RESEARCH ARTICLE

**WILEY**

# GrowliFlower: An image time-series dataset for GROWth analysis of cauLIFLOWER

Jana Kierdorf[1] | Laura Verena Junker-Frohn[2] | Mike Delaney[3] |
Mariele Donoso Olave[1] | Andreas Burkart[3] | Hannah Jaenicke[4] | Onno Muller[2] |
Uwe Rascher[2] | Ribana Roscher[1]

[1]Institute of Geodesy and Geoinformation, Remote Sensing Group, Bonn, Germany

[2]Institute of Bio- and Geosciences, IBG-2: Plant Sciences, Forschungszentrum Jülich GmbH, Jülich, Germany

[3]JB Hyperspectral Devices GmbH, Düsseldorf, Germany

[4]Horticulture Competence Centre, University of Bonn, Bonn, Germany

**Correspondence**
Jana Kierdorf, Institute of Geodesy and Geoinformation, Remote Sensing Group, University of Bonn, Niebuhrstraße 1A, 53113, Bonn, Germany.
Email: jkierdorf@uni-bonn.de

**Abstract**
In this paper, we present GrowliFlower, a georeferenced, image-based unmanned aerial vehicle time-series dataset of two monitored cauliflower fields (0.39 and 0.60 ha) acquired in 2 years, 2020 and 2021. The proposed dataset contains RGB and multispectral orthophotos with coordinates of approximately 14,000 individual cauliflower plants. The coordinates enable the extraction of complete and incomplete time-series of image patches showing individual plants. The dataset contains the collected phenotypic traits of 740 plants, including the developmental stage and plant and cauliflower size. The harvestable product is completely covered by leaves, thus, plant IDs and coordinates are provided to extract image pairs of plants pre- and post-defoliation. In addition, to facilitate classification, detection, segmentation, instance segmentation, and other similar computer vision tasks, the proposed dataset contains pixel-accurate leaf and plant instance segmentations, as well as stem annotations. The proposed dataset was created to facilitate the development and evaluation of various machine-learning approaches. It focuses on the analysis of growth and development of cauliflower and the derivation of phenotypic traits to advance automation in agriculture. Two baseline results of instance segmentation tasks at the plant and leaf level based on labeled instance segmentation data are presented. The complete GrowliFlower dataset is publicly available (http://rs.ipb.uni-bonn.de/data/growliflower/).

**KEYWORDS**
agricultural plant dataset, crop development, crop growth, instance segmentation, machine learning, plant monitoring, UAV

## 1 | INTRODUCTION

Field-grown crops are strongly affected by environmental conditions, thus, to minimize yield losses due to abiotic or biotic stresses, crop production involves careful plant management and complex decisions. Farmers support plant growth and development through irrigation, fertilization, weeding, and pesticides applications; however, these are costly and labor-intensive processes. To optimize plant management and support effective decision-making, farmers rely on frequent crop monitoring; however, this is also time-consuming

process that requires expert knowledge. Typically, farmers and agricultural advisors monitor fields regularly via spot checks of individual plants. Here, remote sensing and analysis methods can help farmers monitor entire fields more comprehensively (Chi et al., 2016; Weiss et al., 2020), and remote sensing data can be acquired at any scale without damaging or impacting the crops. Large-scale observations from satellites or aircraft and medium-scale observations from unmanned aerial vehicles (UAVs) provide an overview of larger agricultural areas (Lillesand et al., 2015). Large-area sensor-based crop monitoring makes it possible to detect heterogeneity in the field and support the farmer's decision-making in terms of field management. With such detailed, area-wide information on biotic and abiotic stress, these factors can be counteracted more selectively to support environmentally friendly plant management. Medium-scale and close-range observations acquired from UAVs and ground robots are beneficial for collecting detailed information and can be used particularly well for phenotyping individual plants. For example, Nock et al. (2016) used optical remote sensing data to define various traits, for example, structural and phenotypical characteristics at all levels, from individual plants to large areas. Other applications using remote sensing data include yield estimation (Chaparro et al., 2018), yield forecasting (Mosleh et al., 2015), and monitoring rapid land surface changes (Verger et al., 2014).

Machine learning (ML) methods have become increasingly important (Lary et al., 2016) in processing and interpreting large amounts of remote sensing data. ML involves learning a predictive function that relates observations to the desired output, and trained models can be designed flexibly relative to the type of observations (Debolini et al., 2015; Reichstein et al., 2019). For example, using ML techniques, plant traits can be identified using remote sensing data (Ali et al., 2015; Verrelst, 2019). A main area of application is plant phenotyping, which can be made more objective and automated using advanced ML methods, for example, deep neural networks. For example, Romera-Paredes and Torr (2016), Ren and Zemel (2017), and Scharr et al. (2016) trained ML models to infer various phenotypic traits, for example, the number of leaves per plant. Similar traits can also be derived using a combination of object and leaf keypoint detection, which facilitates observation of plant growth as done by Weyler et al. (2021). Sa et al. (2016) employed deep convolutional neural networks to detect single fruits, which served as a precursor for subsequent autonomous harvesting (Arad et al., 2020). Drees et al. (2021) used time-series image data of cauliflower and broccoli to predict field growth using conditional generative adversarial networks (Isola et al., 2017). They generated an image of a plant at a later time point and employed the Mask Region-based Convolutional Neural Network (R-CNN; He et al., 2017) to calculate the projected leaf area. Another typical agricultural application is field weed control, where weeds, crops, and soil must be distinguished accurately. Using neural networks, promising results have already been achieved, where the task can be approached using classification (Lottes et al., 2017), detection (Lottes et al., 2018), or semantic segmentation (Ahmadi et al., 2021; Milioto et al., 2018) techniques.

Benchmark datasets with annotations and in-situ measurements are beneficial in terms of facilitating the development of ML methods for plant-specific tasks using remote sensing data. Various benchmark datasets already exist; however, many of these datasets are domain-specific with highly specific objects, for example, buildings (Roscher et al., 2020) and animals (Deng et al., 2009) or other semantics, for example, land cover (Cordts et al., 2015). Generally, such datasets are not suitable for plant applications. The link between ML and plant sciences is becoming increasingly important (Lary et al., 2016), as can be seen from the increasing number of related publications in recent years (Ahmadi et al., 2021; Chebrolu et al., 2017; Förster et al., 2019; Halstead et al., 2020; Kierdorf et al., 2019; Zabawa et al., 2019). Despite increased demand, to the best of our knowledge, only a few publicly available plant-specific datasets are available for ML purposes.

Among the limited number of publicly available datasets or datasets described in the literature, many were acquired in a greenhouse environment (Halstead et al., 2020; Minervini et al., 2016; Mureşan & Oltean, 2018; Scharr et al., 2014) or are based on synthetically generated data (Kierdorf et al., 2022; Ward & Moghadam, 2018), which makes it difficult to apply them to real-world scenarios. In particular, the greenhouse-grown plant *Arabidopsis thaliana* rosettes is frequently used in ML research due to its simple rosette morphology (Scharr et al., 2014). However, the morphologies of agricultural crop plants are more diverse, and their development is affected by changing environmental conditions and both abiotic and biotic stresses. Thus, agricultural datasets that represent real-world field conditions that also cover various challenges, for example, occlusion, shape variability, pose variability, the colors of plants, and plant parts, are required, such as the datasets of Kusumam et al. (2017) and Blok et al. (2021).

Modeling the temporal development of plant growth and plant traits is an active research area, and this requires datasets that monitor plants over time; however, publicly available time-series datasets of plants are rare. One such dataset is the cauliflower (*Brassica oleracea* var. *botrytis*) and broccoli (*Brassica oleracea* var. *italica*) dataset from Bender et al. (2020). The data in this dataset were acquired using a camera-equipped robot that captured close-range images at several time points. However, this dataset is limited to only a few plants and lacks semantic information and accurate georeferencing of single plants.

Cauliflower is a suitable target crop plant to develop ML algorithms because its cultivation, morphology, and economic value give rise to many potential applications in the agriculture digitization context. Cauliflower is a high-value crop that must satisfy various quality criteria. Thus, precise timing of plant management procedures is required to avoid yield losses due to abiotic or biotic stress and produce marketable cauliflowers. Cauliflower harvesting is labor-intensive because each cauliflower must be harvested within approximately 1 week period in which the heads are of sufficient size but are not yet overripe. In addition, cauliflower must be harvested by hand due to within-field variability in plant development. As the head is covered by leaves, each individual cauliflower

head must be touched to determine whether it satisfies the size criteria. After cutting and removing the surrounding leaves, product quality is assessed visually to dismiss heads with discolorations, misshapes, or stress symptoms. Note that cauliflower growth is highly dependent on climate, which makes it difficult to predict the most beneficial harvest time. Depending on the prevailing temperature, irradiance, and soil water availability, plants may develop rather heterogeneously, thus, harvesting of established fields simultaneously can take weeks. Under favorable conditions, plants in sequentially established fields may need to be harvested at the same time, which requires more workers and lowers the price per cauliflower. Early prediction of harvestable plants and harvest time would facilitate better sales planning and provide significant economic advantages to farmers.

