ICPA
17th INTERNATIONAL CONFERENCE ON PRECISION AGRICULTURE
ConBAP 2026
CONGRESSO BRASILEIRO DE AGRICULTURA DE PRECISÃO E DIGITAL
13-16 July 2026 | Porto Alegre, Brazil
ISPA
AsBraAP

Generation of Synthetic High-Resolution Data via Generative Image Translation and Super-Resolution to Support UAV Image Annotation and Model Training

Mauro Alessandro Karasinski[1], Ignácio Lund Gabriel da Silva Carmo[1], Caio Rodrigues Bezerra1, Emerson Matos Macedo[1], Murilo Peixoto Galvão[1], Cirano Cruz Melville[1], Robert Alefe Torres da Costa[1], Shestéfany Valeska Dantas Oliveira[1], and Amaury Burlamaqui Bendahan[1]

[1]LADA – Amazon Digital Agriculture Laboratory, Embrapa Roraima, Brazil

A paper from the Proceedings of the

17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture

13-15 July 2026

Porto Alegre, Brazil

Abstract

Manual annotation of imagery acquired by unmanned aerial vehicles (UAVs) for detection/segmentation tasks is one of the main bottlenecks for deep learning applications in precision agriculture, due to the high cost and the time required to produce consistent labels. In addition, low-altitude flights to obtain ultra–high spatial resolution increase operational complexity and data volume, limiting the scalability of acquisition campaigns. Although neural network–based super-resolution techniques have advanced, the use of super-resolved images as synthetic data to support annotation and training remains underexplored—especially when the goal is to approximate the visual detail of low-altitude flights (e.g., 7 m) from imagery acquired at more operational altitudes (e.g., 30 m), while simultaneously addressing the domain shift between synthetic and real images. This study investigated the potential of generative image translation and super-resolution approaches to reduce the visual gap between UAV imagery acquired at 30 m and 7 m, and assessed whether the resulting synthetic imagery could support image annotation and improve weed detection performance. RGB UAV imagery acquired at 7 m and 30 m flight altitudes was organized into corresponding image patches. ESRGAN and CycleGAN models were trained to enhance image quality and reduce the visual gap between acquisition altitudes. CycleGAN-generated imagery was used to support semi-automatic annotation, combining pseudo-labels from a YOLO detector with mask refinement using the Segment Anything Model (SAM)(Kirillov et al., 2023). Model performance was evaluated through training metrics, qualitative visual assessment, and a weed detection task. Detection performance on real 7 m, real 30 m,

The authors are solely responsible for the content of this paper, which is not a refereed publication. Citation of this work should state that it is from the Proceedings of the 17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture. EXAMPLE: Last Name, A. B. & Coauthor, C. D. (2026). Title of paper. In Proceedings of the 17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture and Brazilian Congress on Precision and Digital Agriculture.

and CycleGAN-translated imagery was compared using Precision, Recall, and F1-score. Both image generation models improved image quality relative to the original 30 m imagery. ESRGAN partially recovered texture and edge information, whereas CycleGAN more effectively reduced the visual gap between the 30 m and 7 m domains. SAM-based segmentation indicated that the translated imagery recovered approximately 77% of the structural richness observed in the 7 m reference imagery. The corrected evaluation confirmed the advantage of CycleGAN-translated outputs over direct 30 m acquisitions. The detector achieved an F1-score of 0.821 on real 7 m imagery, compared with 0.029 on real 30 m imagery and 0.242 on CycleGAN-translated outputs. Precision increased from 0.357 to 0.557, while the number of true detections increased from 10 to 103, representing an 8.3-fold improvement over the 30 m baseline and indicating a substantial reduction in domain shift between acquisition altitudes. Overall, the results demonstrate that generative image translation can effectively reduce the visual and structural gap between UAV imagery acquired at different flight altitudes, improving the detectability of agronomically relevant targets in precision agriculture.

***Keywords*:** UAV imagery; synthetic data; super-resolution; CycleGAN; Segment Anything Model; semi-automatic annotation; object detection; precision agriculture.

Introduction

Unoccupied aerial vehicles (UAVs) have become important tools in precision agriculture due to their ability to provide high-resolution imagery for crop monitoring, weed detection, disease assessment, stand counting, and other site-specific management applications (Zhang and Kovacs, 2012; Tsouros et al., 2019). Combined with advances in deep learning and computer vision, UAV imagery has enabled the development of automated systems capable of identifying agricultural targets with high accuracy (Kamilaris and Prenafeta-Boldú, 2018; Liakos et al., 2018). However, the performance of these systems remains strongly dependent on the availability of large annotated datasets, making image labeling one of the most time-consuming and costly stages of the machine learning workflow.

