

UC Davis

UC Davis Previously Published Works

Title

Standardizing and Centralizing Datasets for Efficient Training of Agricultural Deep Learning Models.

Permalink

<https://escholarship.org/uc/item/22r0s3q4>

Authors

Joshi, Amogh

Guevara, Dario

Earles, Mason

Publication Date

2023

DOI

10.34133/plantphenomics.0084

Peer reviewed

RESEARCH ARTICLE

Standardizing and Centralizing Datasets for Efficient Training of Agricultural Deep Learning Models

Amogh Joshi^{1,2,3}, Dario Guevara^{1,2,3}, and Mason Earles^{1,2,3*}

¹Department of Viticulture and Enology, University of California, Davis, Davis, CA, USA. ²Department of Biological and Agricultural Engineering, University of California, Davis, Davis, CA, USA. ³AI Institute for Next-Generation Food Systems (AIFS), University of California, Davis, Davis, CA, USA.

*Address correspondence to: jmearles@ucdavis.edu

In recent years, deep learning models have become the standard for agricultural computer vision. Such models are typically fine-tuned to agricultural tasks using model weights that were originally fit to more general, non-agricultural datasets. This lack of agriculture-specific fine-tuning potentially increases training time and resource use, and decreases model performance, leading to an overall decrease in data efficiency. To overcome this limitation, we collect a wide range of existing public datasets for 3 distinct tasks, standardize them, and construct standard training and evaluation pipelines, providing us with a set of benchmarks and pretrained models. We then conduct a number of experiments using methods that are commonly used in deep learning tasks but unexplored in their domain-specific applications for agriculture. Our experiments guide us in developing a number of approaches to improve data efficiency when training agricultural deep learning models, without large-scale modifications to existing pipelines. Our results demonstrate that even slight training modifications, such as using agricultural pretrained model weights, or adopting specific spatial augmentations into data processing pipelines, can considerably boost model performance and result in shorter convergence time, saving training resources. Furthermore, we find that even models trained on low-quality annotations can produce comparable levels of performance to their high-quality equivalents, suggesting that datasets with poor annotations can still be used for training, expanding the pool of currently available datasets. Our methods are broadly applicable throughout agricultural deep learning and present high potential for substantial data efficiency improvements.

Introduction

Deep learning models have become standard for modern computer vision-based agricultural tasks. Examples of common standard tasks now largely automated by deep learning include fruit detection [1,2], crop and weed segmentation [3,4], and plant disease classification [5,6]. Certain deep learning models have even been used for tasks beyond standard single-RGB-image predictions, involving hyperspectral and thermal imagery for various analyses [7,8] or tasks involving multiple scales of data, such as spatiotemporal crop yield prediction [9]. Existing approaches to such tasks often involve direct applications of state-of-the-art models developed for other tasks in the field of deep learning, which have been thoroughly evaluated by works such as [10].

A largely prevalent problem in the agricultural domain is a substantial deficiency of task-specific data. Transfer learning [11] is an approach undertaken when training large deep learning models, which attempts to offset this data deficiency, therein transferring knowledge from a source task to the new, reduced size target dataset. In practice, this generally consists of replacing random weight initialization for model parameters with existing pretrained weights from a prior task, converting a

complete training task into a fine-tuning task using existing pretrained models—for instance, many image classification and object detection tasks start off with pretrained weights from the ImageNet [12] and COCO [13] datasets, respectively. However, these datasets are mostly generalized to common objects in environments ranging from inside a home to the streets of a city. Such environments consist of a broad range of objects and conditions, allowing models to locate specific objects with relatively greater ease. In contrast, agricultural environments are highly domain-specific, and these existing pretrained model approaches may not offer as consequential of a knowledge transfer for agricultural domains. An example of the contrast between the aforementioned conditions is shown in Fig. 1. This, in turn, suggests the potential need for alternative pretrained model approaches in order to improve data efficiency when training agricultural deep learning models. Previously, Nowakowski et al. [14], Moon and Son [15], and Shrivastava et al. [16] applied transfer learning to agricultural image classification tasks, demonstrating improved performance using pretrained weights on the ImageNet dataset, citing its large magnitude of data and certain specific classes relevant to agriculture. However, little to no prior work has been conducted

Citation: Joshi A, Guevara D, Earles M. Standardizing and Centralizing Datasets for Efficient Training of Agricultural Deep Learning Models. *Plant Phenomics* 2023;5:Article 0084. <https://doi.org/10.34133/plantphenomics.0084>

Submitted 23 January 2023
Accepted 2 August 2023
Published 6 September 2023

Copyright © 2023 Amogh Joshi et al. Exclusive licensee Nanjing Agricultural University. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY 4.0).

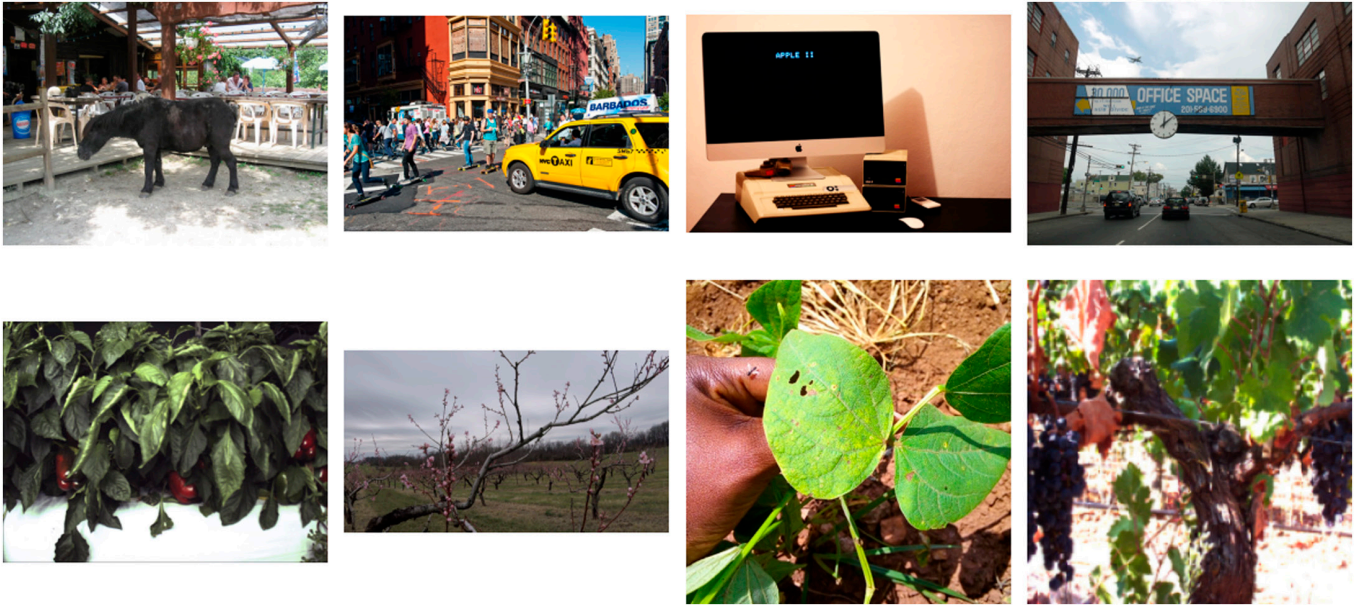


Fig. 1. Sample of images from the COCO dataset (row 1) and from assorted agricultural datasets (row 2), showcasing the contrast between the general environment of COCO imagery and the specific environment of agricultural imagery. These images are displayed in their original aspect ratio.

assessing the viability of transfer learning in other more relevant agricultural tasks, like semantic segmentation and object detection. Furthermore, there does not exist a single centralized repository for agriculture-specific datasets, preventing a large-scale ImageNet-style dataset for agriculture from being developed. While work done by [17,18] have proposed new larger-scale datasets for agriculture, increasing progress toward such a goal, such datasets are usually specific to certain conditions and designed for image classification, leaving other tasks without as much data. As a result, the potential for transfer learning as a viable improvement for data efficiency, at least in agriculture, is diminished.

Another common approach in deep learning that attempts to offset a deficiency of data is the use of data augmentations. For computer vision tasks, this refers to the transformation of an input image along with its corresponding annotations, with the stated intent of increasing diversity of data. Standard augmentations include rotation or flipping of an input image, or adjustment of visual parameters like saturation or brightness. The usefulness of augmentations has been studied in a general case in works such as [19]; however, these applications have largely been focused either on general image classification tasks or for a specific domain that is not similar with agriculture. Agricultural environments are largely different than general environments and are often more complex, involving less quantities of obvious features and more irregular and inconsistent shapes [20]. Certain studies of augmentations have been conducted in works such as [21], where custom color-spaces and vegetation indexes were used when processing input data in order to boost performance on agricultural datasets, but there has not been done any large-scale analysis of augmentation effectiveness for agricultural environments. In fact, a potential benefit of the domain constraints of agricultural data is the prospect for assessing features specific to these environments. This can be used to both develop data processing pipelines specific to agriculture, using such approaches, and adapt existing model architectures to agricultural tasks

using methods like transfer learning. Standing in the way of this development is a lack of a large-scale centralized database for agricultural data.