Thus, in this paper, we present an agricultural dataset, introduced as GrowliFlower, that is suitable for the development of ML approaches. The proposed dataset is intended to address the growth analysis and development of crop plants and the derivation of phenotypic traits relevant for agricultural applications to promote the development of automation in agriculture. The proposed dataset comprises the following.

- RGB and multispectral orthophotos of two different cauliflower fields were acquired over the entire growing period (from planting to harvest time).
- Plant IDs and coordinates, which enables users to extract complete and incomplete time-series of image patches showing individual plants accompanied with insitu reference data captured manually on the field.
- The plant IDs and coordinates also allow users to extract image pairs of plants pre- and post-defoliation accompanied with a time-series of the respective plant to facilitate analysis of the correlation between the external appearance and internal head of the cauliflower plant.
- The proposed dataset's pixel-accurate labeled data are useful for plant and leaf classification, detection, segmentation, instance segmentation, and other similar computer vision tasks.

We also present two baselines demonstrated application examples of plant and leaf instance segmentation using the proposed dataset in a Mask R-CNN (He et al., 2017) application.

## 2 | FIELD DESIGN

Here, we describe the study area. The cauliflower fields used for data acquisition in this study were located on a farm in Western Germany (50°46′6.742″N, 6°58′20.271″O) close to the city of Bornheim, which is 20 km south of Cologne (Figure 1). The mean annual temperature in Bornheim is 14°C, and the mean annual precipitation is 383 mm. This area is dry 142 days a year with an average humidity of 81%. Note that fertile loess soil is available on the farm.

We acquired data for two fields, that is, (1) the field shown in blue in Figure 1 (referred to as field 1 in this paper) in 2020, and (2) the field shown in orange (referred to as field 2 in this paper) in 2021. Note that the cauliflower plants in both fields were planted in rows in a northwest-to-southeast orientation. These fields were designed for sprayers with a working width of 18 m. Before planting, the fields were plowed to prepare the soil. Tractors with 1.8 m track width were used to plant five rows of nursery-grown young cauliflower plants simultaneously, with three rows between the tractor tracks. The distance between the rows was 0.6 m, and
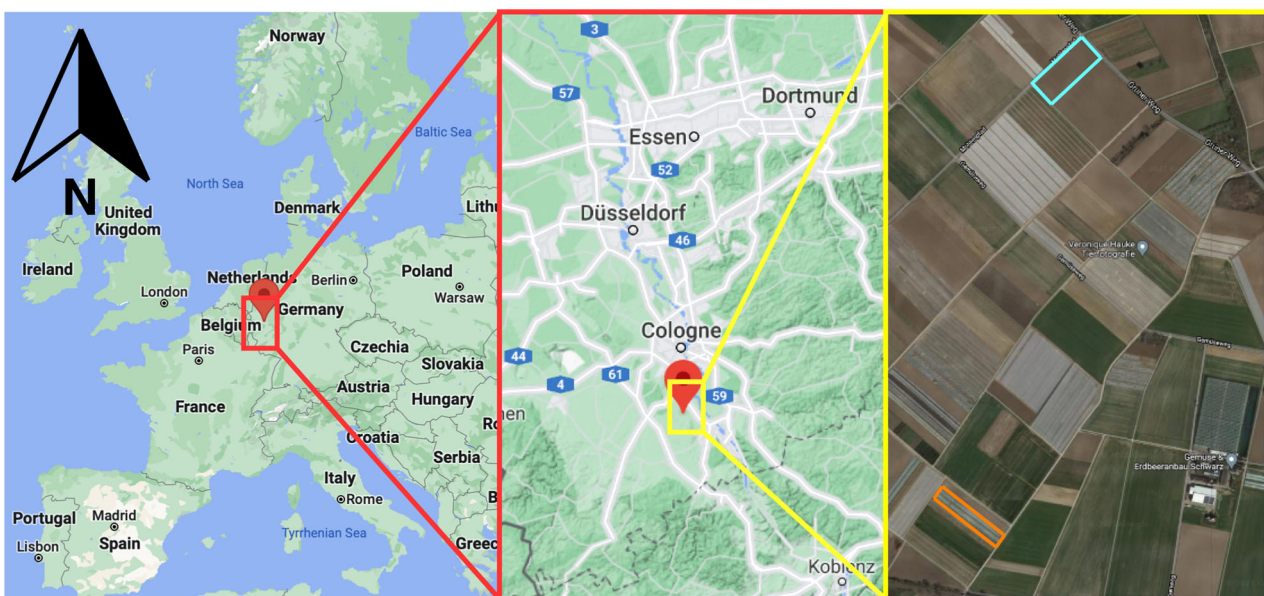


FIGURE 1 Field locations. The fields are located near Cologne, Germany. Blue: field 1 (2020); orange: field 2 (2021). *Map source*: Google Maps.

the distance between the plants in a row was 0.5 m, thereby resulting in a planting density of 33,000 plants/ha. In addition, every 18 m, there was a 2-m wide lane for spraying and irrigation. The fields were subject to conventional farming practices, including hoeing cauliflower plants before canopy closure to reduce weeds and application of pesticides (including herbicides, insecticides, and fungicides). The fields were also irrigated as required using sprinklers. As a result, the abiotic and biotic stresses were rather low in both fields, and the plants developed rather uniformly.

## 2.1 | Field 1

Field 1 has a width of approximately 100 m and a length of 240 m. Thus, the total area of field 1 is approximately 2.4 ha. This field was planted with the Korlanu cultivar (Syngenta). Three-quarters of the field were planted using plants from seedling trays (Figure 2a) on July 28, 2020 from the southwest direction. The remaining northeastern part of the field was planted on July 29, 2020. Note that field 1 was generally free of weeds.

## 2.2 | Field 2

Field 2 field has a width of approximately 55 m and a length of 210 m. Thus, the area is approximately 1.32 ha. This field was planted with the Guideline cultivar (Syngenta). Here, the plants were transplanted from seedling trays on June 15, 2021. Note that field 2 contains more weeds than field 1, especially along the southwestern edge of the field due to previous rhubarb cultivation.

## 3 | DATA COLLECTION

Three types of data were acquired in the data collection process, namely:

1. RGB and multispectral UAV image data with high spatial resolution, which is an indirect measurement of the phenotypic development of the plants.

2. Georeferenced ground control points (GCP) to locate the data in space, spatially arranged according to field size to ensure accurate and robust processing of the orthophotos (Persia et al., 2020).

3. In-situ measurements of phenotypic traits characterizing the development state and stress factors that serve as reference observations.

The different types of data were collected on the same day to synchronize them. However, to ensure that workers were not visible in the image data, data acquisition processes were not conducted at the same time. Data acquisition was conducted once a week during the entire growth period. During the harvest period, data were collected once between two different harvest days and once after the final harvest. Note that drone flights were only performed on sunny or overcast days to ensure stable illumination for the generation of orthophotos without shading effects due to moving clouds. As a result, the time intervals between successive flights vary. Figure 3 shows the data collection dates for fields 1 and 2. As seen in the top timeline, seven orthophotos are only partly available, which is discussed further in Section 4.1. The data collection took a few hours per day, with the in-situ measurements being the most time-intensive. In addition, data collection was adjusted to both field conditions, resulting in adaptations to camera settings, number of GCPs, and flight altitude. In the following subsections, we describe the procedure followed for fields 1 and 2.

