGAN model.

4 Dataset

We conducted experiments on a public dataset of UAV-captured, high-resolution RGB images of maize plants across vegetative stages V5, V6, V8, V10, and V12, containing abnormalities from nutrient deficiencies (e.g., nitrogen) and drought stress [8]. Data were collected in two different fields: Becker and Waseca, MN.

Table 1: Number of randomly extracted image patches in the training and test datasets for the Becker and Waseca fields.

	Growth Stage	Training Set		Testing Set	
		Normal	Abnormal	Normal	Abnormal
Becker	V8	1124	1200	238	247
	V12	2401	3766	909	967
	Total	3525	4966	1147	1214
Waseca	V5	660	668	49	45
	V6	1302	1355	125	110
	V10	1815	1815	211	209
	Total	3777	3838	385	364

Plants in Becker (V8, V12) field showed more severe abnormalities, denser foliage obscuring the background, and distinct soil color compared to Waseca (V5, V6, V10), where abnormalities were milder. Image resolutions differed by stage: 4000×6000 (V6, V8, V10), 6000×4000 (V12), and 4000×3000 (V5). The images were captured using Sentera 65R payload, which is capable of capturing 65MP images with a ground sampling distance (GSD) of 1.5mm at 90 feet altitude, and 0.45cm GSD at the FAA-allowed maximum altitude of 400 feet. This high-resolution capability enables the capture of clear, detailed, and motion-blur-free imagery even at operationally safe altitudes. The dataset lacked abnormality labels so we manually annotated abnormal leaves in Label Studio [21] using bounding boxes—fast to create but limited for fine-grained localization and precise quantification.

The dataset includes 61, 64, 44, 64, and 48 images for V5–V12, totaling 281 images, with 226 used for training and 55 for testing (11, 13, 8, 13, and 10 per stage). Random 250×250 patches were extracted from both sets, a size chosen through preliminary trials. Patches overlapping bounding boxes were labeled "abnormal," otherwise "normal". For regression, the abnormality ratio was computed as the bounding-box area over the patch area, considering only overlaps at boundaries. These labeled patches formed the training and testing data for the ViT_{rg} patch-level model. Table 1 details the distribution of normal and abnormal patches by field and stage, while Figure 2 shows examples. As described in Section 3, the trained ViT_{rg} patch model was extended to whole images using a non-overlapping sliding window to compute AP values. Since a full image covers 110–130 plants, labeling entire images as "normal" or "abnormal" might not be useful. Thus, each of the 55 test images was divided into four quadrants (30–40 plants each), effectively quadrupling the test set size and provides attention to a smaller region.

5 Experimental Results and Analysis

We conducted a comprehensive set of experiments to evaluate the effectiveness of the proposed framework. As a starting point, baseline experiments were per-

(a) (b)

Fig. 2: Randomly selected image patches representing the two classes: (a) normal and (b) abnormal. Abnormal regions are indicated by red boxes. The first row contains patches from Waseca, while the second row contains patches from Becker.

formed to assess the performance of the ViT_{rg} model on randomly extracted image patches. To enhance the generalization capability of ViT_{rg}, we then explored domain adaptation through ensembles of CycleGAN models. By visualizing and analyzing the outputs of individual CycleGANs, we obtained key insights into the relative difficulty of different domain transformations and devised a strategy for discarding poorly adapted images, a process we refer to as selective adaptation. Subsequently, domain adaptation was applied as a data augmentation technique to further improve the performance of the ViT_{rg} patch-level model. The final ViT_{rg} patch-level model was then employed to compute AP values for abnormality detection and quantification at the whole-image level. Across all experiments, each ViT_{rg} patch-level model was trained for 200 epochs using the Mean Squared Error (MSE) loss function, while each CycleGAN model was trained for 400 epochs. The dataset used for these experiments is described in detail in Section 4.

5.1 Domain Adaptation in Maize Fields

In this section, we report domain adaptation experiments between the Becker and Waseca maize fields using CycleGANs, aimed at extracting meaningful insights. Each CycleGAN was trained on image patches sampled from the corresponding classes and fields. A central focus was how the choice of the shared domain D affects performance. In principle, any field could serve as D, though some may yield better results. To study this, we compared the difficulty of converting patches from Becker to Waseca (B2W, with $D = W$) and from Waseca to Becker (W2B, with $D = B$). For each direction, a dedicated CycleGAN was trained and its reconstruction losses analyzed. Figure 3 shows histograms of reconstruction losses for normal and abnormal classes under both transformations.

