USE OF ANNOTATED IMAGE DATA FOR FRUIT DIVERSITY ANALYSIS

Miroslav Jaroš[1], Jiří Podivín[1], Petr Pernes[1], Oldřich Trenz[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

ABSTRACT

This paper deals with a method of development of an annotated image dataset for the detection and classification of plant tissues, aimed at supporting automation in agriculture. The work includes a collection of high-definition image data, their annotation and utility scripts, with the aim of creating a universally accessible dataset for the scientific community. The method is designed to be compatible with off-the-shelf hardware, in order to better support research and development in the field of automated plant identification and plant disease diagnostics. This approach has the potential to significantly improve the efficiency of cultivation processes and support the implementation of advanced technologies in the agricultural sector, along with the automation of this sector.

Keywords: image analysis, plant classification, dataset, learning, annotation, image data

JEL Code: Q16, C55, C88

1 INTRODUCTION

In agricultural engineering, proper fruit detection and counting using image analysis and computer vision algorithms is critical for automating and streamlining processes from measuring phenotypic characteristics to harvesting. The difficulty of this solution lies not only in the availability of the required dataset for learning the model used, but also in the ability of this model to perform its own detection and classification at the level of plants and their tissues. This is quite a challenging discipline, especially due to the diverse structure, color, and considerable number of fruits available, as reported by Ukwuoma, et al. (2022). Until 2018, various machine learning methods were used for detection and classification, and applications using deep learning methods began to gain ground (Zhang, et al., 2014; Then, Kim, 2017), alternatively algorithms published in (Stastny, Skorpil, 2007; Stastny et al., 2021).

Availability of a suitable image dataset is an indispensable prerequisite for training of a model. That is, the availability of a sufficiently comprehensive dataset, consisting of a diverse collection of samples. Considerable number of images is a necessary condition for accurate representation of the actual environment, for example, in orchards and greenhouses, where the fruits may be occluded by the leaves of the plant, overlap, or they may have the same color

https://doi.org/10.11118/978-80-7509-990-7-0084

as the rest of the plant. When an image dataset is available, it is then possible to train a model, or compare existing models and select one that has performance for the selected application deployment (Ukwuoma et al., 2022). Some methods may also include further data augmentation, for example, for better extraction of image features, i.e. the ability to better identify the object being evaluated (Jia et al., 2023), or to increase the number of samples.

Image datasets are created by a sequence of the following steps, which include: image data collection (can involve taking images of selected agricultural crops at different stages of development, at different angles and under different lighting conditions); data labeling and annotation (manual, semi-automatic, or automatic annotation of images, i.e. labeling selected parts of a plant and assigning them to specified classes in order to classify individual crops and plants); splitting the dataset (for purposes of cross validation); optimization and pre-processing of data (adjustment of images with regard to changing the resolution of images in order to train models effectively); incorporation of diversity (extension of the dataset to include a wider range of plants, growing conditions, rotation of images, or use of augmentation to ensure robustness of the model), (Chiu et al., 2020; Wspanialy et al., 2020; Grinblat, et al. 2016). It is worth mentioning that beyond specialized datasets of agricultural crops, there are also general image datasets. Created in order to enable the training of models for the recognition of various objects, and thus also plants. However, there is a limit to the applicability of these resources for training of models for specific use in agriculture (Deng et al., 2009).

In general, the number of available samples (cardinality) and its division into individual classification classes are important for datasets. It is obvious that as the number of classification classes increases, it is necessary to have a larger total number of images. For example, a dataset of weeds, consisting of 14.035 images, was necessary to train a model capable of identifying 25 weed species, with achieved accuracy in the range of 91.8% - 92.4% (Wang et all, 2022). Currently, datasets of various plants and their fruits are available, but the differences are in their robustness, the number of classification classes and the classification ability itself. Other datasets include, for example, the tomato leaf dataset (Chang, 2020) or the tomato fruit dataset (Afonso et al., 2021).

1.1 Description of the created dataset (research content)

As part of our work, we focused on the development of a unique semantic segmentation dataset that is characterized by image resolution and partially automated mask creation. The dataset is characterized by images in 4K resolution, allowing to capture the detailed structure of plants, from leaves to individual fruits (Sapoukhina et al., 2022). This provides a considerable level of detail and allows for deeper analysis and understanding of plant structures.