Recent advances in foundation models have created new opportunities for reducing annotation effort. The Segment Anything Model (SAM) has demonstrated remarkable capability for generating segmentation masks with minimal user interaction, enabling semi-automatic annotation workflows. Nevertheless, the quality of the generated masks remains highly dependent on image characteristics such as spatial resolution, edge definition, texture representation, and object visibility. In agricultural environments, where targets are often small, irregular, and visually complex, images acquired at higher flight altitudes may not provide sufficient detail for accurate automatic delineation.

A common solution is to perform UAV acquisitions at very low altitudes, producing ultra-high-resolution imagery suitable for annotation and object detection tasks. However, low-altitude flights significantly increase mission duration, battery consumption, data volume, and operational complexity, while reducing the area covered per flight. These limitations restrict the scalability of image acquisition campaigns, particularly in large agricultural areas. Consequently, there is increasing interest in computational approaches capable of approximating the visual characteristics of low-altitude imagery while preserving the operational advantages of higher-altitude acquisitions.

Among the available approaches, deep learning–based image generation methods have shown considerable potential. Super-resolution techniques seek to reconstruct high-resolution images from lower-resolution observations by recovering spatial details and enhancing image sharpness (Ledig et al., 2017; Wang et al., 2021). Alternatively, image-to-image translation models such as CycleGAN (Zhu et al., 2017) learn the visual characteristics of a target domain and transfer them to images acquired under different conditions without requiring paired observations. While super-resolution focuses on preserving structural information and reconstructing image details, domain translation aims to reproduce the visual appearance of a target domain, potentially generating imagery that more closely resembles low-altitude acquisitions.

The generation of synthetic imagery has attracted growing interest as a strategy to reduce annotation costs and increase the availability of training data for machine learning applications (Tremblay et al., 2018). However, the practical value of synthetic images depends on their ability to support downstream tasks such as annotation, segmentation, and object detection. Although super-resolution and domain translation have been widely investigated independently, direct comparisons between these approaches for supporting semi-automatic annotation workflows in precision agriculture remain limited. Furthermore, little attention has been given to the combined use of image generation techniques and foundation models such as SAM for accelerating agricultural dataset creation.

Therefore, this study compared two image generation strategies for approximating low-altitude UAV imagery: a supervised super-resolution framework and an unpaired domain translation framework based on CycleGAN. The generated images were used within a SAM-assisted annotation workflow, followed by human validation and correction. The resulting annotations were subsequently employed to train object detection models and evaluate their performance on real UAV imagery.

The central hypothesis was that synthetic imagery generated through super-resolution or domain translation can improve the quality of semi-automatic annotation, reduce manual labeling effort, and contribute to the development of more accurate object detection models. Specifically, this study aimed to: (i) compare super-resolution and CycleGAN-based domain translation for generating synthetic UAV imagery; (ii) evaluate their influence on SAM-assisted annotation; (iii) evaluate their ability to support semi-automatic annotation workflows; and (iv) assess their impact on object detector training and performance under real agricultural conditions.

Methodos

Study Area

The study was conducted in commercial soybean production fields located in the municipality of Alto Alegre, State of Roraima, Brazil, within the northern Amazon region. The area is characterized by extensive mechanized agriculture and represents one of the main soybean-producing regions in northern Brazil.

According to the Köppen climate classification, the regional climate is classified as tropical monsoon (Am), characterized by high annual temperatures, elevated relative humidity, and a marked seasonality between wet and dry periods. Mean annual temperatures typically range between 26 and 28 °C, while annual precipitation exceeds 1,500 mm, with rainfall concentrated between April and September. These climatic conditions favor rapid crop development and promote high spatial variability in vegetation growth, making the region suitable for precision agriculture studies based on unmanned aerial vehicle (UAV) imagery.

Furthermore, the region presents persistent limitations associated with optical satellite monitoring due to frequent cloud cover throughout a large portion of the year, which reduces the availability of cloud-free imagery during critical crop development stages. Consequently, UAV platforms have become an important alternative for acquiring high-spatial-resolution data with greater temporal flexibility, supporting crop monitoring, field scouting, and the development of artificial intelligence applications for agricultural management.