Bringing together datasets from previous studies, we present a novel set of centralized and standardized public agricultural datasets and benchmarks using state-of-the-art deep learning models. This collection of datasets is composed of real data collected from agricultural datasets across 3 different tasks: image classification, semantic segmentation, and object detection. For each task, we develop a standard data format and training pipeline, which we use to generate benchmarks and pretrained models for each dataset. These pipelines and models enable us to conduct a set of experiments in which we adapt existing methods for improving data efficiency, such as transfer learning and image augmentation, as described above, for agricultural environments. We first conduct an assessment of agricultural models for transfer learning by (a) assessing data efficiency and performance when using agricultural pretrained weights for fruit detection models. We apply this study further by (b) assessing data efficiency and performance when using pretrained weights for backbones of agricultural semantic segmentation models for fruit and plant segmentation. These 2 experiments provide us with a quantifiable way to improve data efficiency for agricultural models. Following our analysis of methods to improve models, we continue our study by evaluating methods to improve data, by (c) assessing the effectiveness of standard spatial and visual image augmentations for improving model performance on fruit detection and fruit segmentation tasks. Finally, we note that while our methods suggest methods to improve data, there exists a pool of low-quality data that may potentially be disregarded, and we (d) assess the effect of annotation quality—by observing the performance of models when reducing the quantity of bounding boxes (thus reducing the quality of these annotations)—on fruit detection tasks. Through these experiments, we find a strong potential for improving data efficiency for agricultural deep learning models without considerable modifications to

existing pipelines, with potential implications for a broad range of tasks. Furthermore, we open-source our standardized datasets and pretrained models through our framework AgML (<https://github.com/Project-AgML/AgML>) to facilitate the adoption of our described methods.

Methods and Experiments

To provide a baseline for our experiments, we have collected a set of agricultural datasets and developed a novel set of standard benchmarks and pretrained models on these datasets. These datasets have been standardized by task and are included in the dataset catalog in the open-source agricultural machine learning framework called AgML. We discuss our methodology in developing a standard task-based data processing pipeline for training agricultural machine learning models, and our models' results on these datasets. These pipelines are designed to be adaptable for the experiments conducted later in this research, and serve as our primary tool for them.

Collecting datasets

For the purposes of this work, we restrict datasets to 3 tasks: image classification, semantic segmentation, and object detection. Furthermore, we focus on datasets with only a single 3-channel RGB image input in order to facilitate the creation of benchmarks for a variety of datasets using standard deep learning models without any architectural modifications, allowing easier accessibility for a broader range of future applications. We compile datasets from a number of sources, with the majority coming from [22]. We provide a full listing of each of the datasets used in this dataset as well as key details and references in Table 1, which also contains benchmark performance on metrics as detailed later in the "Models and methods" section.

Data in AgML's collection of public data sources were sourced from each of the 6 human-inhabited continents, including the United States, Brazil, France, Germany, Greece, Uganda, China, and Australia. Furthermore, these data cover a wide variety of conditions, including lightning, camera angle, and amount of obfuscation. We illustrate this diversity in 2 examples. Figure 2 contains an image from each of the object detection datasets used in our research. Each of the datasets contains a varying set of conditions, ranging from environmental differences, such as day versus night, to camera parameters—including ground-level versus drone captures—to the type of fruit (the images shown include apple, grape, mango, and cantaloupe). Similarly, Fig. 3 displays examples of imagery from 3 semantic segmentation datasets used in our research. We note the variation of agricultural tasks within these datasets, with one dedicated to solely weed segmentation, one being a combination of plant and weed segmentation, and another being flower segmentation. Furthermore, the difference in camera angle and environmental conditions is also present in these datasets.

Models and methods

We select our model architectures with 2 primary focuses: state-of-the-art performance and ease of access using existing deep learning frameworks. This approach allows us to construct standard pretrained models with both high-scoring benchmarks as well as ease of reproducibility and application to future tasks. We assess state-of-the-art performance using traditional

performance benchmarks, namely, ImageNet [12] for image classification, CityScapes [23] for semantic segmentation, and COCO test-dev [13] for object detection. Similarly, we define the best-case scenario for ease of access to be the scenario where a model is contained entirely within the PyTorch [24] library family, to reduce external dependencies. We now elaborate on our training procedure for each category of deep learning tasks.

Image classification

When selecting an image classification model, we only evaluate those available in the torchvision library in order to facilitate our goal of ease-of-use and reduction of external dependencies. The EfficientNet family of models has the highest performance out of all models in the library. So, we select EfficientNetB4 as our model for image classification, considering its high ImageNet top-1 accuracy of 82.9% with a similar number of parameters to traditional state-of-the-art models like ResNet50. We select categorical cross-entropy loss as our criterion, defined as

$$L_{CE} = \sum_{i=1}^n y_i \cdot \log \hat{y}_i \quad (1)$$

where n is the number of classes, y_i is the ground truth, and \hat{y} is the prediction. We evaluate model performance using categorical accuracy.

Semantic segmentation

We again evaluate models within torchvision library and select the DeepLabV3 model with a ResNet50 backbone, as described in the original implementation (and as is available in the library). We select a loss function based on the number of classes in the dataset being trained upon. For binary segmentation tasks, involving only a single class, we use binary cross-entropy loss with logits, defined as

$$L_{BCE} = -(y \log \hat{y} + (1 - y) \log (1 - \hat{y})) \quad (2)$$

over each pixel, where y represents the ground truth value and \hat{y} represents the predicted value. For multi-class segmentation tasks, we use dice loss [25]:

$$L_D = 1 - \frac{2y\hat{y} + 1}{y + \hat{y} + 1} \quad (3)$$

where y again represents the ground truth and \hat{y} represents the predicted value. We evaluate model performance using the mean intersection-over-union (mIoU) metric:

$$\text{mIoU} = \frac{1}{n+1} \sum_{i=0}^n \frac{p_{ii}}{\sum_{j=0}^n p_{ij} + \sum_{j=0}^n (p_{ji} - p_{ii})} \quad (4)$$

where n is the number of classes and p_{ij} is the number of pixels of class i predicted to belong to class j .

Object detection

Here, we expand our reach beyond the torchvision library due to the reduced amount of object detection models available in it. We select the EfficientDet family of models—keeping in line with our prior selection of EfficientNet for image

Table 1. A listing of publicly available agricultural datasets used as part of this research (as described in Dataset Name), along with their number of images (Images), agricultural task (Ag. Task), and corresponding benchmark performance for the models described above on a testing set (Benchmark).

Dataset Name	Images	Classes	Ag. Task	Benchmark
Image classification				Accuracy
bean_disease_uganda	1,295	3	<i>B</i>	96.90%
plant_seedlings_aarhus [39]	5,539	12	<i>A</i>	94.39%
soybean_weed_uav_brazil [40]	15,336	4	<i>A</i>	100.0%
sugarcane_damage_usa [41]	153	6	<i>B</i>	100.0%
crop_weeds_greece [42]	508	4	<i>A</i>	100.0%
rangeland_weeds_australia [43]	17,509	10	<i>A</i>	97.94%
leaf_counting_denmark [44]	9,372	9	<i>C</i>	88.90%
plant_village_classification [45]	55,448	39	<i>B</i>	98.91%
plant_doc_classification [46]	2,336	28	<i>B</i>	89.27%
Semantic segmentation				mIoU
carrot_weeds_germany [47]	60	2	<i>D</i>	52.18%
sugarbeet weed segmentation [48]	125	2	<i>D</i>	53.59%
apple_flower_segmentation [49]	148	1	<i>E</i>	68.38%
apple_segmentation_minnesota [50]	670	1	<i>F</i>	79.08%
rice_seedling_segmentation [51]	224	2	<i>D</i>	52.20%
peachpear_flower_segmentation [49]	42	1	<i>E</i>	72.58%
red_grapes_and_leaves segmentation [52]	258	2	<i>F</i>	49.18%
white_grapes_and_leaves segmentation [52]	273	2	<i>F</i>	51.93%
Object detection				mAP@0.5
fruit_detection_worldwide [1]	565	7	<i>G</i>	70.35%
apple_detection_usa [53]	2,290	1	<i>G</i>	94.16%
apple_detection_spain [54]	967	1	<i>G</i>	86.65%
apple_detection_drone_brazil [55]	689	1	<i>G</i>	79.62%
mango_detection_australia [56]	1,242	1	<i>G</i>	95.32%
grape_detection_californiaday [57]	126	1	<i>G</i>	69.01%
grape_detection_californianight [57]	150	1	<i>G</i>	63.99%
Task legend	^{A,B} {Weed, Disease/Damage} Classification, ^C Leaf Counting, ^{D,E,F} {Weed, Flower, Fruit} Segmentation, ^G Fruit Detection			

classification—choosing the EfficientDetD4 model, due to its high performance of 49.3% mAP on the COCO evaluation dataset, with a comparable number of parameters to traditional state-of-the-art models like YOLOv3. We use the open source effdet library (see <https://github.com/rwightman/efficientdet-pytorch>) as our implementation of the EfficientDetD4 model. The loss

function used is focal loss [26] with $\alpha = 0.25$ and $\gamma = 1.5$, as described in the original implementation, and given by:

$$L_F = - (1 - p_t)^\gamma \log(p_t) \quad (5)$$

where $p_t = p$ if $y = 1$; otherwise, $p_t = 1 - p$ (and p is the estimated probability for the class y). We evaluate model performance using