## 3.1 | RGB and multispectral imaging using UAVs

UAV images were captured using a DJI Matrice 600 hexacopter with two mounted cameras (Figure 4). The first camera was a Sony A7 rIII RGB camera with a Zeiss/Batis 2.0 lens (resolution: 47.4 MP). The focal length was 25 mm with a field of view of 71. 5°. A shutter speed of 1/1250th and a floating aperture (highest value: 2.0) were selected. The International Organization for Standardization value was set to automated for field 1 and changed to 50 for field 2 to align our approach with the image-capture settings recommended by Agisoft. The second camera was a MicaSense RedEdge 3 for multispectral image data. It contains five built-in lenses (resolution: 1.2 MP per band). The wavelengths of the five acquired bands and their respective bandwidth were 475 nm (20 nm), 560 nm (20 nm),
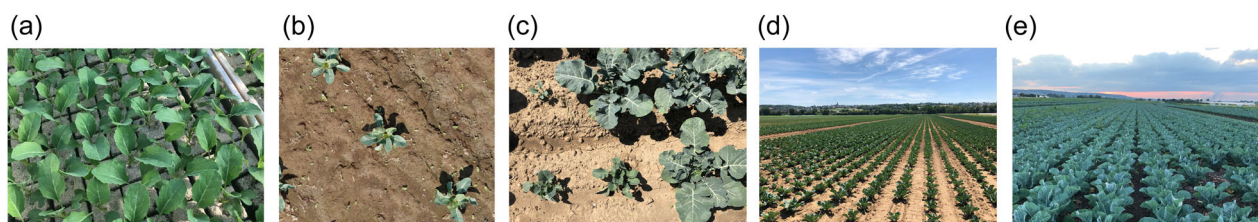


**FIGURE 2** Example field and plant images. (a) Seedling trays before planting. (b) Plants 2 weeks after planting. Images (c, d) were taken 4 weeks after planting and illustrate how different plants develop over. (e) Plants shortly before head formation.
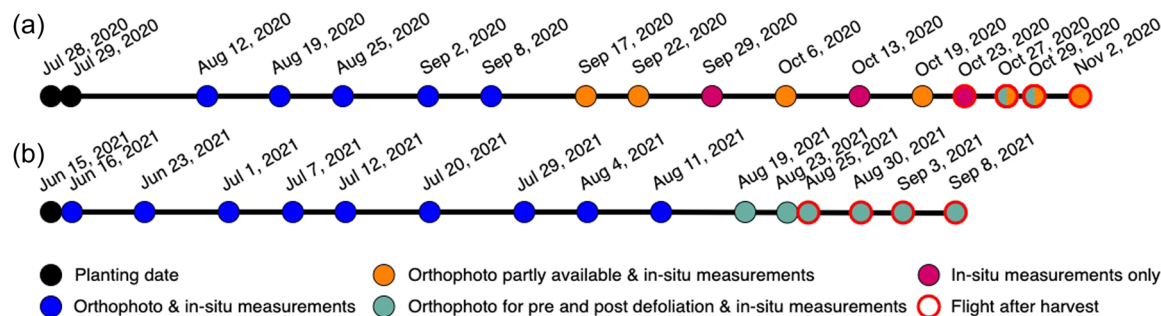
FIGURE 3   Timelines of acquired data for (a) field 1 and (b) field 2. The colors represent the data availability for images and insitu measurements.
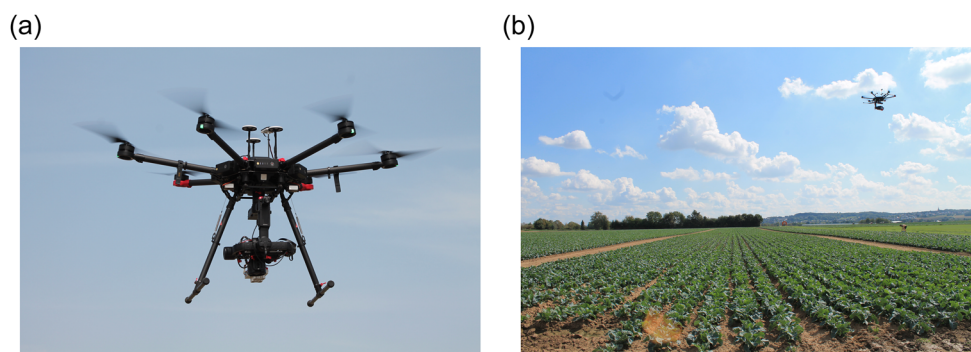


FIGURE 4   DJI Matrice 600 hexacopter for unmanned aerial vehicle image-based measurements.

668 nm (10 nm), 717 nm (10 nm), and 840 nm (40 nm). The focal length of the camera is 5.4 mm. For field 1, an altitude of approximately 10 m and an image overlap of 60/80 were used, and for field 2, an altitude of approximately 16 m and an image overlap of 80/80 were used to optimize the data acquisition process and subsequent image data processing. The following factors were considered in terms of the drone flights. For each flight, no irrigation was permitted in or close to the flight area, the drone was flown at temperatures and wind speeds within the device's safe operating range, and the flights only occurred during periods of no rain.

### 3.1.1 | Time-series flights

On each acquisition date, the drone was flown over a specified area of the field once, which remained the same for the entire growing period. For field 1, this area had a width of 91 m and a length of 62 m, resulting in approximately 0.60 ha. For field 2, the area had a width of 30 m and a length of 131 m, resulting in approximately 0.39 ha.

The cauliflower plant does not necessarily grow straight, thus, the center of the plant in later growing stages does not match the position of the seedling exactly (Grenzdörffer, 2019). Thus, a shift of up to ±10 cm between the center position of the head and the stem position in the early growing stages was observed.

### 3.1.2 | Defoliation flights

In addition to the time-series flights, so-called defoliation flights were conducted. Here, the upper leaf layers covering the cauliflower head were removed manually on individual plants after the time-series flight. This step is referred to as defoliation. Note that we ensured that the defoliated leaves did not affect any neighboring plants. The defoliated plants provided information about the development of the head relative to the plant's outer appearance. By performing another UAV flight after defoliation, a dataset of images showing the time-series of the plant's outer appearance (Figure 5b) and inner head (Figure 5c) on the day of defoliation was acquired.

For field 1, the defoliation of plants was performed over 2 days, that is, October 27 and 29, after harvesting occurred. Thus, the defoliated plants represented plants whose head size did not satisfy the harvest criteria, which generally meant that the head was too small. For field 2, starting on August 19, when most of the cauliflower heads started developing, between 70 and 200 plants, were defoliated weekly. Here, all plants with developed heads were defoliated in rectangular plot regions to minimize the impact of defoliation on the biological growth of neighboring plants. Note that care was taken to not defoliate the reference plants described previously (Section 3.3). A distribution of plots for the first five defoliation time points is shown in Figure 5a. For the final flight (after the last harvest), most remaining plants that had not been harvested
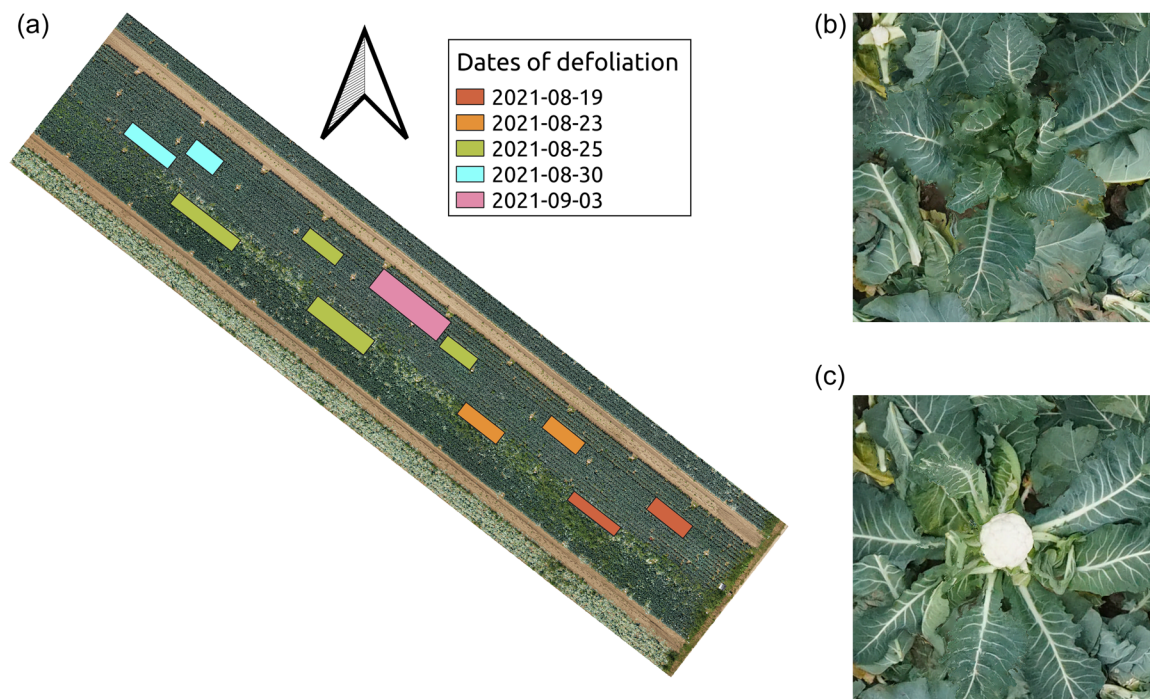
**FIGURE 5** (a) Visual overview of defoliated plant locations for the first 5 weeks of defoliation in field 2. (b, c) Images of a plant pre- and post-defoliation. The locations of random distributed defoliated plants from week six are not shown.

were defoliated, which resulted in random distribution. Thus, this is not shown in Figure 5a.