Table 2 summarizes skewed Gaussian fits to these histograms. For the normal class, distributions were highly similar between B2W and W2B, suggesting com-

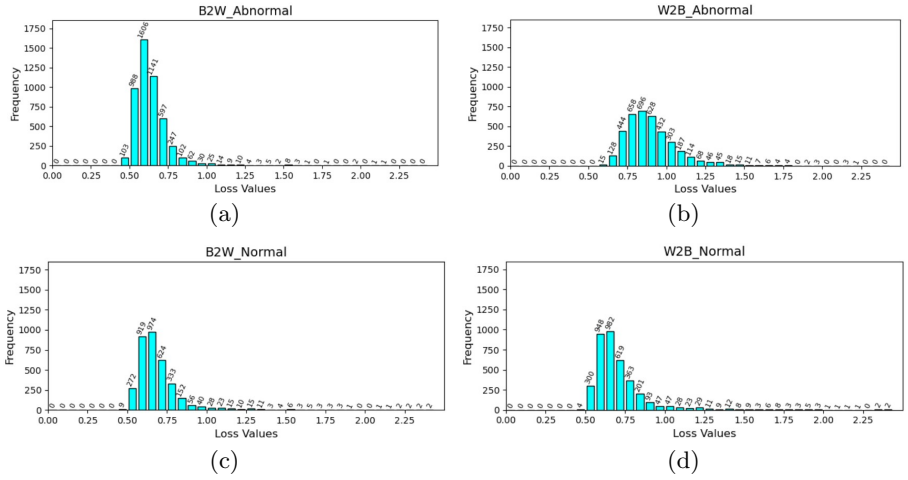


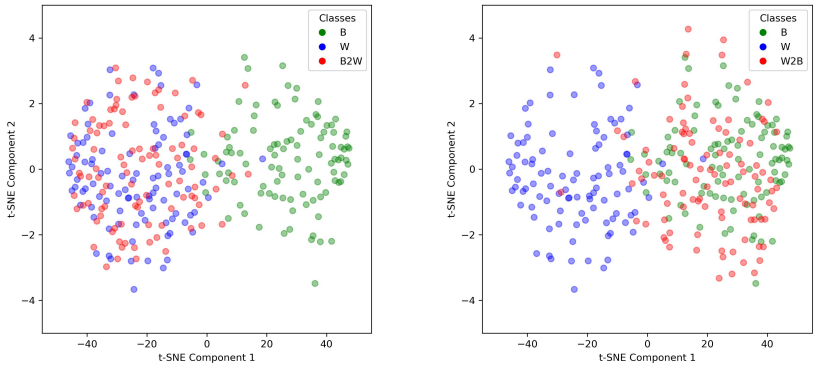
Fig. 3: CycleGAN reconstruction loss histograms. Results are shown for the abnormal class under (a) B2W and (b) W2B, and for the normal class under (c) B2W and (d) W2B.

Table 2: Computed statistics of reconstruction losses for B2W and W2B domain adaptations, shown separately for normal and abnormal classes.

	B2W Abnormal	W2B Abnormal	B2W Normal	W2B Normal
Mean	0.64	0.90	0.69	0.71
Median	0.59	0.84	0.65	0.65
Std	0.12	0.17	0.18	0.20
Skewness	**4.02**	1.79	4.77	4.75
Kurtosis	**31.99**	6.32	35.21	38.87

parable difficulty. For the abnormal class, however, clear differences emerged: skewness and kurtosis for B2W were substantially higher than for W2B. These differences may indicate varying levels of adaptation difficulty, though further validation is needed. The greater challenge for the abnormal class likely stems from its higher variability, since abnormalities in Waseca differ considerably from those in Becker (see Figure 2). Because harder transformations produce more poorly adapted images (see Figure 5), choosing an optimal shared domain D may involve prioritizing easier transformations. Additional discussion is provided in Section 5.3.

We also visualized how domain-adapted images distribute within D. Using t-SNE [22], we compared original source and target images with their adapted versions. Figure 4 presents results for B2W and W2B, based on 300 randomly selected patches (half from each field). Adapted images generally clustered well within the shared domain, though some remained near the source domain. We



(a) (b)

Fig. 4: t-SNE plots of original and domain-adapted patches with (a) $D = W$ and (b) $D = B$. In (a), the B2W-adapted patches (red) cluster more closely with Waseca patches (blue) than with the original Becker patches (green). In (b), the W2B-adapted patches (red) align more closely with Becker patches (green) compared to the original Waseca patches (blue).

attribute this to inherent similarities between Becker and Waseca imagery, which likely led the CycleGAN to generate adapted patches still close to the source distribution.