UAV Image Acquisition

RGB imagery was acquired using an unmanned aerial vehicle equipped with a high-resolution RGB camera. Flights were conducted at two altitudes: 7 m and 30 m above ground level. The 7 m imagery represented the target domain due to its ultra-high spatial detail, whereas the 30 m imagery represented an operational acquisition condition with greater area coverage and lower operational cost. A total of 1,220 images were acquired at 7 m and 371 images at 30 m, all with a native spatial resolution of 640×640 pixels. All images were exported and organized into standardized image tiles to facilitate training and evaluation of the deep learning models.

Dataset Preparation

Because the images acquired at 7 m and 30 m were not perfectly aligned at the pixel level, two distinct dataset preparation strategies were adopted according to the requirements of each image generation approach. For the supervised super-resolution experiments, synthetic low-resolution samples were generated by applying a controlled degradation pipeline to the 7 m imagery, including bicubic downscaling (scale factor 3×), Gaussian blur, luminance noise insertion, and color adjustments. The degradation parameters were empirically calibrated to minimize a composite similarity score between the synthetic degraded images and real 30 m acquisitions, ensuring that the resulting paired dataset closely approximated real acquisition conditions. For the CycleGAN experiments, paired correspondence was not required; therefore, the original 7 m and 30 m image collections were directly employed as independent source and target domains. In both cases, all images were divided into 128×128 pixel patches using a sliding-window approach, yielding approximately 30,500 patches from the 7 m dataset (1,220 images) and 9,275 patches from the 30 m dataset (371 images). For CycleGAN training, 30,000 patches from the 7 m domain were sampled alongside all 9,275 patches from the 30 m domain. Online data augmentation, including random horizontal and vertical flips and 90° rotations, was applied during training to increase sample diversity and improve model generalization. An overview of both dataset preparation pipelines is presented in Figure 1.

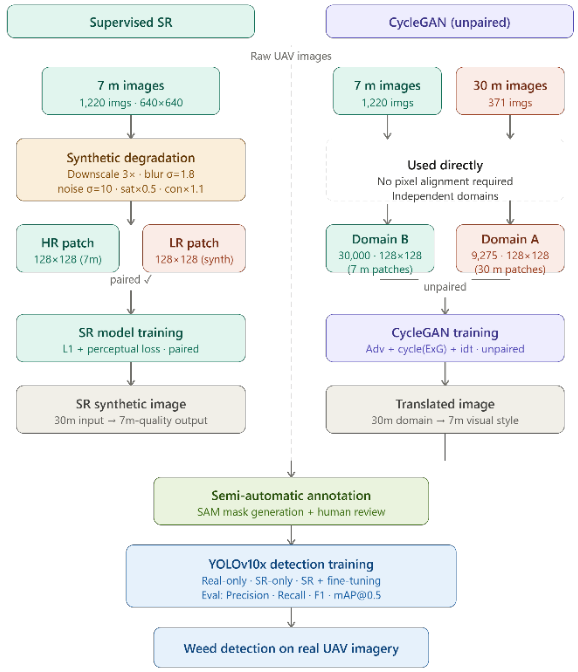

Figure 1. Overview of the two dataset preparation pipelines adopted in this study. The supervised super-resolution pipeline (left) applies a controlled degradation procedure to 7 m imagery to generate synthetic paired low-resolution samples. The CycleGAN pipeline (right) employs the original 7 m and 30 m image collections directly as independent training domains, without requiring pixel-level alignment. Both pipelines extract 128×128 pixel patches and converge into a semi-automatic annotation workflow based on the Segment Anything Model (SAM).

Super-Resolution Framework

A supervised super-resolution model based on ESRGAN (Wang et al., 2018) was trained using paired samples generated from the original 7 m imagery. Synthetic low-resolution images were created through controlled degradation, enabling the model to learn the reconstruction of high-detail images from degraded inputs. The trained model was subsequently applied to 30 m imagery to generate synthetic high-resolution products.

CycleGAN Domain Translation Framework

A CycleGAN-based framework was implemented to learn the visual transformation between the 30 m and 7 m domains without requiring paired observations. To enhance vegetation representation, the Excess Green minus Excess Red (ExGR) index was incorporated as an additional input channel. The trained model generated synthetic images with visual characteristics similar to those observed in the 7 m imagery.

Semi-Automatic Annotation Workflow

The synthetic images were incorporated into a SAM-assisted annotation workflow. Segmentation masks generated by SAM were reviewed and corrected by a human operator, replacing fully manual annotation with a validation-and-correction process. Annotation time was recorded to quantify workflow efficiency.