## 3.2 | GCP points

To localize the image data globally in space, the data were georeferenced with the help of circular 12-bit GCPs with a diameter of approximately 20 cm, as shown in Figure 6. Here, the GCPs were fixed in the ground using plastic pegs, and they were distributed evenly across the field (refer to Figure A1) and positioned on tractor tracks or between plants to avoid displacement by external influences, for example, plowing. In addition, surrounding plants were removed as required to ensure the visibility of GCPs in the image data. We used 21 GCPs in field 1 (35 GCPs/ha) and 44 GCPs in field 2 (113 GCPs/ha) (refer to Figure A1), with each GCP showing a different pattern. The greater number of GCPs in field 2 was due to the fact that they facilitate subsequent image alignment by ensuring that at least three GCPs were present in each captured image, especially for growth stages with a high degree of plant overlap and dense canopies.

As a measuring device for GCP coordiantes, a Trimble R4-Model 3 base station with a horizontal standard deviation of ±5 mm + 0.5 ppm RMS and vertical standard deviation of ±5 mm + 1 ppm RMS was used for both fields. The measured coordinates were acquired in the WGS84/UTM 32N coordinate system. To ensure that the markers for the GCPs were not displaced due to external influences, the GCPs
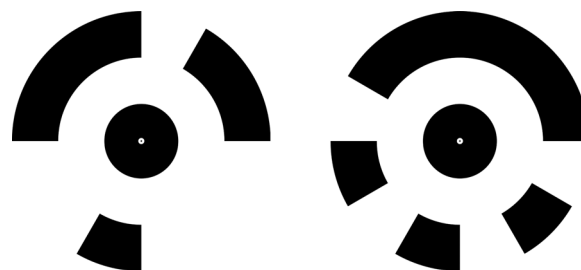


**FIGURE 6** Two ground control point patterns used for acquisition.

were measured at the beginning and end of the data acquisition period to omit displaced GCPs. A third measurement was added for field 2 in the middle of the growing period.

## 3.3 | In-situ measurements of plant development

In each field, so-called reference plots were selected to capture information from reference plants manually. For field 1, four reference plots were assigned (Figure A2a), and each plot comprised three rows with 20 plants each (Figure A2b). Thus, each plot contained 60 plants, for a total of 240 plants in all reference plots. The plots were distributed in the northwestern half of the field along the long side. Five reference plots were assigned for field 2 (Figure 7a). Here, each plot comprised five rows of 20 plants
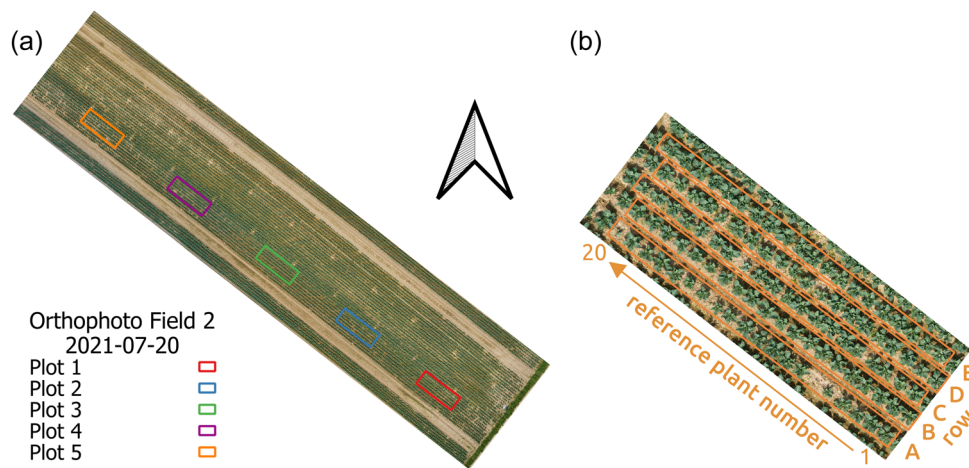
**FIGURE 7**  Visual overview of (a) reference plots for in-situ measurements in field 2 and (b) the design of reference plot 5 (including reference plants and the ordering of reference plant numbers). The plot design is valid for all reference plots in field 2.

(Figure 7c; 100 plants/plot, 500 plants in total). The plots were distributed evenly in the southwestern half of the field along the long side. Thus, the reference data were collected along the entire field. Each reference plant was assigned a specific plant ID identifying the row (field 1: A–C; field 2: A–E) and plant number (Field 1: 1–10, 90–99; Field 2: 1–20).

The following measurements were taken for all reference plants in field 1.

1. Phenological development after BBCH-scale (Biologische Bundes-sanstalt für Land- und Forstwirtschaft, Bundessortenamt und CHemische Industrie)/Feller et al. (1995),
2. height,
3. maximum diameter,
4. other remarks, for example, stress infestation (listed in the attachment in Table A1),
5. head diameter,
6. harvesting status.

Note that the farmer followed a rigorous plant protection schedule, and very few stresses were detected in 2020; thus, information about stresses was not recorded explicitly in 2021. Due to the observed homogeneous development, focus was placed on measurements of BBCH and the height of five representative plants per plot. Here, the head diameter and harvest status were recorded for individual plants.

## 4 | DATASET

The core component of the dataset (Figure 8) comprises both RGB and multispectral orthophotos derived from the captured UAV images. In the orthophotos, single plants are identifiable by their corresponding coordinates and plant IDs. The dataset contains four subsets intended for different ML tasks. The instance

segmentation GrowliFlowerL subset contains patches extracted and processed from the RGB orthophotos, and the remaining three subsets contain time-series data of individual plants. The Growli-FlowerT subset comprises randomly selected time-series data representing a wide variety of cauliflower development. In addition to the time-series data, the GrowliFlowerD subset also contains image pairs of plants before and after defoliation. The GrowliFlowerR subset contains the insitu measurements and the time-series data. For each field, a text file containing the measured GCP coordinates at the beginning and the end of field monitoring is provided. For field 2, the coordinates measured during the growing period are also given.

### 4.1 | Orthophotos (GrowliFlowerO and GrowliFlowerM)

The acquired RGB and multispectral UAV images were aligned to orthophotos using the Agisoft Metashape Professional software to obtain a large-scale overview of the monitored fields. Here, the orthophotos were georeferenced according to the measured GCP coordinates. In addition, the individual orthophotos were exported in the WGS84/UTM 32 coordinate system.

