Data in Brief 55 (2024) 110679

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

OnionFoliageSET: Labeled dataset for small onion and foliage flower crop detection

Juan Felipe Restrepo-Arias[a,*], John W. Branch-Bedoya[b], Paulina Arregocés-Guerra[b]

[a] *Escuela de Ciencias Aplicadas e Ingeniería, Universidad EAFIT, 050022, Medellín, Colombia*
[b] *Facultad de Minas, Universidad Nacional de Colombia, 050041, Sede Medellín, Colombia*

ARTICLE INFO

Article history:
Received 29 April 2024
Revised 10 June 2024
Accepted 19 June 2024
Available online 26 June 2024

Dataset link: OnionFoliageSET (Original data)

Keywords:
Precision agriculture
Smart agriculture
Computer vision
Remote sensing
Green onion
Foliage flowers
Crop detection

ABSTRACT

Digital image datasets for Precision Agriculture (PA) still need to be available. Many problems in this field of science have been studied to find solutions, such as detecting weeds, counting fruits and trees, and detecting diseases and pests, among others. One of the main fields of research in PA is detecting different crop types with aerial images. Crop detection is vital in PA to establish crop inventories, planting areas, and crop yields and to have information available for food markets and public entities that provide technical help to small farmers. This work proposes public access to a digital image dataset for detecting green onion and foliage flower crops located in the rural area of Medellín City – Colombia. This dataset consists of 245 images with their respective labels: green onion (*Allium fistulosum*), foliage flowers (*Solidago Canadensis and Aster divaricatus*), and non-crop areas prepared for planting. A total of 4315 instances were obtained, which were divided into subsets for training, validation, and testing. The classes in the images were labeled with the polygon method, which allows training machine learning

* Corresponding author.
E-mail address: jfrestrep3@eafit.edu.co (J.F. Restrepo-Arias).
Social media: @frestrepoarias (J.F. Restrepo-Arias)

https://doi.org/10.1016/j.dib.2024.110679

algorithms for detection using bounding boxes or segmentation in the COCO format.

Specifications Table

Subject	Artificial Intelligence, Computer Vision and Pattern Recognition, Applied Machine Learning, Agronomy and Crop Science
Specific subject area	Computer vision techniques for detection of crops in small areas
Type of data	Images and annotations
Data collection	For data capture, a DJI Mini2 Unmanned Aerial Vehicle (UAV) was used, with a flight capacity of approximately 25 min per battery. The UAV camera is not interchangeable and has a 3-axis motorized stabilizer 1 / 2.3″ CMOS sensor - f / 2.8 - FOV 83 ° Maximum 4 K video resolution at 30 fps and maximum 12 MP photo resolution 4000 × 3000 pixels RGB. Its Global Navigation Satellite System (GNSS) can use the three available systems: GPS + GLONASS + GALILEO.
Data source location	The region under study was the village of San Cristóbal (6°17′42.68″ N 75°39′25.99″ O), which is part of the rural area in the city of Medellín, Colombia. This area is recognized for its production of vegetables and flowers in small family agricultural units. The farmers in the region usually intersperse these two crops as part of their subsistence economy, to diversify and reduce economic risk.
Data accessibility	Repository name: Zenodo Data identification number: 10.5281/zenodo.10995125 Direct URL to data: https://zenodo.org/records/10995125
Related research article	Restrepo-Arias, J.F., Arregocés-Guerra, P., Branch-Bedoya, J.W. (2023). Crops Classification in Small Areas Using Unmanned Aerial Vehicles (UAV) and Deep Learning Pre-trained Models from Detectron2. In: Zapata-Cortes, J.A., Sánchez-Ramírez, C., Alor-Hernández, G., García-Alcaraz, J.L. (eds) Handbook on Decision Making. Intelligent Systems Reference Library, vol 226. Springer, Cham. https://doi.org/10.1007/978-3-031-08246-7_12

1. Value of the Data

This dataset has several novel aspects compared to existing ones:

