Supplementary

Table S1: Quantitative summary of synthetic-data utility in the reviewed application studies. For each study we report the generative model, the data used to build/train the generator, how synthetic samples were ground-truthed, how synthetic data were integrated into downstream training, and the reported performance against a real-only or competing baseline. "N/A" indicates that the value was not reported in the source. This table provides the underlying evidence base for the task-grouped Table 3, method-family share Table 4, per-study Table 5, strategy-and-compute Table 9, evaluation-framework Table 10, and domain-gap Table 11 tables.

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
2D compositing, copy-paste, and conventional augmentation						
Allken et al. (2019)	Cut-paste compositing (cropped fish on background images)	70 cropped fish images per species (210 total) used to make 5,000-15,000 synthetic	inherited from source (cut-paste); class label kept	synthetic only (compared against real-only baseline)	Real: 71.1% acc. Syn.: 94.1% acc. Δ: +23 pp	pelagic fish species classification (herring/-mackerel/blue whiting)
Giakoumoglou et al. (2023)	DDPM-based object generation + cut-paste-blend composition (Gaussian blurring / Poisson blending)	113 cropped whitefly images for DDPM training; 767 PlantDoc background images, with synthetic whiteflies pasted on 644 images	bounding boxes automatically derived from paste locations and generated object sizes during compositing	detector trained on synthetic composite images; no manually annotated real whitefly detection dataset was used for detector training	Real: N/A Syn.: 0.661 mAP50 with YOLOv8m + Gaussian blurring Δ: +0.034 pp mAP50 over the no-blending baseline (0.627 to 0.661)	whitefly detection on plant leaves

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Gmez-Zamanillo et al. (2026)	Copy-and-paste image composition with multiple blending strategies (Patch, Binary mask, Gaussian, Guided mask, High resolution, Poisson mask, Poisson patch, Mixed). Insect masks generated by XYZ-Z-channel thresholding + dilation + GrabCut refinement. Not a learned generative model.	528 real eggplant-leaf images with whitefly dot annotations + 59 clean eggplant-leaf backgrounds (BASF Spain greenhouse). Whiteflies extracted from a small reduced subset of annotated real images (16/12/8/4/2 images) and pasted onto clean leaf tiles. Real-only training experiments use 16/12/8/4/2 annotated images plus full real training-set reference (OR-CD).	Inherited from source via cut-paste – known dot-annotation locations of extracted insects placed at controlled positions on clean leaf tiles (256x256), with rejected-mask visual inspection (≈10 min per 100-insect image vs ≈30 min full annotation)	Train density-map estimation model on a small set of real images (2/4/8/12/16) plus synthetic 256x256 leaf tiles. Compared to (a) OR-CD = full-real-dataset baseline; (b) OR-RD = same reduced real set without synthetic; (c) reduced real + synthetic with each blending strategy.	Real: R\^2 = 0.52 (OR-RD, 16 real) Syn.: R\^2 = 0.84 (16 real + Patch synthetic) Δ: +0.32 R\^2	Whitefly counting on eggplant leaves via density-map estimation (real outdoor field conditions, 3 cameras)
Guldenring et al. (2021)	cut-paste composition with Poisson blending (programmatic)	as few as 9 manually annotated leaves per field-day session	Synthetic masks are propagated from leaf annotations during cut-paste composition; real-image predictions are refined by GrabCut	synthetic-only training (per field-and-day specific model); GrabCut refinement on real predictions	Real: N/A Syn.: 0.703 Plant IoU (with GrabCut) Δ: +0.07 IoU from GrabCut refinement	grassland weed (Rumex obtusifolius / Cirsium vulgare) semantic segmentation

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Hu et al. (2021)	Cut-paste (color-matched plant instances + Ashikhmin natural texture-synthesized soil backgrounds)	408 plant instances (cut-outs) and 4 soil background templates	Bounding boxes/labels propagated from extracted plant cut-outs (each instance carries label and binary mask) during paste; SSL pseudo-labels for unlabeled real images	synthetic only to synthetic + pseudo-labeled real (iterative SSL)	Real: 50.9 mAP (supervised, real-only baseline) Syn.: 43.6 mAP (synthetic-only teacher); 46.0 mAP (IS-SSL final, 408 plant instances) Δ: +2.4 mAP from SSL over synthetic-only (43.6 to 46.0; -4.9 mAP vs. supervised real-only)	Weed object detection
Jepsen and Mikkel D.B. Jeppesen (2021)	Cut-paste (Dwibedi et al.) with Gaussian and Poisson blending	subset of Chebrolu et al. sugar-beet dataset (used to extract weed/crop instances)	inherited from source (cut-paste preserves instance masks)	synthetic only vs synthetic + real mixed (mixing real degraded performance)	Real: wmIoU=0.634 Syn.: wmIoU=0.767 Δ: +0.133 wmIoU	sugar-beet weed/crop semantic segmentation
Lee and Lee (2025)	Traditional augmentation (geometric + color space transformations); not generative	60,165 images base dataset (PlantVillage + private citrus/kiwi)	inherited from source (augmentation preserves labels)	Real-image training with conventional data augmentation	Geo-only aug.: lower F1 Geo+color aug.: F1 > 98% Δ: up to ∼3% F1	multi-crop disease recognition (24 classes across 5 crops)
Li et al. (2022)	Copy-paste (RLDCP)	450 base annotated images (3 open-source datasets); expanded to 1,350 images via two rounds of RLDCP copy-paste	inherited from source (copy-paste preserves mask)	synthetic + real (augmentation)	Real: MIoU=78.36% (HRNet, no augmentation) Aug.: MIoU=84.70% (HRNet + two RLDCP augmentations) Δ: +6.34 pp MIoU	rice leaf disease (bacterial blight, blast, brown spot) segmentation and severity assessment

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Li et al. (2025b)	Cut-paste (patch-level synthesis)	1540 real WE3DS training images; synthetic generated at 1x-20x scales	inherited from source (mask preserved during paste)	synthetic only and synthetic + real (hybrid)	Real: $\mathrm{mIoU}_{\text{no-soil}}$=0.626 Hybrid: $\mathrm{mIoU}_{\text{no-soil}}$=0.716 (1:10 real:synthetic hybrid) Δ: +0.090 $\mathrm{mIoU}_{\text{no-soil}}$ (+9.0 pp)	species-level crop/weed semantic segmentation (17 classes)
Lin and Chiu (2026)	Copy-Paste image synthesis (Ghiasi et al. [19]); FastGAN was also evaluated but produced low-quality D2 samples and was not used in the final pipeline	177 original D2 (mechanical scratch) images used as the source for manual mask extraction; the 6,000-image defect training set provided clean mango surfaces onto which masks were pasted	Defect masks were manually extracted from the 177 original D2 images, randomly scaled, rotated, and blended onto clean mango surfaces; visually implausible synthetic images (defect outside mango region, unrealistic scale, unnatural appearance) were manually filtered out, expanding D2 from 177 to 672 samples	Augmented (D2-expanded) multi-label dataset used to train ResNet-34 (with self-attention at layer4) end-to-end; non-augmented independent test set used for final evaluation	Real: Macro-F1 68.2% (ResNet34 baseline) Syn.: Macro-F1 78.2% (ResNet34 + Copy-Paste + self-attention, independent test) Δ: +10.0 pp Macro-F1	Multi-label five-class mango surface defect classification (D1 latex, D2 scratch, D3 anthracnose, D4 color inconsistency, D5 black spot)

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Motoi et al. (2025)	Cut-paste with SAM masks + Poisson blending (DCED-guided selection)	529 normal + 166 anomalous real patches as source	Synthetic samples are assigned anomalous labels by construction after DCED-selected anomalous berries are pasted onto normal grape patches	synthetic anomalous samples added to or substituted for real anomalous training samples at 10%, 25%, 50%, and 100%	Real: F1=92.90 (ResNet18 baseline) Syn.: F1=93.70 (addition of 25% synthetic samples, three pasted anomalies) Δ: +0.80 F1	anomaly detection/classification on table grapes
Mujkic et al. (2022)	Cut-paste with object masks + Gaussian blur (for DAE training pairs)	2344 summer-harvest field images + annotated anomaly object set used to generate synthetic anomaly training pairs	Synthetic DAE training pairs are generated by pasting pixel-level annotated anomaly-object masks onto normal field scenes; the paired target is the corresponding normal image	DAE trained on cut-paste synthetic anomaly/normal pairs; compared with AE/VQ-VAE trained on normal-only data, SSAE trained with normal data plus 500 real labeled anomaly samples, and a fully supervised YOLOv5s object-detection baseline	Real: PR AUC = 0.9794 (fully supervised YOLOv5s object detector) Syn.: PR AUC = 0.8861 (DAE trained on cut-paste synthetic anomaly pairs) Semi-sup.: PR AUC = 0.9353 (SSAE with 500 real labeled anomaly samples + max-margin loss) Δ: -0.0933 PR AUC (DAE vs. YOLOv5s); -0.0441 PR AUC (SSAE vs. YOLOv5s)	anomaly/obstacle detection for autonomous agricultural vehicles

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Murean (2022)	Cut-paste compositing (Fruits-360 fruit images onto Internet-scraped tree/foliage backgrounds) with rotation/brightness/cropping augmentation	Fruits-360 source images (90,483 images of 131 fruits/vegetables) and 30 Internet-scraped tree/foliage background images; 3000 generated training images and 600 generated validation images per experiment	Bounding boxes computed automatically from known fruit-paste coordinates (XML/CSV annotation files generated alongside images)	synthetic only (trained on 3000 generated images, tested on 70 real images)	Real: N/A Syn.: AP50=0.750 and F1=0.816 (SSD#2, best reported synthetic-trained run) Δ: N/A	apple detection (object detection)
Najafian et al. (2023)	Cut-paste image synthesis (real and "fake" wheat heads pasted on background frames)	2 manually annotated wheat images + video frames; 10,000 synthesized training images	inherited from source (cut-paste preserves mask)	synthetic only -¿ progressive domain adaptation with augmented real, pseudo-labels, then 2-per-domain real fine-tune	Real: TTA Dice 0.46 (external Γ, Baseline model trained from ImageNet using the small manually annotated real set Θ_t/Θ_v) Syn.+DA: TTA Dice 0.91 (external Γ, Model G, full pipeline; synthetic-only Model S = TTA Dice 0.48) Δ: +0.45 TTA Dice (vs. small-real Baseline)	wheat head semantic segmentation

continued on next page

Table S1 – continued from previous page

Study	**Generative model**	**Generator training data**	**Synthetic ground-truthing**	**Integration with downstream training**	**Reported performance (real-only vs. synthetic-augmented)**	**Downstream task**
Nigam et al. (2025)	Augmentor (geometric augmentation: zoom/flip/rotate); no generative model	5,438 original field images; augmented to 10,252 images using Augmentor	inherited from source (augmentation preserves labels)	augmentation to balance classes (each class to 1000 images)	Original: Accuracy 93.88% Aug.: Accuracy 96.68% Δ: +2.80 pp accuracy	wheat rust disease severity classification (3 rust types x 3 severity levels + healthy)
Nitin et al. (2023)	Manipulation-based data augmentation (geometric + filter transformations); no generative model	Real: baseline CNN (no aug.) Syn.: improved with MBDA Δ: N/A (numerics distributed across tables)	inherited from source (label-preserving transformations)	augmented training set; original images used for validation	Real: baseline CNN (no aug.) Syn.: improved with MBDA Δ: N/A (numerics distributed across tables)	castor insect pest image classification
Sapkota et al. (2022)	Cut-paste from real plant instances + StyleGAN2-ADA for fake plant instances	40-50 plant instances sufficient for synthetic generation; StyleGAN2-ADA trained on clipped real instances	inherited from source (cut-paste preserves instance masks); GAN-derived instances pasted with known masks	synthetic only, real only, and mixed (real + synthetic) compared	Real: mAPm=0.80, mAPb=0.81 Syn.: mAPm=0.60, mAPb=0.64 (real-instance synthetic only); mixed: mAPm=0.80, mAPb=0.83 Δ: -0.20 mAPm / -0.17 mAPb (synthetic-only vs real); +0.00 mAPm / +0.02 mAPb (mixed vs real)	weed (morningglory, grass) detection and segmentation in cotton

continued on next page

Table S1 – continued from previous page

Study	**Generative model**	**Generator training data**	**Synthetic ground-truthing**	**Integration with downstream training**	**Reported performance (real-only vs. synthetic-augmented)**	**Downstream task**
Wang et al. (2026)	Image-compositing pipeline: real single-fish masks transformed (scale 0.5-1.5x, rotation -30 to +30 degrees, translation up to 50% of frame), pasted with morphological erosion + Gaussian blur on real backgrounds; an occlusion rate threshold of 0.5 limits overlap (no deep generative network used)	1,175 manually pixel-annotated single-fish images (from 1,314 SSIM-filtered single-fish frames); 53/66 fish used for training synthesis pool (675 single-fish images for synthesis), 13 fish (302 single-fish images) for test synthesis; 10,000 synthetic train images + 3,000 synthetic test images generated	Each synthetic image's instance masks are derived deterministically from the original LabelMe pixel-level annotations of the source single-fish images, propagated through the geometric transformations (scale/rotate/-translate) used during compositing	Three training settings compared with Mask2Former (ResNet-50, identical schedule): Real-only (real images only), Synthetic-only (synthesized images only), Mixed (full FOD train set = real + synthetic). All evaluated on the same 436 manually annotated real test images.	Real: mAP=0.526 Syn.-only: mAP=0.642 Mixed: mAP=0.687 Δ: +0.116 mAP (Syn.-only vs. Real); +0.161 mAP (Mixed vs. Real)	Fish instance segmentation under occlusion in underwater aquaculture imagery (3 occlusion levels: whole/part/fragment)
Yang et al. (2022b)	Cut-paste / image compositing (real pod scans pasted with random overlap onto background)	Pod scans from 8 soybean accessions (cut-paste source pool)	Auto-generated instance masks during compositing pipeline	synthetic only (no manual real labels for training)	Real: N/A Syn.: AP50 0.96 Δ: N/A	Soybean pod instance segmentation for high-throughput phenotyping

continued on next page

Table S1 – continued from previous page

Study	**Generative model**	**Generator training data**	**Synthetic ground-truthing**	**Integration with downstream training**	**Reported performance (real-only vs. synthetic-augmented)**	**Downstream task**
Yang et al. (2022a)	Cut-paste image generation (in-vitro pods composited at varying overlap degrees)	1600 synthetic in-vitro images (best config); pod scan pool from real samples	Auto-generated instance masks via cut-paste pipeline	pretrain on synthetic, fine-tune on real (COCO -¿ in-vitro synthetic -¿ on-branch real)	Real: AP50=0.773 (COCO → on-branch real, direct adaptation) Syn.+Real: AP50=0.800 (COCO → in-vitro synthetic → on-branch real, two-step transfer) Δ: +0.027 AP50	On-branch soybean pod instance segmentation in mature plants

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Ymous (2026)	Pyramid Blending (PB) and Pyramid Blending + Linear Histogram Color Matching (PB-CMx) for compositing; comparison baselines Stable Diffusion 3.5 Large (SDL3.5) and Flux 2 (image-to-image / text-to-image) generative augmentation	542 real high-resolution training images; low-resolution source set contains 100 weed images and 1800 grass background images, with 89 low-resolution weed images used in the SDL3.5/Flux2 comparison. Synthetic training generated: 3500 PB/PB-CMx images, or 7000 images in the 2x setting; SDL3.5/Flux2 generated 30 samples per 89 low-resolution weed source images, yielding 2670 synthetic images	Each synthetic image inherits the weed-center label of its low-resolution masked weed source; weed positions and quantities controlled by the placement algorithm so heatmap labels are deterministic per pasted weed	Three-stage protocol: (1) train only on synthetic; (2) weighted training on synthetic+real (synthetic weighted 1/4 of real); (3) finetune on real-only after stage 2. ResNet-50 encoder + heatmap decoder, identical training schedules; performance reported on a fixed real Hi-Res test set. "None" = real-only baseline (542 real images, no synthetic).	Previous: 85.50% precision Syn.+Real: 91.63% precision (PB-CMx, 2x synthetic data + additional augmentations + real fine-tuning) Δ: +6.13 pp	Per-pixel weed (dandelion) center heatmap detection in high-resolution lawn/grass imagery for precision-agriculture / golf-course robotic weed control
2D learned generators and image/signal translation						
Abbas et al. (2021)	C-GAN (Conditional GAN)	PlantVillage tomato training split (≈9,600 images, 60% of 16,012)	class label inherited from C-GAN conditioning	synthetic + real (mixed training, ≈4,000 synthetic added)	Real: 94.34% acc. Syn.: 97.11% acc. Δ: +2.77%	tomato leaf disease classification

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Alshammari et al. (2024)	CycleGAN	DiaMOS dataset ≈3,505 images (train split ≈70%)	CycleGAN-generated labelled images for class balancing	synthetic + real (CycleGAN-augmented training set)	Original: 81.42% acc. (ResNet50) CycleGAN-aug.: 86.59% acc. Δ: +5.17 pp	pear disease classification with DiaMOS images (4 classes: healthy, leaf-spot, leaf-curl, slug damage)
Anuradha et al. (2024)	DCGAN (Adam-optimized)	Kaggle satellite image dataset (size unspecified)	No masks; image-level satellite scene/class labels	synthetic + real (mixed training)	Best compared method reported by authors: 96.11% acc. (Deep CNN [21]) Proposed AO-DCGAN+CNN: 98.6% acc. Δ: +2.49 pp	agricultural yield forecasting from satellite imagery
Aslam et al. (2025)	Customized StyleGAN	Public cotton leaf disease datasets; reported sources include Kaggle/Roboflow, with external Mendeley validation	StyleGAN-based synthetic image augmentation for class balancing	synthetic + real (class-balancing augmentation)	Proposed full pipeline: 97% acc. (StackNet); external unseen dataset: 96.4% acc.; Δ: +10.23–12.79% over the best individual classifier	cotton leaf disease classification (7 classes)

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Astuti et al. (2024)	Stable Diffusion 1.5 (with Textual Inversion + Hypernetwork)	Small source image sets for class-specific Hypernetwork generation (62 early blight, 18 late blight, 40 healthy, 36 black leg, 20 leaf roll, and 20 mosaic), plus 200 generic "potato plant" images and 20 "plant at night" images for Textual Inversion; final synthetic CvT-21 training set: 11,121 images	class label inherited from prompt/category used for generation	synthetic-only CvT training; evaluated on a 200-image validation set and 82 real Google images	Real-only baseline: N/A Synthetic-trained CvT: 84% validation acc.; 81.7% Google-image test acc. Δ: N/A	potato leaf disease classification (6 classes: 5 diseases + healthy)
Bai et al. (2024)	Conditional GAN with Case-Based Reasoning (CGANa-CBR)	27,469 64x64 patches from 11 US states, 2003-2015	paired prior real time-series patches; generated images annotated with USDA county-level yield	synthetic + real (CGAN-generated harvest-period images supplement real time series)	Real: baseline CNN RMSE Syn.: -6.3% RMSE avg. Δ: -6.3% RMSE	soybean crop yield prediction from satellite imagery