The ground resolution for the RGB orthophotos of field 1 is 1.65 mm/px for the pixel width and height with a minimum and maximum file size of 1.64 and 6.7 GB, respectively. The ground resolution for field 2 is 3.10 mm/px for the pixel width and height with a minimum and maximum file size of 1.3 and 5.0 GB, respectively. Twelve orthophotos are available for field 1, where five are entirely processed, and seven contain data gaps for small areas where the quality of the UAV-acquired images was insufficient. For field 2, 15 orthophotos are available, as shown in Figure 3b. This set of orthophotos is provided in the GrowliFlowerO subset of the proposed dataset. In addition, the
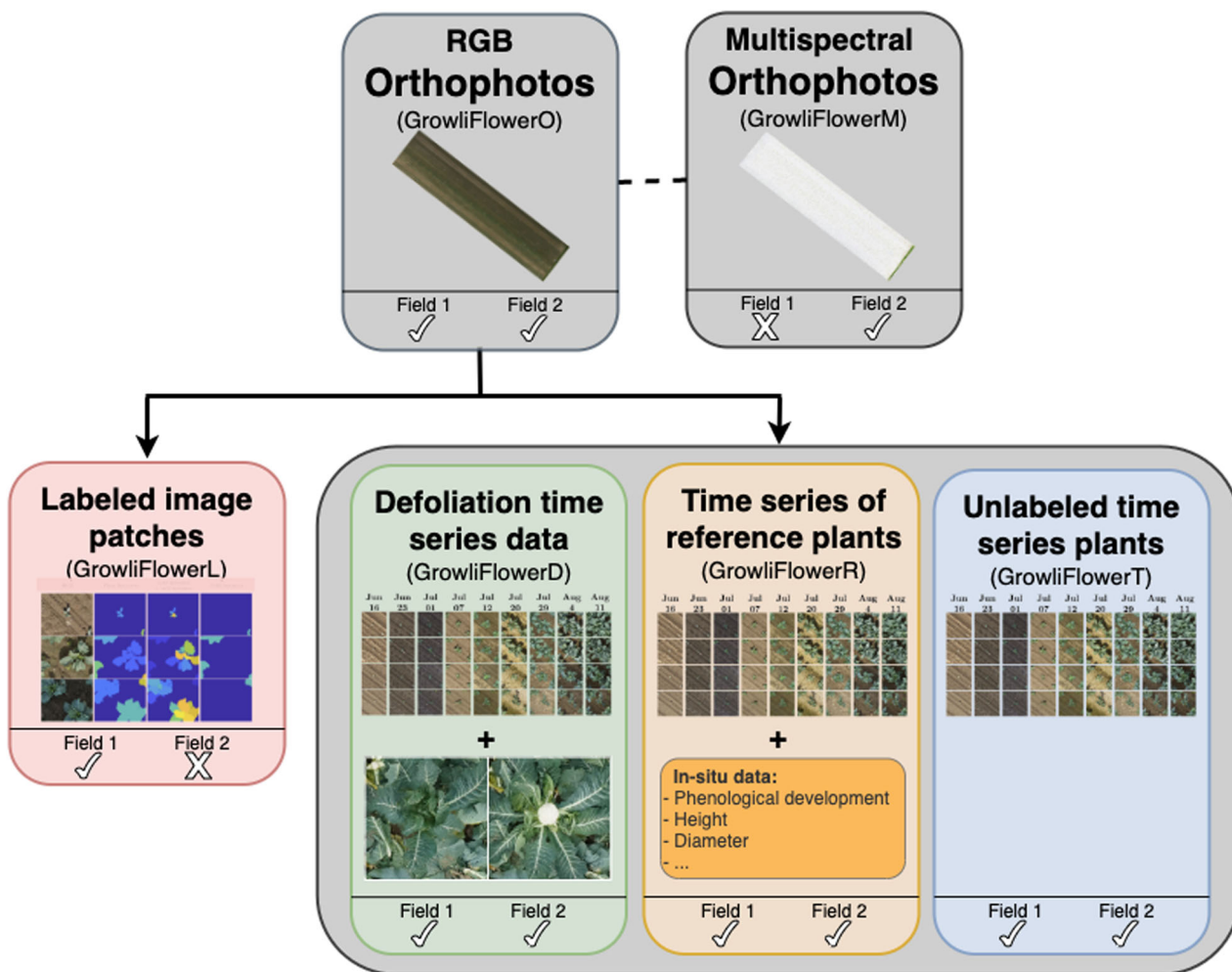
**FIGURE 8** Overview of data in proposed GrowliFlower dataset.

dataset contains multispectral orthophotos for field 2 with a ground resolution of 2.5 cm/px width and length, denoted as the GrowliFlowerM subset.

## 4.2 | RGB image patches

In this section, we describe the data extracted from the RGB orthophotos. Note that the ground resolution of the resulting image patches is the same as that of the respective orthophotos.

Each of the following datasets (excluding the labeled dataset described in Section 4.2.1) contains a text file with global information for each field, containing the image ID, including the plant ID, and corresponding georeferenced UTM coordinates of the plants. Note that the coordinates identify the center of the plants as observed on August 19 for field 1 and July 7 for field 2. In addition, information about the planting day and a proposed assignment as a training, validation, or testing subset are provided as a basis to compare ML methods. To minimize spatial correlation

between sets, the proposed training, validation, and testing subsets are spatially disjoint. However, certain systematic factors from a biological perspective are not excluded. The use of these sets is expected to promote the development of ML methods with high generalizability. For the reference data discussed in Section 4.2.3, the harvesting time is specified, and for the defoliation data discussed in Section 4.2.4, the defoliation date of the plants is specified. In addition, text files with local information for each acquisition date are provided, including the image ID to connect the local information with the global information, and the corresponding local pixel coordinate relative to the respective orthophoto for each data acquisition day. Also note that information about the day after planting (dap) is included.

To use image patches showing single plants, the patches must be extracted from the orthophotos using the plant IDs and coordinates. Here, an image side length and width of at least 490 px for field 1 and at least 256 px for field 2 is recommended to ensure that the entire plant is captured in the image patch regardless of the plant developmental stage.

## 4.2.1 | Labeled image patches (GrowliFlowerL)

This subset, called GrowliFlowerL, comprises pixelwise, manually annotated images, thus, it is well-suited for classification, semantic segmentation, detection, instance segmentation, or stem detection tasks. For this subset, the image patches of four acquisition dates for field 1 are extracted using a sliding window approach. The image patches have a size of 368px × 448px. Here, the size of the patches differs from that of the proposed sizes because only plants from earlier development stages are included. In addition, in this dataset, the focus is not on individual plants but on the variability between images, thus, the plants are not located in the center of the patch.

For each RGB image patch, four annotated masks are provided. These annotated masks contain segmentations of (1) plant instances, (2) leaf instances, (3) void segmentations, and (4) stem positions.

(1) The plant instance mask segments the image in soil and plant pixels with instance information for the plants.
(2) The leaf instance mask segments the plants into single leaves. Note that plants at image borders for which no stem or only one-quarter of the plant is visible are annotated as void and no leaf annotation is applied.
(3) The void segmentation mask is a binary mask where plants located at image borders where no stem is visible are segmented as void. In addition, plants with only a small amount of visible leaf material in the RGB image are also segmented as void.
(4) The stem annotation mask represents the position of the stems of nonvoid plants.

Examples of (1) plant instance masks, (2) leaf instance masks, and (3) void segmentation masks are shown in Figure 9. Two things to note are that weed is not labeled as a plant but as a background and that stem positions are only represented by individual pixels, thus, they are difficult to recognize visually. Therefore, masks that include stem information are not shown in these examples. The annotations are provided with a defined name based on the name of the RGB image patch. Here, each patch contains a maximum of four plants, and there is a number of patches in the dataset that contain no plants (Table 1). This subset is divided into training, validation, and testing sets, and the complete labeled subset is denoted GrowliFlowerL.
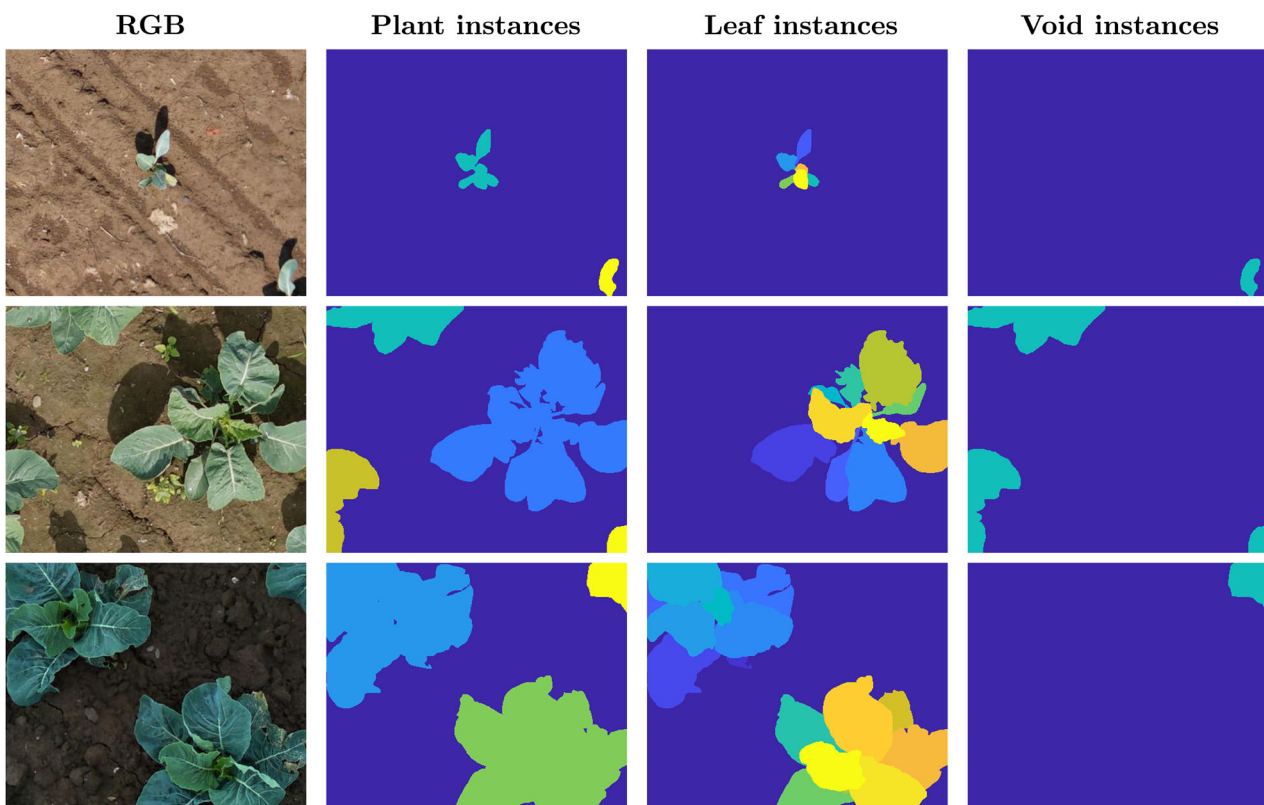


**FIGURE 9** Examples of labeled images for different time points. Column 1 shows that the RGB base for columns 2–4 illustrates corresponding labeled plant instance masks, leaf instance masks, and void segmentation masks. The rows represent different points in time. Dark blue represents the background class, and the other colors represent different (leaf) instances.