5.2 Selective Domain Adaptation

Since our framework uses domain adaptation for data augmentation (Stage 1), it is critical to identify and exclude poorly adapted images from the training sets of the detection and quantification models. Among several metrics evaluated, PSNR proved most effective, as it captures both similarity and image quality. High PSNR values indicate domain-adapted images that closely match the target distribution with minimal distortion, whereas low values reflect strong deviations or artifacts. Formally, let N_S denote the number of source domain images I_i^S, $i = 1, 2, .., N_S$, and N_T the number of target domain images V_j, $j = 1, 2, .., N_T$. For a given source image I_i^S and its domain-adapted version I_i^T, the PSNR score with respect to a target image V_j is defined as:

$$\text{PSNR}_{i,j} = c \cdot \log_{10} \left(\frac{L}{\sqrt{\text{MSE}(I_i^T, V_j)}} \right) \tag{1}$$

Here, L is the maximum pixel value ($L = 1$ in our experiments, as images were normalized to [0,1]), c is a scaling constant ($c = 20$), and MSE is the Mean Squared Error between I_i^T and V_j. For each domain-adapted image I_i^T, we compute $\text{PSNR}_{i,j}$ against all target images V_j, $j = 1, 2, .., N_T$, and calculate the

Table 3: Performance evaluation of selective versus non-selective filtering of poorly adapted image patches. In the selective setting, the ViT_{rg} model is trained on original (last row) and domain-adapted patches after excluding those identified as poorly adapted. In the non-selective setting, training uses both original and all domain-adapted patches without filtering. Quantification results are reported using MSE.

	Trained on Becker+W2B, Tested on Becker		Trained on Waseca+B2W, Tested on Waseca	
	Selective	Non-Selective	Selective	Non-Selective
Wasserstein	8.07	8.17	2.02	2.06
LPIPS	8.07		1.96	
PSNR	8.00		1.97	

(a) (b)

Fig. 5: Examples of poorly adapted image patches identified using PSNR for (a) B2W and (b) W2B transformations. Original patches are shown in the top row, with their domain-adapted versions displayed below.

average score μ_i. We then compute the global average PSNR μ and standard deviation σ across all $N_S \times N_T$ scores. An image I_i^T is flagged as poorly adapted if $\mu_i < \mu - c \cdot \sigma$. Since higher PSNR values correspond to better translation fidelity, this rule identifies images with excessive artifacts or distortions. In our experiments, $c = 1$. We term this filtering process *selective domain adaptation*, which enhances performance as shown in Section 5.3. Figure 5 presents representative discarded examples, including hallucinated green leaves, leaf surfaces resembling soil, and cases where abnormalities disappeared entirely.

To evaluate selective versus non-selective adaptation, we trained a ViT_{rg} patch-level model under two setups: (1) Becker + W2B patches and (2) Waseca + B2W patches. Quantification performance was measured using MSE. We also compared PSNR with other similarity metrics, including Wasserstein distance and Learned Perceptual Image Patch Similarity (LPIPS). As shown in Table 3, selective adaptation consistently outperformed non-selective across both fields. PSNR proved most effective, filtering approximately 7.2% of low-quality W2B and 9.5% of B2W images.

Table 4: Same-field and cross-field patch-level quantification performance. Results are reported for models trained with original data and with selective domain-adapted data used for augmentation.

Trained On	Train MSE	Valid MSE	Test MSE/NMSE (Becker)	Test MSE/NMSE (Waseca)
Becker	7.80	8.28	9.84 / 0.55	2.58 / 0.71
Becker+W2B	6.68	5.87	**8.00 / 0.54**	2.13 / 0.87
Waseca	11.08	10.49	17.46 / 0.97	8.32 / 1.52
Waseca+B2W	11.11	9.80	9.01 / 0.50	**1.97 / 0.57**
Waseca+Becker	6.27	5.93	8.21 / 0.57	3.12 / 0.90

5.3 Data Augmentation Using Selective Domain Adaptation

In this section, we report additional experiments designed to evaluate the effectiveness of selective domain adaptation for data augmentation. To ensure fairness, the same source-domain test sets were used across all experiments. Furthermore, to account for differences in abnormality levels across fields, we also report Normalized MSE (NMSE) scores, which is defined as the MSE divided by the variance of the ground-truth values.