Unlike other available datasets (Gajjar et al., 2022), the dataset created contains diverse image types that include a wide range of agricultural environment scenarios, including complex backgrounds and various stages of plant growth. The dataset was created by a group of students and is freely available to all interested researchers and companies. This approach makes it possible to provide a unique dataset that is not only more extensive than commonly available, but also provides more detailed information for research and applied purposes. Application deployment of such a dataset is possible at the level of learning models for the identification of plants, plant elements, and diseases. (Fenu et al., 2021). The dataset is published under a “Community Data License Agreement – Permissive, Version 2.0” for free download for all for further research and commercial use.

1.2 Aim of the paper

The aim of the paper is to describe a method of dataset creation and provide a comprehensive dataset of agricultural crops for learning classification and detection models usable in the field of agriculture. Emphasis is also placed on the usability of common available off-the-shelf hardware for image data collection, such as mobile phones, and the availability of these image sources, with an intended effect of increasing the efficiency of the cultivation process, identifying plant diseases, strengthening local production, deploying automation in the cultivation process while maintaining sustainability (Hughes et al., 2015).

2 METHODOLOGY AND DATA

The methodology describes the methodological framework for image data collection and the creation of a dataset for learning models for classification and detection, including possible extensions and limitations of this approach.

LabelStudio's annotation tool builds on previous toolchains and methods used to label images and includes several complex annotation workflows. Because agricultural imagery has a large number of objects and with more complex shapes than those in many general data sets, we focused on this tool for this, which allowed us to annotate multiple types of annotations and combine them in diverse ways.

LabelStudio was chosen for its flexibility, support for a variety of data formats, and customizability of annotation processes. To ensure consistency of annotations across the dataset, a manual has been created describing the specifications for tagging individual elements such as Fruit, Stem and Leaf with an XML file definition for LabelStudio, see Figure 1. This definition ensures that all annotators act consistently and with respect to the required quality standards.

2.1 Methodology

The annotation methods used by this tool can be applied separately or in combination with other tools to identify each object in an image. These annotated objects can contain several separate segments, allowing for adaptation to occlusion situations. In other words, even if a part of an object is obscured or not fully visible, the system can identify and annotate those parts of the object that are visible. In this way, it is possible to obtain the most complete and accurate annotation of the object despite the presence of occlusal situations. In addition to the main label, each object can also carry additional metadata that provides further contextual information. As an example, certain fruit may have associated metadata describing its ripeness, a text description, or the ID of the plant in question.

For our purposes we have augmented LabelStudio tools with an integrated model for generating segmentation masks, Segment Anything Model (SAM), further referred to as SAM. This integration allowed us to generate accurate segmentation masks with minimum effort compared to manual approach.

The following is a description of the main tools implemented by LabelStudio and used for our methodology:

2.1.1 Keypoint

Keypoint Annotation in LabelStudio, supported by the SAM, is designed to generate masks for objects using a single point, chosen manually by the user of the LabelStudio. The process begins with the user's selection of a point, to which the SAM model responds by assessing and suggesting the optimal location and shape of the mask. The user can approve this mask or add additional points to increase the annotation's accuracy. The system also allows you to define negative points to exclude some areas from the annotation. Although there is an automatic point confirmation feature, we have found that in order to maintain high accuracy of the annotations it is advisable to disable it. This approach makes LabelStudio augmented with SAM a potent tool for users looking for a combination of efficiency and accuracy in image data annotation.

2.1.2 Rectangle

Rectangle annotation in LabelStudio, supported by the SAM model, is an effective solution for identifying and delineating objects in image data. The process begins with the user's selection of the rectangular box that encloses the object. The SAM model then generates a single mask for the object enclosed. This annotation is useful for cases where the objects are already known and SAM is familiar with them, as it helps the user to better determine the boundaries of the object and be more likely to correctly identify the object. This approach is ideal for projects where accurate localization of known objects is key, and SAM provides important support within the annotation process.

Rectangle annotation can be used with great effect to quickly label large objects within the processed image. But it is not well suited to situations when objects are small relative to the image size, or when their boundaries wouldn't fit within a rectangle selection.