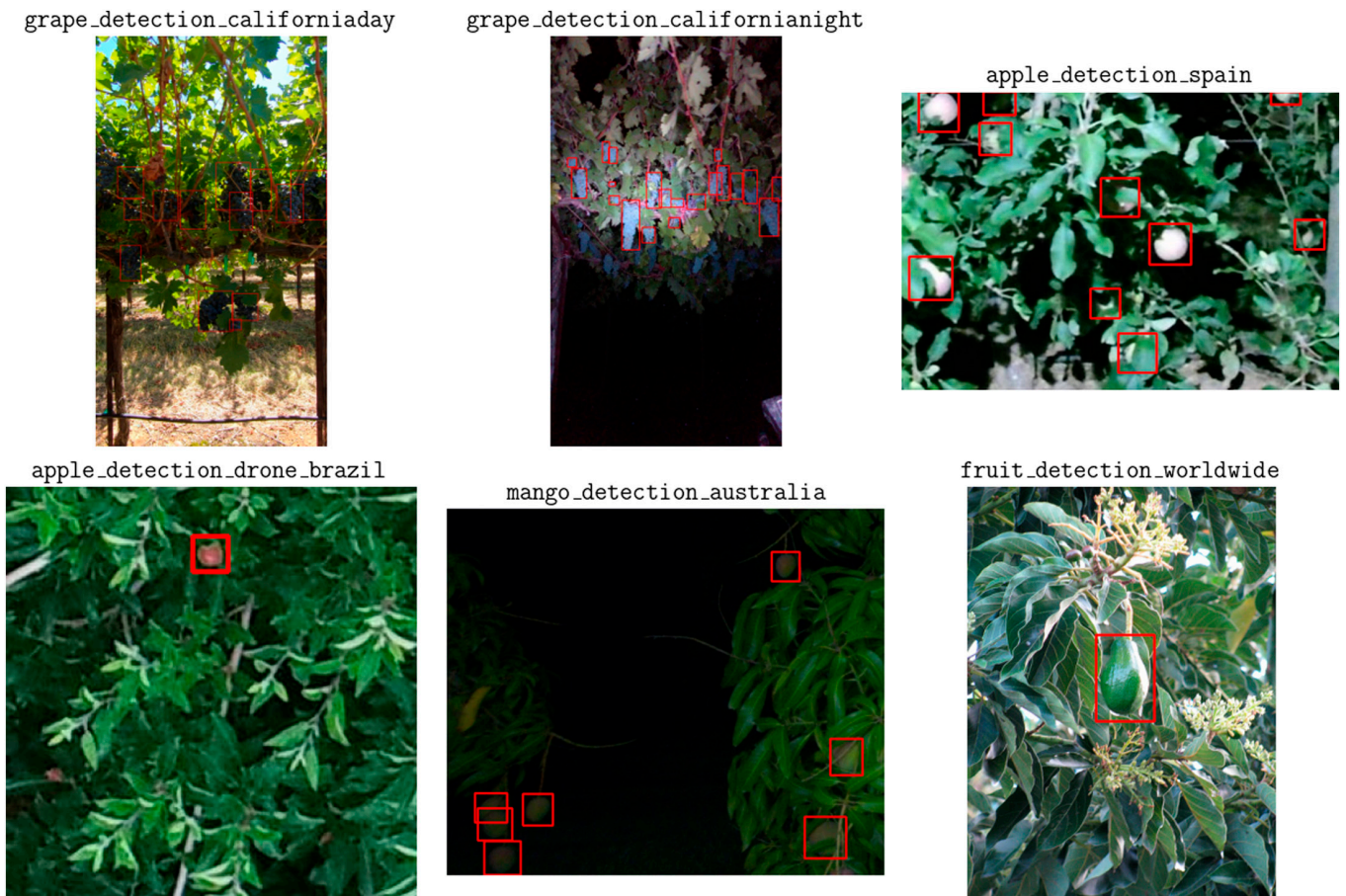


Fig. 2. Sample images from a collection of the object detection datasets used in this research, containing the original image with annotated fruit bounding boxes. The dataset from which each image is sourced from is annotated on top of it. These images are displayed in their original aspect ratio.

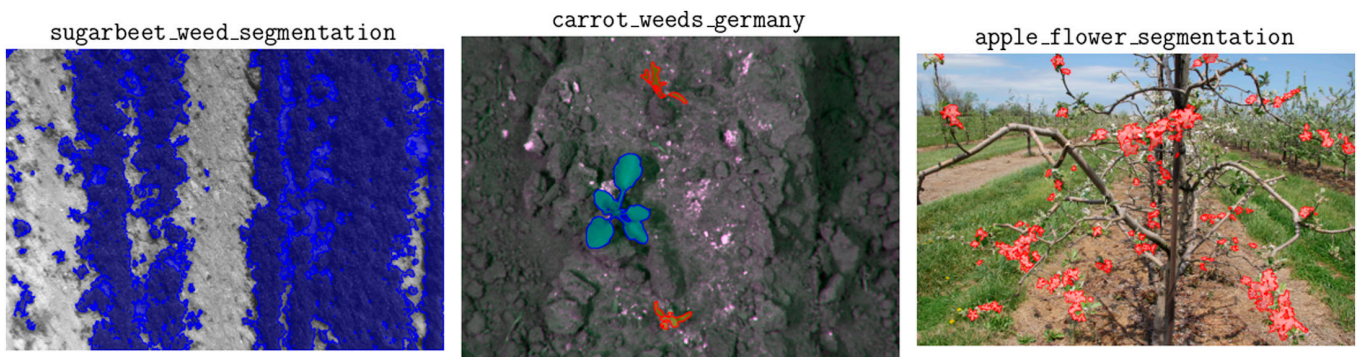


Fig. 3. Sample images from a collection of the semantic segmentation datasets used in this research, containing the original image with annotated segmentation masks. The dataset from which each image is sourced from is annotated on top of it. These images are displayed in their original aspect ratio.

the mean average precision metric at an IoU threshold of 0.5 (which we denote as mAP@0.5).

Model training parameters

For each of the 3 deep learning tasks, we develop training pipelines that are adaptable to each of the different agricultural datasets. These training pipelines are used both for our benchmark and pretrained model development and for our experiments, proving their versatility in adapting to a number of

different agricultural deep learning tasks. We summarize the main training parameters for each task in Table 2. Data pre-processing is minimal; our only step involves resizing images to the image size in Table 2 and model-specific preprocessing, namely, normalization for image classification and semantic segmentation and bounding box format conversion for object detection. In addition, for semantic segmentation and object detection, we decay the initial learning rate by $\gamma = 0.75$ every 8 epochs. All models are trained using the PyTorch Lightning library [27] on an NVIDIA Titan RTX GPU.

Agricultural model weights for object detection

Object detection models are traditionally trained using domain-generic weights: either random initialization or a common benchmark like the COCO dataset. For highly specified environments like agriculture, however, datasets are often of a much smaller magnitude. In many cases, as explored in [22], datasets may only consist of a few hundred images, common in the case of object detection where data annotation takes a substantial amount of time. In turn, using domain-specific agricultural weights as a starting point for deep learning models can significantly boost data efficiency, leading to models attaining high performance with less data and less training time.

To assess the performance of agricultural pretrained weights, we fine-tune an EfficientDetD4 model on the fruit_detection_worldwide dataset and evaluate a set of 5 distinct pretrained model approaches: *COCO*, *GRAPE*, *GRAPENIGHT*, *APPLEDRONE*, and *NONE*. The parameters of these different models are summarized in Table 3. *NONE* and *COCO* serve as our baselines for traditional pretrained weight approaches, while *GRAPE*, *GRAPENIGHT*, and *APPLEDRONE* represent agricultural pretrained models trained for different types of environments—*GRAPE* for daytime, *GRAPENIGHT* for nighttime, and *APPLEDRONE* for drone as opposed to ground imagery.

We then use these pretrained models for 2 experiments involving the fruit_detection_worldwide dataset. We first fine-tune a model on each of its 7 classes (specifically, 7 one-class models) and then one on the entire 7-class dataset (one 7-class

model)—enabling an assessment of the results of pretrained weights on not only fruit localization but also classification. Our experimental method and pipeline is carried from the “Object detection” section, although we restrict our evaluation to the first 50 epochs to obtain a better picture of data efficiency.

Agricultural backbone weights for semantic segmentation

In contrast to the standard available pretrained weights for object detection, tasks like semantic segmentation often do not have even a domain-general set of pretrained weights available for training—especially due to the fact that a network with a certain number of classes cannot properly utilize the pretrained weights from one with a different number of classes, unlike in object detection. In such cases, pretrained weights are only used for the backbone, a feature extraction model that is used as the input stem into a semantic segmentation network. For instance, in our DeepLabV3 model, the backbone was ResNet50, a model otherwise used for image classification tasks. Traditionally, such classifiers are trained on the ImageNet dataset [12]. However, as we intend to demonstrate, using agricultural domain-specific weights for even the backbone can performance gain for the larger network as a whole.

Image classification datasets are largely different from semantic segmentation datasets—while most of our semantic segmentation datasets in Table 1 are for locating fruits or weeds, many of our image classification datasets are for distinguishing between leaves and weeds. For a thorough examination of the viability of agricultural weights, we select 3 different agricultural image classification datasets to use as pretrained models, summarized in Table 4. *VILLAGE* is our largest image classification dataset, with over 55,000 images for species and disease classification, while *COUNTING* is a much smaller dataset focused on a largely different application—leaf counting. We assess these 4 pretrained backbones on 3 different datasets: apple_flower_segmentation, apple_segmentation_minnesota, and rice_seedling_segmentation. Our experimental method and pipeline is the same as in the “Semantic segmentation” section, although we restrict our evaluation to only the first 20 epochs of training.