The results are presented in Table 4. As a baseline, we trained separate ViT_{rg} models for each field using only their respective patch-level images. Using MSE and NMSE as evaluation metrics, same-field performance on Becker was 9.84/0.55; the corresponding performance on Waseca was 8.32/1.52. The corresponding cross-field performances were 2.58/0.71 and 17.46/0.97. We then applied selective domain adaptation for data augmentation in both fields. For Becker, the training set was expanded to include both original Becker data and selectively adapted Waseca images (i.e., W2B). With this augmentation, the ViT_{rg} patch-level model achieved improved scores of 8.00/0.54. For Waseca, performance gains were even more pronounced, with MSE/NMSE dropping to 1.97/0.57. Cross-field performance also showed notable improvements under selective domain adaptation. For comparison purposes, we also trained a model where data augmentation was performed by simply merging the Becker and Waseca data, without first bringing them into a shared domain. In this setting, the model achieved test performance of 8.21/0.57 on the Becker field and 3.12/0.90 on the Waseca field. While the performance on the Becker field improved compared to training only on Becker data, it still fell short of the results obtained with domain-adapted data. For the Waseca field, this approach resulted in substantially worse performance.

5.4 Abnormality Detection and Quantification in Whole Images

In this section, we quantify abnormalities in whole images using the methodology described in Section 3 and Figure 1, with test image details given in Section 4. For each test image, abnormalities are estimated through AP values predicted

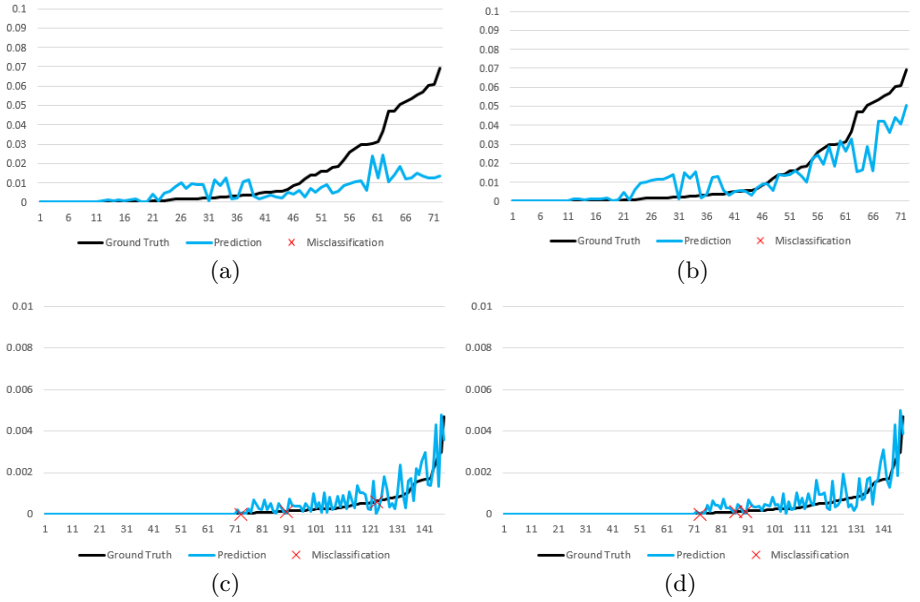


Fig. 6: Same-field evaluation of AP-based predictions. Results are shown for: (a) Becker without data augmentation, (b) Becker with Becker+W2B augmentation, (c) Waseca without data augmentation, and (d) Waseca with Waseca+B2W augmentation. All AP values are scaled between 0 and 100.

Table 5: Performance of AP-based whole-image detection and quantification in same-field and cross-field settings, evaluated on the original test images.