2.1.3 Generating Greyscale masks for semantic segmentation

Machine learning's methodology for generating semantic segmentation masks often uses greyscale images for representing class identification of individual pixels.

In practice, multiple annotations of objects in the same class, such as a leafs, stems, fruits, etc., are merged into a single mask, representing all pixels belonging to a specific class in the original image. (Öztürk, B. and Özkar, M. 2022; Guo, S. et all 2020)

2.2 Data

2.2.1 Data Collection

Data was collected in 4K resolution from a variety of different devices, from mobile phones and digital cameras to drones. This multidisciplinary approach allowed us to capture the image material from different angles and in different lighting conditions, which contributes to greater variability and richness of the dataset. Each image in the dataset contains at least ten annotations, which ensures sufficient diversity of objects to be used for further processing.

2.2.2 Uniqueness of the data

Great emphasis was placed on the uniqueness of the data in order to avoid redundancy and ensure broad representativeness of the samples. Each image is unique and brings new value for research purposes. By carefully selecting and structuring the data, we are able to differentiate between different states and types of plant material, which is crucial for machine learning and computer vision applications in the field of agriculture.

The presented methodology and data collection represent an innovative approach in the field of agricultural research. Detailed resolution and complex annotations allow us to analyze and understand the visual characteristics of plants in depth. The results of this project have the potential to significantly contribute to progress in the automation of agricultural processes and the improvement of decision-making mechanisms in this area.

3 RESULTS

The result of our method is a dataset allowing us to both train new models or existing models using the fine-tuning method, for purposes of semantic segmentation of plant tissues. Its uniqueness lies, firstly, in the size of the sample, and secondly, in the way the input data is acquired using consumer hardware such as mobile phone cameras.

Due to the use of the SAM model with the specific annotation method, it became necessary to modify the original export routine used by LabelStudio in order to allow for local export of annotations on demand. Which can be impossible in Label Studio if the dataset is sufficiently large. Furthermore, Label Studio does not natively allow you to merge masks into one, but generates masks one object at a time, which is unsatisfactory for training purposes. Therefore it was necessary to adjust individual steps of the export process and write a complex script solving the issue. The reason is the use of the SAM model and the annotation method, which the LabelStudio did not expect, but with the advent of SAM, this has fundamentally changed.

Result is a unique dataset of plant organs, with labels for fruit, foliage and stems. With fruit consisting of a mix of apples, rose hips, tomatoes. It is prepared in the expected format and always described in the form of the original file name and then the mask separately for leaves, fruits and stems. Thus, merged masks are the output of multiple annotations of one class by multiple annotators for an individual image.

The use of greyscale mask, combined with advanced machine learning algorithms such as Mask R-CNN, makes it possible to achieve high accuracy in object detection and segmentation in scenarios where it is critical, such as in automated harvesting in agriculture.

This code serves as a template specifying how the SAM model will be integrated into the annotation workflow and LabelStudio GUI. It describes in detail what parameters and settings are used for the correct functioning of the model within the annotation tool, including some specific instructions and rules that determine the behavior of the model when identifying and segmenting objects in images and representation of created masks in LabelStudio GUI.

```
<View>

<Image name="image" value="$image" zoom="true" zoomControl="true" rotateControl="false"

<BrushLabels name="tag" toName="image">

<Label value="Leaf" background="#66ff00"/>

<Label value="Fruit" background="#0000ff"/>

<Label value="Stem" background="#ffff00"/>

<Label value="Flower" background="#ff00d0"/>

</BrushLabels>

<Header value="Please select the KeyPoint"/>
<KeyPointLabels name="tag2" toName="image" smart="true">

<Label value="Leaf" background="#000000" showInline="true" smart="true"/>

<Label value="Fruit" background="#000000" showInline="true" smart="true"/>

<Label value="Stem" background="#000000" showInline="true" smart="true"/>

<Label value="Flower" background="#000000" showInline="true" smart="true"/>

</KeyPointLabels>

<Header value="Please select the Rectangle"/>
<View><Filter toName="tag3" minlength="0" name="filter"/><RectangleLabels name="tag3" t
```

Fig. 1: The image illustrates the XML configuration code used in Label Studio to set up and define SAM for automatic annotation.