Object Detection Experiments

The resulting annotations were used to train YOLO-based object detection models. Models trained with original imagery, synthetic imagery, and combinations of both were evaluated using independent real-image test datasets. Performance was assessed using Precision, Recall, F1-score.

Evaluation Metrics

The quality of the generated images was assessed through both visual inspection and quantitative similarity metrics. Table 1 summarizes the measures computed between synthetic and reference imagery, including domain-specific indices based on vegetation spectral response. For object detection evaluation, standard metrics were adopted: Precision, Recall, F1-score, and mAP@0.5 (mean Average Precision at IoU threshold of 0.5).

Table 1 - Image quality and domain similarity metrics used to evaluate synthetic imagery.

Metric	Formula	Description
Sharpness	$S = \mathrm{Var}(\nabla^2 I)$	Variance of Laplacian — measures edge definition
Contrast	$C = \sigma(I)$	Standard deviation of pixel intensities
Saturation	$Sat = \mu(S_{HSV})$	Mean saturation in HSV color space
ExG	$ExG = 2G - R - B$	Excess Green — vegetation presence
ExGR	$ExGR = ExG - 1.4R + G$	Excess Green-Red — vegetation vs. soil discrimination
Domain Similarity Score	$\delta = \Sigma k \, \lvert m_k synth - m_k real \rvert / m_k real$	Normalized distance across metrics k
PSNR	$PSNR = 10 \log_{10}(MAX^2 / MSE)$	Peak Signal-to-Noise Ratio (paired SR only)
SSIM	$SSIM = ((2\mu x\mu y + c_1)(2\sigma xy + c_2)) / ((\mu x^2 + \mu y^2 + c_1)(\sigma x^2 + \sigma y^2 + c_2))$	Structural Similarity Index (paired SR only)

Results

Both image generation models were trained on a single NVIDIA RTX 6000 Ada GPU using 128×128 pixel patches. The ESRGAN model underwent two training phases: a pixel-wise pre-training stage of 25 epochs using L1 loss with ExG-weighted reconstruction, followed by an adversarial fine-tuning stage incorporating perceptual and discriminator losses, totaling 175 epochs on 9,500 paired HR/LR patches with a batch size of 32. During the pixel phase, the generator loss decreased from 0.164 to 0.144, with validation PSNR stabilizing at approximately 28.4 dB. The adversarial phase introduced discriminator feedback, initially causing PSNR fluctuations before stabilizing as training progressed. Visual inspection of the SR outputs indicated partial recovery of texture detail and edge sharpness relative to the degraded inputs, although residual smoothing artifacts remained in regions containing high-frequency texture.

The CycleGAN model was trained for 71 epochs over approximately 3.6 hours with a batch size of 8, using 30,000 patches from the 7 m domain and 9,275 patches from the 30 m domain. The evolution of the four loss components is presented in Figure 2.

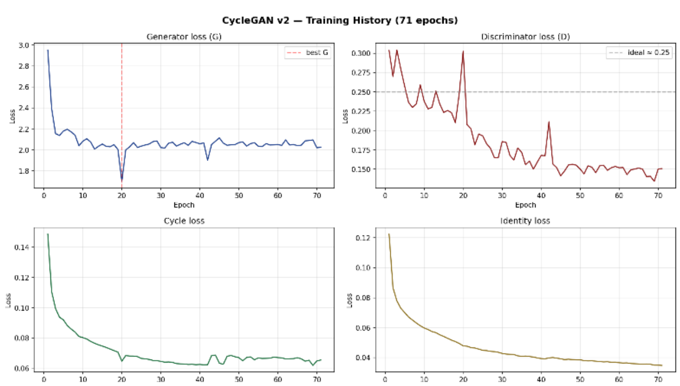

Figure 2. Training loss curves over 71 epochs. Generator loss (G), discriminator loss (D), cycle-consistency loss, and identity loss are shown. The dashed red line indicates the best checkpoint at epoch 20.

The generator loss exhibited rapid convergence during the first 20 epochs, decreasing from 2.95 to a minimum of 1.70 at epoch 20, which was retained as the best checkpoint. Afterward, the loss stabilized around 2.05 with moderate oscillations characteristic of adversarial training. The discriminator loss decreased progressively from approximately 0.30 to 0.15, while the cycle-consistency and identity losses decreased monotonically from 0.149 to 0.065 and from 0.122 to

0.035, respectively. These results indicate that the model maintained structural coherence and domain fidelity despite the increasing dominance of the generators over the discriminators. The best checkpoint was selected at epoch 20, corresponding to the minimum generator loss and a discriminator loss of 0.30, representing a more balanced adversarial state.