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Bamra et al. (2022)	ProjectedGAN, FastGAN, StyleGANv2	961 real microscope images (11 species)	unconditional GAN-based generation of multi-species phytoplankton microscopy images	GANs trained on real images; evaluation focuses on synthetic image quality, with no downstream classifier task	Real: N/A Syn.: FID/KID reported (StyleGANv2 best) Δ: N/A	phytoplankton synthetic image generation for HAB monitoring
Bao et al. (2025)	Diffusion model-based domain adaptation with RT-DETR + DMFB	1,248 labeled source-domain UAV images	layout-conditioned diffusion generation with source label layouts preserved for generated target-style images	pretrain on real source + fine-tune on synthetic target-domain (no real target labels)	Baseline detector: AP reported DASD: +15–23 AP points over baseline Δ: +15–23 AP	wheat scab detection from UAV images
Beltran et al. (2025)	Stable Diffusion / Easy Diffusion + LORA + Kohya	Small image sets of 200–800 samples, varying by fruit and deformity category, were used for LoRA-based generation; final real and synthetic datasets each contained 20,000 images per fruit	class-conditional text/image-to-image generation	pretrain on synthetic, fine-tune on real (Single-Input); also Multi-Input RGB+silhouette	Real: N/A Real+Syn.: 94% acc. (mangoes, Multi-Input MobileNetV2 with RGB+silhouette inputs) Δ: N/A	fruit deformity classification (apples, mangoes, strawberries)

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Benfenati et al. (2022)	Residual VAE + Pix2Pix cGAN	80 grapevine leaf images augmented to 240 were used for L2L generator training; real healthy/diseased cucumber leaf collections were used in the powdery-mildew case study to generate synthetic cucumber leaves, although the paper does not fully specify the retraining/adaptation details	Pix2Pix conditioning preserves leaf shape and venation skeleton	synthetic + real mixed U-Net training; real-only training was infeasible due to too few samples	Real-only: N/A Synthetic realism: AE anomaly-detection AUC=0.25, indicating 75% of synthetic leaves were considered real-like Downstream: >94% powdery-mildew spot segmentation accuracy on real test leaves Δ: N/A	cucumber powdery-mildew segmentation; synthetic leaf generation
Bi and Hu (2020)	WGAN-GP (Wasserstein GAN with gradient penalty) + LSR	873 real images total (1.9% of the 43,843-image PlantVillage set; 38 classes, 10–28 images per class); classic data augmentation applied at training time	Conditional WGAN preserves class label (synthetic images inherit class label from generator condition)	synthetic + real (mixed training; classic augmentation + 30 synthetic images per class per epoch)	Real-only CNN: 60.40% acc. Aug.+Syn.: 84.78% acc. (classic augmentation + WGAN-GP-LSR synthetic images) Δ: +24.4 pp over real-only CNN	

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Bird et al. (2022)	Conditional GAN (cGAN, 2000 epochs)	2690 real lemon images (public dataset)	Conditional GAN preserves class label (healthy/unhealthy) via conditioning	synthetic + real (mixed training augmentation)	Real-only: 83.77% acc. Real+Syn.: 88.75% acc. (with 400 cGAN-generated images; 200 per class) Δ: +4.98 pp	lemon fruit quality classification (good vs. defect)
Careli et al. (2024)	DALL-E, Craiyon, Tess-AI (commercial text-to-image)	N/A (off-the-shelf foundation models, prompted with original images)	synthetic ground-truth protocol not fully specified; IoU was computed by comparing model-predicted masks on generated images with corresponding true labels	real-trained model evaluated on synthetic (proposed for future augmentation)	Real test: IoU=0.762 (YOLOv8x; best IoU=0.763 for Mask R-CNN R101 FPN 3x) Synthetic-image eval.: DALL-E mean IoU=0.698 across 11 models; YOLOv8x on DALL-E images=0.637 Δ: N/A (evaluation only; no synthetic-data retraining)	forest/deforestation segmentation from remote sensing for IoT embedded monitoring
Chen et al. (2024a)	classifier-guided diffusion model (ADM-G, 2D U-Net based) vs BigGAN/ StyleGAN2/ StyleGAN3 baselines	≈4,685 (CottonWeedID10, top 10 classes, ¿=200 images each) and ≈17,509 (DeepWeeds), 90% used for training generators	class-conditional generation (class labels inherited from generator conditioning)	synthetic + real augmentation; also tested 10% real + synthetic	Real: 97.30% acc. Syn.: 98.47% acc. Δ: +1.17%	multi-class weed image classification

continued on next page

Table S1 – continued from previous page

Study	**Generative model**	**Generator training data**	**Synthetic ground-truthing**	**Integration with downstream training**	**Reported performance (real-only vs. synthetic-augmented)**	**Downstream task**
Deng and Lu (2025)	ControlNet-added Stable Diffusion	8436 real images across 10 weed classes, 3 seasons	bounding boxes from ControlNet conditioning (spatial conditioning preserves layout/labels)	synthetic + real (manually selected generated images mixed with real)	Real: 86.9% mAP@50:95 Syn.: 88.3% mAP@50:95 Δ: +1.4% mAP@50:95	multi-class weed detection in field
Drees et al. (2024)	Conditional Wasserstein GAN with gradient penalty (CWGAN-GP) using conditional batch normalization	Arabidopsis 40 train sequences (54,384 images); GrowliFlower 6,572 train sequences (8,522 total); MixedCrop ≈1,555-1,580 train sequences	generated images are conditioned on the input image and target growth conditions; plant traits (PLA/biomass) are derived from generated images using separately trained phenotyping models	generation conditioned on an initial real image and time; for MixedCrop, treatment and process-based simulated biomass can be added as extra conditions. Phenotyping models are trained separately on real images, with segmentation masks for PLA or process-based biomass references for MixedCrop	Real: N/A Syn.: improved MS-SSIM and biomass MAE with treatment+biomass conditioning Δ: qualitative reduction in phenotyping error vs no-condition baseline	crop growth simulation / image generation and biomass/leaf-area estimation for crop mixtures

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Du et al. (2025)	ADM-DDIM diffusion model fine-tuned on pest images; compared with traditional generative baselines	IP102 large-scale pest dataset (¿75,000 images, 102 classes) with a pronounced long-tailed distribution; pest images from the training set were used to fine-tune the diffusion model for generating underrepresented categories	class label inherited from the target pest category used for generation	synthetic + real (synthetic images added to underrepresented/tail categories to rebalance the long-tailed distribution)	Real: baseline accuracy on imbalanced IP102 Real+Syn.: +5.3% average classification accuracy after diffusion-based augmentation Δ: +5.3% accuracy averaged across six classification networks	long-tailed pest image classification (102 classes, IP102)
Duan et al. (2022)	StackGAN-v2-based GAN (3 generators / 3 discriminators, conditioned on 16-18 phenotypic trait vector)	20,066 panicle images; 11,218 rice plant; 10,081 maize plant; 6,680 cotton plant	condition vector = phenotypic traits extracted via LabVIEW from real images	generation/evaluation only; no downstream classifier comparison	Real: N/A Syn.: SSIM 0.83 (panicle) Δ: N/A	crop visualization (rice panicle, rice plant, maize plant, cotton plant from phenotypic traits)
Egusquiza et al. (2025)	Latent Diffusion Model with DiT backbone, conditioned on leaf segmentation mask + regression disease label (DiffusionPix2Pix); GAN baselines Pix2Pix-G, Pix2Pix-GD	17,306 wheat leaf images (80%/10%/10% train/val/test split, ≈13,845 training)	paired leaf segmentation masks + expert-annotated disease percentage labels	image-generation benchmark only; downstream data augmentation potential discussed but not evaluated	Real-image reference: 78.57% expert “real” rate Syn.: 71.3% expert “real” rate (DiffusionPix2Pix) Δ: -7.27 pp in expert “real” rate	winter-wheat leaf image synthesis with graded yellow-rust disease severity

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Fawakherji et al. (2024)	DCGAN (shape) + SPADE conditional GAN (style/texture for crop and background, RGB and NIR)	1,000 training images from Bonn sugarbeet dataset	DCGAN-generated synthetic crop mask drives SPADE generator and replaces the real crop patch/mask	synthetic + real mixed (only minority-class crop patches replaced); tested with RGB and RGB+NIR	Real: 0.70 mIoU (RGB) Real+Syn.: 0.78 mIoU (RGB shape+style) Δ: +0.08 mIoU	sugarbeet crop/weed semantic segmentation in multispectral precision farming
Gheorghiu et al. (2024)	pix2pix for leaf-mask segmentation + CycleGAN healthy-to-diseased image translation for offline class-balancing augmentation	RoCoLe dataset (training count not stated; rebalanced for class augmentation)	CycleGAN translates healthy images into target diseased classes; synthetic labels are assigned according to the target disease domain, while pix2pix masks are used to segment the foreground leaf	synthetic + real mixed (offline class-balancing augmentation) plus online augmentations (MixUp, CutMix, Cutout, FMix)	Real: ViT-small non-augmented baseline, Acc/F1 = 64.7/46.3 Mixed real+CycleGAN: Acc/F1 = 73.7/57.2 Δ: +9.0 Acc. points and +10.9 F1 points under the matched dropout + lr=0.0002 configuration	Robusta coffee leaf disease classification (healthy / red spider mite / rust low/medium/high)
Haruna et al. (2023)	StyleGAN2-ADA (SG2-ADA) with Laplacian-variance blur filter for image curation	BB 1,584 / TG 1,308 / BS 1,440 / RB 1,600 real images for SG2-ADA training; 26,694 synthetic images produced	class label inherited from SG2-ADA training class; manual annotation of synthetic bounding boxes	synthetic + real mixed (GAN-based augmentation vs standard augmentation)	Real: 0.84 mAP (Faster-RCNN) Syn.: 0.93 mAP (Faster-RCNN GAN-aug) Δ: +0.09 mAP	rice leaf disease detection (bacterial blight, tungro, brown-spot, rice-blast)

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Hu et al. (2024b)	DMFGAN (DCGAN-based with multi-feature extraction block, MFLoss)	300 real images per category from four PlantVillage grape classes, 1200 images in total	class-specific synthetic images assigned the corresponding PlantVillage class label; no image-to-image label transfer	synthetic + real (mixed training of classifier)	Real: 79.44% Acc. (None + ResNet50) Mixed real+DMFGAN: 95.82% Acc. (DMFGAN + ResNet50) Δ: +16.38 Acc. points	grape leaf disease classification
Huo et al. (2025)	StyleGAN3	1200 real tomato growth-stage images across 4 stages (per-category StyleGAN3 training for 1000 epochs; 2723 synthetic generated)	inherited class label per StyleGAN3 model trained per growth-stage category	synthetic + real (combined training)	Real: 94.58% acc (ViT-Base) Syn.: 98.39% acc (ViT-Base + StyleGAN3) Δ: +3.81% acc	tomato growth-stage classification
Jiang and Ostadabbas (2023)	VAE pose generation + ControlNet (Stable Diffusion + HED boundary, Bi-ControlNet)	3D animal meshes; synthetic animal renderings conditioned by HED boundary maps; 300 real natural background scenes for stylization	automatic labels from the 3D pose pipeline via bone vector transformations; pose structure maintained during HED-guided ControlNet stylization	synthetic + real training (99 real + 3000 SPAC synthetic images per animal)	Zebra: R(99) PCK@0.05=78.7%, R(99)+SPAC(3K)=96.3%, Δ=+17.6; Rhino: R(99)=88.3%, R(99)+SPAC(3K)=95.9%, Δ=+7.6	zebra and rhino 2D pose estimation

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Kang et al. (2025)	RipenessGAN (U-Net encoder–decoder generator with class embedding Rˆ32 + Growth Day Embedding Rˆ16, self-attention at 48x48; PatchGAN discriminator with spectral normalization; hinge adversarial loss + Gaussian-weighted temporal-consistency loss with sigma=7 days). Compared with standard CycleGAN (9 ResNet blocks + PatchGAN) and class-conditional DDPM (UNet2DCondition, 1000 timesteps, 50 sampling).	Open-access jujube dataset (Ban et al. 2023) with 5000 original images across five maturity stages in the source dataset (700 train + 200 val + 100 test per class); this study models four ripeness classes (green / white-ripe / semi-red / fully-red), with images resized to 192×192 and augmented to 20000 via rotation, flipping, and brightness adjustment.	Class label from condition (class c in {0,1,2,3} + growth day d in [0,56] supplied as input embeddings to generator); discrete maturity class label inherited from the conditioning input.	ResNet-18 classifier trained under 4 settings: (1) Original only; (2) Original + CycleGAN-augmented; (3) Original + DDPM-augmented; (4) Original + RipenessGAN-augmented. Same train/val/test split.	Real: 97.75% accuracy (Original only) Mixed real+RipenessGAN: 98.75% accuracy Δ: +1.00 pp	Jujube ripeness classification (4 stages: green / white-ripe / semi-red / fully-red) via ResNet-18

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Karam et al. (2022)	CPB+GAN semi-automated augmentation pipeline (web-interface-based Copy-Paste-Blend tool; DCGAN augments the whitefly source-mask/object pool)	home-collected Bemisia tabaci whitefly dataset: 770 images with ≈3400 labeled objects; key low-data training subset of 140 images augmented to 560 images	whitefly labels are inherited from the source object class; CPB creates bounding-box annotations at the pasted locations, while DCGAN only diversifies the source whitefly mask/object pool	synthetic + real (dataset increased from 140 to 560 images via cut-paste, then GAN diversification)	Real: recall@IoU0.5= 54.4% (PestNet trained on 140 real images) Mixed real+CPB+GAN: recall@IoU0.5= 94.6% (140 images augmented to 560) Δ: +40.2 recall points	whitefly (Bemisia tabaci) object detection

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Khalil et al. (2026)	Dual metaheuristic GAN with Chaoborus generator loss + Australian Crayfish discriminator loss + identity-block-based ResNet-style skip connections for thermal-signature preservation; compared with StyleGAN2 / Progressive GAN / BigGAN baselines.	445 thermal training images from a 636-image FLIR-E8 paddy-leaf dataset (Vellore Institute of Technology, Tamil Nadu, India; 6 classes: Bacterial Leaf Blight 220, Blast 67, Leaf Spot 80, Leaf Folder 34, Hispa 142, Healthy 93). For each original training image, five GAN-generated synthetic variations were produced, yielding 2670 training images including originals.	Disease labels are inherited from the original source images used for augmentation; identity blocks preserve disease-specific thermal signatures rather than class labels, and GAN-generated thermal variations are labeled with the corresponding source disease class	Train classifiers (ViT, ResNet-50, EfficientNet-B7, DenseNet-201) on (a) Original 445-image set; (b) Standard augmentation; (c) Proposed dual-metaheuristic GAN augmentation (2670 images). Validation/test sets held with original images only (95 val, 96 test).	Real: ViT 83.45% accuracy on the original training set Mixed real+proposed GAN: ViT 97.89% accuracy Δ: +14.44 pp	Paddy leaf disease classification (6 classes including healthy) on thermal imagery

continued on next page

Table S1 – continued from previous page

Study	**Generative model**	**Generator training data**	**Synthetic ground-truthing**	**Integration with downstream training**	**Reported performance (real-only vs. synthetic-augmented)**	**Downstream task**
Krosney et al. (2023)	CUT GAN (contrastive unpaired translation)	unpaired indoor/-composite and outdoor single-plant images across four species; the larger all-species CUT setting used 512 input-domain composites and 512 outdoor target images, while downstream detector datasets contained 80000 synthetic images each for Baseline/-Composite/GAN and 160000 images for Merged	plant labels and bounding boxes are inherited from the inserted indoor plant images/masks; CUT translation largely preserves plant position, enabling reuse of the source bounding boxes after synthetic field-style translation	synthetic only (training); evaluated on 253 real multi-plant field images	Baseline synthetic blue-screen: mAP@0.5=0.010 Merged synthetic Composite+GAN: mAP@0.5=0.528 Δ: +0.518 mAP@0.5	single-class plant detection with YOLOv5 nano; generator/training data include four species (canola/oat/-soybean/wheat), while downstream evaluation uses real canola and soybean field images
Li et al. (2024a)	CA-GAN (channel-attention conditional GAN)	sPSD (9 plant seedling classes) and ISAS dataset (5 weed species); 1000 synthetic samples per class generated	class-conditioned generation; synthetic samples are assigned the class specified by the input class code, rather than preserving a source-image label	synthetic images evaluated for class discriminability using a ResNet-56 auxiliary classifier trained on real images; no downstream synthetic+real classifier-training experiment	Aux. ResNet on real sPSD: 99.22% accuracy Aux. ResNet on CA-GAN synthetic sPSD: 82.63% accuracy Δ: -16.59 pp	weed species recognition / synthetic weed image generation

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Li et al. (2026b)	PMT-GAN (improved CycleGAN with multi-scale feature extraction module, Swin Transformer blocks in the generator, and SSIM-based cycle-consistency loss; max 512 channels, intermediate 32-channel layer, 512x512 input)	2061 manually curated 512x512 pineapple-instance images (Dataset 1: 681 immature, 361 early-maturity, 349 mid-maturity, 670 full-maturity), pre-augmented to Dataset 2 (Dori+aug = 4039 images) which is then used to train PMT-GAN; original 1061 field images	Class label assigned according to the target maturity-stage domain (cycle-consistent translation between four maturity-stage domains; bounding-box locations can be reused when image geometry is preserved)	Three datasets compared on identical YOLOv5n/v8n/v10n/v11n/v12n training: Dori (original, 2061 images), Dori+aug (traditional augmentation), Dori+gan (PMT-GAN-augmented, 7307 images total balanced across stages: Stage 1: 1787, Stage 2: 1847, Stage 3: 1848, Stage 4: 1825). Test set held constant.	Dori: YOLOv8n mAP@0.5=96.7% Dori+gan: YOLOv8n mAP@0.5=99.0% Δ: +2.3 pp	Pineapple-maturity object detection (4-stage classification: Immature, Early-maturity, Mid-maturity, Full-maturity) for harvesting robots

continued on next page

Table S1 – continued from previous page

Study	**Generative model**	**Generator training data**	**Synthetic ground-truthing**	**Integration with downstream training**	**Reported performance (real-only vs. synthetic-augmented)**	**Downstream task**
Liu et al. (2024b)	StyleGAN2-ADA	8920 real maize leaf-spot images used to train StyleGAN2-ADA; expanded to 14,600 images	LabelImg manual bounding-box annotation on real images; generated images inherit class label from StyleGAN2-ADA (FID 27.33)	real only vs real + StyleGAN2-ADA-generated synthetic (combined training)	Real: F1 0.85 Syn.: F1 0.89 Δ: +0.04 F1	Maize leaf-spot disease region detection
Liu et al. (2024a)	CycleGAN with self-attention (SM-CycleGAN)	Tobacco: 165 normal-color + 165 color-distorted training images; tea: 64 images per class/category for training across five tea categories	GAN preserves source class (image-to-image translation between healthy/diseased domains)	synthetic image generation; evaluated by image-quality metrics (PSNR/SSIM), not downstream classifier	Baseline: CycleGAN PSNR/SSIM SM-CycleGAN: PSNR +3.55% and SSIM +2.48% on tea-leaf disease images Δ: improved image-quality metrics, not downstream detection/classification accuracy	crop image data enhancement (tobacco-roasting color restoration, tea leaf disease)

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Meekeren et al. (2023)	Stable Diffusion 2.1-base text-to-image generation with positive/negative prompt engineering	670 generated images; no task-specific Stable Diffusion training/fine-tuning; pretrained model used with prompts	final automatic annotations generated by a custom YOLOv5m detector trained on MinneApple; Mask R-CNN was first tested but was unsuitable because it also labeled ground apples	synthetic-only vs. real-only training; separate YOLOv5m models, both evaluated on the MinneApple test set	Real-only: AP@50 0.70 Synthetic-only: AP@50 0.61 Δ: -0.09 AP@50	apple detection in orchard images / MinneApple benchmark
Miski (2026)	OpenAI Sora diffusion-transformer for video synthesis	Structured prompts plus 16 PlantVillage reference images; Sora was used zero-shot / not fine-tuned. The extracted and curated frames formed a 1,467-image synthetic training set for the downstream classifier	Prompt-derived binary labels (healthy/damaged), with manual curation to remove blurry or low-quality frames; no pixel-level or per-frame manual annotation was reported	Synthetic-only classifier training; real images used for validation and final testing on 618 real images	Real-only: N/A Synth-trained → real test: Accuracy 98.71%, F1-score 98.71% for ResNet-18 Δ: N/A	strawberry leaf binary health classification (healthy vs. damaged/stressed)