Effectiveness of augmentations for generating diverse data

A standard procedure for improving the performance and, in particular, generalizability of deep learning models is to apply

Table 2. Summary of core training parameters for each task.

Task	Image size	Optimizer	Learning rate	Epochs
Image classification	224	Adam	0.001	100
Semantic segmentation	512	NAdam [12]	0.005	50
Object detection	512	AdamW [31]	0.0002	100

Table 3. Summary of pretrained model approaches for single-class detection.

Model name	Datasets pretrained on
<i>NONE</i>	None
<i>COCO</i>	COCO [13]
<i>GRAPE</i>	grape_detection_californiaday
<i>GRAPENIGHT</i>	grape_california_night
<i>APPLEDRONE</i>	apple_detection_drone_brazil

Table 4. Summary of pretrained backbones for semantic segmentation. Note that in contrast to object detection, the number of classes of the image classification models is not relevant, as they are being used as feature extractors.

Model name	Datasets pretrained on
<i>NONE</i>	None
<i>IMAGENET</i>	ImageNet [12]
<i>VILLAGE</i>	plant_village_classification
<i>COUNTING</i>	leaf_counting_denmark

visual and spatial augmentations to the input data, with the intended goal of increasing the diversity of the input data. Standard spatial augmentations include rotation, distortion (such as affine transforms), and cropping, while standard visual augmentations range from saturation and brightness contrast to Gaussian noise addition. For agricultural deep learning models, generalizability is a crucial goal, as environmental conditions vary considerably in different scenarios.

In turn, we conduct a robust analysis of various augmentations across a number of datasets in order to collect insight into the most viable augmentations for improving generalizability for agriculture. Performance is assessed on a collection of different augmentations on a set of different datasets across our 3 different tasks—image classification, semantic segmentation, and object detection—to provide a greater insight into each augmentation’s effectiveness. Our analysis only involves the usage of a single augmentation per trial, without combinations, in order to provide an insight into the capability of each augmentation independently, relevant to our goal of assessing augmentations in an agricultural context. Additionally, we collect final metric scores early in training, after only 20 epochs. As noted from our prior observations, most models begin approaching the relative peak of their performance at this threshold. So, a better analysis of the effectiveness of augmentations is done by analyzing their performance improvements earlier in training. We use the albumentations [28] library for our augmentations, noting its seamless integration into our existing PyTorch training pipeline. The augmentations we use, alongside model performance corresponding to the datasets we select, are summarized

in Table 5 to eliminate redundancy. Furthermore, we provide a visual example of each of the augmentations we use in Fig. 4.

Effects of annotation quality on learning

Data annotation is a heavily time-consuming task within deep learning. In the field of agriculture, this is an especially prevalent issue, due to the present nuance and even simply the quantity of objects such as fruits that need to be annotated in fruit detection tasks. Oftentimes, the amount of effort needed to annotate a certain dataset can result in that dataset being generated with imperfect and even obviously inaccurate annotations, potentially leading to them being unusable for deep learning tasks. However, with the already present scarcity of agricultural data, any loss of existing data has larger impacts on the open-source data community. Thus, we now assess how the quality of annotations on agricultural data affects models’ overall performance. In particular, we aim to observe whether models are capable of achieving high performance without perfect annotations, potentially leading to less strict annotation requirements for deep learning models in the future.

We use 3 fruit detection datasets for this stage, namely, `grape_detection_californiaday`, `apple_detection_drone_brazil`, and `fruit_detection_worldwide`. For each dataset, we set aside a common test set. For the remaining data, we create a set of datasets with reduced quality annotations, removing a certain percentage of bounding boxes for each image. This procedure is conducted for retention of only 30% of annotations, to 90% of annotations, with 10% increments. Using our object detection pipeline from the “Object detection” section, we train for

Table 5. Summary of performance of different augmentations on the aforementioned semantic segmentation and object detection datasets. Bold indicates the highest performance for a certain dataset.

Augmentation	A	B	C	D	E	F
	mAP@0.5			mIOU		
original (0)	64.08%	53.62%	75.11%	69.78%	78.52%	51.04%
horizontal-flip (1)	62.76%	58.32%	74.46%	66.97%	79.19%	51.35%
vertical-flip (2)	68.45%	56.12%	74.48%	62.48%	75.35%	49.19%
shear (3)	63.31%	58.28%	80.96%	65.61%	77.54%	51.04%
rotate (4)	60.77%	61.06%	84.42%	64.51%	73.96%	49.75%
translate (5)	65.53%	53.05%	86.60%	66.26%	79.22%	48.57%
brightness (6)	59.01%	52.69%	78.80%	66.45%	77.72%	44.50%
hsv-shift (7)	56.51%	49.32%	82.39%	51.34%	73.12%	42.95%
gaussian-blur (8)	65.17%	57.09%	84.60%	69.25%	78.61%	50.96%
rain (9)	60.44%	42.25%	81.20%	69.49%	73.09%	50.11%
fog (10)	65.35%	48.66%	78.81%	62.73%	70.81%	44.31%
sun-flare (11)	65.92%	55.87%	80.72%	49.86%	74.70%	43.63%
Dataset legend	^A grape_detection_californiaday ^B grape_detection_californianight ^C apple_detection_drone_brazil ^D apple_segmentation_minnesota ^E apple_flower_segmentation ^F rice_seedling_segmentation					

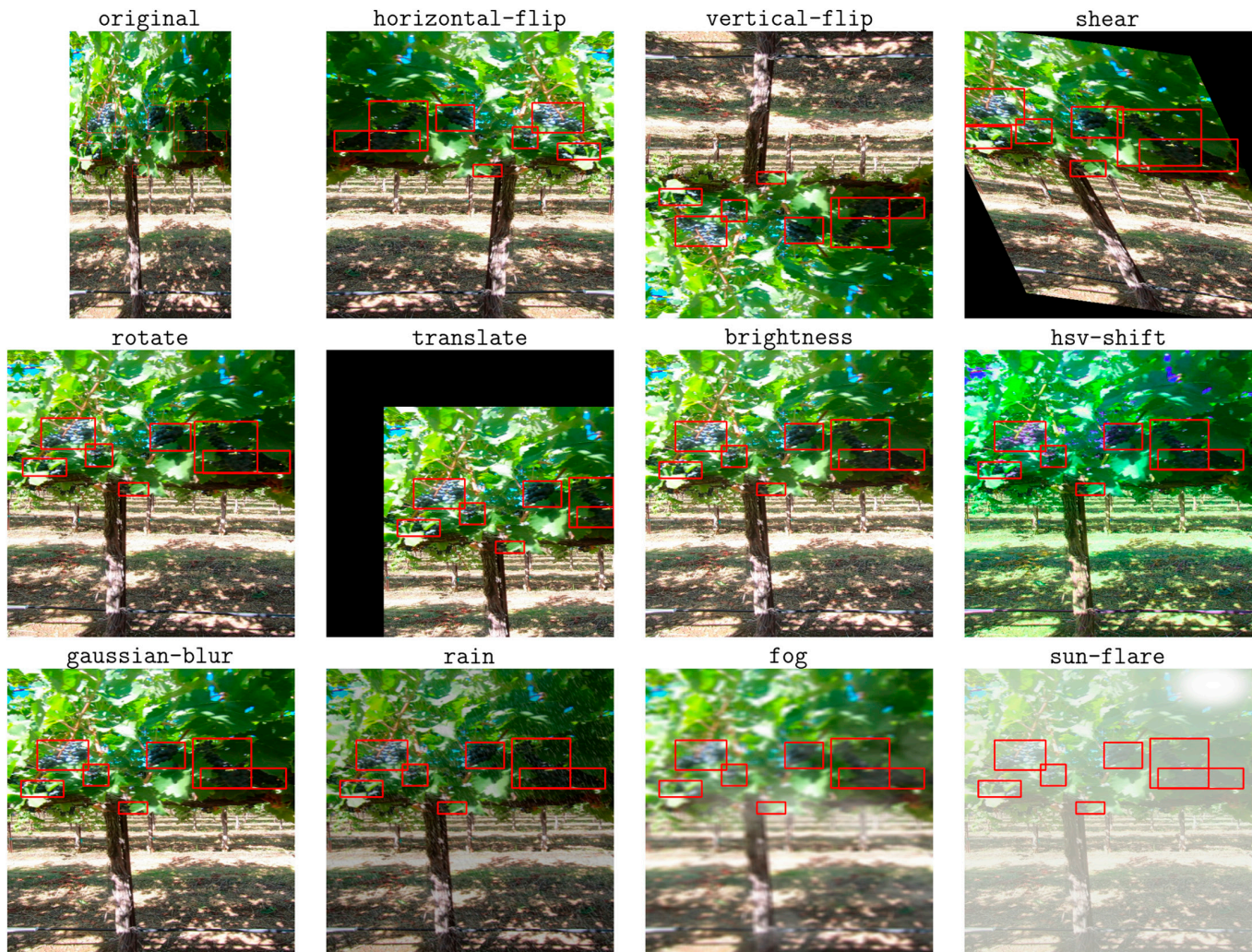


Fig. 4. An example image from `grape_detection_californiaday` with each of the augmentations used in this research applied to it. The original image is not resized, while the augmented images are (as they would be in an actual training pipeline).