- The dataset has images of onion and foliage crops at different stages, with their respective labels, and can be used to inventory crops and areas prepared for cultivation in small areas.
- The dataset is publicly accessible and can be used by researchers or professionals in training their own machine-learning algorithms.
- The dataset can be used by public or private institutions to monitor the establishment of new areas with small green onion and foliage crops.
- The images are taken at a lower altitude compared to images in existing datasets for small-scale crops, allowing for easier identification of crop types.
- It contains high-resolution images, enabling several types of feature extraction.
- Foliage flowers and onions are crops for which there is scarce or no information in existing datasets.
- The labels allow for training segmentation algorithms, which can be applied in various ways.
- The applications of this dataset can help establish methodologies for conducting crop inventories in hard-to-reach areas with challenging topography, where updated satellite images are unavailable or unable to provide reliable information.

2. Background

The detection and quantification of small-scale crops has been an area of growing interest in the scientific community due to the great difficulty that exists in collecting information about the volumes of agricultural products that are grown in a given area, spread among many producers [1–4]. Each country has its particularities in distribution, size, types of crops, cultural practices, and support systems for small producers. Efforts have already been made to simplify the collection of data regarding small producers for the granting of subsidies based on satellite images [5]. This complexity makes having specific datasets by crop type and region necessary to train more specialized detection algorithms. Machine learning techniques, specifically deep neural networks, are those that are currently being used with the best results [1,3,6,7].

Publicly available datasets in the field of PA remain scarce, and most focus mainly on weed detection or fruit counting on crops in small areas, consisting of satellite images of large areas of land or regions [8]. Some datasets detected in the literature for the detection of crop types are based on satellite images and for areas and crops in different contexts [9–11]. Our dataset's main contribution is that it was made from drone images taken at very close distances, which may allow more precise identification of green onion and foliage flowers at different stages of development. Furthermore, no dataset with these characteristics exists for the area under study.

3. Data Description

The final dataset consisted of the labeling of three classes: (1) foliage flowers (label: foliage), (2) Green onion (label: onion) and (3) areas without crop (label: no crop) (Fig. 1). Areas labeled "no crop" are areas prepared for planting.

After labeling, the images were randomly divided into a training set of 201 images, validation set of 22 images and test set of 22 images. Tables 1, 2, and 3 show the number of instances

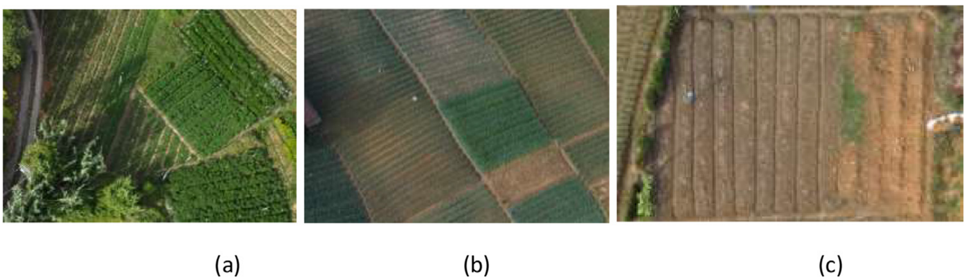

Fig. 1. Examples of the three selected classes: (a) foliage, (b) onion, (c) no crop.

Table 1
Instances distribution in train set for 201 images.

Classes	Onion	Foliage	No crop	Total
Instances	1148	975	1155	3278

Table 2
Instances distribution in validation set for 22 images.

Classes	Onion	Foliage	No crop	Total
Instances	183	180	208	571

Table 3
Instances distribution in test set for 22 images.