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Modak and Stein (2024a)	Stable Diffusion 1.5 fine-tuned with SAM-derived masks	real pseudo-RGB sugar-beet/weed images with manual annotations	YOLOv8x-based model-guided annotation of synthetic images	real + synthetic mixed training, with synthetic share varied from 10% to 200%; compared with conventional augmentation	YOLOv9t scratch: mAP50 0.608 → 0.878; Δ +0.270	sugar-beet crop/weed detection
Moreno et al. (2023)	Stable Diffusion 2.1, with 512px and 768px models fine-tuned on cropped weed instances	30 real cropped bounding-box images per weed species	Automatic ExGR-based segmentation and OpenCV contour extraction for synthetic bounding boxes	synthetic-only, real-only, and mixed real + synthetic training	Mixed training improved mAP@0.5 from 0.962 to 0.992 under isolated conditions and from 0.849 to 0.930 under inter-weed/overlap conditions for YOLOv8l; RetinaNet improved from 0.960 to 0.989 and from 0.837 to 0.928, respectively	weed object detection

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Muhammad et al. (2023)	RePaint, a DDIM/DDPM-based diffusion inpainting method, compared with InstaGAN	PlantVillage subset with tomato and grape leaf images: 9 tomato disease classes, 3 grape disease classes, and healthy tomato/grape classes	Mask-guided synthesis/inpainting using leaf segmentation masks; U-Net masks were used for guidance, with split/inverted masks for RePaint	Synthetic image generation was evaluated using FID and KID; no downstream classifier or detector accuracy was reported	Image-quality comparison: average FID decreased from 206.02 with InstaGAN to 138.28 with RePaint, and average KID decreased from 0.159 to 0.089; these are generative-quality metrics rather than real-only vs. synthetic-augmented downstream results	plant disease image augmentation for tomato and grape leaves
Mullins et al. (2025)	DALL-E 2 (image variations from real seed images)	real seed images sampled from 66 ripe blueberry, 66 red leaf, and 100 hair fescue ground-truth images	qualitative human screening of DALL-E 2 variations; accepted generated images were manually annotated with bounding boxes in Roboflow, while images with morphology/artifact errors were discarded	ground-truth only, generated only, and combination dataset with no more than 40% generated images	Real-only: mAP50 0.615 (red leaf) Real+Syn.: mAP50 0.848 (red leaf, combination) Δ: +0.233 mAP50	object detection (ripe fruit, weed, and disease)

continued on next page

Table S1 – continued from previous page

Study	**Generative model**	**Generator training data**	**Synthetic ground-truthing**	**Integration with downstream training**	**Reported performance (real-only vs. synthetic-augmented)**	**Downstream task**
Myers et al. (2024)	Modified CycleGAN (segmentation-mask preserving)	synthetic dataset Δ_S; a single manually annotated wheat-field frame plus video frames	GAN preserves source mask (modified CycleGAN enforces segmentation consistency)	synthetic only to fine-tune with pseudo-labeled real data	Real: N/A (no real-only baseline; cut-paste-synthetic Δ_S baseline Dice 0.709 internal) Syn.: Dice 0.834 (internal, final) Δ: +0.125 Dice (over cut-paste-synthetic Δ_S, not real-only)	wheat head semantic segmentation
Rai et al. (2025a)	Stable Diffusion 3.5M fine-tuned with DreamBooth and LoRA	≈30 real images per disease/status category were used for SD fine-tuning (32 anthracnose, 35 downy mildew, 34 healthy, and 30 mosaic virus images); 750 real images total across three classes were used as the H0 real-only baseline	class labels inherited from class-specific SD fine-tuning and prompt-conditioned generation; real-field test images were expert-validated	H0: real only; H1: synthetic only; H2: 1:1 real-to-synthetic; H3: 1:10 real-to-synthetic; H4: H3 plus an unknown ImageNet class	Real-only H0: weighted F1 = 0.65; synthetic-only H1: weighted F1 = 0.74; hybrid H2: weighted F1 = 0.92; hybrid H3: weighted F1 = 1.00; H4: weighted F1 = 0.99. Compared with H0, H3 improved by +0.35 weighted F1	watermelon disease/status classification into fungal, healthy, and virus classes; H4 additionally included an unknown class
Rai et al. (2025b)	Stable Diffusion variants (SDXL, SD3.5M, SD3.5L) + DreamBooth + LoRA	36 real anthracnose images in the reported experiments: 12 field-collected + 24 open-access IPM images; 500 synthetic images generated for benchmarking	Prompt-driven class conditioning using disease-specific identifiers and prompt weights; no pixel/box/keypoint ground-truthing reported	No downstream classification or detection training reported; study focuses on synthetic image generation and computational benchmarking	Real-only vs. synthetic-augmented: N/A SD3.5M: ∼18 GB inference memory, 1.02 kWh/500 images, LPIPS ∼0.36 Δ: N/A	Synthetic watermelon disease image generation, mainly anthracnose

continued on next page

Table S1 – continued from previous page

Study	**Generative model**	**Generator training data**	**Synthetic ground-truthing**	**Integration with downstream training**	**Reported performance (real-only vs. synthetic-augmented)**	**Downstream task**
Rana and Gatti (2025)	Modified WGAN-GP with transposed convolutions, dropout, and selective batch normalization	876 RGB and 850 IR leaf-bound clipped samples of *Raphanus raphanistrum* (155×115 px), split 80/20 for training/testing	No pixel-level or bounding-box synthetic labels were generated; the study focused on image-level synthetic sample generation and SSIM-based similarity evaluation	No downstream detector or segmentation model was trained with the synthetic images; integration was discussed as a potential augmentation use case	No real-only vs. synthetic-augmented downstream baseline was reported. Image-quality evaluation showed improved SSIM for modified WGAN-GP over the unmodified WGAN-GP: RGB SSIM 0.5364 → 0.6615 (Δ=+0.1251), and IR SSIM 0.3306 → 0.4154 (Δ=+0.0848)	Synthetic RGB and IR weed image generation for *Raphanus raphanistrum*
Velázquez-González et al. (2024)	Stable Diffusion + VGG-16 supervisory classifier	15–25 Internet images for new crop-node training	prompt labels with classifier-based filtering at 90% confidence and feedback-loop supervision	synthetic-only training; real-image evaluation	N/A for real-only comparison; +23.85% precision and +10.8% recall over CNN baselines using the same synthetic dataset	stressed fern/tree classification
Sapkota and Karkee (2025b)	DALL-E 2 text-to-image generation	pretrained DALL-E 2; no task-specific generator training; 501 synthetic orchard images generated, with 489 retained after quality filtering for YOLO training	manual bounding-box annotation of synthetic images	synthetic-only training; models trained entirely on LLM-generated images and qualitatively tested on real orchard images	Real-only: N/A Syn.-only: mAP@50=0.978, precision=0.916, recall=0.969 (YOLOv12n, synthetic test set) Δ: N/A	apple detection in orchards

continued on next page

Table S1 – continued from previous page

Study	**Generative model**	**Generator training data**	**Synthetic ground-truthing**	**Integration with downstream training**	**Reported performance (real-only vs. synthetic-augmented)**	**Downstream task**
Sapkota et al. (2024)	OpenAI DALL-E text-to-image generative model	N/A; pretrained DALL-E was used without task-specific generator training	Manual bounding-box annotation of LLM-generated apple images	Synthetic-only training/-validation; real-camera testing on commercial Scilate orchard images	Real-only baseline: N/A Synthetic-only trained: mAP@50=0.91 on real Nikon images (YOLO11x); mAP@50=0.855 on real Azure Kinect images (YOLO11x) Δ: N/A	Apple detection in commercial orchards (object detection)
Sapkota et al. (2025)	DALL-E text-to-image generation; YOLO11 + SAMv2 for automatic annotation	Not reported; pretrained DALL-E was used to generate synthetic apple-orchard images from text prompts	automatic mask annotation using zero-shot YOLO11 bounding boxes followed by SAMv2 mask generation	synthetic-only training; YOLO11-seg models trained on automatically annotated synthetic images, with real orchard images used only for validation	Real-only: N/A Syn.-only: mask P=0.902 and mask mAP@50=0.833 on real orchard validation images (YOLO11m-seg) Δ: N/A	apple instance segmentation in commercial orchards
Tan et al. (2025a)	SD-v1.4 + DreamBooth + image-guided diffusion	Captioned indoor/outdoor plant images; PlantVillage/CropDisease; ≈68k indoor soybean images	Inherited source class labels; compositing-derived detection labels for soybean–weed scenes	Real + synthetic/translated images under varying augmentation ratios	Tomato acc.: 0.779 → 0.843 at 400% synthetic augmentation (+6.4 pp); weed detection precision/recall also improved	Phenotype classification; soybean–weed detection/classification

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Ullah et al. (2025)	FastGAN + Pix2Pix	FastGAN: 300 barley, 120 Arabidopsis, and 120 maize RGB images; Pix2Pix: 100 barley RGB–mask pairs and 80 RGB–mask pairs each for Arabidopsis and maize	Synthetic RGB images were generated by FastGAN, and corresponding binary masks were predicted by a Pix2Pix RGB-to-mask model trained on manually annotated real RGB–mask pairs; a subset of synthetic outputs was manually annotated for Dice-based validation	Not used to train a downstream segmentation model; synthetic RGB images served as inputs for Pix2Pix mask generation, and the generated masks were evaluated against manual annotations	Real-only vs. synthetic-augmented: N/A Pix2Pix mask accuracy on synthetic images: Dice=0.96 barley, 0.94 Arabidopsis, 0.95 maize at 1024×1024 Δ: N/A	generation of RGB–binary mask pairs for greenhouse plant shoot segmentation
Wang et al. (2025b)	Video diffusion model (DynamiCrafter prior) + custom agro-controllable parameter injection	≈5000 real crop growth images captured over 4-5 months (multi-camera fixed-position)	Self-supervised video generation (no per-frame labels); evaluated against real held-out frames	synthetic only (video generation task; compared with open-domain video models)	Real: N/A Syn.: FVD 342.1 Δ: N/A	Crop full-growth-cycle video synthesis / phenotype simulation

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Wang et al. (2023)	Conditional GAN (pix2pix-style image translation)	4,500 synthetic strawberry instances generated by image compositing	GAN preserves source mask (synthetic input has known segmentation)	synthetic only (cGAN trained on synthetic; tested on real)	Real: N/A Syn.: qualitative success Δ: N/A	strawberry ripe-fruit detection and stem cutting in vertical-growing greenhouse
Wang et al. (2025a)	Dual-discriminator GAN (FHWD with FFT + wavelet losses)	PlantVillage (10 crop diseases used for training)	inherited from source (class-conditional generation)	synthetic + real (data augmentation for classification)	Real: baseline accuracy Syn.: +7.25% accuracy Δ: +7.25%	tomato leaf disease classification (9 disease types)
Wu and Xu (2021)	Improved Adversarial-VAE (two-stage VAE-GAN + VAE) with multi-scale residual learning and dense connections	10,892 real tomato leaf images from PlantVillage training split (60% of 18,162)	Class labels inherited from the source class; one generative model trained per tomato leaf class	Synthetic + real mixed training; training set doubled from 10,892 to 21,784 images	Real-only: 82.87% Synthetic-augmented: 88.43% Δ: +5.56 percentage points	Tomato leaf disease/health classification (10 classes)
Xiang et al. (2025)	DODA: latent diffusion with LI2I layout-image-to-image conditioning and ViT-B/CLIP domain embeddings	GWHD labeled detection data, plus unlabeled GWHD/GWHD 2+ images for diffusion pre-training	Input layouts provide bounding-box ground truth, preserved by LI2I generation	Synthetic domain-specific detection data generated from target-domain reference images; detector fine-tuned on generated data for adaptation	DODA adaptation improved AP by 7.5 on average across GWHD domains, with a maximum AP gain of +15.6	wheat head detection/domain adaptation

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Xu et al. (2025a)	CycleGAN (unpaired sunny¡-¿cloudy domain translation)	Real UAV soybean field images under sunny/overcast conditions (817 sunny + 364 cloudy)	Existing pixel-level weed masks transferred via CycleGAN translation; SSIM 0.702 between real and synthetic	synthetic + real (two-stage: CycleGAN-augmented training set)	Real: IoU 0.812 Syn.: IoU 0.921 Δ: +0.109 IoU	UAV soybean weed semantic segmentation under varying weather
Yu et al. (2025)	Improved CycleGAN (Tea CycleGAN with SKConv multi-scale fusion)	2000 real images (1000 each Longjing 43 / Zhongcha 108, day + night)	GAN preserves source bounding-box layout via restoration-paste strategy	synthetic + real (data augmentation)	Real: mAP 73.94% Syn.: mAP 83.54% Δ: +9.6% mAP	tea shoot detection (cross day-night, cross cultivar)
Zhang et al. (2025)	DADM (Denoising Diffusion Probabilistic Model with spatial+channel attention block in U-Net)	Corn/rice/soybean grain image datasets (1500 images per grain type, 750 per variety; collected from harvester)	DADM-generated images manually screened for label accuracy; original labels reused for similar shapes	synthetic + real (DADM enhancement vs original vs traditional augmentation)	Real: MIoU 75.36% (DeepLabV3+, corn breakage grain, no enhancement) Syn.: MIoU 83.62% (DeepLabV3+ + DADM enhancement) Δ: +8.26 pp MIoU for this model; +5.74 pp average over seven models	Semantic segmentation of grain breakage and impurities for harvest quality monitoring

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Zhang et al. (2024b)	Guided-GAN / Across-CycleGAN (multi-dimensional phenotypic feature extraction)	pear source domain images + target domain images (unsupervised cross-domain)	GAN preserves source mask/bounding boxes (automatic cross-species labeling)	synthetic only (auto-labeled target-domain dataset for detector training)	Pitaya: EasyDAM_V2 mAP 82.1% → Proposed mAP 87.8% Eggplant: synthetic-only mAP 69.12% → Proposed mAP 87.0% Cucumber: synthetic-only mAP 64.8% → Proposed mAP 80.7%	cross-species fruit detection auto-labeling (pear -¿ pitaya/eggplant/cucumber)
Zhang et al. (2024c)	StyleGAN	497 original training images	manual annotation of generated samples (LabelImg)	synthetic + real (sample balancing across flight heights)	Real: 95.2% Syn.: 95.6% Δ: +0.4%	sandalwood tree detection (UAV remote sensing)
Zhang (2025)	Improved CycleGAN (with feature-map loss for fine-grained texture/style)	96 real loquat images across 3 classes (anthracnose, fruit splitting, normal); expanded to 1720 with augmentation	Class labels preserved across CycleGAN domain translation; manual augmentation also applied	synthetic + real (combined augmented training set vs original 96 images)	Real: 93.75% Syn.: 97.42% Δ: +3.67%	Loquat disease image classification (3 classes)

continued on next page

Table S1 – continued from previous page

Study	**Generative model**	**Generator training data**	**Synthetic ground-truthing**	**Integration with downstream training**	**Reported performance (real-only vs. synthetic-augmented)**	**Downstream task**
Zhang et al. (2025b)	Supervised conditional GAN for leaf completion, with pix2pix using a U-Net generator and PatchGAN discriminator; compared against pix2pixHD, pix2pix-Turbo, CycleGAN, U-GAT-IT, and CycleGAN-Turbo.	The dataset included 110 lettuce canopy images for instance segmentation and 782 detached-leaf images for completion-model development. Canopy images were split 7:3 for segmentation evaluation, and detached-leaf images were split 8:2 for completion training/testing. The completion models were trained on paired incomplete-leaf and complete-leaf images resized to 256×256 and placed on a 1024×1024 black background.	Each in-vivo segmented incomplete leaf was paired with its corresponding ex-vivo detached complete leaf through labeled canopy/leaf ID matching. ExG–Otsu segmentation was used to extract the detached complete leaf as the completion target, providing paired image-level supervision for supervised CGAN training.	A single supervised image-to-image translation framework was used for leaf completion. Evaluation was conducted on the held-out 20% detached-leaf test split using R^2 and RMSE for extracted leaf traits, including length, width, perimeter, and area, together with SAMScore for semantic similarity. Unsupervised unpaired translation models such as CycleGAN and U-GAT-IT were used as baselines to contrast paired supervised CGANs with unpaired CGANs on the same completion task.	No real-only vs. synthetic-augmented downstream-training comparison was reported. Unpaired image-translation baseline: CycleGAN leaf-area $R^2 = 0.067$ Paired supervised CGAN: pix2pix leaf-area $R^2 = 0.948$ Δ: +0.881 R^2	Leaf completion under canopy occlusion for butterhead-lettuce phenotyping, followed by extraction of leaf morphology traits including length, width, perimeter, and area.