As illustrated in Figure 3, the CycleGAN model successfully translated imagery from the 30 m domain toward the visual characteristics of the 7 m domain, generating outputs with enhanced spatial detail, improved texture definition, and more accurate representation of crop rows and vegetation structure. Visual inspection across the evaluated sample pairs indicated that the translated imagery more closely resembled the high-resolution reference imagery than the original 30 m inputs, suggesting a substantial reduction in the domain gap between acquisition altitudes.

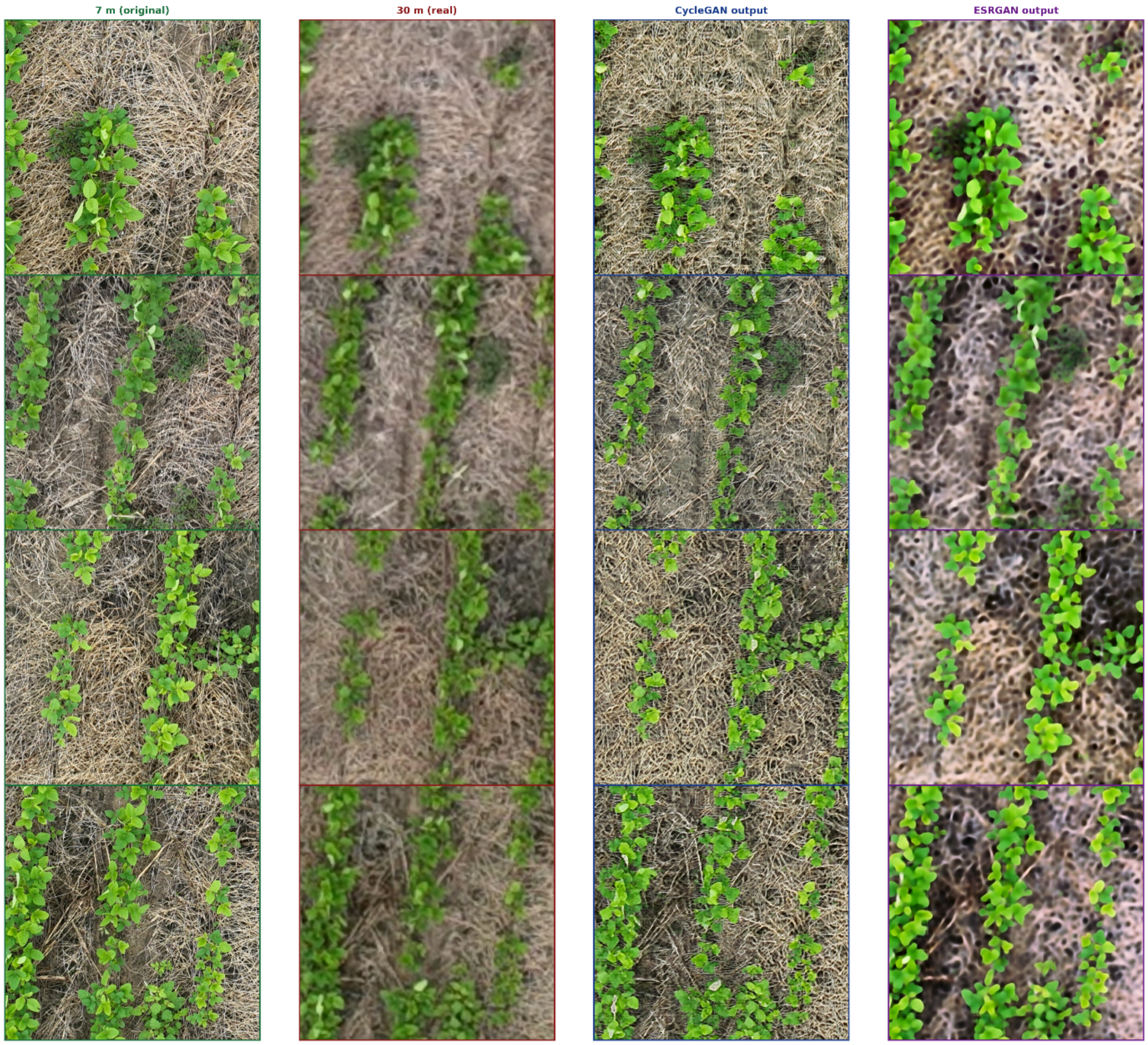

Figure 3. Qualitative comparison of CycleGAN domain translation results. Each row shows a representative sample: original 7 m imagery (left), real 30 m imagery (center), and the corresponding CycleGAN-translated output (right).