Classes	Onion	Foliage	No crop	Total
Instances	175	98	193	466

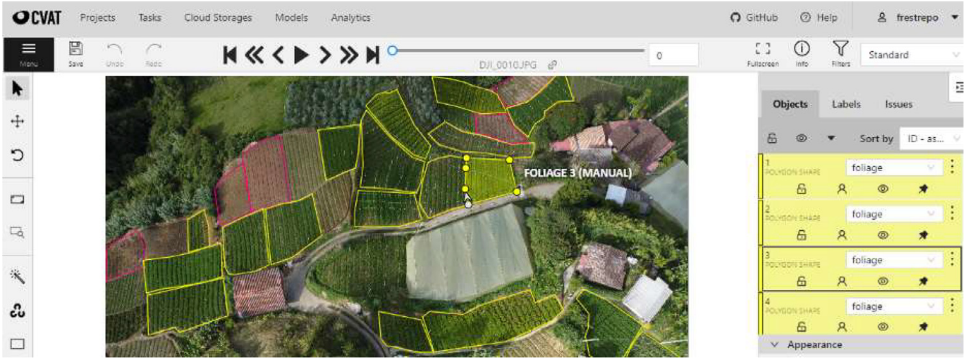

Fig. 2. Example of data labeling.

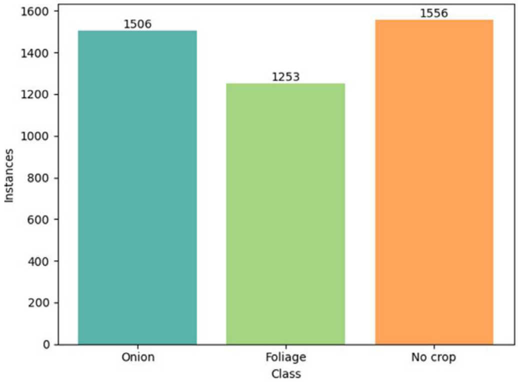

Fig. 3. Distribution of classes in the dataset.

of each class that contributed to the training, validation, and test sets, respectively. The labeling was carried out with the polygon labeling method with CVAT (Computer Vision Annotation Tool, https://cvat.org) (Fig. 2).

The overall distribution of instances shows that this is a balanced dataset, with an average of approximately 17 instances per image. The distribution of instances and the percentages for training, testing, and validation can be seen in Figs. 3 and 4, respectively.

The annotation method used to create polygons around the crops in the images was manual, following the contours of the plots. No AI-assisted techniques were used. The annotators were two of the authors of this work, who have worked with these types of crops for several years, so the criteria were based on their accumulated experience. The labels were reviewed three times to correct potential errors and ensure accuracy and consistency in the annotations.

4. Experimental Design, Materials and Methods

The study area is in Colombia, in the department of Antioquia, very close to the urban area of the city of Medellín, at an average altitude of 2250 m above sea level. This is an area with an agricultural tradition in the production of vegetables and flowers. Producers in this area base their production decisions on market supply and demand (Fig. 5).

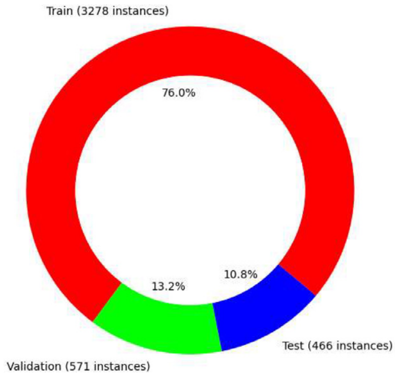

Fig. 4. Train, validation, and test distribution.

Fig. 5. General overview of the study area. Source: Google Earth 08/17/2021.

The images in the dataset were taken between 10 a.m. and 2 p.m. to utilize the best lighting and avoid shadows. No images were captured on rainy days or with foggy conditions. To ensure variability in the dataset, photos include combinations of crops with houses, roads, forests, uncultivated land, and even other types of crops not represented in the primary dataset. The images have a resolution of 72 ppi (pixels per inch) both vertically and horizontally.