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
3D procedural modelling, rendering engines, and simulators						
Aghamohammadesmaeil-ketabforoosh (2024)	Blender procedural 3D rendering (Cycles)	N/A (Blender procedural, not learned from real images)	automatic masks from Blender during rendering	synthetic-only training with domain confusion, tested on real webcam images	Real: N/A Syn.: DSC=98% (ripe, val); 92% (real-time test) Δ: N/A	strawberry disease+ripeness classification and ripe/unripe segmentation
Angarano et al. (2024)	Blender procedural (3D plant models)	N/A (procedural)	automatic from 3D simulator	synthetic only training, tested on real (Sim2Real DG)	No real-only vs. synthetic-augmented comparison. DG/Sim2Real benchmark: Ours 72.45% mean IoU vs. ERM synthetic-training baseline 69.83% (Δ = +2.62 IoU points).	multi-crop semantic segmentation / domain generalization
Barth et al. (2018)	Blender procedural (3D plant model)	21 plant parameters from 115 positions on 15 stems; 750 empirical images of 50 plants (used as reference, not for training)	automatic from 3D simulator (per-pixel class + depth labels)	synthetic dataset release (companion paper handles training)	No explicit real-only vs. synthetic-augmented mean-IoU delta reported	sweet pepper plant-part semantic segmentation (dataset paper)
Barth et al. (2019)	Blender procedural rendering (3D plant model from prior work)	10,500 synthetic, 50 empirical (model from Barth 2018)	automatic from 3D simulator	pretrain on synthetic, fine-tune on real	Real: lower IoU than bootstrapped Syn.: mean IoU=0.40 (syn. bootstrap + 30 empirical FT) Δ: stabilizes ≈3,000 syn. images	sweet pepper plant-part semantic segmentation (7 classes)

continued on next page

Table S1 – continued from previous page

Study	**Generative model**	**Generator training data**	**Synthetic ground-truthing**	**Integration with downstream training**	**Reported performance (real-only vs. synthetic-augmented)**	**Downstream task**
Carbone et al. (2022)	Unity game engine (procedural fields with extracted real textures, NIR simulated via aligned textures)	34 real sunflower training images (105 real sunflower images for evaluation; sugar-beet prior variant: 1034 real images, 820 for evaluation)	automatic from Unity simulator (per-pixel labels from simulation)	synthetic-only, real-only, and synthetic+real training (Unity, Reduced, Augmented, Reduced Augmented)	Real: 69.98% mIoU (sunflower, RGBN real-only training) Best Aug.: 78.98% mIoU (Reduced Augmented, RGBN; 34 synthetic + 34 real) Δ: +9.0 pp mIoU vs real-only	sunflower-weed semantic segmentation, with comparison to prior sugar beet-weed segmentation
Carbone et al. (2020)	Unity game engine with simulated NIR channel	1034 real sugar-beet images used as texture/asset basis	automatic from Unity simulator	synthetic only, real only, synthetic + real (mixed)	Real: 78.56% mIoU (RGBN) Syn.: 70.38% mIoU (Unity+Real) Δ: -8.18% mIoU	sugar beet plant-weed semantic segmentation
Chuquimarca et al. (2023)	Unreal Engine (3D banana model with 4 ripeness levels)	3,495 real banana images and 161,280 synthetic	automatic from Unreal Engine simulator (synthetic ripeness label assigned)	pretrain on synthetic, fine-tune on real (transfer learning)	Real: 0.872 acc. Syn.: 0.917 acc. Δ: +4.5%	banana ripeness classification (4 levels)
Cicco et al. (2017)	Unreal Engine 4 (procedural agricultural scene generation with real-world plant/soil textures)	≈1600 real images (Real A: 700; Real B: 900) used as texture/reference basis	automatic from Unreal simulator (per-pixel labels)	real only, synthetic only, real-augmented (200 real + Synthetic D)	Real: IoU=83.1 (Real B, Test B) Syn.: IoU=76.2 (Real-Augmented); 61.1 (Syn. D) Δ: syn.-only IoU below real	pixel-wise crop/weed/soil segmentation (sugar beet)

continued on next page

Table S1 – continued from previous page

Study	**Generative model**	**Generator training data**	**Synthetic ground-truthing**	**Integration with downstream training**	**Reported performance (real-only vs. synthetic-augmented)**	**Downstream task**
Doig et al. (2025)	Blender + Infinigen procedural generation (3D modelling with Volume Absorption/Scattering shaders)	N/A (procedural generation, no GAN training)	automatic from 3D simulator (bounding boxes / segmentation masks generated during render)	synthetic only + unlabelled real (UDA: Mean Teacher / ALDI++); also synthetic + labelled real augmentation	Real: AP50=86.9 (UDD oracle) Syn.: AP50=75.0 (Mean Teacher, UDD) Δ: -11.9 AP50 vs UDD oracle	sea urchin (black spiny) detection in benthic underwater imagery
Dolata et al. (2021)	3D simulation/rendering (custom physics-based potato simulation environment with Python integration, built on Blender)	N/A (procedural 3D modelling parameterized by statistical shape distribution; segmentation network uses real annotations)	automatic per-pixel masks and exact physical dimensions from simulation	real images for instance segmentation; synthetic images only for training the size-regression model (knowledge transfer to real)	Real-data instance-seg model PQ 0.889 (RQ 0.976, SQ 0.911) Synthetic-data model PQ 0.876 (RQ 0.948, SQ 0.922; differences not significant) Synthetic-trained size regressor R^2=0.869 (range 0.829–0.890) on simulated validation; real-harvester test had no quantitative reference	potato instance segmentation + minimal-diameter size estimation for on-line yield monitoring
Duboudin et al. (2019)	Blender procedural (Weber-Penn tree model add-on, HDRI lighting, domain randomization)	N/A (procedural; no learned generator)	automatic semantic-segmentation maps from 3D simulator	synthetic only with embedded domain randomization	Real: N/A Syn.: N/A (no quantitative evaluation) Δ: N/A	fruit-tree image generation for semantic segmentation / object recognition in robotic harvesting

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Feng et al. (2025)	Blender (Cycles physically-based rendering) of 3D chicken carcass models	N/A (procedural Blender; up to 1000 synthetic images generated)	automatic per-instance ID masks from Blender render	synthetic + real mixed at ratios 0:60, 250:60, 500:60, 750:60, 1000:60	Real: 0.784 segm mAP (YOLOv11-seg) Syn.: 0.833 segm mAP (YOLOv11-seg, Syn-1000) Δ: +0.049 segm mAP	chicken-carcass instance segmentation in poultry-processing imagery
Grski (2024)	Unity game engine procedural composition (cutout/-paste with randomization)	dissected from limited natural NOAA rockfish dataset (fish cutouts + backgrounds)	automatic bounding-box labels from Unity Perception package	synthetic-only and synthetic+natural background variants compared with natural-only	Real: 20.51% accuracy (natural-only) Syn.: 79.19% accuracy (Unity2; highest frame coverage) Δ: +58.68 pp accuracy	fish detection in underwater video

continued on next page

Table S1 – continued from previous page

Study	**Generative model**	**Generator training data**	**Synthetic ground-truthing**	**Integration with downstream training**	**Reported performance (real-only vs. synthetic-augmented)**	**Downstream task**
He et al. (2025a)	Unreal Engine 5 (UE5) procedural simulator + scanned 3D leaf textures	N/A (simulator-based; not data-driven training of generator)	automatic from 3D simulator (per-pixel masks, healthy/ disease/ background/ fruit)	pretrain/train on synthetic, UDA to unlabeled real target domain (no real labels used)	Real: N/A (no real labels used; UDA from synthetic to *unlabeled* real — no supervised real-only baseline exists in source) Syn.: synthetic-trained UDASeg reaches mIoU 81.96% on real field test (best, with depth-denoise + bottom-extension test-input preprocessing); 41.10% on raw test images Δ: N/A vs real baseline (the +40.86 mIoU is a test-input-preprocessing effect on the same model, Table 3 RI vs IEPD, not a synthetic-vs-real gain)	tomato foliar-disease semantic segmentation / severity quantification
Ishiwaka et al. (2021)	Foids 3D simulation platform (custom bio-inspired Boids-style simulator + 3D rendering)	N/A (procedural simulation, no data-driven generator training); 10,000 synthetic input-output pairs generated	automatic from 3D simulator (2D/3D bounding boxes, silhouettes from mesh geometry)	synthetic only (YOLOv4 trained purely on Foids synthetic data, tested on real videos)	Real: N/A Syn.: 264 fish (vs 272 human count) Δ: N/A (synthetic-only)	fish detection and counting in salmonid sea cages

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Ivanovic et al. (2022)	Blender procedural 3D rendering (Cycles)	N/A (procedurally generated; references prior synthetic capsicum dataset)	Automatic bounding boxes from 3D scene geometry in Blender	pretrain on synthetic, fine-tune on small real labeled set	Real: N/A Syn.: yield within 10% margin (sim) Δ: N/A	Sweet pepper detection and fruit counting / yield estimation
Katyara et al. (2021)	Blender 3D vine models (three hand-designed vine models + domain randomization)	N/A	automatic from 3D simulator	synthetic only (sim2real transfer)	Real: N/A Syn.: Faster R-CNN (ResNet-101) spur-detection mean precision 0.679; pruning skill validated sim2real qualitatively Δ: N/A	grape vine spur detection for robotic pruning
Khan et al. (2025b)	L-systems (procedural plant model)	N/A (procedural; 1350 maize synthetic images and 1200 canola synthetic images per calibration variant, five variants C1S–C5S)	automatic from procedural model	synthetic only and synthetic + real (mixed)	Real: maize MAE=0.66, canola MAE=1.06 (all real images, tested on 100 real) Syn.: maize MAE=1.05, canola MAE=1.59 (synthetic-only, best canola variant C4S; tested on 100 real) Δ: synth-only is 1.59× worse for maize, 1.5× worse for canola	maize leaf counting and canola inflorescence branch counting
Khan et al. (2024)	L-systems (procedural plant model)	N/A (procedural; 1350 synthetic maize and 1200x5 synthetic canola images)	automatic from procedural model	synthetic only, synthetic + real, varied amounts	Real: MAE=0.48 (maize, 8 real) Syn.: MAE=0.42 (maize, 8 real + syn) Δ: -0.06 MAE	maize leaf counting and canola inflorescence branch counting

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Kitzler et al. (2026)	Procedural / image-composition (not a learned generative model): Compose3D pastes plant snippets onto background images with distance-aware overlay, edge blurring, HSV jitter, depth-Gaussian sensor noise, and per-snippet rotation/flip. Compared with Copy-Paste (Ghiasi 2021) extended to D-channel.	1540 WE3DS RGB-D training images preprocessed into 249 background images + 5122 plant snippets, plus 66 additional soil-only images recorded with the same stereo system (315 backgrounds total)	Foreground plant snippets carry their pixel-level segmentation masks; class labels inherited from source training image. Compose3D specifies target crop cover 1.37% and weed cover 0.72% (matching average WE3DS training cover) with uniform class sampling.	ESANet RGB-D segmentation network (ResNet34 backbone) trained from scratch or pretrained for 500 epochs with each augmentation method; switch-off epochs at 100/200/300/400 explored. Tested on held-out WE3DS eval (1028 images) and unseen SOY20 (176 images) and SBT21 (432 images) datasets; generalization performance = mean of test-SOY20 and test-SBT21 mIoU.	Real: generalization mIoU 53.7% Syn.: generalization mIoU 64.5% (Compose3D) Δ: +10.8% mIoU	RGB-D semantic segmentation of soil / crop / weed in nadir-view crop-farming scenes (selective weed control)

continued on next page

Table S1 – continued from previous page

Study	**Generative model**	**Generator training data**	**Synthetic ground-truthing**	**Integration with downstream training**	**Reported performance (real-only vs. synthetic-augmented)**	**Downstream task**
Klein et al. (2024)	Procedural 3D plant model + PBR rendering (computer graphics)	N/A (synthetic-only training)	automatic from procedural model	synthetic only (no real images used in training)	Real: N/A Syn.: ≈89.6% accuracy Δ: N/A	tomato leaf disease (healthy vs infected) classification
Lei et al. (2024)	Helios 3D radiative transfer simulator (procedural)	N/A (procedural simulation)	automatic from 3D simulator	synthetic + real (mixed) and synthetic-only (unsupervised)	Real: mAP50=0.685 (bean, 35 real) Syn.: mAP50=0.775 (bean, 35 real + 105 syn) Δ: +0.090 mAP50	bean leaf detection and strawberry detection
Li and Xiang (2024)	Unreal Engine 5 (game-engine real-time ray-traced/Lumen rendering; hand-built digital-twin scene, Quixel Megascans assets)	N/A (simulator, not training a downstream perception model in this paper)	automatic from simulator	synthetic only (simulator demonstration)	Real: N/A Syn.: UE5↔ROS end-effector L2 error 0.021mm avg (max 0.211mm) measured 1 s after a setpoint is reached; 1.769mm avg (max 19.148mm) instantaneous at setpoint; two UR10 arms, 120 setpoints Δ: N/A	photorealistic robotic simulation for agricultural image data generation
Li et al. (2025c)	Unreal Engine 5 photorealistic simulator (Quixel Megascan plant assets); synthetic stereo + point-cloud dataset	96 seeded UE5 environments; 15,349 partial/complete point-cloud pairs and 15,349 stereo disparity pairs from synthetic simulation; 88 plant models across 12 categories	Automatic depth, semantic, and instance ground truth from UE5 simulator; partial point clouds from controlled camera trajectories	synthetic only (training); real-world deployment for inference (sim-to-real)	Real: Bad-0.5=52.59 (Selective-IGEV trained on real ETH3D stereo data) Syn.: Bad-0.5=17.47 (same model trained on the synthetic stereo set) Δ: -35.1 Bad-0.5 (lower is better)	3D plant reconstruction / next-best-view planning with point-cloud completion and stereo matching

continued on next page

Table S1 – continued from previous page

Study	**Generative model**	**Generator training data**	**Synthetic ground-truthing**	**Integration with downstream training**	**Reported performance (real-only vs. synthetic-augmented)**	**Downstream task**
Li and Kasaei (2024)	Ignition Gazebo simulator (domain randomization)	Straw6D synthetic dataset (procedural; 2,000 images, 80/20 train/test split)	automatic from 3D simulator	synthetic only (sim-to-real transfer; backbone pretrained on real StrawDI)	Real: N/A Syn.: 84.77% AP @ 3D IoU Δ: N/A	strawberry 6DoF pose and 3D size estimation for robotic harvesting
Liu et al. (2023)	Blender procedural 3D rendering (corn-field model with parametric row/plant spacing)	Synthetic dataset across 4 collection configurations (orthophoto + multi-angle + spiral); ≈100 manually labeled real UAV images for comparison	Pixel-perfect masks from flat-color labeled twin model in Blender (corn leaf = white, sky = green, ground = black)	synthetic only, real only, synthetic + real (combined)	Real-only (U-Net++, >100 manual UAV imgs): IoU 0.6535 Synthetic-only: IoU 0.4669/0.4964/0.5887 (Synthetic-1/2/3; all below real-only) Synthetic+real combined (best): IoU 0.6675 Δ: +0.014 IoU = combined vs real-only (synthetic-only alone underperforms real-only)	Corn leaf semantic segmentation (and LAI correction)

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Martini et al. (2024)	Blender procedural 3D plant models + Gazebo simulation	≈5k–19k synthetic images per crop (e.g., zucchini 19200, lettuce/chard 4800, lavender 5260, vineyard 13840, apples 15280, pear 7980)	automatic from 3D simulator (perfect masks)	synthetic only vs. synthetic + 100 real images (mixed training/fine-tune)	Real: N/A (no real-only baseline; only ≤100 real imgs added as fine-tune) Syn.: Apples synthetic-only IoU 0.4575 on real test; synthetic + 100 real (fine-tune) IoU 0.8398 on real test (source Table 4) Δ: +0.382 IoU (+38 pp) from adding 100 real fine-tune imgs — not synthetic-vs-real	crop-row binary semantic segmentation for autonomous robot navigation
Mayanja et al. (2025)	Helios 3D plant biophysical model (ray-traced thermal camera simulation)	≈40,000 synthetic thermal images (1331 per of 30 measurement periods)	automatic from 3D simulator (known gs parameters per simulation)	trained only on synthetic, evaluated on real thermal imagery	Real: R^2=0.7 vs porometer Syn.: R^2=0.89 (synthetic validation) Δ: N/A	stomatal conductance parameter estimation from thermal imagery in cowpea
McMillen and Yilmaz (2025)	Unreal Engine 5 procedural biome generation	1169 synthetic samples (proof-of-concept)	automatic from 3D simulator (pixel-perfect labels)	synthetic only (train + test split 80:20 within synthetic)	Real: N/A Syn.: mIoU 85.83% (FuseForm RGB) Δ: N/A	multimodal forest semantic segmentation (6 evaluated classes: Ground/Beech/Rocks/Maple/Dead/Alder; source abstract states seven)

continued on next page

Table S1 – continued from previous page

Study	**Generative model**	**Generator training data**	**Synthetic ground-truthing**	**Integration with downstream training**	**Reported performance (real-only vs. synthetic-augmented)**	**Downstream task**
Moonen et al. (2026)	CAD2Render (Unity) with 3D models from Meshy/Trellis/Einscanner HX/Blender	1000 synthetic images per dataset (900 train / 100 val); ≈5–6 3D models per class	automatic from 3D simulator (BOP format)	synthetic only (pretrain) and synthetic + 4 real images (fine-tune)	Real-only baseline (random-init, 4 real imgs, no synthetic pretrain): mAP50 0.781 Syn.: synthetic-pretrain + same 4 real (Image-to-3D, best): mAP50 0.956 Δ: +0.175 mAP50 = benefit of synthetic pretraining (both fine-tuned on the same 4 real imgs; source Table 1)	potato vs stone object detection (bin-picking)
Ranario et al. (2025)	Stable Diffusion + ControlNet with optimized text embeddings and cross-attention guidance; source synthetic images from Helios 3D plant biophysical modeling framework	N/A (uses pretrained diffusion models; AgML public synthetic+real grape/flower datasets)	Labels carried over from source synthetic domain via attention guidance preserving object semantics	synthetic-source translated to target domain (training data); evaluated on real-target object detection	Real: N/A Syn.: AP=0.33 (SynGrape to Borden Day) Δ: +0.05 AP vs CropGAN	cross-domain plant trait identification (object detection on grape clusters / flowers)

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Rodrguez-Gmez et al. (2019)	Unreal Engine 4 + Blender procedural sunflower scene rendering (with infected/healthy variants, dwarfism, color changes)	N/A (custom Blender base sunflower model, photo-realistic textures)	Synthetic frames automatically labeled (soil/vegetation, healthy/infected) via Photoscan annotation pipeline	soil-crop SVM: synthetic + real (combined); healthy-infected SVM: real-only (final-stage infected 3D model unavailable)	Real: healthy-infected SVM (real-only) 89.2–94.6% (94.58/91.25/93.15/89.24) accuracy across four weather/time conditions Syn.: soil-vegetation classifier 97.4% accuracy with synthetic-augmented training Δ: N/A (no like-for-like real vs. synthetic-augmented comparison for the disease SVM)	infected-crop classification / disease mapping (sunflower downy mildew)
S. et al. (2021)	L-System (L+C language) procedural plant modeling via LStudio + RPA pipeline	N/A (parametric L-system model; no real training images used to fit generator)	Auto-generated bounding-box labels from procedural L-System rendering pipeline	Three settings: synthetic only, mixed (synthetic+real), real only; YOLOv3 detector	Real: mIoU=0.7896 Syn.: mIoU=0.5517 (synthetic-only) Δ: -0.238 mIoU (syn-only vs real)	Artichoke seedling detection (YOLOv3)
Sapoukhina et al. (2022)	procedural/parametric synthetic fluorescence image generator (statistics-derived, not GAN/diffusion)	synthetic dataset generated under derived statistical conditions from empirical CFI dataset	automatic annotation (lesions placed programmatically with known masks)	synthetic only (sim2real); transfer to empirical dataset	Real: N/A Syn.: recall=0.793 (U-Net on empirical) Δ: N/A	plant disease lesion segmentation in chlorophyll fluorescence imaging (Arabidopsis-Pseudomonas)

continued on next page

Table S1 – continued from previous page

Study	**Generative model**	**Generator training data**	**Synthetic ground-truthing**	**Integration with downstream training**	**Reported performance (real-only vs. synthetic-augmented)**	**Downstream task**
Selvaraj and Farzan (2024)	Blender-designed apple-tree assets (Sapling Gen) rendered in the Gazebo physics simulator (Husky + dual UR5 setup)	N/A (simulation-based; 130 synthetic images: 100 train / 30 test)	Manual pixel-level labeling of synthetic images via Segments.ai (binary apple masks)	U-Net (and U-Net + ImageNet-pretrained ResNet-50/EfficientNet encoder) trained on synthetic only; tested on real images for sim-to-real domain adaptation	Vanilla U-Net: Dice=0.843 (syn-trained, on real) +ResNet-50 encoder: Dice=0.9111 (syn-trained, on real) Δ: +0.068 Dice (ResNet-50 encoder vs vanilla U-Net; both syn-only training)	Apple semantic segmentation for robotic harvesting
Steele et al. (2024)	3D computer simulation (virtual fish schooling, Blender-style)	N/A (asset-based simulation; 700 virtual images generated)	automatic from 3D simulator	synthetic + real (mixed training; virtual-to-real ratios tested)	Real: ∼95.5% mAP (above 95% per paper); F1=0.91 Syn.: mAP=91.8% (M6 mixed: 630 virtual + 70 real) Δ: −3.7 mAP (mixed M6 91.8% ¡ real)	fish detection in recirculating aquaculture systems
Tabak et al. (2023)	Blender procedural generation pipeline (custom-designed)	N/A (procedural generation; uses real photographed leaf textures applied to 3D plant models)	Auto-generated pixel-perfect labels from Blender rendering (leaf class)	synthetic only (DeepLabv3+ trained exclusively on synthetic, tested on real)	Real: N/A Syn.: IoU=0.63 (PP best) Δ: +0.16 IoU (texture match vs wrong texture)	Sweet pepper leaf semantic segmentation (for 6DOF pose estimation in robotic harvesting)