25 epochs on each dataset and record their $mAP@0.5$ on the aforementioned test sets.

Results and Discussion

Standard benchmarks for agricultural datasets

The performance of each of our task-based models on the datasets we have collected is summarized in Table 1. Our standard pipelines enable our models to achieve comparable performance with existing benchmarks on our collected datasets, in certain cases even exceeding them. As mentioned in the “Models and methods” section, these datasets, along with pretrained models that achieved these benchmarks, are available through the open-source framework AgML, enabling further research on developing even more efficient data and model pipelines.

Performance of agricultural pretrained weights for object detection

A summary of the results of our experiments using agricultural pretrained weights for object detection is shown in Fig. 5, with

each of the fruits representing the 7 one-class models, and complete representing the one 7-class model. For all of the 7-class models, the pretrained agricultural models substantially outperform the *COCO* and *NONE* baselines. We observe that the agricultural models tended to plateau at their maximum mean average precision before 10 epochs, for most fruits—some, like strawberry, avocado, and mango, see this plateau as early as 5 epochs. On the other hand, *COCO*, the standard baseline, usually takes between 25 to 40 epochs to reach its own maximum value, while the *NONE* model fails to even break zero mean average precision for 5 out of 7 fruits, and only reaches a comparable score to *COCO* for one. For the 7-class model, we see that the agricultural pretrained models follow a similar trajectory to *COCO*, taking around or over 30 epochs to plateau at their maximum mean average precision.

A number of observations can be drawn from these results. Pretrained weights are clearly a better approach when training object detection models, as they converge markedly faster than random weights, saving time and resources in training. The poor performance of the *NONE* model demonstrates the value of transferring prior knowledge when training on new tasks. Furthermore, for agricultural domain-specific detection tasks,

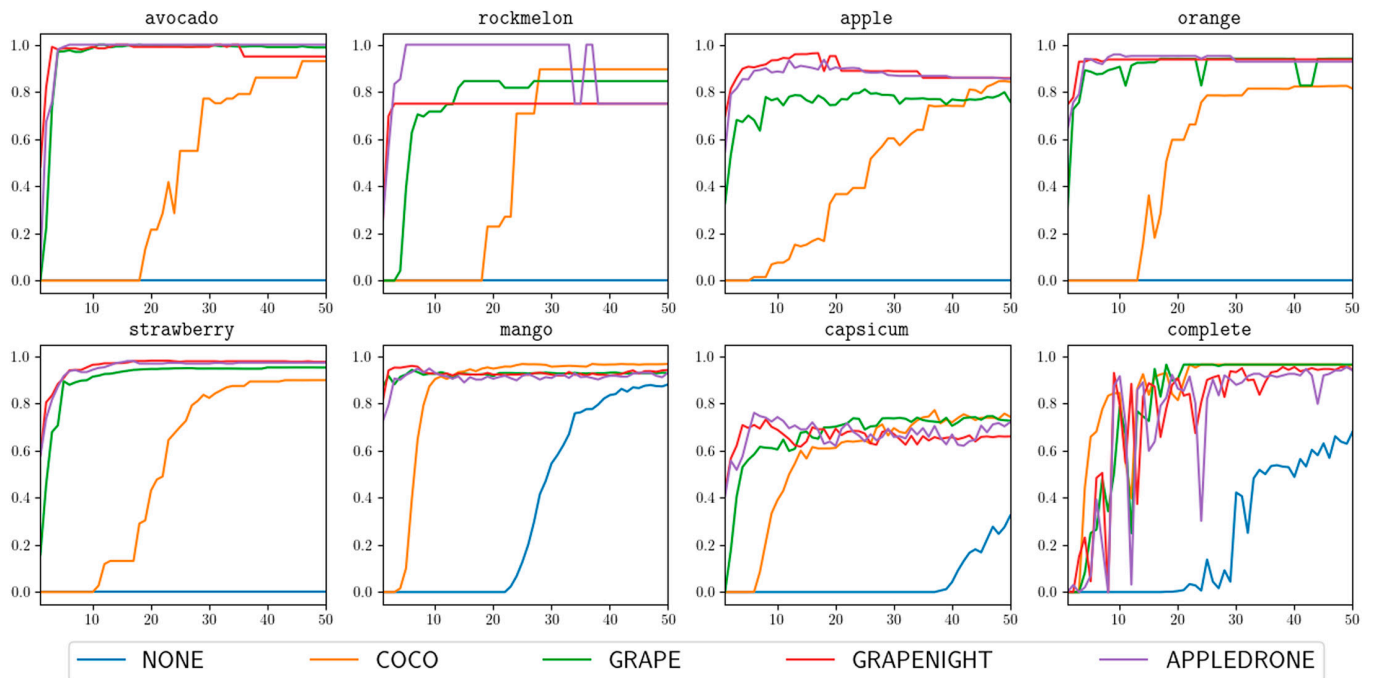


Fig. 5. Comparison of mAP@0.5 values on 7 different individual fruits from the fruit detection worldwide dataset, alongside the complete dataset (listed as complete, the bottom right plot).

using agricultural pretrained weights can provide, at minimum, comparable performance to *COCO*, or other standard weights, but at best, even better performance and much faster convergence. Our results are consistent with previous observations for other tasks, such as [29], where agricultural pretrained weights were found to increase performance. Our results are further consistent with previous observations specifically for object detection such as in [30], where using pretrained fruit detection weights from similar environments resulted in a considerable performance increase. A key insight from our work, however, is the fact that agricultural environment played little role in performance improvement—all 3 models, despite their distinct environments, achieved similar performance on a dataset in yet another noticeably different environment. Furthermore, each agricultural pretrained model followed a similar performance trajectory, corroborating our observations that using agricultural pretrained weights of any form results in not only higher performance but also faster convergence. In turn, this approach, which only requires a slight modification to existing pipelines, can notably improve training results for future agricultural deep learning work.

Performance of agricultural backbone weights for semantic segmentation

Similar to the previous section, we summarize the results of our semantic segmentation models in Fig. 6. We find that our models with agricultural pretrained backbones, *VILLAGE* and *COUNTING*, tend to outperform models with the existing *NONE* and *IMAGENET* backbones to a measureable degree. For the datasets *apple_segmentation_minnesota* and *rice_seedling_segmentation*, our agricultural pretrained models reach their maximum mIoU almost within the first couple of epochs, while the other 2 models either take between 10 and

15 epochs to reach a similar degree of performance or do not reach it at all. For *apple_flower_segmentation*, our models reach similar levels of performance as *COCO*, but maintain a more consistent trajectory of performance increase, while *COCO* and *NONE* see a spike in performance followed by a quick drop, suggesting relative inconsistency.

Semantic segmentation is a unique task due to the required level of data annotation required for it—each pixel in an image must be assigned to a certain class. This is reflected in the reduced size of segmentation datasets, an example of which can be seen in the image counts of Table 1. As a result, there is potentially not enough available data to generate a large-scale pretrained network for semantic segmentation, especially when noting the potential difference in classes between a pretrained network and an applied, task-specific network—resulting in loss of knowledge in weights that cannot be transferred. This is noted in works like [31,32], where standard domain-general weights are used in lieu of their agricultural equivalents. However, image classification datasets are in relative surplus in comparison to semantic segmentation, and most segmentation models use classification models as their feature extractors, representing the encoder module for a typical encoder–decoder architecture. For most tasks, a standard approach is to freeze weights for this feature extraction module, the encoder, and only update parameters for the segmentation head, the decoder. This is inspired by the fact that the feature extraction model obtains the most relevant information regarding the location of the objects being segmented—however, for agricultural networks, this approach seems largely counterintuitive, as the feature extractor is usually trained on domain-general images and this unable to recognize the most relevant features for plants. Using agricultural pretrained weights, on the other hand, can improve the localization performance of the feature extractor, in turn improving the network’s performance as a whole. Our results show that using

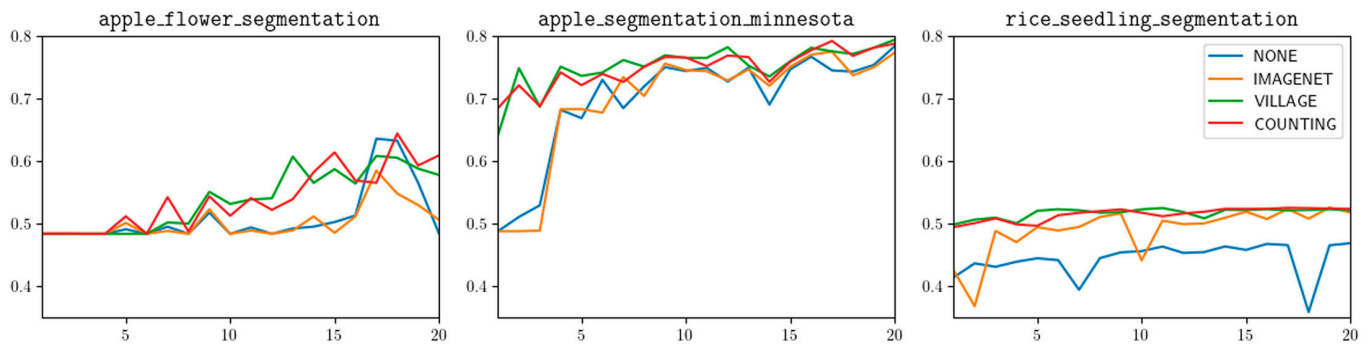


Fig. 6. Comparison of mIoU values for 4 different pretrained backbones on 3 different evaluation datasets. The y axis is on a consistent scale, although each dataset has its models converge to a different range of values.

agricultural weights for feature extractors serves as a potential workaround for the lack of a standard semantic segmentation pretrained model, still enabling improved performance and reduced training time.