	Test Acc(%) (Becker)	Test MSE/NMSE (Becker)	Test Acc(%) (Waseca)	Test MSE/NMSE (Waseca)
Becker	100%	0.0001 / 0.0074	93.91%	1.9e-07 / 0.0002
Becker+W2B	100%	**0.0001 / 0.0073**	94.59%	1.2e-07 / 0.0001
Waseca	98.61%	0.0003 / 0.015	97.97%	1.8e-07 / 0.0002
Waseca+B2W	98.61%	9.9e-05 / 0.005	97.97%	**1.6e-07 / 0.0002**
Waseca+Becker	96.43%	0.0009 / 0.05	91.28%	1.4e-05 / 0.02

by the ViT_{rg} patch-level model. To aid interpretation, images are ordered by ground truth AP values, ranging from low (left) to high (right). Figures 6 (a) and (c) present predicted (blue) versus ground truth (black) AP values for Becker and Waseca under same-field training/testing without augmentation. The model generally overestimates AP at low abnormality levels and underestimates at high levels. Augmentation improves predictions in both fields, as shown in Figures 6 (b) and (d). Because Waseca abnormalities are less severe, its predictions appear noisier. Quantitative results in Table 5 show NMSE values are far lower in Waseca (0.0002/0.0002) than in Becker (0.0074/0.0073) for same-field evaluation.

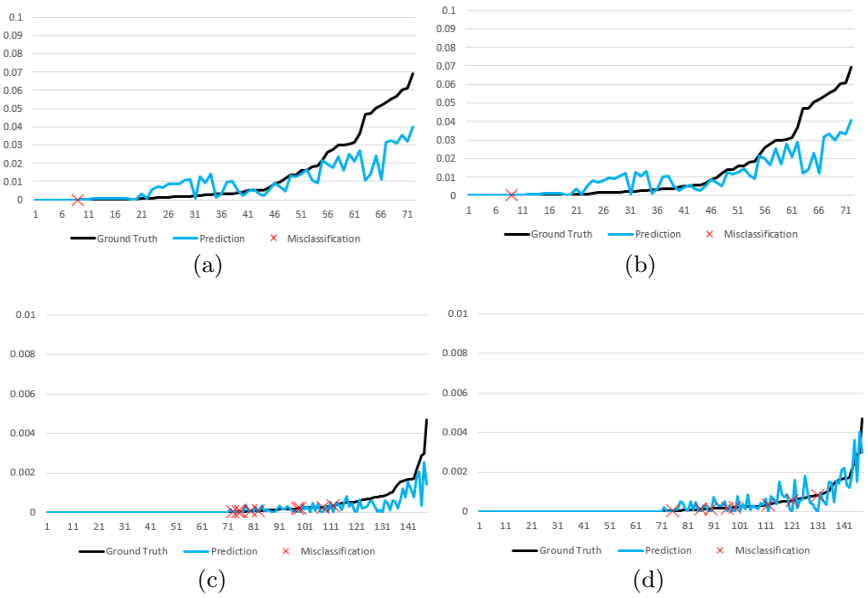


Fig. 7: Cross-field evaluation of AP-based whole-image predictions. Results are shown for: (a) trained on Waseca without augmentation and tested on Becker, (b) trained on Waseca with Waseca+B2W augmentation and tested on Becker, (c) trained on Becker without augmentation and tested on Waseca, and (d) trained on Becker with Becker+W2B augmentation and tested on Waseca. All AP values are scaled between 0 and 100.

Cross-field results are shown in Figures 7 (a) and (c). As in same-field tests, AP values tend to be overestimated for low-abnormality and underestimated for high-abnormality images. Augmentation again improves predictions, supported by Figures 7 (b) and (d) and the MSE/NMSE values in Table 5. Waseca predictions remain noisier, largely due to the different visualization scale required for its lower abnormality levels. Notably, NMSE dropped by more than 50% (0.015 $\rightarrow$ 0.005) when training on Waseca and testing on Becker.

Beyond quantification, AP values also function as abnormality detectors, since non-zero values imply abnormalities. To test this, we used the smallest AP value among abnormal training samples as a threshold T_q: a test image was classified abnormal if its AP exceeded T_q, normal otherwise. Misclassified images are marked with red crosses in Figures 6 and 7. In same-field experiments, augmentation did not affect detection accuracy. In cross-field settings, accuracy improved from 93.91% to 94.59% when training on Becker and testing on Waseca. With training on Waseca and testing on Becker, accuracy was unchanged, though NMSE again decreased by more than 50%, as noted earlier. For comparison, we evaluated the patch-level ViT_{rg} model trained directly on both Becker and Waseca field data without applying domain adaptation to bring

them into a shared domain. When assessed on whole images, the model exhibited poor performance on both quantification and detection tasks across both fields.

