Data in Brief 48 (2023) 109165

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Strawberry-DS: Dataset of annotated strawberry fruits images with various developmental stages

Esraa Elhariri [a,*], Nashwa El-Bendary [b,*], Samir Mahmoud Saleh [c]

[a] Faculty of Computers and Artificial Intelligence, Fayoum University, Fayoum, Egypt
[b] College of Computing and Information Technology, Arab Academy for Science, Technology and Maritime Transport (AASTMT), Aswan, Egypt
[c] Central Laboratory for Agricultural Climate, Agricultural Research Center, Giza, Egypt

ARTICLE INFO

Article history:
Received 30 December 2022
Revised 17 March 2023
Accepted 12 April 2023
Available online 20 April 2023

Dataset link: Strawberry_DS_Dataset (Original data)

Keywords:
Strawberry maturity classification
Ripeness
Agricultural crops
Deep learning
Machine learning
Object detection

ABSTRACT

Strawberry (Fragaria X ananassa) is one of the most popular fruits cultivated around the world. That is owing to its unique flavor and nutritious properties in addition to the wide usage utility in fresh or processed condition. Strawberry fruits also have a significant economic importance around the world with strong potential as an export commodity. As a matter of fact, investigation and assessment of various strawberry fruit characteristics at different developmental stages is crucial for candidate cultivars selection in fruit plantation as well as fruit yield prediction. Strawberry fruits developmental stage is conventionally applied visually based on expert knowledge, which is a time and labor exhaustive process. Thus, this paper presents a dataset, namely Strawberry-DS (Strawberry-Developmental Stages) dataset, consisting of strawberry fruits (Festival CV type) expert-annotated images at various developmental stages. Data collection was performed on site during the period between January and March from a greenhouse located in the Central Laboratory for Agricultural Climate (CLAC) at the Agricultural Research Center of the Ministry of Agriculture and Land Reclamation in Giza, Egypt. The dataset comprises 247 high-resolution RGB (.jpg) images annotated manually, using Roboflow Annotate annotation tool with reference to ground

* Corresponding authors.
E-mail addresses: emh00@fayoum.edu.eg (E. Elhariri), nashwa.elbendary@aast.edu (N. El-Bendary).

https://doi.org/10.1016/j.dib.2023.109165
2352-3409/© 2023 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

truth Region of Interest (RoI), and presented in YOLO (.txt files) format. The presented Strawberry-DS dataset can be generally used for developing various automated models of strawberry fruits detection, fruits maturity stage classification, as well as visual counting, through taking into account the visual features such as shape, color, size, and texture of strawberry fruits. Strawberry-DS is freely available at: https://data.mendeley.com/datasets/z6dtfdpzz8/1.

© 2023 The Author(s). Published by Elsevier Inc.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Specifications Table

Subject	Computer Science, Machine Intelligence
Specific subject area	Computer vision; Image classification; Pattern recognition; Automated harvesting; Autonomous robot; Visual yield estimation
Type of data	2D RGB digital images (.jpg) Annotations: YOLO format (.txt)
How the data were acquired	Real RGB digital images of strawberry fruits (Fragaria X ananassa, Festival CV) in different ripeness stages were collected using Sony Xperia Z2 LTE-A D6503 smartphone 20.7 MP camera with a CMOS sensor system and resolution of 3840 x 2160 pixel (Mpix). The dataset images, which contain both fully-visible strawberry fruits and partially-visible strawberry fruits concealed by leaves or by other fruits, were manually annotated, using Roboflow Annotate annotation tool.
Data format	Raw Data
Description of data collection	In-field collection of strawberry fruits images was achieved inside a greenhouse during the period between January and March at day-light. Images are captured from the top view of the fruits and also from different view angles. Strawberry-DS.zip dataset archive consists of 247 .jpg images of strawberry fruits in six ripeness stages, namely Green, White, Early-Turning, Turning, Late-Turning, and Red.
Data source location	A greenhouse located in the Central Laboratory for Agricultural Climate (CLAC) at the Agricultural Research Center of the Ministry of Agriculture and Land Reclamation in Giza, Egypt
Data accessibility	The dataset is freely accessible at: https://data.mendeley.com/datasets/z6dtfdpzz8/1 DOI: 10.17632/z6dtfdpzz8.1