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Thayananthan et al. (2025)	Gazebo + Blender procedural simulation	N/A (virtual cotton-field assets; 193 augmented scene-segmentation images, 40 cotton-boll images)	manual annotation in Roboflow (not 3D-auto-export); YOLOv8n-seg and YOLOv8n trained separately on the two tasks	synthetic only (training and testing in simulation)	Real: N/A Syn.: mAP=85.2% (YOLOv8n-seg, 3-class farm-scene instance segmentation: sky / cotton plants / ground); cotton-boll detection mAP ≈92.7% (YOLOv8n, separate model) Δ: N/A	farm-scene instance segmentation (sky / plants / ground) for vision-guided autonomous navigation, plus cotton-boll detection in a virtual cotton field
Toda et al. (2020)	Domain randomization (synthetic compositing of scanned seed cutouts on virtual canvas)	N/A (synthetic images composited from scanned seed textures)	automatic (annotations generated during compositing)	synthetic only	Real: N/A Syn.: AP=95% (recall=96%) Δ: N/A	barley/crop seed instance segmentation (phenotyping)
Wang et al. (2022)	Blender procedural plant modelling	100 synthetic models generated in Blender	automatic from Blender (per-leaf labels)	synthetic + real (fused dataset of 620 point clouds)	Real: N/A Syn.: AP 97.2% @IoU0.25 Δ: N/A	lettuce 3D leaf instance segmentation
Ward et al. (2019)	Blender procedural (domain randomization)	N/A (procedural; uses 30 leaf textures and 4 background pots from CVPPP)	automatic from 3D simulator (Blender renders pixel-level leaf masks)	synthetic + real (mixed training, 50/50 per batch)	Real: SBD 71 Syn.: SBD 81 Δ: +10 SBD	leaf instance segmentation (Arabidopsis/tobacco, CVPPP LSC)
Ward and Moghadam (2020)	Blender procedural / domain randomization (UPGen)	N/A (uses 17,957 leaf geometries + 40,552 leaf textures from public datasets)	automatic from 3D simulator	synthetic only (and synthetic + few real)	Real: transfer-learning baseline Syn.: SBD 88% Δ: +26.6%	leaf instance segmentation across multiple plant species (CVPPP LSC + new species)

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Yoncalik et al. (2026)	LLM-driven Unreal Engine procedural generation (RAG + GPT-4)	N/A (LLM pipeline; 672 asset path combinations)	N/A (no downstream detector trained; environment generation only)	synthetic only (synthetic environment generation)	Real: N/A Syn.: N/A Δ: N/A	3D agricultural simulation environment generation (apples, vegetables) for synthetic data pipelines
3D reconstruction and neural rendering pipelines						
Chopra et al. (2024)	NeRF (cross-spectral, multi-modal)	self-collected real-world orchard/garden datasets (apple, peach, eggplant, tomato, pepper sequences) using RGB+thermal+event sensor suite	NeRF-rendered novel views; YOLOv8 + AnyLabeling used for object detection labels	NeRF reconstructed views used as input for YOLOv8 fruit detector (compared against Nerfacto and EvDeblurNeRF baselines)	Real: mAP50=53.3 (Nerfacto RGB) Syn.: mAP50=77.2 (Cross-Spectral) Δ: +44.8% mAP50	fruit detection from NeRF-reconstructed orchard/garden scenes
Hu et al. (2024a)	NeRF (Instant-NGP, Instant-NSR)	per-scene multi-view captures (100 images per dataset, fixed for fair comparison)	N/A (reconstruction-quality study)	N/A (no downstream task model trained on synthetic vs real)	Real: N/A Syn.: 31.37 dB PSNR (Instant-NGP) Δ: N/A	novel-view synthesis and 3D reconstruction of crop plants
Huang et al. (2025)	3D Gaussian Splatting (3DGS)	multi-view images of oilseed rape plants (3DGS reconstruction per-plant)	manual annotation of reconstructed point clouds for segmentation training	3DGS used as point-cloud acquisition; segmentation trained on reconstructed clouds (no real-vs-synthetic comparison)	Real: N/A Syn.: 96.01% mIoU (CKG-PointNet++) Δ: N/A	oilseed rape point-cloud organ segmentation and phenotypic-trait extraction

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Jiang et al. (2025)	3D Gaussian Splatting (3DGS); compared with NeRF	≈100 video frames per plant from smartphone; 50 plants total; 157 annotated images for YOLOv11x (109 train / 48 val)	Roboflow manual polygon annotation for 2D boll masks (YOLOv11x vs. SAM compared as mask predictors); LiDAR point clouds as 3D ground truth	synthetic 3D reconstruction used directly to derive phenotypes (not augmenting a 2D classifier)	Real: NeRF PSNR 27.08 Syn.: 3DGS PSNR 33.99 Δ: +6.91 PSNR	3D plant reconstruction and cotton boll counting/phenotyping
Ku et al. (2025)	NeRF (stationary camera variant)	N/A (per-scene rotation video, ≈30s)	N/A (3D reconstruction, not labeled detection task)	N/A (reconstruction task)	Real: N/A Syn.: F-score ≈100.00 Δ: N/A	3D point cloud reconstruction of plants (apricot, banana, bell pepper, maize ear, succulents)
Plum et al. (2023)	Unreal Engine 5 procedural generation (3D models from scAnt photogrammetry)	10,000 synthetic annotated images per dataset (generated from 100-individual digital populations)	automatic from 3D simulator (bounding boxes, key points, segmentation masks, depth/normal passes)	synthetic only, and synthetic + real (mixed sb1: 10,000 syn + 100 real)	Real: mAP 0.878 Syn.: mAP 0.951 (mixed sb1) Δ: +0.073 mAP	animal (insect) detection, tracking, pose-estimation, semantic segmentation

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Polic et al. (2022)	Blender procedural (3D pepper photogrammetry-Meshroom model + sapling tree generator + leaf/flower planes) with morphology/color augmentation	N/A (3D pepper model derived from 200 smartphone photos via Meshroom; 540 synthetic images rendered)	Procedural Blender rendering produces automatic bounding boxes and semantic-segmentation labels	COCO-pretrained MobileNet-SSD trained on 540 synthetic images, then fine-tuned on ∼40 manually labeled real pepper images	Real: N/A Syn.: ≈5 mm mean position dissipation Δ: N/A	sweet-pepper 6DOF pose detection for robotic harvesting
Shrestha et al. (2025)	3D Gaussian Splatting (3DGS)	RGBD video of real flowers used to fit 3DGS scene (single scene reconstruction)	automatic from 3D scene (poses projected onto rendered images; SAM for masks)	synthetic only (knowledge distillation from teacher trained on synthetic to student)	Real: prior StickBug 49.00% (same-platform baseline) Syn.: FloPE 61.36% (StickBug) / 78.75% (UR5 arm) Δ: +12.36 pp (same-platform StickBug)	flower 6-DoF pose estimation for precision pollination
Hybrid 3D-to-2D and simulator-to-real refinement pipelines						
Aghamohammadesmaeilketabforoosh et al. (2025)	Blender procedural (geometry nodes) + Deep Domain Confusion	0 real images for generator (purely procedural Blender); 10,000 synthetic rendered	automatic from 3D simulator (Blender compositing masks)	synthetic only (with DDC domain adaptation), tested on real	Real: N/A Syn.: DSC=94.8% (ripe) Δ: +36.8% (200 to 10k imgs)	strawberry ripe/unripe semantic segmentation
Anagnostopoulou et al. (2023)	3D procedural (Open3D) + ControlNet (Stable Diffusion)	N/A (procedural; 100 3D scenes, 3,000 base images)	automatic from 3D simulator (rendered masks, COCO+BOP)	synthetic only (compared simple synthetic vs realistic synthetic)	Real: N/A Syn.: mAP=89.5% (realistic) Δ: +7.8% mAP (vs simple syn.)	mushroom detection / instance segmentation / 3D pose estimation

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Barth et al. (2020)	CycleGAN (translates rendered synthetic to empirical)	10,500 synthetic + 50 empirical sweet pepper images	GAN preserves source mask (cycle-consistent translation)	pretrain on synthetic-translated, fine-tune on empirical	Real: IoU=0.48 (syn. bootstrap + empirical FT baseline) Syn.: IoU=0.52 (translated bootstrap + empirical FT) Δ: +8% IoU	sweet pepper plant-part semantic segmentation
Cieslak et al. (2024)	Procedural Blender + L+C (L-system) plant modeling, optional CUT GAN domain adaptation	not used to train generator (procedural); 12,000 synthetic images generated, compared with 12,000 real	automatic from Blender (semantic labels for crops/weeds)	synthetic only, real only, mixed at varying ratios	Real: 0.3519 mIoU(weed) Syn.: 0.3673 mIoU(weed) Δ: +0.0154 mIoU(weed)	crop/weed semantic segmentation (soybean primary; cotton out-of-distribution)
Fei et al. (2021)	semantically constrained CycleGAN driven by 3D Helios crop model	synthetic 3D rendered grape images (Helios) + paired real images (M-k unlabeled + k labeled, varying k=2-50)	source labels from 3D model preserved through semantic constraint of GAN	pretrain on synthetic, fine-tune on few labeled real, augment with semGAN-generated labeled images in target domain	Real: vanilla fine-tune AP@0.3/AP@0.5 baseline Syn.: SemGAN + fine-tune outperforms vanilla at all k Δ: positive AP gains (largest at low k)	grape fruit detection in vineyards (domain adaptation across day/night/cultivar/camera)

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Geng et al. (2025)	Score-distillation from 2D diffusion models + ControlNet + Neural Templates + K-Planes hybrid 4D representation/PBR renderer	N/A (distilled from pre-trained 2D diffusion models, no agricultural training set)	N/A (no labeled real downstream task)	synthetic-only (4D asset generation)	Real: N/A Syn.: N/A (qualitative only) Δ: N/A	temporal 4D object-intrinsics generation (rose lifespan, plant sprouting, banana rotting, etc.) - tangential agricultural relevance via plant lifespans but task is graphics asset generation
Hartley and French (2021)	Blender + L-system (synthetic) + CycleGAN (sim2real domain adaptation, with heatmap-regression support and clustering)	5,000+ synthetic Blender images with 100,000+ annotations; CycleGAN trained on these vs GWHC images	automatic bounding boxes from Blender camera + world coordinates; CycleGAN preserves geometry	synthetic only / synthetic + real mixed; with/without heatmap-support / clustering	Real: 0.8262 Mean IoU Syn.: 0.8642 Mean IoU (HM-support 4-clusters + Real) Δ: +0.038 Mean IoU	wheat-head detection on Global Wheat Head Challenge (GWHC) dataset
Hartley et al. (2024)	Stable Diffusion + LoRA (low-rank adapter) + ControlNet, with Blender L-Systems for source synthetic plants	small target-domain real datasets used for LoRA fine-tuning (datasets like CVPPP A1/A2/A3, Komatsuna; A2/A3 contain only 10 training images)	instance segmentation masks generated from 3D Blender models; preserved through ControlNet conditioning	synthetic only (LoRA-Synth), synthetic + real (LoRA-All), compared against CycleGAN baseline and real-only training	Real: real-trained MaskRCNN MaP baseline Syn.: LoRA-All MaP up to 0.669 Δ: surpasses real-only on 10-image A2/A3 sets	leaf instance segmentation on rosette plants

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Hartley et al. (2025)	Stable Diffusion 2.1 + depth ControlNet + LoRA + 3D Gaussian Splatting (SDS optimization)	30 plant images per species for LoRA fine-tuning (bean/kale/mint)	N/A (no semantic labels); L-system mesh or real point cloud as 3DGS initialization; species-specific LoRA conditions style	3D asset generation; supports synthetic L-System initialization and real point-cloud enhancement	Baseline: 11.01 dB PSNR-masked (GaussianDreamer, real point clouds) Method: 16.12 dB PSNR-masked (PlantDreamer, real point clouds) Δ: +5.11 dB PSNR-masked (method vs baseline)	3D plant model generation (bean, kale, mint) for phenotyping data augmentation
Jiang et al. (2022)	VAE-based pose generator + 3D animal mesh + style transfer (StyTr2)	animated 3D animal models + a few hundred reference poses	automatic from 3D simulator (keypoints attached to skeleton)	synthetic + real (99 real images + 3000 synthetic SynAP)	Real: PCK@0.05=78.7% Syn.: PCK@0.05=92.4% Δ: +13.7% PCK@0.05 (HRNet-w32)	zebra/quadruped 2D pose estimation
Jung et al. (2024)	3D rendering (Blender Sapling Tree Gen + Unreal Engine 5) + I2I translation (CUT, QS-Attn, Decent)	Real PWD set 4037 cropped images (UAV) drives I2I; 9052 synthetic images generated	Auto-labeled bounding boxes via projecting 3D tree object bounds to screen coords (suspected/infected/dead)	synthetic only, real only, and ensemble (real + synthetic), plus pretrain-on-synthetic + fine-tune on real	Real: F1 90.22% Syn.: F1 92.88% (R+S) Δ: +2.66 F1	Pine Wilt Disease image classification (suspected/infected/dead)

continued on next page

Table S1 – continued from previous page

Study	**Generative model**	**Generator training data**	**Synthetic ground-truthing**	**Integration with downstream training**	**Reported performance (real-only vs. synthetic-augmented)**	**Downstream task**
Kauny et al. (2024)	Procedural leaf/millimeter-paper model + Stable Diffusion ControlNet (Canny edges) + filtering	procedural model (no training data); ControlNet uses pretrained Stable Diffusion	automatic from procedural model (semantic masks, surface-area labels); filtering by segmentation consistency	synthetic + real (1.7K real + 5K synthetic) and synthetic-augmented variants	Real: MRE=12.9% (test) Syn.: MRE=9.2% (test) Δ: -3.7% MRE	leaf surface-area regression and leaf semantic segmentation (beech/oak)
Masuzawa et al. (2023)	Blender CG with PlantFactory parametric cucumber plant models + adversarial image conversion (pix2pix, CycleGAN, CyCADA)	50 plant nodes measured (5 plants) used to fit parametric model; 214 CG-generated images; 40 real images for comparison/conversion training	Pixel-perfect semantic labels from 3D Blender scene (fruit/stem/flower/leaf/other)	synthetic only, synthetic + I2I conversion, vs real-only	Real: mIoU 0.644 Syn.: mIoU 0.447 (CyCADA) Δ: -0.197 mIoU but +7/63 fruit detection recall at 13 vs 61 hours annotation	Cucumber plant semantic segmentation and fruit detection
Meyer et al. (2024)	NeRF (Nerfacto-based semantic NeRF) + Blender procedural fruit trees (XFrog assets) for synthetic dataset	N/A (NeRF optimized per-scene from posed images; 300 synthetic images per tree, ≈350 real images per tree)	Blender-rendered ground-truth segmentation masks for synthetic; Grounded-SAM auto-masks and U-Net masks for real (manual fruit counts as GT)	synthetic + real (synthetic for benchmarking counting, real for apple validation)	Seg. backbone: SAM-mask F1 0.88 vs. GT (Blender-rendered) mask F1 0.95 (avg. over synthetic fruit types, Δ +0.07); real Fuji-apple holdout F1 0.919 (SAM) / 0.962 (U-Net)	3D fruit counting / yield estimation

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Olatunji et al. (2020)	CGAN (image-to-image translation on 2D shape maps); training data synthesized via Blender + Monte-Carlo procedural shape equation for kiwifruit	large synthesized dataset of computational kiwifruit shapes (no real images used to train generator); 56 real kiwifruit empirically scanned for validation	Procedurally synthesized 3D kiwifruit meshes voxelised with known full surface as ground truth; partial-occlusion shape maps as input	synthetic only for training; real-empirical scan data only for validation	Real: N/A Syn.: MAPE ¡ 5% (empirical weight) Δ: N/A	3D fruit shape reconstruction / weight estimation
Qiu et al. (2025)	Diffusion model (Zero123++) + Large Reconstruction Model (LRM/triplane); Real2Sim procedural Blender generator	5,000 procedurally generated synthetic apple tree models (4,970 for training + 30 R2S-Val; R2S-Diffusion, R2S-LRM, R2S-NeRF datasets)	automatic from procedural simulator (mesh-sampled point clouds as ground truth)	synthetic only (training); zero-shot deployment on real field trees	Real: baselines (Nerfacto/InstantMesh/ COLMAP/sensor) Syn.: DATR (MM-Recon, synthetic-trained) Δ: ≈360x throughput vs TLS; lowest CD on synthetic and best field-tree trait MAE (trunk 0.68 cm/ branch 3.67), JSD comparable across methods	3D apple tree reconstruction from sparse-view images for digital twin/agricultural robotics
Saraceni et al. (2024)	Unity 3D simulator + cut-paste of real instances (with PCA alignment)	2,400 pseudo-labeled real images + 1,889 Unity-simulator synthetic images (Synthetic-Pasted + Pseudo combined = 4,289 training images)	automatic from 3D simulator (synthetic masks) + real instances pasted with their SAM-derived masks	synthetic + pseudo-real (mixed training)	Real: F1=0.233 (pseudo baseline CloseUp1) Syn.: F1=0.45 (Synthetic-Pasted+Pseudo) Δ: +0.22 F1	table-grape detection in vineyards

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Schauer et al. (2023)	Blender 3D model rendering (model-based) and Stable Diffusion text-conditioned inpainting (with full SD v1.5 fine-tuning) for synthetic ragwort generation	110 J. Vulgaris + background images (256$\times$256 px), cropped from the existing object-detection dataset, for SD v1.5 fine-tuning; Blender path model-based (custom meshes, no training images)	Blender renders give pixel-perfect segmentation masks and bounding boxes automatically; SD inpainting uses inpaint masks as bounding-box proxies	real only (ELAN-Real), synthetic only (ELAN-Synth), and mixed (ELAN-Mix); evaluated on real validation set	Real: F1=0.950 (ELAN-Real) Syn.: F1=0.859 (ELAN-Synth); F1=0.917 (ELAN-Mix) Δ: -0.091 F1 (syn-only vs real); -0.033 F1 (mixed vs real)	ragwort weed detection in grassland (binary classification / heatmap localization)

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Ushiroji and Fujinaga (2026)	3D Gaussian Splatting (3DGS) reconstruction (Scaniverse "Splat" mode) for 3D model creation; Unreal Engine 5 (UE5) for synthetic image rendering with random fruit position, maturity (RED/-GREEN/MIDDLE), and lighting (1500/3000/6000 lx)	Cultivation-bench scan ≈20 min and per-fruit-cluster scan ≈10 min using iPhone 16 Pro + LiDAR / Scaniverse Splat mode; from these 3DGS models the pipeline rendered 120, 360 or 1200 synthetic training images (with 30, 90 or 300 validation images, 8:2 split) at 1080x1920 resolution, downsampled to 480x640.	YOLO-format bounding-box annotations are automatically generated by projecting the 8 vertices of each tomato's 3D bounding box (with assigned class identifier 0=RED, 1=GREEN, 2=MIDDLE) from world coordinates through the UE5 view-projection pipeline (Eq. 1-7); no manual annotation required.	YOLOv8s transfer-learning trained on synthetic-only datasets generated under each condition (100 epochs, batch 16); evaluated on a fixed real test set of 50 manually-annotated greenhouse images masked with a 0.8 m depth threshold.	Real: N/A Syn.: mAP@50=0.768 (Condition H, best) Δ: +0.444 mAP@50 (3DGS fruits vs sphere)	Real-time three-class tomato fruit object detection (RED/-GREEN/MIDDLE maturity) for a tomato-harvesting robot
Vanherle et al. (2025)	3D Gaussian Splatting (foreground) + cut-paste compositing	≈1 minute video per object	automatic (opacity map from GS render gives mask)	synthetic only (compared to Cut-and-Paste and Diffusion-based baselines)	Real: N/A Syn.: ≈80 mAP Δ: +30 mAP	instance segmentation of household/specific objects (proof of concept; non-agricultural household objects)