Augmentation effectiveness

We summarize the performance of each of the augmentations used in this work, on 6 different datasets (*A, B, C* object detection, and *D, E, F* semantic segmentation), in Table 5. When using a single augmentation, as in our experiments, observe that spatial augmentations—namely, augmentations (1) to (5) as denoted in the table, when used independently—generally outperform visual augmentations—(6) to (11)—for all of the datasets used in this experiment. The augmentations that provide the largest performance increase are all spatial, while the visual augmentations, in many cases, fail to even provide any performance increase whatsoever. For our subsequent analysis, we refer to each dataset using its key in the table (*A* refers to grape_detection_californiaday, and so on).

Our next key observation regards the effect of different augmentations on different environments and conditions. For instance, datasets *A* and *B* consist of the same fruit in the same environment, simply in day conditions versus night conditions. While certain visual augmentations, such as rain and fog, provide a performance boost for *A*, they actually considerably reduce performance for *B*, indicating that certain factors, such as rain and fog, are less valuable information in nighttime environments as opposed to daytime environments. Another example of the effects of different augmentations, this time for spatial augmentations with reference to camera positioning, is found when observing the results for *C*. While *A* and *B*, the other object detection datasets, consist of ground cameras capturing images of plants, *C* consists of aerial imagery captured by a drone. In turn, augmentations like rotation and translation provide a better opportunity for the model to generalize to different aerial camera positions, as opposed to less distortive augmentations like horizontal and vertical flips. For semantic segmentation tasks, we observe a less obvious boost in performance from augmentations. In fact, for *D*, no augmentations provide the best result, while for *E* and *F* the performance boost is within 1% of no augmentations. This potentially stems from the different conditions of semantic segmentation tasks as opposed to object detection tasks—semantic segmentation tasks, especially for fruit or leaf segmentation, like *D* and *E*, can come down to individual pixels, and such precision may likely be distorted by spatial augmentations. In all tasks, while

visual augmentations appear to provide less of an obvious benefit as opposed to spatial augmentations, our results still provide some insight into their potential viability. In particular, augmentations that affect the hue of an image, such as HSV shift and brightness or contrast shifts, tend to actually decrease performance, as they result in a model adapting to those input colors. On the contrary, Gaussian blurring provides a consistent, though sometimes minute, performance increase relative to the input. This suggests that a reduction of features, as done by adding a slight blur to the input, can in fact increase performance by potentially having the model learn more general features in the input environment.

A major benefit of augmentations is their ability to generate data potentially modified for a large variety of conditions, enabling greater generalizability for agricultural models. Existing state-of-the-art work done for domain transfer often involves the usage of generative adversarial networks (GANs), such as [33], who developed a GAN for transferring sample imagery between day and night domains, and [34], who used a CycleGAN network to edit the fruits present in imagery while maintaining the environmental conditions. While augmentations may not necessarily be able to provide an entire domain transfer as in the prior methods, they can still provide a generalization of conditions—for instance, rain and fog augmentations, as used in our research, can generalize data to a broader range of common conditions across the world. Other augmentations, like sun flare, can reduce the impact of edge cases where images are obscured. We note that certain augmentations expand beyond a single data point, potentially involving multiple images being transformed together. This is the case in common augmentations such as mosaic and patch, as explored in their applications for agricultural robotics in [35]. While we do not include these augmentations in our research, as they involve larger modifications to our pipelines, they do present potential further performance improvements for agricultural models. If used properly based on the input environment, augmentations provide an efficient way to produce a larger set of environments for agricultural data, improving generalizability of models.

Annotation quality

The performance of each of our annotation quality models is recorded in Fig. 7, which displays the performance of models relative to the quality of annotations on the datasets they are trained on. For the datasets grape_detection_californiaday and apple_detection_drone_brazil, we find a consistent increase in performance as the quality of annotations increases. Notably,

we do not observe a one-to-one correspondence between quality of annotations and level of performance: e.g., 30% of annotations actually correspond to over 50% of peak performance, suggesting that models trained on lower quality annotations are, at least to some extent, able to localize similar objects to a slightly higher than expected degree. We provide a sample of predicted bounding boxes for each grape_detection_californiaday iteration in Fig. 8. This observance, however, does not hold for

fruit_detection_worldwide, which has a distinction of being a multi-class dataset in contrast to the prior 2. Surprisingly, for this dataset, we find that our models achieve maximum performance with only 50% of annotations. This example, in turn, provides an insight into the hazardous potential of using lower-quality annotations. At 50% of annotations, our model may have only learned general features, thus allowing it to still perform much better on other sample images. However, higher

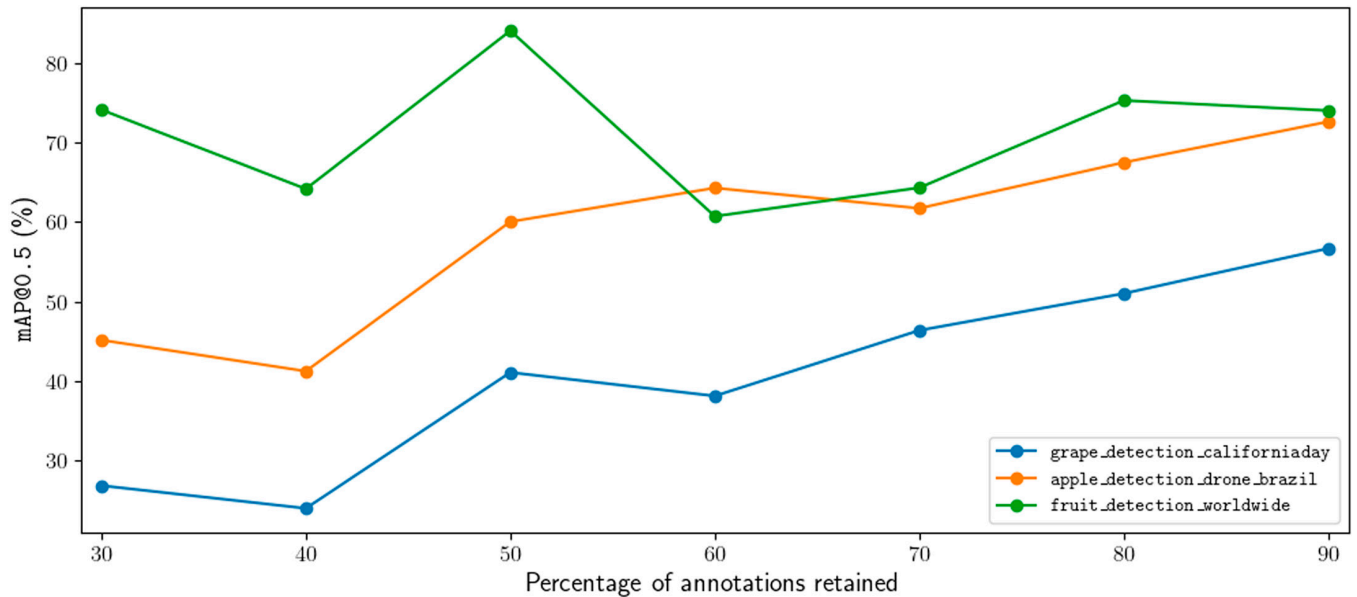


Fig. 7. Performance of models on 3 different datasets dependent on the percentage of bounding box annotations retained per image.