6 Conclusions

In this study, we proposed a deep learning framework for detecting and quantifying maize field abnormalities from UAV imagery. To address limited training data, we proposed transforming images from different fields into a shared domain in order to remove irrelevant variations in the data, strengthening training and improving generalization to unseen data. We also applied selective domain adaptation to filter poorly adapted samples, ensuring higher-quality data in training and evaluation. The framework's feasibility was validated through extensive experiments across two fields with diverse conditions. This work has several limitations. First, the absence of a larger multi-field dataset restricted broader evaluation and exploration of new research directions. Still, experiments across varying stages, soils, and lighting demonstrate the practicality of the approach. Second, training separate adaptation models for each domain pair is computationally costly and complicates adapting new data to the shared domain D. A more scalable solution would be a single model handling multiple source domains. Finally, more advanced architectures such as CycleGAN-Turbo [23] and CUT [24] may further enhance adaptation performance. As part of future work, we plan to transform unseen test data from different domains into a shared domain for quantification and detection for the purpose of domain generalization. Moreover, we aim to broaden the applicability of our framework beyond agriculture, particularly in medical imaging, where domain adaptation has the potential to greatly enhance model generalization across diverse imaging modalities and conditions.

Acknowledgment: This work was supported by the National Institute of Food and Agriculture/USDA, Award No. 2020-67021-30754.

References

1. U. S. D. of Agriculture, "Usda forecasts us corn production down, soybean and cotton production up from 2023," *National Agricultural Statistics Service*, 2024.
2. S. Asseng and F. Asche, "Future farms without farmers," *Science Robotics*, vol. 4, 2019.
3. A. Pretto and et al., "Building an aerial–ground robotics system for precision farming," *IEEE Robotics and Automation Magazine*, vol. 28(3), pp. 29–49, 2020.
4. F. Zhuang and et al., "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109(1), 2021.
5. M. Xu and et al., "A comprehensive survey of image augmentation techniques for deep learning," *Pattern Recognition*, vol. 137, 2023.
6. J.-Y. Zhu and et al, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017.
7. A. Dosovitskiy and et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*, 2021.

8. D. Zermas and et al., "A methodology for the detection of nitrogen deficiency in corn fields using high-resolution rgb imagery," *IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 4, pp. 1879–1891, 2020.
9. P. Pandey and et al., "Synthetically labeled images for maize plant detection in uas images," in *International Symposium on Visual Computing*, pp. 543–556, Springer, 2023.
10. P. Lottes and et al., "Joint stem detection and crop-weed classification for plant-specific treatment in precision farming," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems*, 2018.
11. F. Magistri and et al., "From one field to another—unsupervised domain adaptation for semantic segmentation in agricultural robotics," *Computers and Electronics in Agriculture*, vol. 212, p. 108114, 2023.
12. Z. Fei and et al., "Enlisting 3d crop models and gans for more data efficient and generalizable fruit detection," in *IEEE/CVF International Conference on Computer Vision Workshops*, pp. 1269–1277, 2021.
13. D. Gogoll and et al., "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*, 2020.
14. Y.-C. Chen and et al., "Crdoco: Pixel-level domain transfer with cross-domain consistency," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
15. R. Bertoglio and et al., "A comparative study of fourier transform and cyclegan as domain adaptation techniques for weed segmentation," *Smart Agricultural Technology*, vol. 4, 2023.
16. M. V. Giuffrida and et al., "Leaf counting without annotations using adversarial unsupervised domain adaptation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019.
17. E. Bellocchio and et al., "Combining domain adaptation and spatial consistency for unseen fruits counting: A quasi-unsupervised approach," *IEEE Robotics and Automation Letters*, vol. 5(2), pp. 1079–1086, 2020.
18. S. Marino and et al., "Unsupervised adversarial deep domain adaptation method for potato defects classification," *Computers and Electronics in Agriculture*, vol. 174, 2020.
19. P. Sermanet and et al., "Overfeat: Integrated recognition, localization and detection using convolutional networks," 2014.
20. O. Patashnik and et al., "Balagan: Cross-modal image translation between imbalanced domains," in *Computer Vision and Pattern Recognition Workshops*, 2021.
21. M. Tkachenko and et al., "Label Studio: Data labeling software," 2020-2024. Open source software available from https://github.com/HumanSignal/label-studio.
22. V. der Maaten and et al., "Visualizing data using t-sne.," *Journal of machine learning research*, vol. 9, no. 11, 2008.
23. G. Parmar and et al., "One-step image translation with text-to-image models," *arXiv:2403.12036*, 2024.
24. T. Park and et al., "Contrastive learning for unpaired image-to-image translation," in *European Conference on Computer Vision*, 2020.