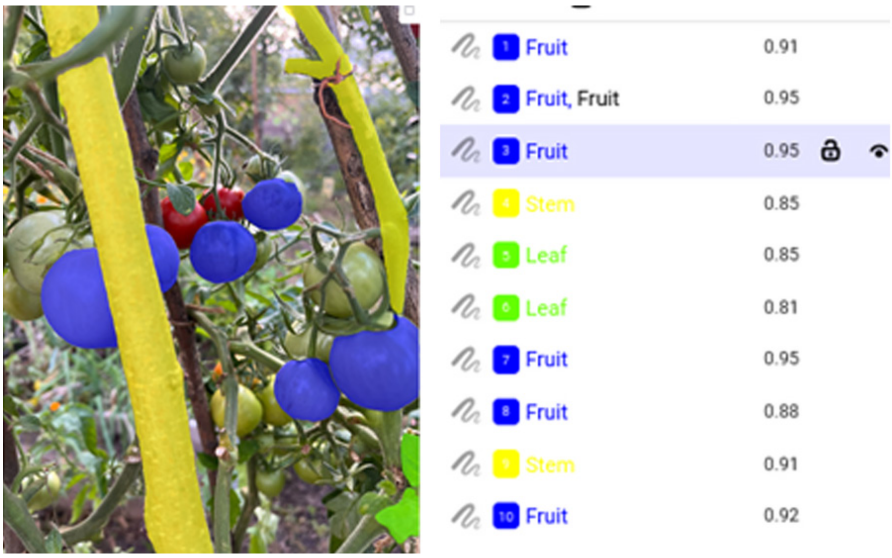

Fig. 2: An example of the Label Studio interface used to create segmentation masks – the masks were generated by the SAM model. On the right is the estimated accuracy of each mask.

For a deeper understanding of how this XML configuration and definition for the SAM model are used in LabelStudio, it is useful to look at the relevant documentation or examples that may be available online. You can visit the LabelStudio website or browse the SAM documentation directly for more information on exactly how this configuration and definition is applied in practice.

Fig. 3: Overlayed with masks. The colors of each mask class have been changed for better visibility.

The use of greyscale images to indicate the affiliation of pixels to individual classes is one of the basic methods of computer vision, specifically in solving visual data segmentation tasks. The advantages of the method include easy coding of the resulting masks, and unambiguous interpretation of the model.

The disadvantage is the possibility of assigning each pixel to exactly one class. This makes it impossible to hierarchically classify objects or recognize individual instances of objects of the same class.

4 DISCUSSION AND CONCLUSIONS

In our efforts to optimize data handling and enhance machine learning applications in agriculture, we evaluated various data export formats offered by Label Studio. The internal representation of segmentation masks in Label Studio uses Run-Length Encoding (RLE) to minimize storage demands. Because of this choice, export of the data, particularly in formats like COCO, places significant computational demands when handling high-resolution images. This challenge has sometimes prevented the complete export of annotated datasets. To address this, we developed utility software that allows for direct data export from the internal format used by Label Studio, thus circumventing the less efficient default export mechanisms.

Creating segmentation masks in Label Studio has demonstrated considerable advantages for agricultural machine learning and computer vision. The interface is intuitively designed, accessible even to non-experts, and is highly adaptable to specific project requirements. However, the substantial volume and high resolution of the processed data pose challenges. Despite these, the potential for significant automation and process enhancement in agriculture through our methodology is vast. Future work will aim to expand this approach to additional plant types and tissues and integrate other technologies such as spectral analysis to further enhance object detection and classification capabilities.

Our dataset presents a significant evolution over traditional datasets like plant village (Hughes et al., 2015), which primarily focus on segmenting specific plant parts or tomato image dataset (Afonso et al., 2021) In contrast, our dataset encompasses a comprehensive scope, capturing entire plant structures – leaves, fruits, and stems–in a single high-resolution image. This extensive segmentation capability facilitates a more detailed and holistic analysis of plant features, crucial for advancing agricultural research and applications.