The number of segments detected by SAM was used as a proxy for the spatial richness of each image domain. The 30 m real imagery yielded a mean of 73 segments per image, reflecting the limited structural detail available at operational acquisition altitude, whereas the 7 m reference imagery produced a mean of 251 segments per image. The CycleGAN-translated outputs achieved a mean of 194 segments per image, representing an increase of approximately 168% relative to the 30 m inputs and recovering approximately 77% of the segmentation richness observed in the 7 m reference domain. In the best-performing sample, the translated image achieved 92% of the 7 m segmentation density, suggesting near-complete recovery of fine-scale spatial structure in scenes with high vegetation coverage. These results indicate that the CycleGAN domain translation substantially enhanced the detectability of agronomically relevant structures, including individual weed leaves, straw elements, and crop row boundaries, bringing the translated imagery significantly closer to the visual and structural characteristics of the high-resolution reference domain, as illustrated in Figure 4.

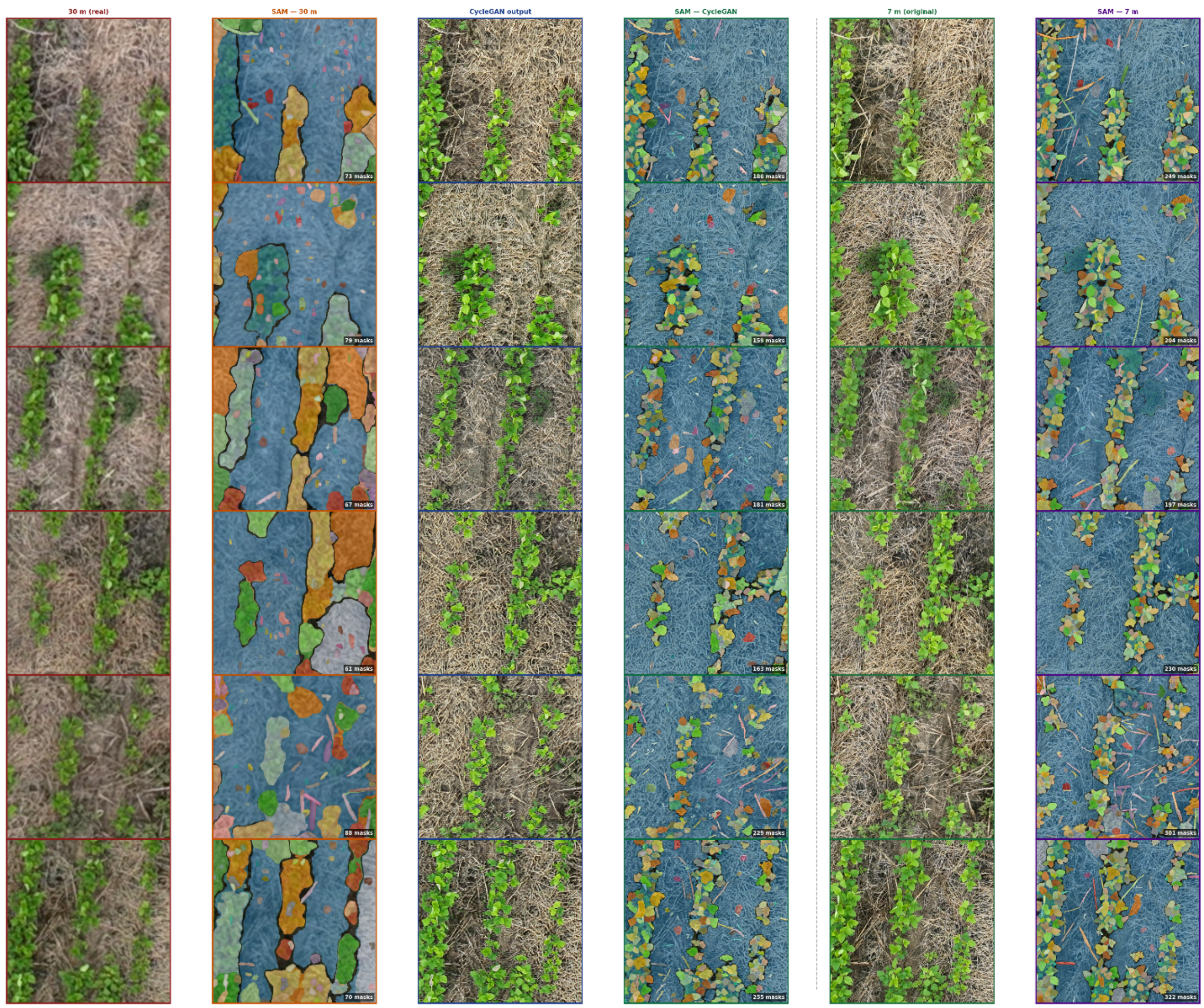

Figure 4. SAM segmentation results for six representative samples across three image domains: 30 m real (left), CycleGAN-translated output (center), and 7 m reference (right). The number of detected segments per image is shown in the lower-right corner of each segmentation panel.