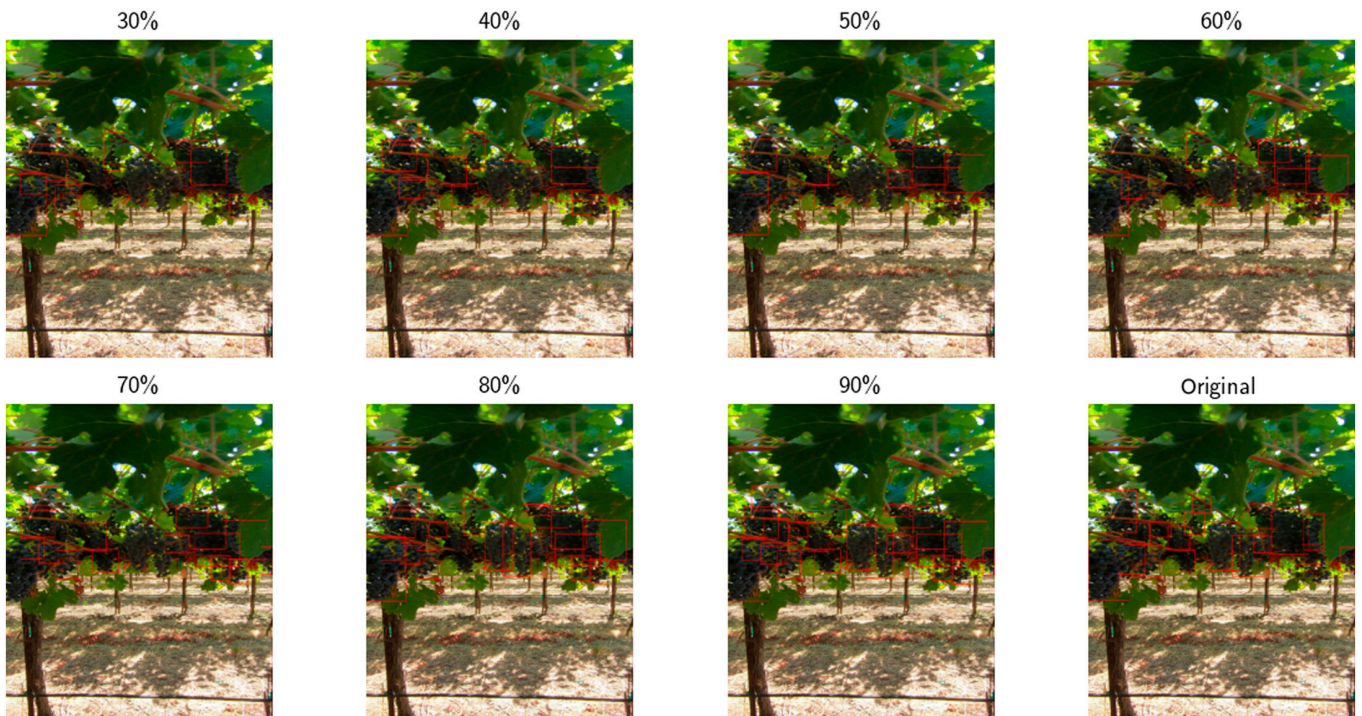


Fig. 8. Sample predictions by models trained on different levels of annotation quality. The percentages above each image reflect the amount of bounding boxes retained for the data used in training that specific model.

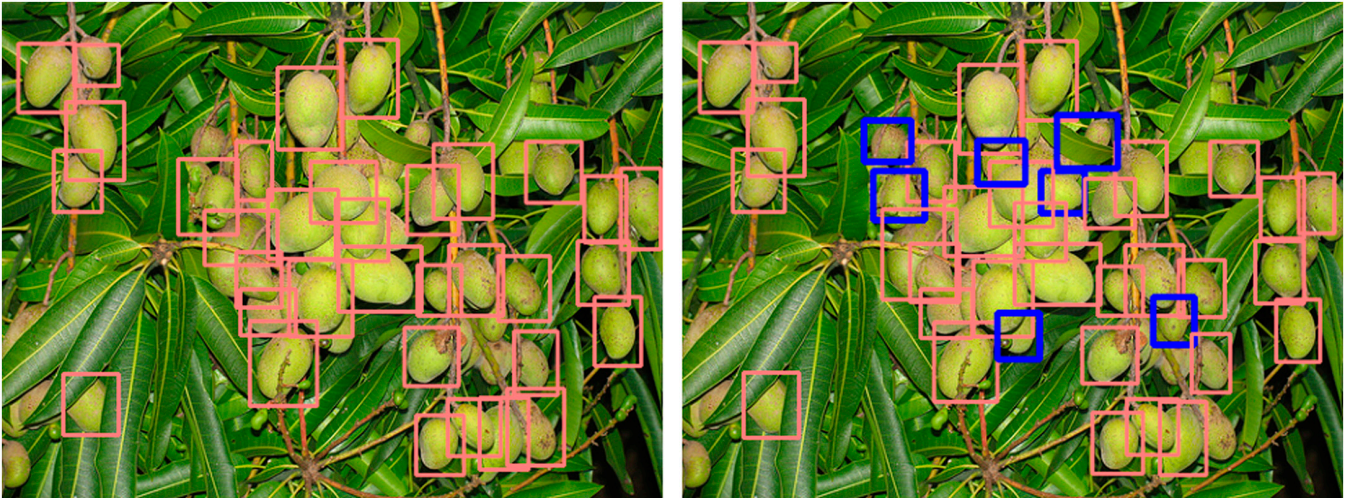


Fig. 9. Example ground truth and predicted annotations on the dataset `fruit_detection_worldwide` by its corresponding benchmark model. The images on the left contain ground truth annotations, while the images on the right contain predicted annotations, with orange representing correctly predicted annotations and blue representing accurately predicted fruits that are not included in the annotations.

percentages of still low-quality annotations may result in a model learning to distinguish between different instances of the same fruit by extremely nuanced features, overfitting on the samples on which it is trained. This serves as an interesting case study in and of itself—for instance, on sample images from the `fruit_detection_worldwide` dataset, the corresponding benchmark model is able to predict fruits that are not annotated, as displayed in Fig. 9.

A number of methods have been proposed in recent work to try and automate the data annotation process, which is one of the major bottlenecks for agricultural deep learning. Some approaches, such as [36], use robotic systems to capture data from a wider range of angles and automatically annotate bounding boxes knowing the location of plants in controlled environments. Other approaches [37,38] involve semi-supervised learning, taking advantage of highly precise data to train high-performance models that in turn improve predictions on unlabeled images. Such approaches show potential for not only making new high-quality agricultural datasets but also even potentially improving upon existing datasets, annotating missed fruits or other objects that may have been missed out on. Nevertheless, our work demonstrates that fruit detection models can still obtain relatively high performance with some fruits unannotated. In turn, potential new approaches may involve using lower-quality datasets in coordination with better datasets, expanding the pool of available data for agriculture and further boosting model performance.

Conclusion

In our work, we have developed a novel set of standardized and centralized agricultural datasets, alongside benchmarks and pretrained models using state-of-the-art models. Our custom pipelines achieve comparable performance with existing benchmarks using no extensive data or architectural modifications, making them widely applicable to a variety of agricultural deep learning tasks. We have also assessed a number of existing methods for improving model performance domain-specific to agriculture, including using agricultural pretrained model weights and image augmentations. Furthermore, we have even

explored the viability of traditionally overlooked lower-quality data, potentially expanding the data pool for agriculture. Our results demonstrate that slight training modifications can substantially boost model performance and result in shorter convergence time. We have open-sourced our standardized and centralized versions of the datasets used in our work, alongside our pretrained models and benchmarks, to guide easier adoption of our described methods.

Acknowledgments

Funding: This project was partly supported by the USDA AI Institute for Next Generation Food Systems (AIFS), USDA award number 2020-67021-32855. **Author Contributions:** Amogh Joshi. Conceptualization, Methodology, Software, Writing- Original Draft. Dario Guevara Conceptualization, Methodology, Software, Writing - Review & Editing. J. Mason Earles Conceptualization, Methodology, Resources, Funding Acquisition, Writing - Review & Editing. **Competing interests:** The authors declare that they have no competing interests.

Data Availability

All of the datasets that have been used in this paper can be found at AgML (<https://github.com/Project-AgML/AgML>), using the `agml.data` module and catalog. Datasets are stored in standard formats for image classification, semantic segmentation, and object detection, and can be either downloaded raw or loaded into an `AgMLDataLoader`, provided in the library. Benchmark weights for each of the models can also be found in AgML, using the `agml.models` module.

References

1. Sa I, Ge Z, Dayoub F, Upcroft B, Perez T, McCool C. DeepFruits: A fruit detection system using deep neural networks. *Sensors*. 2016;16:1222.
2. Bargoti S, Underwood J. Deep fruit detection in orchards. Paper presented at: 2017 IEEE International Conference on