The flights were scheduled to follow zigzag patterns in straight line, with a spacing of every 10 m and a minimum altitude of 50 m, never exceeding 100 m in height. This range was chosen to achieve optimal resolution of the crops, facilitating their identification. Figs. 6 and 7 show in detail the characteristics of the flight pattern used to take the images.

The flights were conducted over a period of two weeks with daily flight frequencies. However, flights were not conducted on 4 days as scheduled due to rainy or foggy conditions. A collection of 462 images was captured. Images of green onion and foliage crops were taken at different stages of development. The crops are surrounded by structures like houses and roads, as can be seen in the examples in Fig. 8. After debugging and selecting the images, 245 were selected that made up the final dataset.

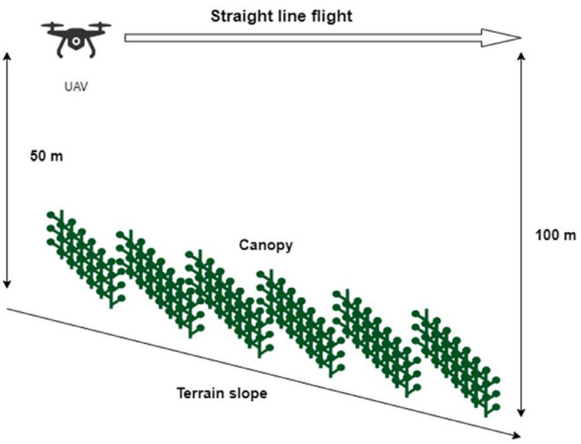

Fig. 6. General overview of the flight pattern with respect to the terrain slope.

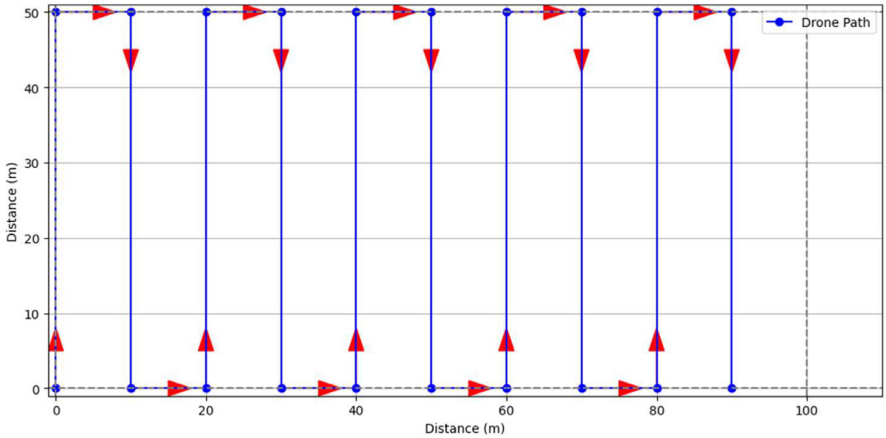

Fig. 7. Example of drone zigzag flight pattern for agricultural field imaging.

5. Examples of Practical Applications

In this context, aerial images are used to monitor and manage crop health, growth stages, and yield predictions across large agricultural areas. By using machine learning algorithms trained on the dataset, farmers can identify specific crop types, detect different growth stages, and use the results to predict yields or combine these results with data from other sources to detect signs of disease or pest infestation, and assess the overall health of their fields.

For example, the dataset can be used to train a convolutional neural network (CNN) to differentiate between onion, foliage flowers an no crops. Once trained, this model can analyze new aerial images to automatically identify the type and distribution of crops in a given region. This information can help farmers make data-driven decisions about where to apply fertilizers or pesticides, optimize irrigation schedules, and plan harvests more efficiently, ultimately leading to increased productivity and sustainability in agriculture [12–14].