The resulting annotations were used to train a YOLO-based weed detector, whose performance was evaluated on independent test imagery. As shown in Table 2, the corrected evaluation using domain-specific ground-truth labels confirmed the advantage of CycleGAN-translated imagery over direct 30 m acquisitions. The detector achieved an F1-score of 0.821 on real 7 m imagery, 0.029 on real 30 m imagery, and 0.242 on CycleGAN-translated outputs, corresponding to an 8.3-fold improvement over the 30 m baseline. The translated imagery also improved precision and increased true detections from 10 to 103.

Table 2. Detection performance across real and CycleGAN-translated imagery using corrected ground-truth annotations.

Domínio	Precision	Recall	F1	TP	FP	FN
7 m real	0.835	0.806	0.821	537	106	129
30 m real	0.357	0.015	0.029	10	18	656
CycleGAN	0.557	0.155	0.242	103	82	563

The results obtained in this study highlight the potential of unpaired domain translation as a practical strategy for bridging the visual gap between UAV imagery acquired at different altitudes. The CycleGAN model successfully transferred the visual characteristics of 7 m imagery to 30 m acquisitions without requiring spatially aligned image pairs, recovering approximately 77% of the SAM segmentation richness of the reference domain and increasing the number of detected segments by 168% relative to the original 30 m inputs. Notably, the ExGR-weighted cycle-consistency loss adopted in this study contributed to prioritizing the recovery of vegetation-related features, which are agronomically most relevant for weed detection and annotation tasks. The qualitative comparison presented in Figure 5 further confirmed that the translated outputs closely approximated the texture definition, crop row delineation, and leaf-level detail characteristic of low-altitude acquisitions, while preserving the spatial layout of the original 30 m scenes.