## 4.2.2 | Time-series for plant data (GrowliFlowerT)

For each field, the plant coordinates are provided to allow users to extract time-series plant images. This data is denoted GrowliFlowerT. The time-series data of field 1 comprise the early plant developmental stages and the harvest dates, but lack dates, when the canopy around the cauliflower head was closed. The time-series data of field 2 comprise all growth stages.

For field 1, the coordinates for approximately one-third of the plants are determined (3804 plants in total). The distribution of the location of the extracted data is visualized in Figure A3a. The selected plants are distributed along the southeastern edge of the field due to the availability of data for most time points and the ability to determine the harvest window of individual plants. The subset is divided into training, validation, and testing sets, as shown in Figure A3a. In addition, cauliflower planted on July 28 or July 29 is included in all three sets to ensure that the variability within the

**TABLE 1** Overview of distribution of labeled images acquired on different dates.

| Definition | All images | Images with plants [Train/Val/Test] | Images without plants [Train/Val/Test] |
|---|---|---|---|
| 2020_08_12 | 844 | 745 [521/110/112] | 99 [71/15/15] |
| 2020_08_19 | 892 | 781 [547/117/117] | 111 [78/16/17] |
| 2020_08_25 | 383 | 367 [257/55/55] | 16 [12/2/2] |
| 2020_09_08 | 79 | 79 [56/11/12] | 0 [0/0/0] |

sets is guaranteed. Note that the orthophotos do not overlap entirely; thus, image data are not available for all plants at all times, which results in temporally incomplete time-series data. For field 2, 8736 plant coordinates were extracted and distributed evenly over the field. The subset is divided into training, validation, and testing sets, as shown in Figure A3b. Here, all plant coordinates are provided as georeferenced UTM coordinates.

To use individual plant images, the user must crop the patches around the local plant coordinates determined in the subset. In addition to all global plant coordinates, this subset contains the local coordinates of the patches for each acquisition date, which at a size of 490 px × 490 px for field 1 and 256 px × 256 px for field 2 lie completely within the orthophoto and are not showing spatial data gaps, as patches shown in Figure 12b. Five examples of the time-series data are shown in Figure 10 for field 1, and four examples are shown in Figure 11 for field 2. Due to the spatial data gaps, the number of coordinates per date for field 1 varies, which leads to temporal gaps in the time-series data. The largest set of time-series that includes equal time steps consists of 3611 time-series based on eight time points, including the five time points up to day after planting 42 (September 8), and all three time points from day after planting 91 (October 27). In addition to the file that contains all UTM coordinates, a text file containing the UTM coordinates for this set is also provided; thus, the user can extract the time-series data for the selected plant IDs. After removing the patches with spatial data gaps, we retained 8348 complete time-series images for field 2. Due to the heterogeneous weed occurrence in field 2, the patches contain different amounts of weed, as shown in Figure 12a. Due to the given UTM coordinates, it is possible to extract the complete time-series set of local coordinates for both fields if required.
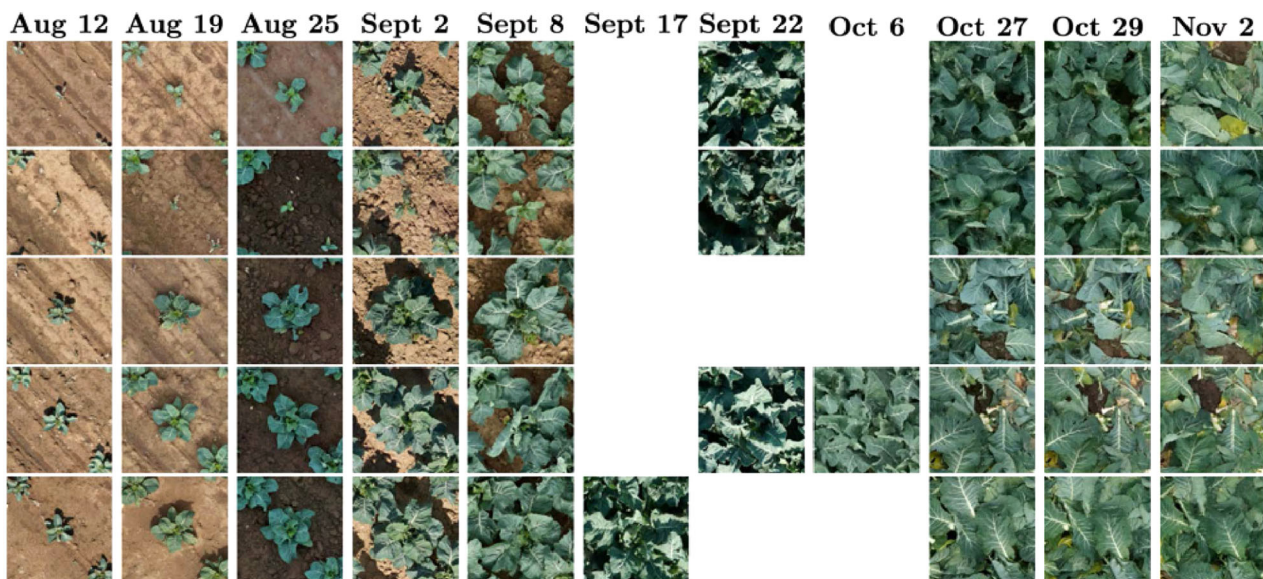


**FIGURE 10** Time-series illustration of five different plants in GrowliFlowerT subset in field 1. All rows represent time-series of plants containing temporal data gaps due to the poor image quality of the corresponding UAV images (indicated by omitted images). The columns represent the recording days and show the five representative plants captured at the same time on that day. UAV, unmanned aerial vehicle.
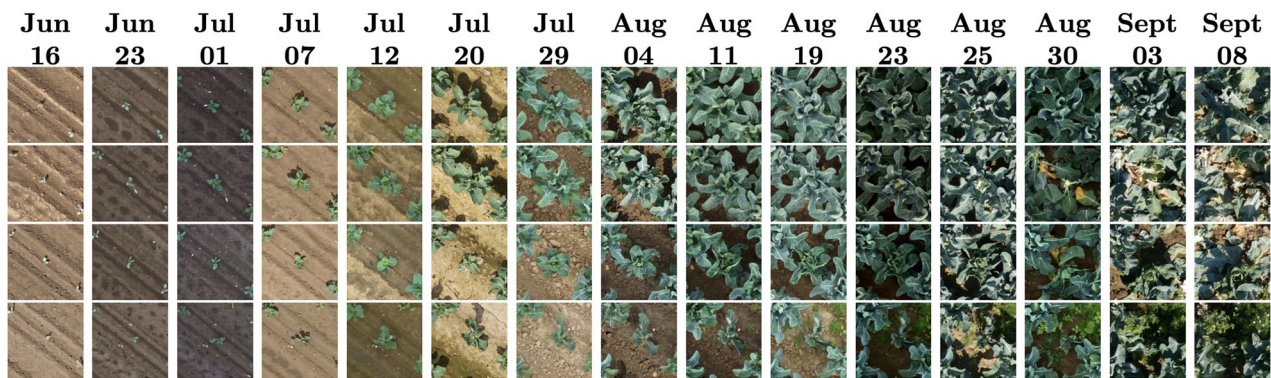
**FIGURE 11**    Four plant time-series from field 2. A row represents a time-series. The columns represent the acquisition dates.



**FIGURE 12**    Data gaps and different amounts of weed occurrences in image data during different stages of growth. (a) Different amounts of weed occurrence on the acquisition date of August 11 and (b) data gap occurrence.