The diversity of our dataset is also augmented by the variety of data collection devices employed, ranging from mobile phones and digital cameras to drones. This multidisciplinary approach not only diversifies the perspectives and lighting conditions under which the images are captured but also greatly enhances the dataset's variability. This is crucial for developing robust machine learning models as each image is annotated with at least ten distinct labels, ensuring a rich diversity conducive to intricate object detection and segmentation tasks.

The uniqueness of our dataset is underscored by its high resolution and comprehensive annotations. We have meticulously curated our dataset to eliminate redundancy and ensure a wide representativeness of agricultural samples. Each unique image enriches agricultural research, distinguishing between various states and types of plant materials effectively.

Our dataset supports both the development of new models and the fine-tuning of existing ones for the semantic segmentation of plant tissues. Modifications were made to the standard export routines in Label Studio to facilitate on-demand local annotation export, addressing the limitations of handling extensive datasets and merging annotations for effective training.

This innovative dataset not only propels the automation of agricultural processes but also enhances decision-making mechanisms within the sector. Using a greyscale mask in conjunction with advanced algorithms like Mask R-CNN achieves high precision in object detection and segmentation, critical for applications such as automated harvesting. Prepared in expected formats with clearly identified and separated masks for different plant parts, our dataset is poised for immediate application in cutting-edge agricultural research, marking a significant contribution to the field and promising to drive substantial advancements in agricultural technology and plant science.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

Thanks to the students from the course Fundamentals of Artificial Intelligence for data collection and annotation of dataset data, specifically the following students: Belek Omurzakov, Daniil Astapenko, Jakub Hanuška, Laura Molnárová, Martin Navrátil, Matěj Pour, Jakub Malák, Ondřej Duda, Jakub Ossendorf, Libor Randula, Radek Růžička, Lukáš Kleveta, Martin Weiss, Lucie Suchánková, Filip Kosar, Filip Adámek, Adam Lněnička, Filip Jahn, Josef Pavlíček, Adam Smekal, Tomáš Zatloukal, Martin Kellner, Maxim Kipila, David Vlastník.

Published dataset: https://huggingface.co/datasets/farmaieu/plantorgans (FarmAi u.z., 2024)

REFERENCES

AFONSO, M., FONTEIJN, H., POLDER, G., WEHRENS, R., LENSINK, D., FABER, N. and MOOIJ, M. 2021. *Rob2Pheno Annotated Tomato Image Dataset (Version 3)* [online]. Available at: https://data.4tu.nl/articles/dataset/Rob2Pheno_Annotated_Tomato_Image_Dataset/13173422/3

CHANG, Y.-H. 2020. *Dataset of Tomato Leaves* [dataset online]. https://data.mendeley.com/datasets/ngdgg79rzb/1

CHIU, M. T., XU, X., WEI, Y., HUANG, Z., SCHWING, A., BRUNNER, R., KHACHATRIAN, H., KARAPETYAN, H., DOZIER, I., ROSE, G., WILSON, D., TUDOR, A., HOVAKIMYAN, N., HUANG, T. S. and SHI, H. 2020. *Agriculture-Vision: A Large Aerial Image Database for Agricultural Pattern Analysis (Version 2)* [online]. https://arxiv.org/abs/2001.01306

DENG, J., DONG, W., SOCHER, R., LI, L.-J., KAI, L. and LI, F.-F. 2009. ImageNet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition* [online]. https://ieeexplore.ieee.org/document/5206848

FARMAI, U. Z. 2024. plantorgans (Version 6bdf588) [dataset]. *Hugging Face*. https://doi.org/10.57967/HF/2023

FENU, G. and MALLOCI, F. M. 2021. DiaMOS Plant: A Dataset for Diagnosis and Monitoring Plant Disease. *Agronomy*, 11(11), p. 2107. MDPI AG. https://www.mdpi.com/2073-4395/11/11/2107

GAJJAR, V. K., NAMBISAN, A. K. and KOSBAR, K. L. 2022. Plant Identification in a Combined-Imbalanced Leaf Dataset. In: I*EEE Access*. Vol. 10, pp. 37882–37891. Institute of Electrical and Electronics Engineers (IEEE). https://ieeexplore.ieee.org/document/9751066