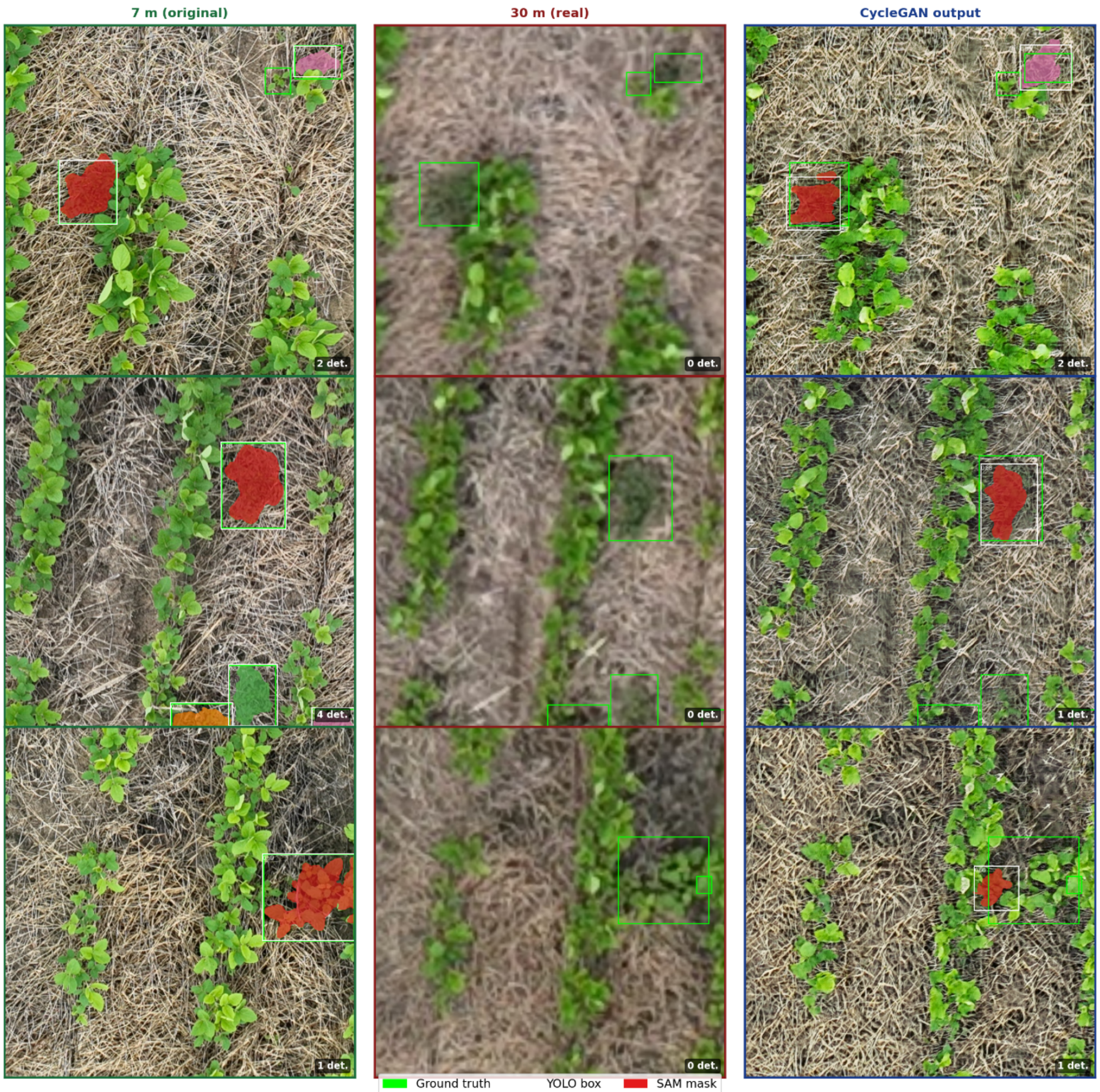

Figure 5. YOLO detection and SAM segmentation results across three image domains: 7 m reference (left), 30 m real (center), and CycleGAN output (right). Green boxes indicate ground truth, white boxes indicate YOLO detections, and colored regions indicate SAM masks generated within each detected bounding box.

Conclusion

This study demonstrated that CycleGAN-based domain translation is a viable approach for generating synthetic UAV imagery approximating the visual quality of low-altitude acquisitions without requiring paired or spatially aligned training data. The translated outputs recovered approximately 77% of the SAM segmentation richness observed in the 7 m reference domain, increasing detected segments by 168% relative to the original 30 m inputs. Compared to supervised super-resolution methods such as ESRGAN, which depend on synthetic degradation pipelines that may introduce domain gap between simulated and real image degradation, CycleGAN operates directly on real multi-altitude image collections, making it particularly suitable for scenarios where co-registered datasets are unavailable. The integration of translated imagery into a SAM-assisted annotation workflow further confirmed the potential of this approach to support semi-automatic annotation workflows in precision agriculture applications. Future work will expand the evaluation to larger datasets, additional flight altitudes, and different crop conditions.

Acknowledgments

We thank CITS-Amazonas, SUFRAMA, and Foxconn for their financial support, and Embrapa Roraima for its logistical support during field data collection, all of which were essential for carrying out this study. This is Contribution No. 1 from the Simids team.

References

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147, 70–90.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., et al. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4681–4690).

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18(8), 2674.

Tremblay, J., Prakash, A., Acuna, D., Brophy, M., Jampani, V., Anil, C., et al. (2018). Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 969–977).

Tsouros, D. C., Bibi, S., & Sarigiannidis, P. G. (2019). A review on UAV-based applications for precision agriculture. *Information*, 10(11), 349.

Wang, X., Xie, L., Dong, C., & Shan, Y. (2021). Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*.

Wilson, G., & Cook, D. J. (2020). A survey of unsupervised deep domain adaptation. *ACM Transactions on Intelligent Systems and Technology*, 11(5), 1–46.

Zhang, C., & Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: A review. *Precision Agriculture*, 13(6), 693–712.