**TABLE 2**    Number of reference plant image patches per acquisition date for field 1 (2020).

| Date | Aug 12 | Aug 19 | Aug 25 | Sept 2 | Sept 8 | Sept 17 | Sept 22 | Oct 06 | Oct 19 |
|---|---|---|---|---|---|---|---|---|---|
| # images | 239 | 239 | 239 | 239 | 239 | 239 | – | – | 193 |
| Date | | | Oct 27 (Post) | | | Oct 29 (Post) | | | Nov 2 |
| # images | | | 119 | | | 119 | | | 12 |

### 4.2.3 | Time-series for reference plant data (GrowliFlowerR)

For each field, the subset includes the plant IDs and coordinates, which allows the user to extract an image time-series set of monitored reference plants that appear similar to those described in Section 4.2.2. The time-series data for field 1 comprise the early plant developmental stages and the harvest dates, but lack dates, when the canopy was closed. The time-series data for field 2 comprise all growth stages (see Figure 11). Table 2 shows the distribution of available plant IDs and the number of images of plants per time point for field 1. Note that the pre-defoliation orthophotos of October 27 and October 29 do not overlap the reference plots due to the low quality of the corresponding UAV images. Here, the reference plants were not defoliated, thus, the orthophotos of defoliation flights are used to extract images of these days to acquire a reference time-series. For field 2, all local coordinates are given for all acquisition dates, which allows the user to extract complete image time-series. Here, the data are divided into training, validation, and testing set for both fields. In addition, the plants in each plot are

**TABLE 3**    Number of defoliated plants per acquisition date for field 2 (2021).

| Date | Aug 19 | Aug 23 | Aug 25 | Aug 30 | Sept 3 | Sept 8 |
|---|---|---|---|---|---|---|
| # images | 110 | 115 | 251 | 116 | 71 | 54 |

presented in each set. The visual distribution for both fields is shown in Figure A4.

### 4.2.4 | Time-series for defoliated plant data (GrowliFlowerD)

For field 1, the GrowliFlowerD subset contains a total of 130 plant IDs and coordinates for defoliated plants (30 for October 27 and 100 for October 29). For field 2, the subset contains a total of 717 plant IDs and coordinates for defoliated plants. The coordinates allow the user to extract the time-series of defoliated plants. Table 3 presents an overview of how many plants were defoliated on different acquisition days. In addition to the time-series data, pairs of pre- and post-defoliation images
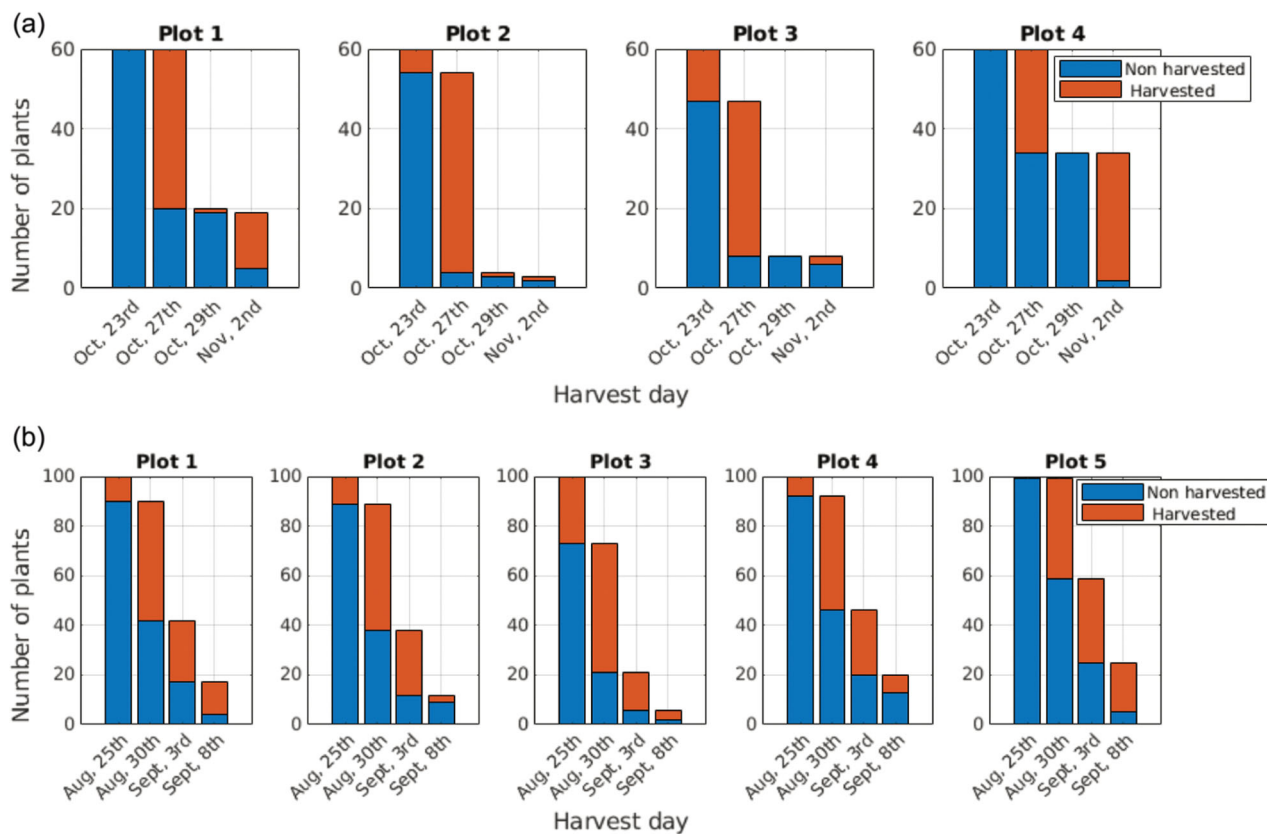
**FIGURE 13** Overview of harvested and nonharvested plants per reference plot per day.

are provided in the subset. The data are divided into training, validation, and testing sets for both fields, and each defoliation day is presented in each set. The visual distribution of both fields is shown in Figure A5.

## 4.3 | Insitu data

Two CSV files are provided, that is, one for each field, and these files contain the plant ID and the measurements described in Section 3.3 for each data acquisition day. The measured values correlate with the images in the GrowliFlowerR subset. Figure 13 shows the distribution of the number of harvested plants in the reference plots per acquisition date for fields 1 and 2.

## 5 | BASELINE FOR INSTANCE SEGMENTATION APPLICATION

### 5.1 | Experimental setup

We describe two possible applications of the proposed dataset by creating baselines using the labeled GrowliFlowerL subset and

the Mask R-CNN (He et al., 2017), which is a state-of-the-art instance segmentation method. *We use the pytorch implementation available at* https://pytorch.org/tutorials/intermediate/torchvision_tutorial.html.

We consider plant instance and leaf instance segmentation tasks. Thus, we use the mask and bounding boxes derived from the plant instance mask as the target for the first baseline. The mask and bounding box derived from the leaf instance mask are used as the target for the second baseline. For the leaf instance segmentation baseline, the given void instances are used as the background because only leaves that do not belong to void plants are labeled. Note that the estimation of semantic masks for individual instances enables the derivation of phenotypic traits. Here, we applied a random horizontal flipping data augmentation technique with a probability of 0.5.

We trained the Mask R-CNN on a computer with Intel Core i7-6850K 3.60 GHz processor and a GeForce GTX 1080Ti GPU with 11 GB RAM. The network was pretrained on the COCO dataset (Lin et al., 2014), and training was performed over 100 epochs with a learning rate of 0.001 and batch size of 2. We used a stochastic gradient descent optimizer, and ResNet-50 was used as the backbone network.

## 5.2 | Evaluation metrics

The first evaluation metric is the Intersection over Union (IoU), which is calculated as follows:

$$IoU = \frac{\text{Area of Intersection}}{\text{Area of Union}} = \frac{TP}{TP + FP + FN}, \quad (1)$$

following the evaluation metrics of the COCO dataset (Lin et al., 2014), where TP represents true positives, FP represents false positives, and FN represents false negatives. We also consider two additional metrics, that is, precision $p$ and recall $r$, which are defined as follows.

$$p = \frac{TP}{TP + FP}, \quad r = \frac{TP}{TP + FN}. \quad (2)$$

The F1 score summarizes and balances precision and recall by calculating a harmonic mean. The F1 score is defined as follows.

$$F1 = \frac{2 \cdot p \cdot r}{p + r}. \quad (3)$$

We compute precision, recall, and F1 score relative to the single object class cauliflower plant and calculate the scores for the IoU thresholds $t_{IoU} = 0.50$ and $t_{IoU} = 0.75$. In addition, we determine the average precision (AP), average recall (AR), and average F1 (AF1) scores over all IoUs in the interval $0.50 - 0.95$ with a step size of 0.05 as for the COCO benchmark. This is indicated by $(\cdot)@0.5 - 0.95$. For the leaf instance segmentation baseline, we reduce the evaluation on recall, as we do not want to penalize predictions on void pixels. The consequence of penalizing predictions on void pixels would be to penalize the model for identifying leaves that were simply not labeled as such.