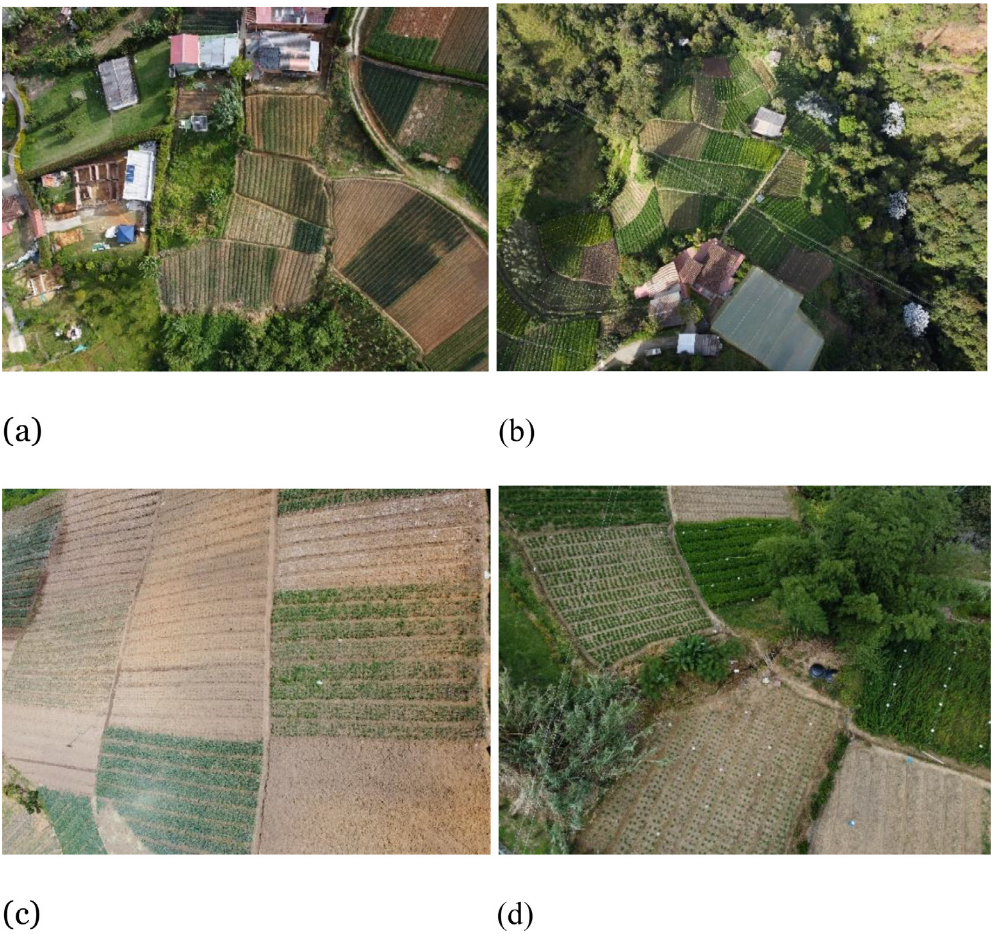

Fig. 8. Examples of images from dataset: (a) green onion growing in areas with structures around, (b) green foliage growing in areas with structures around (c) green onion in different stages, (d) foliage in different stages.

Limitations

This dataset has limitations in terms of the number of possible crops, only two. The number of images should also be increased in future works. The difference in heights in the flights can also be a limitation for area measurements. For future datasets, it is recommended that flights be performed at the same height.

In the captured images, bias may exist due to the inherent conditions of the terrain such as topography, vegetation, and infrastructure, which are unique to the study area. On the other hand, the sizes of the sampled crops are representative of all small-scale crops in the surrounding region. These plots typically vary in size, ranging from approximately 300 to 1000 m^2s.

The images may also be constrained by the specifications of the drone camera used, as it adheres to lower specifications compared to most drones used for capturing images in agricultural contexts. Therefore, it is advisable to use cameras with higher resolution in future work to assess how it can enhance the performance of the detection models to be trained.