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Vierbergen et al. (2023)	Photogrammetry + Unity flower-field simulator	29 photogrammetric Calendula 3D models; 15,000 synthetic images for training	automatic from 3D simulator (keypoint heatmaps)	synthetic only (sim-to-real)	Real: N/A Syn.: F1 86% Δ: N/A	Calendula flower detection and 3D localisation for automated harvesting
Vuleti et al. (2022)	Blender procedural generation + photogrammetry-derived pepper/peduncle 3D models	≈200 photos per pepper for photogrammetry-based 3D model reconstruction; 51 real smartphone images manually labeled for real-only training/fine-tuning/evaluation	Auto-generated bounding-box and semantic-segmentation labels via multi-layer scene rendering with depth filtering	synthetic only vs. real only vs. synthetic + real (fine-tune); MobileNet-SSD via transfer learning	Real: mAP 17.4 Syn.+Real (fine-tune): mAP 20.8 Δ: +3.4 mAP (vs real-only; syn-only=15.9)	Sweet pepper fruit/flower object detection (for robotic harvesting/yield estimation)

continued on next page

Table S1 – continued from previous page

Study	**Generative model**	**Generator training data**	**Synthetic ground-truthing**	**Integration with downstream training**	**Reported performance (real-only vs. synthetic-augmented)**	**Downstream task**
Wok et al. (2026)	(1) Stable Diffusion Inpainting (pretrained, no weight tuning) constrained by SegFormer-derived forest-floor masks; (2) Unreal Engine 5 (v5.5) photorealistic simulator with domain randomization (lighting, viewpoint, vegetation density, occlusion)	64 real UAV RGB images (selected from 125 after deduplication and quality filtering); 44 real images per fold used as inpainting backgrounds and YOLOv8x training set, plus added synthetic data	Inpainting: SegFormer mask constrains insertion locations; the 674 inpainted boulder images were manually annotated in CVAT. Simulator: automatic annotations exported from the UE5 scene state.	Real-only baseline (44 real images per fold) vs Real + 674 inpainted synthetic boulder images vs Real + 192 simulator (SimForest) images containing annotated boulders; same YOLOv8x training schedule (250 epochs) and identical real-image validation set (20 images per fold) across all three settings	Real: mAP50 0.579 Syn.: mAP50 0.647 Δ: +0.068 mAP50	2D bounding-box detection of forest-floor obstacles (boulders) in low-altitude UAV RGB imagery for autonomous-forestry path planning
Tabular, spectral, and multimodal sensor synthesis						

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Cao et al. (2026)	AE-GAN (autoencoder-based GAN with encoder-decoder generator and a fully-connected discriminator with sigmoid output). A standard GAN was also evaluated for comparison.	AE-GAN was trained on the original training subset of the 180-sample dataset (split 7:1:2 -¿ 126 training samples) using a 1$\times$1$\times$176-d (hyperspectral) and 1$\times$1$\times$178-d (fluorescence) "spectrum-NCUI" joint vector as input/reconstruction target	NCUI label is treated as the dimension-matched, inherently paired component of the generator's input/output joint vector (synchronous generation of spectral curve and corresponding NCUI). No manual labelling of synthetic samples; PCA-derived NCUI carried implicitly through reconstruction.	Synthetic samples added only to the 126-sample training subset; test set kept entirely as unaugmented raw samples for objective evaluation. A sensitivity sweep added 0/180/360/540/720 synthetic samples; adding 180 gave the best balance.	Real: Rˆ2P=0.7988 (CNN) Syn.: Rˆ2P=0.8876 (CNN+AE-GAN) Δ: +0.0888 Rˆ2P	Regression of pepper-leaf Nitrogen Comprehensive Utilization Index (NCUI) from fused hyperspectral + fluorescence-hyperspectral data
Lan et al. (2025)	Gaussian distribution sampling + VAE	based on existing cotton yield records; sampling new vectors from Gaussian fit	synthetic samples drawn to match original yield-data distribution	synthetic + real (data augmentation for regression)	Real: N/A Syn.: RMSE=58.4 lbs/acre (Bahawalnagar) Δ: N/A (delta not in abstract)	cotton yield prediction (cross-year/cross-district regression)

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Nithya and Latha (2025)	PrivTab-GAN (DP-SGD-based tabular GAN with domain-adversarial NN)	crop recommendation dataset (10,440 preprocessed samples; 35,505 raw records 1961-2022)	tabular synthetic generation conditioned on attributes	train-on-synthetic/test-on-real and train-on-real/test-on-synthetic scenarios	Real: Accuracy 0.97 (Crop Rec) Syn.: Accuracy 0.96 (Crop Rec, S1) Δ: -0.01 accuracy	tabular agricultural classification (crop recommendation, weather prediction, irrigation)
Razavi et al. (2024)	Variational Autoencoder (VAE) for tabular data	real crop-yield dataset 2016-2020 from Senegal MAER; ≈38 spatiotemporal/-physiographical features	Synthetic tabular records inherit feature distributions; statistical similarity (Pearson, CDF) used to validate fidelity; 2500 synthetic samples generated	real only and synthetic-augmented (VAE samples added to training set)	Real: RMSE=0.294 Syn.: improved nRMSE/MASE on augmented training Δ: stability/robustness gains (per-crop deltas in supplementary)	crop yield prediction (regression for 5 crops: groundnut, maize, pearl millet, rice, sorghum)

continued on next page

Table S1 – continued from previous page

Study	**Generative model**	**Generator training data**	**Synthetic ground-truthing**	**Integration with downstream training**	**Reported performance (real-only vs. synthetic-augmented)**	**Downstream task**
Tan et al. (2025c)	ACGAN (Auxiliary Classifier GAN) - Spe-GAN for 1D spectra enhancement and Spa-GAN for 2D image enhancement (after ResNet18 dimensionality reduction)	365 cotton-plant samples (185 healthy + 180 early-stage VW); training set 292 samples (148 healthy + 144 early-stage) and test set 73 samples (37 healthy + 36 early-stage), with paired hyperspectral curves and corresponding RGB images. Each training run uses both raw real samples and synthetic samples produced by the ACGAN's generator during adversarial training.	Class-label conditioning via the auxiliary classifier head of ACGAN; generator produces spectra/images conditioned on the disease class label (healthy vs. early-stage VW). Quality verified by RMSE/mean (similarity) and t-SNE/variance/std (diversity) of generated vs real samples.	Synthetic samples generated by Spe-GAN/Spa-GAN are used to enrich the training set during adversarial training; the same ACGAN model performs the early-detection classification (the auxiliary classifier provides the label prediction). Test set is 73 real samples (37 healthy + 36 early-stage), unaugmented.	Real: LSTM OA=89.04% (spectral baseline) Syn.: Spe-GAN OA=94.52% Δ: +5.48 pp OA (spectral)	Binary classification (healthy vs early-stage Verticillium wilt) of cotton plants from hyperspectral cubes (spectral or image features)
Tan et al. (2025b)	Regression GAN (RGAN), variant of GAN extended to regression (joint spectrum+target generation)	108 Hami melons (3 dose groups, ≈68 used as real training); RGAN generates additional spectrum-residue pairs (e.g., +30, +500 generated samples)	RGAN generator outputs both spectrum and corresponding (synthetic) residue value; reliability validated against held-out real data	synthetic + real (real data augmented with RGAN-generated samples for SVR/PLSR training)	Real: SWIR-SVR R2p=0.8352 Syn.: SWIR-SVR R2p=0.8714 (+500 samples) Δ: +0.0362 R2p	regression of acetamiprid pesticide residue content in Hami melons from VNIR/SWIR hyperspectral data

continued on next page

Table S1 – continued from previous page

Study	**Generative model**	**Generator training data**	**Synthetic ground-truthing**	**Integration with downstream training**	**Reported performance (real-only vs. synthetic-augmented)**	**Downstream task**
Tulu et al. (2026)	Pix2PixGAN (paired conditional GAN, U-Net generator + 70x70 PatchGAN discriminator) and CycleGAN (unpaired cycle-consistency variant, applied here to the paired RGB-thermal set)	≈3,400 raw UAV images collected over three growing seasons (2020-2023) from 32 plots (16 sweet corn, 16 green beans), processed into 256x256 plot patches; after augmentation 68,628 spatially-consistent RGB-thermal image pairs. Train/val/test split was reported in two ways in the paper: 80%/20% in the abstract and 70/15/15 in Section 2.3.1, stratified by crop and irrigation treatment and constrained by flight date.	Pixel-aligned (paired) co-registered RGB-thermal patches obtained via Pix4D radiometric calibration, ArcMap image-to-image alignment to a GCP-georeferenced base map, and AgriSenAI plot-level extraction; thermal floating-point temperatures normalised to 8-bit grayscale per Eq. (1) before training.	Generators trained adversarially with Adam (lr 2e-4, β1=0.5, β2=0.999), batch size 1, 500 epochs, _L1=100 for Pix2Pix; trained models then used to generate synthetic thermal mosaics from RGB inputs at inference, from which CWSI was computed using ten Tcold/Thot estimation methods (M1-M10).	Real: Pearson r>0.98 (Pix2Pix sweet corn) Syn.: PSNR=42.98 dB (Pix2Pix) Δ: +2.42 dB PSNR (Pix2Pix vs CycleGAN)	RGB-to-thermal image translation supporting plot-level Crop Water Stress Index (CWSI) estimation for irrigation scheduling in sweet corn and green beans

continued on next page

Table S1 – continued from previous page

Study	Generative model	Generator training data	Synthetic ground-truthing	Integration with downstream training	Reported performance (real-only vs. synthetic-augmented)	Downstream task
Vidal-Domnguez et al. (2026)	Convolutional Variational Autoencoder (VAE-CNN) and conditional GAN (pix2pix-based GAN, U-Net generator + PatchGAN discriminator)	Per-survey training subsets (80% of each of 6 surveys): S1 surveys 11/12/13 = 5,952 / 6,192 / 7,704 augmented patches (248/258/321 =training captures, 80% of 310/324/403 raw); S2 surveys 21/22/23 = 8,448 / 8,448 / 10,344 augmented patches (352/352/431 training captures, 80% of 440/440/539 raw); pooled site-level VAES2 / GANS2 use the union of three S2 surveys	Paired RGB-NIR images acquired with a MicaSense RedEdge multispectral camera mounted on a Tarot T18 octocopter; aligned by manufacturer's tools and used as ground-truth NIR for reconstruction	Each generative model trained per-survey or per-site to predict NIR from RGB; evaluated on 10% real test split per survey under matched and mismatched shadow categories. Generation-quality evaluation in an applied agricultural/forestry remote-sensing context (no separate downstream perception task).	Real: GANS2 PSNR 34.34 dB (moderate-shadow max; median 28.13) Syn.: VAES2 PSNR 35.06 dB (moderate-shadow max, paper headline; median 27.29) Δ: +0.72 dB on peak PSNR only — by median GANS2 is higher; VAE's reported advantage is stability (narrower IQR, higher min, lower NRMSE dispersion)	RGB-to-NIR multispectral band reconstruction for crop/forest health monitoring under variable illumination

Table S2: Reviewed papers that do not present an agricultural synthetic-data application study (reviews, foundational generative-model and 3D-vision papers, simulators/tools, dataset releases, and adjacent methodology). These works are summarised qualitatively because they do not report a like-for-like real vs. synthetic-augmented downstream comparison.

Reference	Category	Contribution / why excluded from quantitative table
Review and survey literature		
Afonso and Giufrida (2023)	Review/Survey	Editorial introducing a Frontiers Research Topic on synthetic data for computer vision in agriculture. Editorial / opinion piece, no original empirical experiments.
Akkem et al. (2024)	Review/Survey	Titled as a review but primarily an original study proposing and evaluating VAE/GAN generation of synthetic *tabular* data (soil/weather features) for crop recommendation in smart farming; generates structured, non-image data rather than computer-vision imagery.
Bac et al. (2014)	Review/Survey	Review article on state-of-the-art harvesting robots for high-value crops. Literature review, not a synthetic data study.
Bengesi et al. (2024)	Review/Survey	Comprehensive review of generative AI: GANs, GPT, autoencoders, diffusion models, and transformers. General GAI survey, not agriculture-specific or SDG empirical study.
Bhujel et al. (2025)	Review/Survey	Systematic survey of public computer vision datasets for precision livestock farming. Literature review/PRISMA survey of 67 public livestock CV datasets – does not generate or train on synthetic data.
Bongomin et al. (2024)	Review/Survey	Review of UAV image acquisition and processing for high-throughput phenotyping. Literature review on UAV phenotyping; no synthetic data experiments.
Chen et al. (2024b)	Review/Survey	Survey of end-to-end autonomous driving systems (270+ papers). Autonomous-driving survey, not agricultural.
Chuquimarca et al. (2025)	Other	assesses DL model robustness for banana ripeness classification under varying illumination using LIME gamma-based image enhancement (photometric augmentation only). photometric augmentation (gamma correction via LIME) of real images, not synthetic image generation by a generative model; the data are not "synthetic" in the systematic-review sense (no GAN/diffusion/cut-paste generation; the 3D-rendered images are reused from Chuquimarca et al. 2023). Generic augmentation/robustness study.
Gavai et al. (2025)	Review/Survey	Systematic literature review on privacy-preserving data platforms for agricultural/-food supply chains; no synthetic-data experiments.
George et al. (2014)	Review/Survey	Review article on field phenotyping platforms and long-term experiments for crop genotype \$\times\$ environment \$\times\$ management interactions; no synthetic-data generation.

continued on next page

Table S2 – continued from previous page

Reference	Category	Contribution / why excluded from quantitative table
Ghazal et al. (2024)	Review/Survey	literature survey on computer vision in smart agriculture and precision farming covering image acquisition, stitching, analysis, decision-making and treatment. review article (no original empirical synthetic-data experiments)
He et al. (2026)	Review/Survey	Book-chapter review on high-throughput field-based phenotyping platforms and their application to tomato improvement (fixed in-field cameras, gantries, PhenoPoles, sensor calibration, long-term temporal coverage). Book chapter; no synthetic-data generation.
Hu et al. (2025a)	Review/Survey	Comprehensive review of diffusion models in smart agriculture. Survey/review article.
Hu et al. (2025b)	Review/Survey	Review of diffusion models for hyperspectral image processing. Review article.
Huihui et al. (2023)	Review/Survey	Review of image motion-deblurring techniques in precision agriculture. Review article.
Jiang et al. (2024)	Review/Survey	Survey of robotic perception of transparent objects covering datasets, segmentation, depth reconstruction, and pose estimation. Review article, not an empirical agricultural synthetic-data study; not agriculture-specific.
Kroly and Galambos (2022)	Review/Survey	Tutorial/example paper showing how to build a Blender-based synthetic mushroom-segmentation dataset with domain randomization. Tool/methodology demonstration; no quantitative model performance or downstream comparison reported.
Kaur et al. (2022)	Review/Survey	Policy/practice review on farmers' data privacy and confidentiality in precision agriculture. Policy/governance review, not an empirical SDG study.
Khan et al. (2025a)	Review/Survey	Review of AI applications in food security and sustainable agriculture. Review article; no empirical SDG experiment.
Kingma and Welling (2019)	Review/Survey	Tutorial/introduction monograph to Variational Autoencoders. Methodological introduction/tutorial; not agricultural.
Kumar Kasera et al. (2024)	Review/Survey	Comprehensive survey of IoT/AI applications across pre-, during-, and post-harvest smart agriculture activities, proposing a 5G-based framework. Literature review/survey paper; no synthetic data generation or empirical model results.
LeCun et al. (2006)	Review/Survey	Foundational tutorial on Energy-Based Models for structured prediction. Foundational ML methodology paper, no agricultural application.
Li et al. (2023)	Review/Survey	Comprehensive review of label-efficient learning methods (active, semi-, weakly-, self-, unsupervised) in agriculture. Survey/review article.
Li et al. (2025a)	Review/Survey	Survey of 3D reconstruction techniques (classical, NeRF, 3DGS) in plant phenotyping. Comprehensive review/survey paper.

continued on next page

Table S2 – continued from previous page

Reference	Category	Contribution / why excluded from quantitative table
Lu et al. (2022)	Review/Survey	Systematic review of GAN-based image augmentation in agriculture and food (59 articles, 2017-2022). PRISMA-style literature review/survey; no original empirical results.
Min et al. (2025)	Review/Survey	Review/survey of computer vision and generative AI in agriculture (PRISMA methodology); literature review.
Mirbod et al. (2023)	Other	Stereo vision system with deep learning occlusion handling for on-tree apple fruit size estimation in orchards. Empirical agricultural vision study but does not generate or use synthetic training data; uses real stereo imagery with Faster R-CNN / Mask R-CNN trained on real annotated images. Out of scope for synthetic-data review.
Montoya-Cavero et al. (2022)	Review/Survey	Survey/review of vision systems for harvesting robots covering produce detection and localization sensors and methodologies. Literature review paper, not an empirical synthetic-data study.
Nikolenko (2019)	Review/Survey	Comprehensive survey of synthetic data for deep learning across computer vision, simulation environments, NLP, bioinformatics, and privacy applications. Survey/review paper (arXiv:1909.11512), not an empirical agricultural study.
Okura (2022)	Review/Survey	Cross-cutting review of 3D modeling and reconstruction of plants/trees across computer graphics, vision, and plant phenotyping. Invited review paper (Breeding Science).
Page et al. (2021)	Review/Survey	PRISMA 2020 statement - updated reporting guideline for systematic reviews. Methodology/reporting-guideline paper for systematic reviews.
Pallottino et al. (2025)	Review/Survey	Review of generative AI applications and perspectives in agriculture covering GANs, VAEs, transformers, and diffusion in precision/animal farming and food chain. Literature review/survey paper.
Paulin and IvasicKos (2023)	Review/Survey	Review and analysis of synthetic dataset ("synthsets") generation methods/techniques in computer vision domain. Literature review of synthetic dataset generation methods.
Qin et al. (2025)	Review/Survey	Systematic review of deep-learning applications in lettuce cultivation (pest/disease, monitoring, yield, weeds, irrigation). Systematic review paper.
Raghunathan (2021)	Review/Survey	Annual Review article on synthetic data approaches (full/partial/selective synthesis), differential privacy, and statistical inference. Review article on statistical synthetic data, not agricultural CV.
Rahman et al. (2024)	Review/Survey	PRISMA-based review of GAN-based image augmentation in farming (livestock and crop), analyzing 128 publications across plant disease/weed/animal applications. Systematic literature review.

continued on next page

Table S2 – continued from previous page

Reference	Category	Contribution / why excluded from quantitative table
Rugji et al. (2025)	Review/Survey	Critical review of AI applications in food safety, agriculture, and food security covering ML, computer vision, predictive analytics, and supply-chain optimization. Literature review paper.
Schieber et al. (2024)	Review/Survey	Systematic PRISMA review of indoor synthetic data generation methods for object detection, 6D pose estimation, and scene understanding. Literature review (34 included studies); not an empirical agricultural application.
Shahriar et al. (2025)	Review/Survey	Comprehensive review of generative AI (GANs, VAEs, LLMs) applications in digital agri-food sector covering productivity, quality, safety, sustainability, and ethics. Literature review/survey; not an empirical study.
Soni et al. (2023)	Review/Survey	Review of Blender 3D software versions, interfaces, and capabilities for general 3D modeling/animation. Software/tool review paper; not tied to an agricultural application.
Soualiou et al. (2021)	Review/Survey	Review of Functional-Structural Plant Models (FSPMs) and their applications in crop science. Literature review, not an empirical SDG study.
Szeliski (2006)	Review/Survey	Tutorial on image alignment and stitching algorithms (motion models, direct/feature-based registration, blending). Foundational computer-vision tutorial; not agricultural application.
Tang et al. (2023)	Review/Survey	Review of optimization strategies for fruit detection in unstructured orchard environments (pre/post image-sampling). Literature review; not an empirical study.
Wang et al. (2024)	Review/Survey	Review of the GreenLab functional-structural plant model (FSPM), its theory, applications, and software. Review article on a process-based plant simulation framework; not an empirical SDG-for-detection study.
Wu et al. (2023)	Review/Survey	Review of remote-sensing-based crop monitoring challenges and opportunities. Review article, not an empirical SDG study.
Xu et al. (2024)	Review/Survey	Perspective on plant-disease recognition datasets in the deep-learning era; proposes a taxonomy. Perspective/position article, not empirical SDG.
Yang et al. (2025)	Review/Survey	Comprehensive survey of diffusion-model methods and applications. Methodological survey on diffusion models; not agricultural-application study.
Dataset releases and benchmark papers		
Baker et al. (2024)	Tool/Platform	Introduces Synavis pipeline coupling CPlantBox FSPM with Unreal Engine for synthetic dataset generation. Technical advance / pipeline paper; no downstream DL model performance comparison reported.