Value of the Data

- Strawberry-DS dataset is recommended to be considered for developing robotic harvesting systems that are supposed to contribute significantly to increase yield and decrease production costs and time. Accordingly, it can be counted as a supporting contribution to the research field of advanced agriculture automation.
- Strawberry-DS dataset is essential for building reliable and robust machine vision models for supporting various decisions related to pre-harvesting, harvesting, and post-harvesting processes of strawberry fruits. That contribution should allow farmers and agriculture companies to widely adopt and deploy automated harvesting recommendation systems based on visual recognition of strawberry fruit developmental stages. In addition, automated counting systems of strawberry fruits based on the visual features are supposed to ensure strawberry fruit quality while being packed.
- Strawberry-DS dataset can be utilized by researchers through implementing various ML and DL feature selection and feature extraction methods in order to generate numerous sets of descriptors for each one of the different strawberry fruits developmental stages. Moreover,

the dataset can also be used by researchers for developing, training, testing, and validating several Convolutional Neural Networks (CNN) classification and recognition models. Also, various challenges facing fruit detection, segmentation, and developmental stage recognition could be of interest to many researchers, such as addressing the problems of overlapped fruits, color similarity between unripe fruits and leaves and stem, occlude fruits with the same ripeness stage, etc.

- Several applications could be developed, based on trained models with the Strawberry-DS dataset, in order to support farmers through using their smartphones along with other advanced technological instruments, such as drones, to monitor and predict the yield effectively in addition to accurately evaluating the harvesting time of strawberry fruits.
- Strawberry-DS dataset can be utilized by researchers through implementing various data augmentation techniques, such as spatial augmentation (rotation, shift), intensity augmentation (blurring, noises), and Generative Adversarial Network (GAN) based augmentation methods, in order to enlarge dataset size and investigate the effect of applying different augmentation techniques on ML systems performance.

1. Objective

The rationale behind generating the proposed dataset of strawberry fruits (Festival CV type) expert-annotated images at various developmental stages; namely, Green, White, Turning, and Red, originally focuses on enabling training and validation of Deep Learning (DL) models for image detection, segmentation, and classification. That target approach should support developing vision-based autonomous robotic harvesting systems by utilizing several trained Convolutional Neural Network (CNN) DL models. The main goal of developing such robotic based harvesting systems is to contribute significantly to increase yield and decrease production costs and time through automatically detecting, recognizing the growth stage, and accordingly collecting fruits with the correct ripeness stage. Moreover, unlike other online related datasets, such as the ones presented in [1–3], that have just addressed ripe/unripe classification and fruit segmentation challenges, the noticeable strength point of our proposed Strawberry-DS dataset is that it contains images covering six strawberry fruit developmental stages namely, Green, White, Early-Turning, Turning, Late-Turning, and Red, with an advantage that facilitates more accurate estimation of fruit harvesting time.

2. Data Description

Strawberry is a microclimatic fruit that is grown all over the world in a variety of soils and climates. Strawberry fruits have shown an increasing consumption growth rate with a significantly increased production over the last two decades [4,5]. According to the research studies proposed in [6,7] as well as other related studies, the typical classes of strawberry fruit phenological developmental stages are reported to be four stages, namely Green, White, Turning, and Red. In the proposed paper, through consulting an expert to develop more precise clustering, the developmental stages of strawberry ripening are divided into six different ripeness stages, through dividing the Turning stage into three developmental sub-stages. Accordingly, the 247 annotated raw RGB strawberry fruit images in Strawberry-DS dataset are divided into six ripeness stages, namely Green, White, Early-Turning, Turning, Late-Turning, and Red. Fig. 1 shows sample images of strawberry fruit phenological developmental stages in Strawberry-DS dataset. Table 1 lists the number of fruits in each maturity stage included in the Strawberry-DS dataset.

The RGB digital images of Strawberry-DS dataset were acquired for strawberry fruits (Fragaria X ananassa, Festival CV) in different ripeness stages inside a greenhouse during the period between January and March at day-light. The images have been captured from the fruit top view considering different view angles using Sony Xperia Z2 LTE-A D6503 smartphone 20.7 MP camera with a CMOS sensor system and resolution of 3840 x 2160 pixel (Mpix). The data source

Fig. 1. Illustrative images of strawberry fruit six developmental stages from Strawberry-DS dataset.