Ethics Statements

The authors have read and follow the ethical requirements for publication in Data in Brief and confirming that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

CREDIT Author Statement

J.F. Restrepo-Arias: Conceptualization, methodology, image capture, labeling, data curation, writing, **J.W. Branch-Bedoya:** Supervision, original draft preparation, reviewing. **P. Arregocés-Guerra:** labeling, data curation.

Data Availability

OnionFoliageSET (Original data) (Zenodo).

Acknowledgements

Funding: This work was supported by the Universidad Nacional de Colombia Sede Medellín.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R. Yang, Z.U. Ahmed, U.C. Schulthess, M. Kamal, R. Rai, Detecting functional field units from satellite images in smallholder farming systems using a deep learning based computer vision approach: a case study from Bangladesh, Remote Sens. Appl. 20 (2020), doi:10.1016/j.rsase.2020.100413.
[2] Z. Du, J. Yang, C. Ou, T. Zhang, Smallholder crop area mapped with a semantic segmentation deep learning method, Remote Sens. 11 (7) (2019), doi:10.3390/RS11070888.
[3] Y. Lv, et al., The delineation and grading of actual crop production units in modern smallholder areas using RS Data and Mask R-CNN, Remote Sens. 12 (7) (2020), doi:10.3390/rs12071074.
[4] S. Wang, F. Waldner, D.B. Lobell, Unlocking large-scale crop field delineation in smallholder farming systems with transfer learning and weak supervision, Remote Sens. 14 (22) (2022), doi:10.3390/rs14225738.
[5] J. Schmedtmann, M.L. Campagnolo, Reliable crop identification with satellite imagery in the context of Common Agriculture Policy subsidy control, Remote Sens. 7 (7) (2015) 9325–9346, doi:10.3390/rs70709325.
[6] G. Li, et al., Multi-year crop type mapping using sentinel-2 imagery and deep semantic segmentation algorithm in the Hetao Irrigation District in China, Remote Sens. 15 (4) (2023), doi:10.3390/rs15040875.
[7] J. Pena, Y. Tan, W. Boonpook, Semantic segmentation based remote sensing data fusion on crops detection, J. Comput. Commun. 07 (07) (2019) 53–64, doi:10.4236/jcc.2019.77006.
[8] Y. Lu, S. Young, A survey of public datasets for computer vision tasks in precision agriculture, Comput. Electron. Agric. 178 (August) (2020) 105760, doi:10.1016/j.compag.2020.105760.
[9] A. Maulit, A. Nugumanova, K. Apayev, Y. Baiburin, M. Sutula, A Multispectral UAV imagery dataset of wheat, soybean and barley crops in East Kazakhstan, Data 8 (5) (2023), doi:10.3390/data8050088.
[10] L.V. Oldoni, I.D.A. Sanches, M.C.A. Picoli, R.M. Covre, J.G. Fronza, LEM+ dataset: for agricultural remote sensing applications, Data Br. 33 (2020), doi:10.1016/j.dib.2020.106553.
[11] R. Rustowicz, R. Cheong, L. Wang, S. Ermon, M. Burke, and D. Lobell, “Semantic Segmentation of Crop Type in Africa: A Novel Dataset and Analysis of Deep Learning Methods.” [Online]. Available: https://github. URL:https://github.
[12] I. Teixeira, R. Morais, J.J. Sousa, A. Cunha, Deep learning models for the classification of crops in aerial imagery: a review, Agriculture 13 (5) (2023) MDPI, doi:10.3390/agriculture13050965.

[13] I. Kalita, G.P. Singh, M. Roy, Crop classification using aerial images by analyzing an ensemble of DCNNs under multi-filter & multi-scale framework, Multimed. Tools Appl. 82 (12) (2023) 18409–18433, doi:10.1007/s11042-022-13946-1.

[14] A. Bouguettaya, H. Zarzour, A. Kechida, A.M. Taberkit, A survey on deep learning-based identification of plant and crop diseases from UAV-based aerial images, Cluster Comput. 26 (2) (2023) 1297–1317, doi:10.1007/s10586-022-03627-x.