continued on next page

Table S2 – continued from previous page

Reference	**Category**	**Contribution / why excluded from quantitative table**
Choi et al. (2024)	Dataset paper	introduces DAVIS-Ag synthetic plant dataset (502K images, 632 orchards) built using AgML and Helios 3D simulator for active vision benchmarking. dataset/-platform release paper; provides baseline benchmarks but the main contribution is a synthetic dataset for agricultural active-vision research, not an empirical performance study with synthetic-vs-real comparison
Demieville et al. (2025)	Dataset paper	Phenomics dataset note for EMS-mutagenized sorghum population (Field Scanner data, 194 TB). Real-data dataset paper; no synthetic data generation.
Greff et al. (2022)	Tool/Platform	introduces Kubric, an open-source Python framework on PyBullet+Blender for generating photo-realistic synthetic datasets at scale. simulator/tool/platform paper (not agricultural-specific); generic synthetic data generator demonstrated on 13 datasets covering NeRF, optical flow, etc.
Mertolu et al. (2024)	Dataset paper	Introduces PLANesT-3D real-plant point-cloud annotated dataset (34 SfM/MVS plant models) and SP-LSCnet semantic segmentation method; dataset paper, no synthetic data generation.
Ojo et al. (2024)	Dataset paper	Splanting: 3D Gaussian Splatting capture pipeline for plant shoot architecture; introduces splant capture procedure and exemplar dataset across plant species. Tools/dataset paper introducing 3DGS for plant phenotyping; no synthetic-vs-real model performance comparison.
Sarzaeim et al. (2023)	Dataset paper	Geospatial climate + multi-omics database (CLIM4OMICS) for maize phenotype prediction with QC/CC analytics. Database/dataset paper; no synthetic image generation.
Schunck et al. (2021)	Dataset paper	Releases Pheno4D, a multi-temporal labeled 3D point-cloud dataset of maize and tomato plants for phenotyping. Dataset-only paper; provides annotated data, no synthetic generation comparison.
Sun et al. (2023)	Dataset paper	Constructs an annotated 3D MVS soybean point-cloud dataset (Soybean-MVS) covering whole growth period and tests RandLA-Net/BAAF-Net. Dataset paper (real MVS-reconstructed scans), no synthetic data generation.
Tabaa and Caro (2025)	Dataset paper	GreenhouseSplat: a Gaussian-Splatting framework and dataset of 82 photorealistic cucumber plants across 8 row ends (4 rows scanned at both ends) for ROS-based mobile-robot simulation. Tools/dataset paper (simulation framework + dataset for robotics evaluation, no downstream detection-task comparison).
Wiesner-Hanks et al. (2018)	Dataset paper	Dataset paper releasing 18,222 maize NLB-annotated field images (handheld/-boom/drone). Pure dataset paper, no synthetic-data generation.
Tools, simulators, and modelling platforms		

continued on next page

Table S2 – continued from previous page

Reference	Category	Contribution / why excluded from quantitative table
An et al. (2026)	Tool/Platform	Proposes AgriGaussian, a 3D Gaussian Splatting-based reconstruction method for high-fidelity field-scale phenotyping of chili pepper plants using a phenotyping robot. Pure 3D reconstruction / phenotyping pipeline (3DGS + Depth-Anything v2 + chunked training); produces reconstructed point clouds and PSNR/SSIM/LPIPS metrics for reconstruction quality, not synthetic 2D images for downstream model training. No generative model for data augmentation and no real-vs-synthetic downstream comparison.
Buckley et al. (2023)	Other	Plant physiology simulation study using Helios ray-tracing of canopy sunfleck dynamics for stomatal optimality models. Uses Helios as a biophysical simulator; not generating training images or training a CV model.
Calera et al. (2023)	Other	Vision-based under-canopy navigation system for the TerraSentia agricultural rover using only real camera data. Real-data-only navigation pipeline; no synthetic data generation or training.
Cheng et al. (2026)	Tool/Platform	BlenderProc-based synthetic image generation pipeline for AI-based instance segmentation in robotic assembly. Industrial/manufacturing application (robotic assembly with CAD models); not agricultural.
Du et al. (2026)	Tool/Platform	Plant-to-Camera 3D reconstruction pipeline (SegFormer foreground segmentation + SRCI calibration-constrained sparse reconstruction + CL-MVSNet dense reconstruction) for indoor rapeseed and other-species phenotyping. Real-world 3D reconstruction / phenotyping pipeline producing point clouds; no generative model and no synthetic-augmentation downstream comparison.
Emmi et al. (2013)	Tool/Platform	Tools/platforms paper describing SEARFS, an open-source 3D simulation environment for fleets of agricultural robots; no synthetic-data generation for training a perception model.
Griwodz et al. (2021)	Tool/Platform	Tools paper introducing AliceVision Meshroom, an open-source photogrammetry / 3D reconstruction pipeline; general-purpose, not agricultural and not synthetic-data generation.
Gutirrez Cejudo et al. (2024)	Tool/Platform	Tools/platforms paper proposing Agri-RO5, a multi-agent Unity3D + ROS architecture for agrirobot fleet simulation; no DL training or synthetic-data generation comparison.
Heidsieck et al. (2025)	Tool/Platform	Software/tool paper introducing point-cloud handling in the GroIMP FSPM platform. Tools/platform paper; no empirical agricultural ML task or synthetic-data comparison.

continued on next page

Table S2 – continued from previous page

Reference	Category	Contribution / why excluded from quantitative table
Joshi et al. (2026)	Tool/Platform	YOLOv8n-based detection of nutritional deficiencies (Healthy/N/K/N_K) on tomato leaves using a "layered" classical augmentation pipeline (flip, rotation, saturation, exposure, brightness) on the OLID I dataset; deployed as the AgriCure Android app. Uses only conventional pixel-level augmentation (flip, rotation, saturation 20%, brightness +/-25% on N class, exposure +/-15% on N_K class); no generative model (GAN/diffusion/copy-paste) and no synthetic-image generation. mAP@0.5 rises from 0.611 (no aug) -¿ 0.762 (standard aug) -¿ 0.927 (layered aug + class balancing), but this is a class-balancing/traditional-aug study.
Lei et al. (2021)	Tool/Platform	Encapsulated bioreactor system for rapid wheat plant-development data collection plus AI modelling; mentions a Blender synthetic wheat-leaf dataset for LAI estimation but only as work-in-progress. System/platform paper (work-in-progress); no quantitative downstream performance with vs. without synthetic data reported.
Li et al. (2024c)	Tool/Platform	Unreal Engine 5 + ROS photorealistic robot-arm simulator (AgriRoboSimUE5) generating RGB/segmentation/depth for 10 Quixel plant assets, with a texture-based multi-label segmentation method, plus NeRF reconstruction. Simulator/tool paper; evaluates simulator quality (segmentation pixel counts, point-cloud counts, NeRF PSNR/SSIM) but does not train/compare a downstream agricultural recognition model on real vs synthetic data.
Ma et al. (2026)	Tool/Platform	Single-image 3D botanical-tree reconstruction pipeline that uses a Pix2Pix-based conditional GAN with species conditioning to predict 2D skeleton depth maps from segmented RGB images, followed by procedural modelling for branch and leaf generation. All three networks (YOLOv7 segmentation, CNN species classifier, cGAN skeletonisation) are trained exclusively on procedurally generated synthetic data; the paper reports only depth-error against synthetic ground truth (avg 0.07) plus qualitative comparisons on real images and gives no real-vs-synthetic-trained downstream comparison. The application target is virtual-environment tree generation (computer graphics) rather than an empirical agricultural-vision task with quantitative real-world metrics.
McDonald et al. (2022)	Tool/Platform	ImageJ macro for color-space thresholding and lesion counting of soybean frogeye leaf spot; image processing pipeline, no synthetic data or ML.
Mitchell et al. (2019)	Foundational method	Model Cards framework for ML model documentation/transparency; not agricultural, not synthetic data.
Pavlenko et al. (2024)	Other	Case study using Farming Simulator video game with precision-farming DLCs as an educational/awareness tool. Gamification/education case study, not a synthetic-data-for-CV agricultural study.

continued on next page

Table S2 – continued from previous page

Reference	Category	Contribution / why excluded from quantitative table
Pradal et al. (2008)	Tool/Platform	OpenAlea - visual programming and component-based software platform for plant modelling. Software platform/tool paper for FSPM, not synthetic-data CV.
Sheffer et al. (2026)	Tool/Platform	NeRF-based pipeline (Instant-NGP backbone with low-light/RAW loss and per-ray segmentation/height-based canopy peeling) for reconstructing the terrain and stems beneath dense forest canopies from RGB drone imagery, with downstream evaluations on person detection and tree-stem counting. Forest/SAR application rather than crop-agriculture. A Blender-rendered synthetic CAD forest scene is used only as ground-truth reference for measuring reconstruction quality (M-SSIM 0.65/0.44 with/without canopy versus 3DGS 0.64/0.40); it is not used to train a downstream agricultural model. No real-vs-synthetic-trained downstream comparison.
Tsolakis et al. (2019)	Tool/Platform	AgROS, a ROS-based emulation tool for evaluating agricultural UGVs. Tools/platform paper; no SDG-based ML training.
Uchiyama et al. (2017)	Tool/Platform	Compact indoor 3D phenotyping platform (KOMATSUNA dataset) with annotation tool for leaf segmentation/counting. Dataset/platform release paper; no synthetic data generation or model performance comparison.
Vaillant et al. (2022)	Tool/Platform	Jupyter-based virtual modelling environment for FSPMs, applied to V-Mango tree development model. Tools/modelling-platform paper, not SDG for ML.
Wang et al. (2025c)	Tool/Platform	VIO-MPC fusion framework for autonomous orchard headland turning under GNSS denial; field-tested in banana orchards. Robotics navigation/control study; no synthetic-data generation for vision model training.
Wu et al. (2020)	Tool/Platform	Portable low-cost MVS-based 3D reconstruction phenotyping platform for maize shoots. Hardware/software platform paper using multi-view stereo for 3D reconstruction; not synthetic data for a CV detector.
Xiang et al. (2023)	Tool/Platform	PhenoBot 3.0 stereo-vision robot + AngleNet CNN for field-based maize leaf-angle measurement. Field robotics + keypoint CNN study; uses real stereo imagery, no synthetic-data generation pipeline.
Xu et al. (2025b)	Tool/Platform	Physics-based RL teacher-student framework for whole-body human-object interaction imitation. Not agricultural; humanoid motion imitation paper. Out of scope.
Zhou et al. (2020)	Tool/Platform	CPlantBox whole-plant modelling framework for water/carbon flow simulation (root + shoot architecture). Plant-physiology simulation framework; not synthetic image generation for CV.
Foundational computer-vision, generative-AI, and 3D methods		

continued on next page

Table S2 – continued from previous page

Reference	Category	Contribution / why excluded from quantitative table
Arshad et al. (2024)	Foundational method	Evaluates NeRF techniques (Instant-NGP, NeRFacto, TensoRF) for 3D plant geometry reconstruction vs LiDAR ground truth. Methodology evaluation paper for 3D reconstruction; not synthetic data for downstream model training.
Bailey (2019)	Foundational method	Presents Helios, a 3D plant and environmental biophysical modeling framework. Foundational tool/framework paper for biophysical simulation.
Baker et al. (2025)	Foundational method	Extends Synavis framework to couple FSPM photosynthesis with Unreal Engine ray tracing for synthetic data production. Technical advance / methodology paper; no downstream classifier/detector benchmark.
Bay et al. (2008)	Foundational method	Original SURF (Speeded-Up Robust Features) detector and descriptor paper. Foundational computer vision methodology paper, not agriculture-specific or SDG-related.
Betker et al. (2023)	Foundational method	DALL-E 3 technical report on training text-to-image models with synthetic image captions for improved prompt following. foundational generative-model methodology paper; not an agricultural application; uses synthetic captions (text), not synthetic images, to train the generator
Bochkovskiy et al. (2020)	Foundational method	YOLOv4 - foundational object detector paper proposing CSP, Mosaic augmentation, and bag-of-freebies on MS COCO. Generic object-detection methodology paper, not tied to an agricultural application or synthetic-data generation.
Borkman et al. (2021)	Foundational method	Unity Perception package – open-source toolset for synthetic computer-vision data generation. Foundational tool/platform paper (Unity engineering team); demonstrated on generic 2D object detection (groceries-style), not agricultural.
Bubencek (2022)	Foundational method	Unity-based plugin for generating synthetic image data (RGB, depth, normals, optical flow, segmentation) for ML. Generic methodology/tool paper without an agricultural application or quantitative SDG comparison.
Dellmann and Berns (2022)	Foundational method	Position/overview chapter on building realistic Unreal-Engine simulations for agricultural robots. Methodology/position paper; no quantitative SDG-vs-real comparison or trained model evaluation.
Denninger et al. (2019)	Foundational method	BlenderProc – modular Blender-based pipeline for procedural photorealistic synthetic data generation. Foundational synthetic-data tool paper, not tied to agricultural application.
Denninger et al. (2023)	Foundational method	BlenderProc2 – JOSS software paper for an updated Python-API procedural rendering pipeline. Foundational tool/software paper.
DeVries and Taylor (2017)	Foundational method	Cutout regularization technique for CNNs evaluated on CIFAR-10/100 and SVHN. Generic ML regularization methodology paper, not agricultural.

continued on next page

Table S2 – continued from previous page

Reference	Category	Contribution / why excluded from quantitative table
Dinh et al. (2015)	Foundational method	Foundational paper introducing NICE (Non-linear Independent Components Estimation), a normalizing-flow generative model evaluated on MNIST/TFD/SVHN/CIFAR-10; not agricultural.
Du and Mordatch (2020)	Foundational method	Foundational ML methodology paper proposing energy-based models (EBM) with Langevin-dynamics MCMC sampling for ImageNet/CIFAR generation; not agricultural.
Espejel Flores et al. (2025)	Foundational method	Tools/methodology paper comparing Unity vs Gazebo simulators for digital twins of robotic tomato harvesting; visual-fidelity / integration analysis, no DL training comparison.
Gebru et al. (2021)	Foundational method	Methodology/position paper proposing "datasheets for datasets" documentation framework; not agricultural and not synthetic-data-generation research.
Goodfellow et al. (2014)	Foundational method	Foundational paper introducing Generative Adversarial Networks (GANs) on MNIST/CIFAR/TFD; not agricultural.
Habaragamuwa et al. (2022)	Foundational method	Methodology paper using disentangled VAE for explainable plant disease classification on PlantVillage; VAE generates feature-variation visualizations for explanation, not synthetic training data; no real-vs-synthetic-augmented comparison.
Hemmerling et al. (2008)	Foundational method	Foundational paper on the XL language and GroIMP modeling environment with example tree-competition simulation. Methodological/platform paper for FSPM; not a synthetic-data utility study.
Heusel et al. (2017)	Foundational method	Foundational paper introducing TTUR for GANs and the FID metric. Generative-model methodology paper; not agricultural.
Ho et al. (2020)	Foundational method	Foundational paper on Denoising Diffusion Probabilistic Models (DDPM). Generative-model methodology paper; not agricultural.
Isinkaye et al. (2025)	Foundational method	Hybrid VAE + Vision Transformer model for multi-class plant disease classification on PlantVillage (corn/potato/tomato). VAE used for dimensionality reduction/feature extraction (not for synthetic image generation); only on-the-fly augmentation (flip/rotate/zoom) was applied. No generative synthetic dataset is created or evaluated for downstream training, so this is a methodology paper rather than a synthetic-data study.
Isola et al. (2017)	Foundational method	Foundational pix2pix paper on conditional adversarial nets for image-to-image translation. Methodology paper, not agricultural.
Karras et al. (2021)	Foundational method	Foundational StyleGAN3 paper introducing alias-free generators with translation/rotation equivariance. Generative-model methodology paper, not agricultural.

continued on next page

Table S2 – continued from previous page

Reference	Category	Contribution / why excluded from quantitative table
Karras et al. (2020)	Foundational method	StyleGAN2 - foundational GAN methodology paper analyzing and improving StyleGAN image quality. Methodological paper introducing StyleGAN2; not tied to an agricultural application.
Karras et al. (2019)	Foundational method	Original StyleGAN paper proposing a style-based generator architecture for GANs. Foundational generative model methodology paper; not agricultural.
Kerbl et al. (2023)	Foundational method	Original 3D Gaussian Splatting paper for real-time radiance field rendering. Foundational 3D representation methodology paper; not agricultural.
Kingma and Welling (2022)	Foundational method	Original Variational Autoencoder (VAE) paper introducing the SGVB estimator and AEVB algorithm. Foundational generative model methodology paper.
Kniemeyer et al. (2007)	Foundational method	Introduces GroIMP platform and XL/RGG language for functional-structural plant modeling. Tools/platform paper; foundational FSPM software.
Lee and Park (2025)	Foundational method	WGAN-GP-based conditional GAN with an Extreme Critic head (multi-head Critic that explicitly classifies extreme ¿=20 mm/hr precipitation) for downscaling ERA-5 (≈25 km) coarse meteorological fields to MRMS 1-km precipitation in the Illinois Corn Belt. Generative-model methodology paper for meteorological precipitation downscaling; the ”synthetic” output is the high-resolution precipitation field itself (regression target = MRMS), evaluated by MSE/FSS/temporal correlation against MRMS ground truth, not used to train an agricultural CV model. No real-vs-synthetic downstream agricultural-vision comparison; the application context (agricultural water management) is motivational rather than empirical.
Lei and Bailey (2025)	Foundational method	Soil Optics Generative Model (SOGM) using a denoising diffusion probabilistic model with text-based property inputs to simulate soil reflectance spectra (400-2499 nm). Generative methodology paper for radiative-transfer/remote-sensing inputs; not an agricultural vision/object-detection study with real-vs-synthetic downstream model comparison.
Li et al. (2026a)	Foundational method	Object-centric 3D Gaussian Splatting (3DGS) framework for strawberry plant reconstruction and phenotyping using SAM-2-based foreground masking and RGBA-loss-masked optimization, evaluated on 15 real plants for plant height and canopy width estimation. Methodology paper for neural rendering reconstruction; no synthetic image generation for downstream model training and no real-vs-synthetic comparison. The 3DGS reconstructs from real iPhone-captured video; trait estimation uses DBSCAN+PCA on the reconstructed point cloud rather than a model trained on synthetic data.