Table 1
Strawberry-DS dataset description.

Class		No. of fruits at each ripeness stage	
Green		455	
White		257	
Turning	Early-Turning	28	117
	Turning	35	
	Late-Turning	54	
Red		233	

location is a greenhouse located in the Central Laboratory for Agricultural Climate (CLAC) at the Agricultural Research Center of the Ministry of Agriculture and Land Reclamation in Giza, Egypt. The dataset images, which contain both fully-visible strawberry fruits and partially-visible strawberry fruits concealed by leaves or by other fruits, were manually annotated, using Roboflow Annotate annotation tool. The data formats of files in Strawberry-DS dataset are RGB digital images (.jpg) and their corresponding YOLO format (.txt) annotation files. Fig. 2 shows example images from Strawberry-DS dataset with their corresponding annotations.

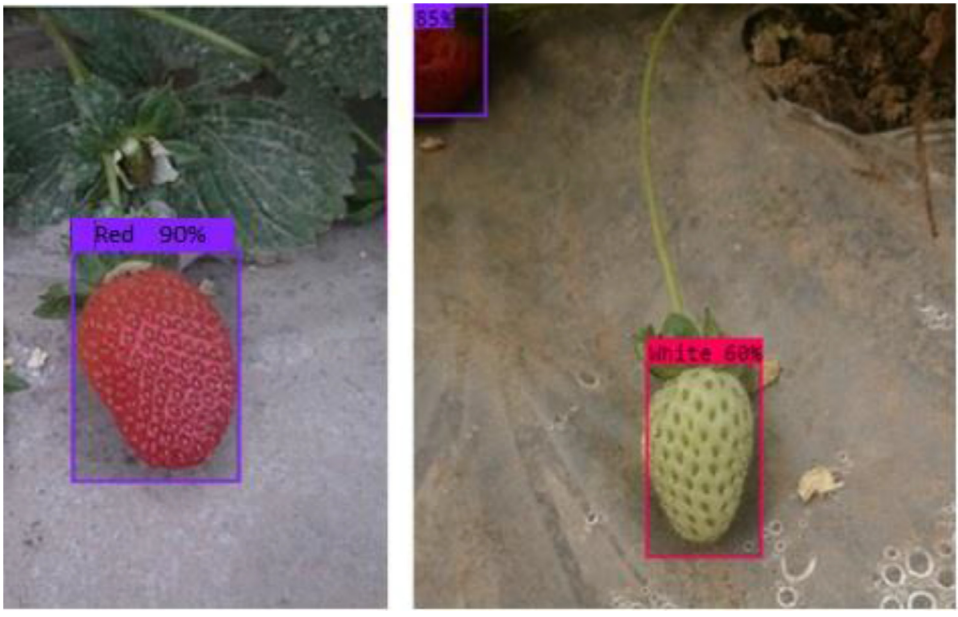

Fig. 2. YOLO detection of strawberry fruits.

3. Experimental Design, Materials and Methods

3.1. Image Capturing

The strawberry images were taken in an orchard of the Central Laboratory for Agricultural Climate (CLAC), Agricultural Research Center, Cairo - Egypt located in [30°02′45.5"N 31°12′17.7"E]. The collected dataset was captured using Sony Xperia Z2 LTE-A D6503 smartphone 20.7 MP camera with a CMOS sensor system and resolution of 3840 × 2160 pixel (Mpix). The collected dataset contains several challenges such as fruit occlusion and/or overlapping, in addition to the noisy background problem, as shown in Fig. 3.

3.2. Image Annotation

The dataset, which has been uploaded to the Roboflow Annotate website, is arranged for a total of 247 original raw Strawberry images .jpg. The dataset consists of six classes: (a) Green, (b) White, (c) Early-Turning, (d) Turning, (e) Late-Turning, and (F) Red. Label files .txt format containing the class names (0→Green, 1→White, 2→Early-Turning, 3→Turning, 4→Late-Turning, and 5→Red) with associated ground truth bounding boxes were annotated into the format required by YOLO. The dataset contains 1062 ground truth bounding box annotations.