- Robotics and Automation (ICRA); 29 May 2017–3 June 2017; Singapore.
3. Jeon HY, Tian LF, Zhu H. Robust crop and weed segmentation under uncontrolled outdoor illumination. *Sensors (Basel)*. 2011;11:6270–6283.
 4. Milioto A, Lottes P, Stachniss C. Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs. Paper presented at: IEEE International Conference on Robotics and Automation (ICRA); 2018 May 21–25; Brisbane, QLD, Australia.
 5. Sharada PM, Hughes DP, Salathé M. Using deep learning for image-based plant disease detection. *Front Plant Sci*. 2016;7:1419.
 6. Ümit A, Uçar M, Akyol K, Uçar E. Plant leaf disease classification using EfficientNet deep learning model. *Ecol Inform*. 2021;61:101182.
 7. Elsherbiny O, Zhou L, Feng L, Qiu Z. Integration of visible and thermal imagery with an artificial neural network approach for robust forecasting of canopy water content in Rice. *Remote Sens*. 2021;13(9):1785.
 8. Wang C, Liu B, Liu L, Zhu Y, Hou J, Liu P, Li X P. A review of deep learning used in the hyperspectral image analysis for agriculture. *Artif Intell Rev*. 2021;54:5205–5253.
 9. Nevavuori P, Narra N, Linna P, Lipping T. Crop yield prediction using multitemporal UAV data and Spatio-temporal deep learning models. *Remote Sens*. 2020;12(23):2000.
 10. Li M, Zhang Z, Lei L, Wang X, Guo X. Agricultural greenhouses detection in high-resolution satellite images based on convolutional neural networks: Comparison of faster R-CNN, YOLO v3 and SSD. *Sensors (Basel)*. 2020;20:4938.
 11. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng*. 2010;22:1345–1359.
 12. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. Paper presented at: IEEE Conference on Computer Vision and Pattern Recognition; 2009 Jun 20–25; Miami, FL, USA.
 13. Lin T-Y, Maire M, Belongie SJ, Bourdev LD, Girshick RB, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL. Microsoft COCO: Common Objects in Context. 2014. CoRR abs/1405.0312.
 14. Nowakowski A, Mrziglod J, Spiller D, Bonifacio R, Ferrari I, Mathieu PP, Garcia-Herranz M, Kim D-H. Crop type mapping by using transfer learning. *Intl J Appl Earth Observ Geoinform*. 2021;98:102313.
 15. Moon T, Son JE. Knowledge transfer for adapting pre-trained deep neural models to predict different greenhouse environments based on a low quantity of data. *Comput Electron Agric*. 2021;185:106136.
 16. Shrivastava VK, Pradhan MK, Thakur MP. Application of pre-trained deep convolutional neural networks for Rice Plant disease classification. Paper presented at: International Conference on Artificial Intelligence and Smart Systems (ICAIS); 2021 Mar 25–27; Coimbatore, India.
 17. Zheng Y-Y, Kong J-L, Jin X-B, Wang X-Y, ST-L, Zuo M. CropDeep: The crop vision dataset for deep-learning-based classification and detection in precision agriculture. *Sensors (Basel)*. 2019;19(5):1058.
 18. Sahili ZA, Awad M. The power of transfer learning in agricultural applications: AgriNet. *Front Plant Sci*. 2022;13:992700.
 19. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *J Big Data*. 2019;6:60.
 20. Wang L, Wang J, Liu Z, Zhu J, Qin F. Evaluation of a deep-learning model for multispectral remote sensing of land use and crop classification. *The Crop J*. 2022;10(5):1435–1451.
 21. Moreira G, Magalhães SA, Pinho T, dos Santos FN, Cunha M. Benchmark of deep learning and a proposed HSV colour space models for the detection and classification of greenhouse tomato. *Agronomy*. 2022;12(2):356.
 22. Lu Y, Young S. A survey of public datasets for computer vision tasks in precision agriculture. *Comput Electron Agric*. 2020;178:105760.
 23. Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, Benenson R, Franke U, Roth S, Schiele B. The Cityscapes dataset for semantic urban scene understanding. 2016. CoRR abs/1604.01685.
 24. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, et al. PyTorch: An imperative style, high-performance deep learning library. Paper presented at: Proceedings of the 33rd International Conference on Neural Information Processing Systems; 2019 Dec 8; Red Hook, NY, USA.
 25. Sudre CH, Li W, Vercauteren T, Ourselin S, Jorge Cardoso M. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. Paper presented at: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. 2017 Sep. p. 240–248.
 26. Lin T-Y, Goyal P, Girshick RB, He K, Dollár P. Focal loss for dense object detection. 2017. CoRR abs/170802002.
 27. Falcon W. Pytorch lightning. GitHub. Note: <https://github.com/PyTorchLightning/pytorch-lightning> 3. 2019.
 28. Buslaev A, Iglavikov VI, Khvedchenya E, Parinov A, Druzhinin M, Kalinin AA. Alumentations: Fast and flexible image augmentations. *Information*. 2020;11(2):125.
 29. Too EC, Yujian L, Njuki S, Yingchun L. A comparative study of fine-tuning deep learning models for plant disease identification. *Comput Electron Agric*. 2019;161:272–279.
 30. Bresilla K, Perulli GD, Boini A, Morandi B, Grappadelli LC, Manfrini L. Single-shot convolution neural networks for real-time fruit detection within the tree. *Front Plant Sci*. 2019;10:611.
 31. Sodjinou SG, Mohammadi V, Mahama ATS, Gouton P. A deep semantic segmentation-based algorithm to segment crops and weeds in agronomic color images. *Inform Process Agric*. 2022;9(3):355–364.
 32. Peng Y, Wang A, Liu J, Faheem M. A comparative study of semantic segmentation models for identification of grape with different varieties. *Agriculture*. 2021;11(10):997.
 33. Fei Z, Olenskyj A, Bailey BN, Earles M. Enlisting 3D crop models and GANs for more data efficient and generalizable fruit detection. Paper presented at: IEEE/CVF International Conference on Computer Vision Workshops ICCVW; 2021. p. 1269–1277.
 34. Zhang W, Chen K, Wang J, Shi Y, Guo W. Easy domain adaptation method for filling the species gap in deep learning-based fruit detection. *Horticul Res*. 2021;8:1–13.
 35. Su D, Kong H, Qiao Y, Sukkarieh S. Data augmentation for deep learning based semantic segmentation and crop-weed classification in agricultural robotics. *Comput Electron Agric*. 2021;190:106418.
 36. Beck MA, Liu C-Y, Bidinosti CP, Henry CJ, Godee CM, Ajmani M. An embedded system for the automated

- generation of labeled plant images to enable machine learning applications in agriculture. *PLOS ONE*. 2020;15: e0243923.
37. Fatima T, Mahmood T. Semi-supervised learning in smart agriculture: A systematic literature review. Paper presented at: 2021 6th international multi-topic ICT conference (IMTIC); 2021 Nov 10–12; Jamshoro & Karachi, Pakistan.
 38. Ciarfuglia TA., Motoi IM, Saraceni L, Nardi D. Pseudo-label generation for agricultural robotics applications. Paper presented at: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops; 2022 Jun 19–20; New Orleans, LA, USA.
 39. Giselsson TM, Jørgensen RN, Jensen PK, Dyrmann M, Midtby HS. A public image database for benchmark of plant seedling classification algorithms. arXiv . 2017. <https://doi.org/10.48550/arXiv.1711.05458>
 40. Santos Ferreira, Alessandros, Daniel Matte Freitas, Gercina Goncalves da Silva, Hemerson Pistori, and Marcelo Theophilo Folhes. Weed detection in soybean crops using ConvNets. *Comput Electron Agric*. 2017;143:314–324.
 41. Alencastre-Miranda M, Davidson JR, Johnson RM, Waguespack H, Krebs HI. Robotics for sugarcane cultivation: Analysis of billet quality using computer vision. *IEEE Robot Autom Lett*. 2018;3(4):3828–3835.
 42. Espejo-Garcia B, Mylonas N, Athanasakos L, Fountas S, Vasilakoglou I. Towards weeds identification assistance through transfer learning. *Comput Electron Agric*. 2020;171:105306.
 43. Olsen A, Konovalov DA, Philippa B, Ridd P, Wood JC, Johns J, Banks W, Girgenti B, Kenny O, Whinney J, et al. DeepWeeds: A Multiclass Weed Species Image Dataset for Deep Learning. *Sci Rep*. 2019;9:2058.
 44. Teimouri N, Dyrmann M, Nielsen PR, Mathiassen SK, Somerville GJ, Jørgensen RN. Weed growth stage estimator using deep convolutional neural networks. *Sensors*. 2018;18(5):1580.
 45. Hughes DP, Salathè M. An open access repository of images on plant health to enable the development of mobile disease diagnostics through machine learning and crowdsourcing. 2015. CoRR abs/1511.08060.
 46. Singh D, Jain N, Jain P, Kayal P, Kumawat S, Batra N. PlantDoc: A Dataset for Visual Plant Disease Detection. Paper presented at: Proceedings of the 7th ACM IKDD CoDS and 25th COMAD: Association for Computing Machinery; 2020 Jan 15; New York, NY, USA.
 47. Haug S, Ostermann J. A crop/weed field image dataset for the evaluation of computer vision based precision agriculture tasks. In: Agapito L, Bronstein MM, Rother C, editors. *Computer Vision—ECCV 2014 Workshops. ECCV 2014. Lecture Notes in Computer Science*. Cham: Springer International Publishing; 2015. p. 105–116.
 48. Sa I, Chen Z, Popović M, Khanna R, Liebisch F, Nieto J, Siegwart R. weedNet: Dense semantic weed classification using multispectral images and MAV for smart farming. *IEEE Robot Autom Lett*. 2018;3(1):588–595.
 49. Dias PA, Tabb A, Medeiros H. Multispecies fruit flower detection using a refined semantic segmentation network. *IEEE Robot Autom Lett*. 2018;3:3003–3010.
 50. Häni N, Roy P, Isler V. MinneApple: A benchmark dataset for apple detection and segmentation. *IEEE Robot Autom Lett*. 2019;5(2):852–858.
 51. Khan A, Ilyas T, Umraiz M, Mannan ZI, Kim H. CED-net: Crops and weeds segmentation for smart farming using a small cascaded encoder-decoder architecture. *Electronics*. 2020;9(10):1602.
 52. Kalampokas T, Tziridis K, Nikolaou A, Vrochidou E, Papakostas GA, Pachidis T, Kaburlasos VG. Semantic segmentation of vineyard images using convolutional neural networks. In: Iliadis L, Angelov PP, Jayne C, Pimenidis E, editors. *Proceedings of the 21st EANN (Engineering Applications of Neural Networks) 2020 Conference*. Cham: Springer International Publishing; 2020. p. 292–303.
 53. Karkee M, Bhusal S, Zhang Q. Apple dataset benchmark from orchard environment in modern fruiting wall; 2019.
 54. Gené-Mola J, Vilaplana V, Rosell-Polo JR, Morros J-R, Ruiz-Hidalgo J, Gregorio E. KFuji RGB-DS database: Fuji apple multi-modal images for fruit detection with color, depth and range-corrected IR data. *Data Brief*. 2019;25:104289.
 55. Santos TT, Gebler L. A methodology for detection and localization of fruits in apples orchards from aerial images. 2021. CoRR abs/2110.12331.
 56. Koirala A, Walsh KB, Wang Z, McCarthy C. Deep learning for real-time fruit detection and orchard fruit load estimation: Benchmarking of ‘MangoYOLO’. *Precis Agric*. 2019;20:1107–1135.
 57. AI, Plant and Biophysics Lab, Grape Detection 2019 Day. 2019.