GRINBLAT, G. L., UZAL, L. C., LARESE, M. G. and GRANITTO, P. M. 2016. Deep learning for plant identification using vein morphological patterns. Computers and Electronics in Agriculture, 127, pp. 418–424. *Elsevier BV* [online]. https://www.sciencedirect.com/science/article/pii/S0168169916302242

GUO, P., XUE, Z., LONG, L. R. and ANTANI, S. 2020. Cross-Dataset Evaluation of Deep Learning Networks for Uterine Cervix Segmentation. *Diagnostics*, 10(1), p. 44. MDPI AG [online]. https://www.mdpi.com/2075-4418/10/1/44

HUGHES, D. P. and SALATHE, M. 2015. An open access repository of images on plant health to enable the development of mobile disease diagnostics (Version 2). *arXiv* [online]. https://arxiv.org/abs/1511.08060

JIA, W., XU, Y., LU, Y., YIN, X., PAN, N., JIANG, R. and GE, X. 2023. An accurate green fruits detection method based on optimized YOLOX-m. *Frontiers in Plant Science*, 14. Frontiers Media SA [online]. https://www.frontiersin.org/articles/10.3389/fpls.2023.1187734/full

ÖZTÜRK, B. and ÖZKAR, M. 2022. Creating and Using Mask Images for Segmentation in Point Cloud Data. In: The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences: Vol. XLIII-B2-2022 (pp. 1133–1138). *Copernicus GmbH* [online]. https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLIII-B2-2022/1133/2022/

PAK, M. and KIM, S. 2017. A review of deep learning in image recognition. In: 2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT). *IEEE* [online]. https://ieeexplore.ieee.org/document/8320684

SAPUKHINA, N., BOUREAU, T. and ROUSSEAU, D. 2022. Plant disease symptom segmentation in chlorophyll fluorescence imaging with a synthetic dataset. *Frontiers in Plant Science*, Vol. 13. Frontiers Media SA [online]. https://www.frontiersin.org/articles/10.3389/fpls.2022.969205/full

STASTNY, J. and SKORPIL, V. 2007. Analysis of Algorithms for Radial Basis Function Neural Network. In: *Personal Wireless Communications*. Springer New York, vol. 245, pp. 54-62, ISSN 1571-5736, ISBN 978-0-387-74158-1, WOS:000250717300005

STASTNY, J., SKORPIL. V., BALOGH, Z. and KLEIN, R. 2021. Job shop scheduling problem optimization by means of graph-based algorithm. Applied Sciences, 11(4), 1921. ISSN 2076-3417. https://doi.org/10.3390/app11041921

UKWUOMA, C. C., ZHIGUANG, Q., BIN HEYAT, M. B., ALI, L., ALMASPOOR, Z. and MONDAY, H. N. 2022. Recent Advancements in Fruit Detection and Classification Using Deep Learning Techniques. In: KHAN, D. M. (Ed.). *Mathematical Problems in Engineering*. pp. 1–29. Hindawi Limited [online]. https://www.hindawi.com/journals/mpe/2022/9210947/.

WANG, P., TANG, Y., LUO, F., WANG, L., LI, C., NIU, Q., and LI, H. 2022. Weed25: A deep learning dataset for weed identification. *Frontiers in Plant Science*, 13, Frontiers Media SA [online]. https://www.frontiersin.org/articles/10.3389/fpls.2022.1053329/full

WSPANIALY, P., BROOKS, J. and MOUSSA, M. 2020. An Image Labeling Tool and Agricultural Dataset for Deep Learning (Version 1). *arXiv* [online]. https://arxiv.org/abs/2004.03351

ZHANG, Y., WANG, S., JI, G. and PHILLIPS, P. 2014. Fruit classification using computer vision and feedforward neural network. J*ournal of Food Engineering*, 143, 167–177. Elsevier BV. https://doi.org/10.1016/j.jfoodeng.2014.07.001

Contact information

Miroslav Jaroš: e-mail: qqjaros7@mendelu.cz
Jiří Podivín:e-mail: qqpodivin@mendelu.cz
Petr Pernes: e-mail: qqpernes@mendelu.cz
Oldřich Trenz: e-mail: oldrich.trenz@mendelu.cz