With reference to an agricultural expert, the annotation of strawberry fruits ripeness stages in the Strawberry-DS dataset images has been applied as follows: Green annotation, refers to fruits where the surface is completely “green”, while the shade of green may vary from light to dark. White annotation refers to fruits where the surface is definitely “white” with no or insignificant green or red shades. For the three turning stages, the red color ranges from 10% to 90% of the whole fruit surface. Early-Turning annotation refers to fruits where 10% to 30% of the surface shows a change from white to tannish-red color. Turning annotation, refers to fruits where 30% to 60% of the surface shows pink red color. Late-Turning annotation refers to fruits where 60%

Fig. 3. Sample images from Strawberry-DS dataset with various challenges of fruit occlusion, overlapping, and noisy background.

Fig. 4. Strawberry-DS sample image illustrating three objects with the corresponding annotation data.

to 90% of the surface shows pinkish-red or red color. Finally, for fruits with more than 90% of the surface is red, that refers to Red annotation.

For YOLO annotation, each image has a corresponding .txt file with the same name. In other words, if the image file name is DSC_1359.JPG the annotation file name in the labels directory must be DSC_1359.txt. Each annotation file contains the annotated bounding boxes for all target objects in the corresponding image with a line consisting of object class, object coordinates, height and width for each ground truth object in the image that looks like:

<object-class> <center_x> <center_y> <width> <height>

For example, the .txt annotation file for an image having only three objects, as shown in Fig. 4, will be as follows:

5 2547.5 1366.5 369 609 → strawberry fruit in Red ripeness stage, with coordinates of x=2547.5, y=1366.5, and a bounding box with width=369, height=609

0 429.5 946 323 194 → strawberry fruit in Green ripeness stage, with coordinates of x=429.5, y=946, and a bounding box with width=323, height=194

0 2898.5 1010.5 203 231→ strawberry fruit in Green ripeness stage, with coordinates of x=2898.5, y=1010.5, and a bounding box with width=203, height=231

Ethics Statements

This research and analysis did not involve the use of human subjects or animal experiments.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

Data Availability

Strawberry_DS_Dataset (Original data) (Mendeley Data).

CRediT Author Statement

Esraa Elhariri: Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft; **Nashwa El-Bendary:** Conceptualization, Data curation, Formal analysis, Methodology, Writing – review & editing; **Samir Mahmoud Saleh:** Data curation, Formal analysis.

Acknowledgements

This work was partially funded by the research grant from L'Oréal-UNESCO for Women in Science Levant & Egypt fellowship awarded to Nashwa El-Bendary, the second and corresponding author of this research.

References

[1] Pastell, M., Madis, L., & Bloch, V. (2022, February 17). Strawberry dataset for object detection. Zenodo. (Last accessed: March 13, 2023), https://zenodo.org/record/6126677#.ZA9d1ZFBy3A.

[2] Pérez, I., Gegúndez Arias, M. E., Cortés-Ancos, E., & Marin, D. (n.d.). StrawDI: The Strawberry Digital Images Data Set. StrawDI. . (Last accessed: March 13, 2023), https://strawdi.github.io/.

[3] Gao, K. (2022). Strawberry object detection dataset. Roboflow. . (Last accessed: March 13, 2023), https://universe.roboflow.com/kanke-gao/strawberry-xfxpu.

[4] A.N. Saber, F. Malhat, C. Anagnostopoulos, K.M. Kasiotis, Evaluation of dissipation, unit–unit-variability and terminal residue of etoxazole residues in strawberries from two different parts in Egypt, J. Consum. Protect. Food Saf. 15 (3) (2020) 229–236.
[5] X.Q. Yue, Z.Y. Shang, J.Y. Yang, L. Huang, Y.Q. Wang, A smart data-driven rapid method to recognize the strawberry maturity, Inf. Process. Agric. 7 (4) (2020) 575–584.
[6] Hayat Topcu, Ipek Degirmenci, Duygu Ayvaz Sonmez, Aibibula Paizila, Harun Karci, Salih Kafkas, Ebru Kafkas, Sezai Ercisli, Aishah Alatawi, Sugar, invertase enzyme activities and invertase gene expression in different developmental stages of strawberry fruits, Plants 11 (4) (2022) 509.
[7] C.E. Basson, J.H. Groenewald, J. Kossmann, C. Cronjé, R. Bauer, Sugar and acid-related quality attributes and enzyme activities in strawberry fruits: Invertase is the main sucrose hydrolysing enzyme, Food Chem. 121 (4) (2010) 1156–1162.