continued on next page

Table S2 – continued from previous page

Reference	Category	Contribution / why excluded from quantitative table
Mildenhall et al. (2020)	Foundational method	Foundational paper introducing Neural Radiance Fields (NeRF) for novel-view synthesis from posed images using an MLP and volume rendering. Foundational generative/rendering methodology paper; not an agricultural application. Establishes the NeRF technique used by many downstream agricultural studies.
Mirza and Osindero (2014)	Foundational method	Foundational/methodological paper introducing Conditional GANs (cGAN) on MNIST; not agricultural.
Modak and Stein (2024b)	Foundational method	Pipeline architecture combining SAM and Stable Diffusion (Dreambooth) to synthesize weed-infested field images for object detection training. Methodology/pipeline paper proposing synthetic image generation; reports a qualitative and quantitative CLIP-IQA comparison with real weed images and identifies PSNR and Inception Score as additional metrics for planned future quantitative evaluation; it does not train a downstream detection model or report task performance with vs. without synthetic data. Future-work scope.
Morrical et al. (2021)	Foundational method	NViSII Python-based path-tracing renderer tool (NVIDIA OptiX); methodological/tool paper, not an agricultural application.
Nichol et al. (2022)	Foundational method	GLIDE foundational paper on text-guided diffusion image generation; methodological, not agricultural.
Northcutt et al. (2021)	Foundational method	Study of pervasive label errors in test sets of 10 ML benchmarks (ImageNet, CIFAR, etc.) and consequences for model selection. ML methodology / benchmark-analysis paper, not an agricultural synthetic-data study.
Pan et al. (2025)	Foundational method	TokenHSI - unified transformer-based policy for physics-based human-scene interactions (sitting, climbing, carrying). Physics-based character animation methodology, not agricultural.
Peng et al. (2023)	Foundational method	GNSS-free end-of-row detection and headland maneuvering for orchard navigation using a depth camera. Robot navigation/sensor methodology paper, not synthetic-data-driven CV.
Prez et al. (2023)	Foundational method	Foundational Poisson image-editing paper introducing seamless cloning via guided interpolation by solving Poisson PDE with Dirichlet boundary conditions. Foundational image-processing methodology paper (2003 Microsoft Research) widely cited for cut-paste compositing in synthetic-data pipelines; not agricultural.
Podell et al. (2023)	Foundational method	SDXL - improved latent diffusion model for high-resolution text-to-image synthesis with multiple novel conditioning schemes and refinement model. Foundational generative model paper (Stable Diffusion XL); not an agricultural application.
Porter and Duff (1984)	Foundational method	Foundational paper on compositing digital images with alpha channel and matte algebra. Classic computer graphics methodology paper (alpha compositing).

continued on next page

Table S2 – continued from previous page

Reference	Category	Contribution / why excluded from quantitative table
Qu et al. (2026)	Foundational method	Multi-Scale NeRF (MSNeRF) for low-quality-image rice 3D reconstruction together with VRKGNet, a multimodal point-cloud + image segmentation network, evaluated on the newly released MMR rice dataset for semantic and instance segmentation. The "synthetic" data referred to in the paper are pseudo-ground-truth point clouds derived from real multi-view image acquisition (NeRFacto-trained, denoised, expert-refined). No generative model produces training images for a downstream comparison and no real-vs-synthetic ablation exists - the paper is a 3D-reconstruction + segmentation methodology paper with all training/testing on real-acquired data.
Radford et al. (2016)	Foundational method	DCGAN - foundational paper introducing deep convolutional GAN architectural constraints for stable training and unsupervised representation learning. Foundational generative model paper (DCGAN).
Ramesh et al. (2022)	Foundational method	unCLIP/DALL-E 2 - hierarchical text-conditional image generation using CLIP latents with two-stage prior + diffusion decoder. Foundational generative model paper (DALL-E 2/unCLIP).
Rezende et al. (2014)	Foundational method	Foundational ML paper introducing stochastic backpropagation and deep latent Gaussian models (DLGMs) for variational inference. Methodological/foundational generative model paper, not tied to agriculture.
Rezende and Mohamed (2016)	Foundational method	Foundational ML paper introducing normalizing flows for variational inference. Methodological/foundational generative model paper, not tied to agriculture.
Rippel and Adams (2013)	Foundational method	Introduces the deep density model (DDM) for high-dimensional probability estimation via bijective deep transformations. Foundational ML methodology paper on density estimation, not agricultural.
Rombach et al. (2022)	Foundational method	Introduces Latent Diffusion Models (LDMs / Stable Diffusion) for high-resolution image synthesis. Foundational generative-model paper (Stable Diffusion); not an agricultural application study.
Saleem et al. (2024)	Foundational method	Class-specific data augmentation policy search via genetic algorithm for soybean stress classification (uses standard transforms, not synthetic image generation). Uses augmentation transforms (rotation/flip/etc.), not generative synthetic data; methodology focused on policy selection.
Schnberger and Frahm (2016)	Foundational method	Introduces COLMAP, an incremental Structure-from-Motion pipeline for 3D reconstruction. Foundational 3D reconstruction tool/methods paper, not agricultural.
Shah et al. (2017)	Foundational method	Introduces AirSim, an Unreal Engine-based high-fidelity simulator for autonomous vehicles. Foundational simulator/tools paper, not agricultural application.

continued on next page

Table S2 – continued from previous page

Reference	Category	Contribution / why excluded from quantitative table
Sohl-Dickstein et al. (2015)	Foundational method	Foundational paper introducing diffusion probabilistic models via nonequilibrium thermodynamics. Foundational generative-model methodology paper, not agricultural.
Song and Ermon (2020a)	Foundational method	Foundational paper introducing score-based generative modeling via score matching with Langevin dynamics (NCSN), evaluated on MNIST/CelebA/CIFAR-10. Methodological/foundational generative-model paper; not an agricultural application.
Song and Ermon (2020b)	Foundational method	Methodological improvements to score-based generative models (NCSNv2) scaling to higher-resolution images. Foundational generative-modeling paper; no agricultural task.
Song et al. (2021)	Foundational method	Foundational ICLR paper unifying score-based and diffusion modeling via stochastic differential equations. Foundational generative-modeling methodology, not agricultural.
Song et al. (2022)	Foundational method	Score-based generative models applied to inverse problems in CT/MRI medical imaging. Methodological/medical imaging paper, not an agricultural SDG application.
Tancik et al. (2023)	Foundational method	Nerfstudio modular Python framework for NeRF development. Foundational tools/methodology paper, not agricultural.
Tian et al. (2025)	Foundational method	Comparative analysis of NeRF, 3DGS, photogrammetry, and TLS for 3D reconstruction of urban forest stands and tree-parameter extraction. Methodological comparison of 3D-reconstruction techniques; no SDG-trained downstream detection model.
Uyehara et al. (2025)	Foundational method	AgRowStitch: open-source pipeline using SuperPoint + LightGlue for high-fidelity image stitching of ground-based agricultural row images. Tools/methodology paper for image stitching; no SDG.
van den Oord et al. (2016)	Foundational method	Foundational paper introducing Conditional Gated PixelCNN for image generation. Foundational generative-model paper; not agricultural.

continued on next page

Table S2 – continued from previous page

Reference	Category	Contribution / why excluded from quantitative table
Xiang and Ge (2026)	Foundational method	Presents a parametric 3D triangle-mesh model of the maize leaf using hyperbolic spiral midrib, tractrix-curve cross-section, and Sanderson blade-contour curves, with applications including geometric correlation analysis, radiation-capture simulation, and synthetic dataset generation for phenotyping. Methodology/tool paper. The maize-leaf parametric model is presented and its outputs (RGB image, binary mask, leaf segmentation, leaf-angle/length annotation, point cloud, depth map) are demonstrated as candidate synthetic data for phenotyping pipelines, but the paper does NOT train or evaluate any deep-learning model on real-vs-synthetic data and reports no downstream task metrics. Reported quantitative results concern leaf surface-area correlations (e.g., r=0.040 for M1 vs leaf area, p¡0.005) and radiation-capture simulations (GHI 893.27, 924.81, 952.68 W/mˆ2 across three latitudes), not perception accuracy.
Yun et al. (2019)	Foundational method	CutMix regularization/augmentation strategy for training stronger CNN classifiers (ImageNet, CIFAR, Pascal VOC). Generic ML methodology paper; not tied to an agricultural application.
Zhang et al. (2021)	Foundational method	Deep image compositing (multi-stream fusion network) for automatic portrait foreground+background blending. Generic image-compositing methodology; not agricultural.
Zhang and Sabuncu (2018)	Foundational method	Generalized cross-entropy loss for training DNNs robust to noisy labels (CIFAR, FASHION-MNIST). Foundational ML methodology paper; synthetic noise on labels but not agricultural application.
Zhang et al. (2018)	Foundational method	Original mixup paper proposing convex-combination data augmentation for general deep-learning generalization. Foundational/methodological paper; not agricultural application.
Zhao et al. (2019)	Foundational method	Saliency-map-aided GAN for cross-camera RAW-to-RGB image mapping (proposed model "RAW2RGB-GAN": U-Net generator + PatchGAN discriminator; auxiliary saliency maps produced by a separate pre-trained SalGAN). Generic image-translation methodology; not agricultural application.
Zhong et al. (2020)	Foundational method	Original Random Erasing data-augmentation method (general image classification, detection, re-ID). Foundational/methodological paper, not agricultural SDG.
Zhu et al. (2017)	Foundational method	Original CycleGAN paper for unpaired image-to-image translation with cycle-consistency loss. Foundational generative-model paper; not agricultural application.
Adjacent empirical studies and out-of-scope applications		

continued on next page

Table S2 – continued from previous page

Reference	Category	Contribution / why excluded from quantitative table
Bailey (2025)	Other	Derives geometric model for sunlit/shaded leaf area fraction in heterogeneous canopies; evaluates against 3D leaf-resolving simulation. Biophysical/canopy model paper, not deep-learning synthetic data study.
Cao et al. (2022)	Other	VAE-based multiple instance regression for corn yield prediction from remote sensing. Uses VAE for anomaly detection/feature learning on remote-sensing variables; no synthetic image generation for CV training.
Deane et al. (2021)	Other	DynaDog+T parametric canine model for synthetic dog images and segmentation. Pet/animal computer-vision domain, not agricultural (livestock or crop) – dogs are not livestock.
Dutagaci et al. (2020)	Other	Datasets-only paper releasing ROSE-X, an X-ray CT 3D rosebush dataset with organ-level annotations and four baseline leaf/stem segmentation methods; benchmark resource, not synthetic-data generation.
Goff et al. (2025)	Other	Self-driving research paper from comma.ai using world-model and reprojective simulation for end-to-end driving policy training; not agricultural.
He et al. (2024a)	Other	Cross-regional crop classification on Sentinel-2 using a hypothesis testing distribution method, no synthetic data generation. Uses real Sentinel-2 imagery and ML classifiers; not a synthetic-data study.
He et al. (2025b)	Other	lightweight YOLOv5n + Swin Transformer for cottonseed defect detection and two-stage classification on a 915-cottonseed dataset. empirical agricultural vision study but uses only real images with traditional augmentation (rotation/scaling/cropping/noise); no synthetic image generation via GAN/diffusion/3D-render/cut-paste, so out of scope
He et al. (2024b)	Other	Swin Transformer + Circular Smooth Label improvements to YOLOv5 for oriented bounding box detection of maize leaves to estimate azimuth angles from UAV images. empirical agricultural study but uses only real UAV-collected maize images (180 images, ≈40,000 leaves) with no synthetic data generation; out of scope
Hoppe et al. (2024)	Other	Study of transferability of ML crop classifiers on Sentinel-2 imagery. Real remote-sensing study; no synthetic data generation.
Huang et al. (2024)	Other	data article releasing AgroSegNet, a 50,000-pair Blender-rendered synthetic shadow dataset with physics-accurate shadow masks for agricultural settings. dataset-only release (data-in-brief article); does not train or evaluate downstream models for an empirical agricultural task
Jarmusch et al. (2025)	Other	Microbenchmark analysis of NVIDIA Blackwell GPU architecture. GPU/HPC paper; not agricultural.

continued on next page

Table S2 – continued from previous page

Reference	Category	Contribution / why excluded from quantitative table
Jimnez and Decaestecker (2024)	Other	Study of imperfect annotations on CNN training for nuclei segmentation/classification in digital pathology. Digital pathology, not agricultural.
Jordon et al. (2022)	Other	Royal Society explainer on synthetic data: what, why, and how, focused on privacy. Position/explainer paper, not agricultural.
Kusrini et al. (2020)	Other	Conventional image augmentation (noise/blur/contrast/affine) for VGG-16 multi-class mango pest classification. Uses only classical geometric/photometric augmentation (no generative model, no 3D rendering, no cut-paste); not a synthetic-data study under the inclusion criteria.
Le et al. (2021)	Other	EDEN multimodal synthetic dataset of enclosed garden scenes (¿300K images, 100+ Blender Cycles garden models) with semantic/depth/normal/flow/intrinsic annotations. Synthetic dataset release paper; benchmarks downstream tasks (semantic segmentation, monocular depth) but its purpose is dataset provision rather than an applied agricultural vision study.
Li et al. (2024e)	Other	FSPM (GroIMP) study comparing tomato planting orientations and row spacings for canopy light interception and yield in solar greenhouses. FSPM-based agronomic simulation study, not synthetic image data for training a perception model.
Li et al. (2024b)	Other	FSPM (GroIMP) study optimizing daylily planting patterns for bud yield via physiological/source-sink modeling. FSPM-based plant-physiology simulation, not SDG for ML perception.
Liu et al. (2022b)	Other	3D light-environment simulation of Chinese solar greenhouses using GroIMP and ray tracing to determine optimal greenhouse orientation; not a synthetic-data-for-ML study.
Liu et al. (2022c)	Other	Hardware development of an impact-type grain-yield monitoring system with mass-flow/elevator/ground-speed/moisture sensors for winter wheat combines. Sensor/system development paper; no computer-vision model and no synthetic data.
Liu et al. (2022a)	Other	3D in-silico light simulation of reflective films inside Chinese solar greenhouse with tomato canopy using GroIMP; not synthetic-data-for-ML.
Martinelli et al. (2025)	Other	Digital reconstruction of hazelnut tree (laser scan + photogrammetry) for VR/haptic pruning training; not ML synthetic-data utility study.
Mostofi et al. (2024)	Other	VAE integrated with multi-head graph attention network to address class imbalance in construction-productivity prediction datasets. Application is construction management, not agriculture. Out of scope.
Nagasubramanian et al. (2021)	Other	Compares active learning strategies (DBAL, entropy, least confidence, core-set) for soybean stress and weed classification; no synthetic data generation.

continued on next page

Table S2 – continued from previous page

Reference	Category	Contribution / why excluded from quantitative table
Nawaz et al. (2022)	Other	ResNet-34 + CBAM Faster-RCNN for tomato leaf disease classification on PlantVillage; uses real data only with traditional augmentation, no synthetic data generation.
Niu et al. (2023)	Other	UAV multispectral time-series imagery and machine-learning regression (PLSR/SVM/GBM/RF) for predicting needle N and NSC content in slash pine for breeding selection. Empirical agricultural study but uses only real UAV imagery and chemical reference; no synthetic-data generation or use.
Qin et al. (2024)	Other	Unity 3D synthetic dataset of 1.2K virtual plant-root models (apple seedling, sweet potato) with 80K rendered 2D views for 3D-reconstruction benchmarking. Dataset-only release / benchmark paper; tests existing 3D-reconstruction algorithms but does not train an agricultural perception model with synthetic-vs-real comparison.
Sapkota and Karkee (2025a)	Other	Demonstrates DALL-E text-to-image and image-to-image generation for agricultural visualization; compares image quality (PSNR/FSIM) but does not train a downstream task model. Image-quality / communication study using DALL-E; no downstream detection/classification performance comparison.
Schaudt et al. (2023)	Other	Medical-imaging study comparing five generative models (a StyleGAN-based WGAN-GP GAN, a DDPM, and three Stable-Diffusion fine-tuning variants: standard fine-tuning, LoRA, and DreamBooth) for pneumonia chest X-ray synthesis and downstream classification on 1,082 CXR images. It is non-agricultural and therefore excluded from the agricultural quantitative table.
Sferrazza et al. (2024)	Other	Simulated humanoid robot benchmark (HumanoidBench) for whole-body locomotion and manipulation. Robotics simulation benchmark, not agricultural application.
Shi et al. (2025)	Other	Vision + 2D LiDAR fusion-based navigation line extraction for autonomous robots in pomegranate orchards. Sensor fusion / autonomous navigation paper using YOLOv8-ResCBAM trained on real images; no synthetic data generation.
Sivakumar et al. (2021)	Other	CropFollow vision-based under-canopy navigation system using monocular RGB and MPC for row-following in corn/soybean. Real-data deep learning navigation paper; no synthetic data generation.
Sivakumar et al. (2024)	Other	Field-deployment lessons from CropFollow++ keypoint-based under-canopy navigation system. Deployment study using real field data; no synthetic data generation.

continued on next page

Table S2 – continued from previous page

Reference	Category	Contribution / why excluded from quantitative table
Wang and Liu (2024)	Other	TomatoDet: Swin-DDETR-based detector with IBiFPN for greenhouse tomato disease detection; uses only conventional augmentation (no SDG). Uses only classical data augmentation (horizontal/vertical flip, brightness adjustment, Gaussian noise), no synthetic data generation.
Wei et al. (2024)	Other	NeRF-based (GS-NIR) implicit-surface 3D reconstruction of pepper plants for in-situ phenotyping. Uses NeRF for 3D reconstruction/phenotyping accuracy comparison; does not generate synthetic 2D training data for a downstream detector.
Xie et al. (2022)	Other	UAV-RGB SfM-based estimation of potato canopy coverage and plant height to evaluate spraying volume. Empirical UAV remote sensing study using SfM 3D point clouds; no synthetic training data for a learned model.
Zhai et al. (2025)	Other	CropCraft proposes inverse procedural modeling for 3D reconstruction of soybean and maize canopies from multi-view images by fitting a NeRF, then optimizing parameters of a procedural plant morphology model via Bayesian optimization with a histogram-statistics depth-map loss; validated on a real multi-view field dataset paired with manual LAI/leaf-angle measurements. Applied agricultural 3D-reconstruction paper. The procedural model produces synthetic depth maps used inside the optimization loop and is also used to train an MLP baseline on 20K synthesized (histograms, parameters) pairs, but the paper's evaluation is reconstruction-quality of canopy structure (LAIE, LAIPE, AME, ASDE) against ground-truth field measurements (LAI-2200 scanner, protractor) on 18 soybean scenes and 5 maize scenes - not a downstream perception model trained with vs without synthetic data on a real test set. CropCraft's results: soybean LAIE 0.69, LAIPE 0.15, AME 12.07 deg, ASDE 7.39 deg; maize LAIE 0.97, LAIPE 0.26. The MLP baseline (trained on 20K synthetic pairs) gives soybean LAIE 0.92 / LAIPE 0.24 / maize LAIE 2.79, illustrating a sim-to-real domain gap rather than a synthetic-augmentation benefit.
Zhang et al. (2024d)	Other	Bionic hexapod robot with adaptive gait/clearance for agricultural field scouting. Hardware/robotics development paper; no synthetic data generation.
Zhang et al. (2024a)	Other	Wheat Teacher: one-stage anchor-based semi-supervised wheat-head detector with pseudo-labeling and consistency regularization. Semi-supervised learning method, not generative synthetic data; no SDG.
Zheng et al. (2024)	Other	Tomato-NeRF for high-resolution 3D reconstruction of tomato plants from smartphone video. NeRF reconstruction for digital twin; geometry-quality study, no synthetic 2D training data for a detector.

continued on next page

Table S2 – continued from previous page

Reference	Category	Contribution / why excluded from quantitative table
Ziamtsov and Navlakha (2019)	Other	Machine-learning methods for lamina/stem classification, lamina counting, stem skeletonization on 3D point clouds of tomato/Nicotiana. Empirical phenotyping study using real 3D scans (54 plants); no synthetic data generation.