## 5.3 | Results

We calculate the metrics relative to the detected bounding boxes and the segmented masks of the respective objects. The segmented masks provide information about the cumulative number of correctly classified pixels and thus, the more accurate shape of the object. The bounding box enables derivation of the detection accuracy and thus, the localization of the object.

Table 4 summarizes the results of the plant instance segmentation task for the baseline method. As can be seen, 95% at IoU ≥0.5 are predicted correctly. In addition, precision at the bounding box and pixel levels are greater than 80% for all IoU thresholds ≤0.8

(Figure 14a). At an IoU value of ≥0.85, precision decreased rapidly. This trend is also observed for both recall (Figure 14b) and the F1 score (Figure 14c). For higher IoU values, we found that prediction at the pixel level is less accurate than at the bounding box level because slight changes in segmentation generally result in more errors in the segmentation mask than in the bounding box (Figure 15). An overview of the results is given in Table 4.

We found that many objects and masks are estimated accurately (Figure 16a). The results show all predictions with a confidence score greater than a threshold of 50%. Precise contours are estimated, and in the earlier development stages, the instances are well separated spatially. Note that the model does not predict the ground as an object in any case. In addition, smaller weeds that can be seen in some patches are also not considered objects, which is desirable, because in this way we identify that the model distinguishes cauliflower from weeds. We found that inaccuracies occur with plants that lie at the edge of the image patches. In such cases, only small parts of the plant are visible, thus, the leaves are not adjacent to each other, as shown in Figure 16b (top left and bottom left). We also found that errors occur in later developmental stages because the plants overlap (Figure 16b bottom right), which represents a more challenging scenario than well-separated plants. In particular, for overlapping plants, it is even difficult for the human eye to assign leaves to individual instances. In addition, compared with the earlier stages of development where no overlap occurs, fewer training images were available for the later stage of development. The small number of images means that less variability in the data is captured, making prediction on new unknown data more difficult.

Another distinctive feature involves plant objects from which leaves fall or plants that are impaired in their growth and thus decay. In such cases, it is difficult for the model to distinguish whether one or more plants are represented (Figure 16b top right).

For the leaf instance segmentation task, which is a more difficult task compared with plant instance segmentation, we achieve a very good recall result of 74% at the bounding box level and 77% at the pixel level. The distinction between individual leaf instances is more complex than the distinction between plant instances. In addition, here, we assign the void-labeled objects to the `background` class for this baseline rather than the `leaf` class because individual void plants can contain several leaves, however, such leaves were not labeled individually. Note that the calculated values for recall are similar both the pixel and bounding box levels.

We can find explanations for the recall values in the visual consideration of the results, even though these results show predictions with a confidence score greater than a threshold of

**TABLE 4** Plant instance segmentation results: precision, recall, and F1 score for predicted bounding boxes (BBox) and segmentation masks (pixel) for the class `plant`

| Evaluation | Global metrics | | | Precision | | Recall | | F1 | |
|---|---|---|---|---|---|---|---|---|---|
| | AP | AR | AF1 | P@0.5 | P@0.75 | R@0.5 | R@0.75 | F1@0.5 | F1@0.75 |
| BBox | 0.917 | 0.933 | 0.843 | 0.952 | 0.899 | 0.965 | 0.913 | 0.958 | 0.906 |
| Pixel | 0.844 | 0.858 | 0.770 | 0.954 | 0.902 | 0.963 | 0.913 | 0.959 | 0.908 |

Abbreviations: AP, average precision; AR, average recall; AF1, average F1.
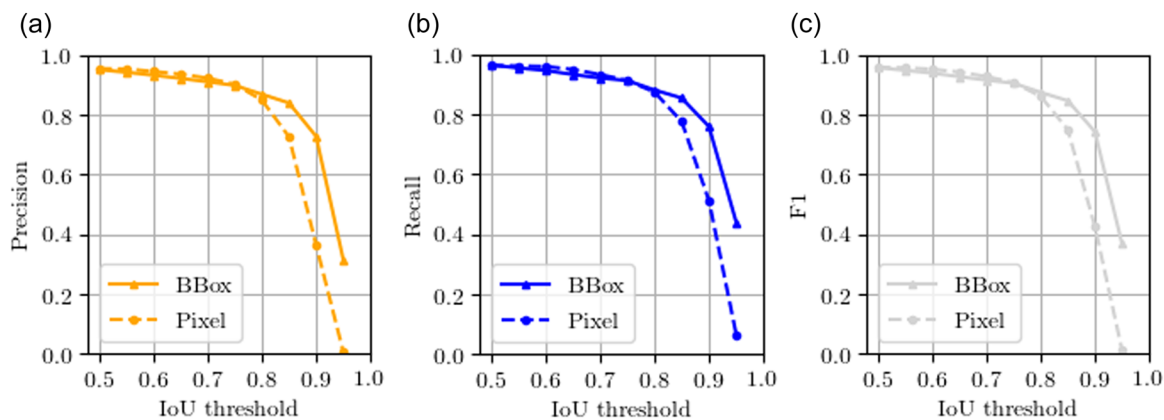
**FIGURE 14** Representation of (a) precision, (b) recall, and (c) F1 score for class cauliflower `plant`. The graphs show the evaluation at different IoU thresholds on the bounding box (BBox), thus object (solid-line), and pixel (dashed-line) level.
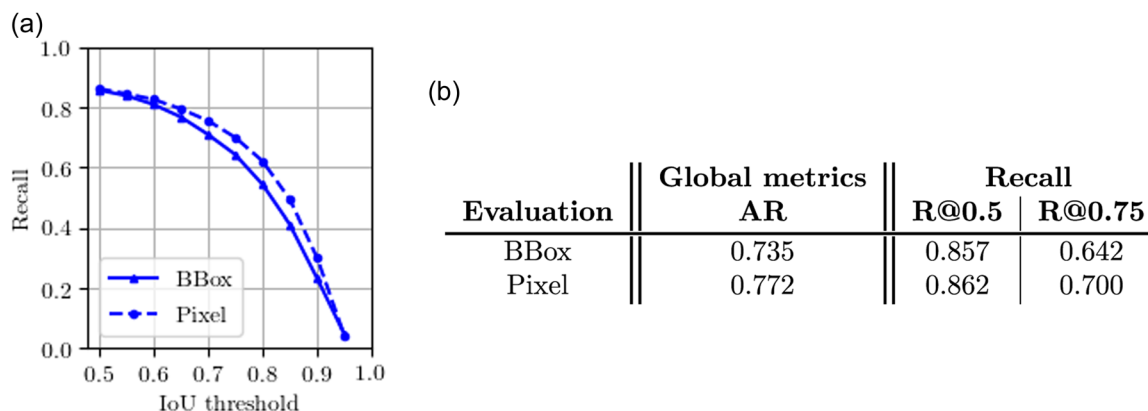


| Evaluation | Global metrics AR | Recall | |
|---|---|---|---|
| | | R@0.5 | R@0.75 |
| BBox | 0.735 | 0.857 | 0.642 |
| Pixel | 0.772 | 0.862 | 0.700 |

**FIGURE 15** Recall results for leaf instance segmentation task. (a) Evaluation at different Intersection over Union (IoU) thresholds on the bounding box (BBox), that is, object (solid-line), and pixel (dashed-line) level. (b) shows the respective average recall (AR), *R@0.5*, and *R@0.75* values.

50%. By defining void instances as the background, the model is challenged to predict the leaves belonging to void instances not as *leaf* objects, as shown in Figure 16d (top left and bottom left). It is difficult for the model to distinguish whether plants at the edge of the patches are void instances or leaf instances. Thus, either leaves are predicted that are not in fact present in the target (representing low precision) or no leaves are predicted even they are present in the target (representing low recall). For plants that are completely visible in the patch, the model demonstrates better prediction performance. Another source of error is the prediction of several instances on a single leaf, as shown in Figure 16d (top right and bottom right) because the model is required to learn various features, for example, leaf structure and size, because such features play a crucial role in distinguishing different leaves.

To sum up, we observe that our instance segmentations, plant instance as well as leaf instance, perform and can be used for different growth stages of the cauliflower plants.

## 6 | REFLECTION AND DISCUSSION

The results of this study provide insight into the acquisition of image time-series under field conditions. We observe that the flight altitude of the UAV must be adjusted depending on the characteristics and the height of the cultivated plants to capture images with high quality in terms of resolution and gapless spatial data. However, to obtain accurate image data, enough keypoints must be distributed in the field. Our work shows that GCPs are suitable as keypoints because they help align and georeference the orthophotos more accurately. For simplified use of the data for ML approaches, the data should have similar characteristics as exposure over time. For this purpose, data must be recorded under consistent weather conditions. However, we note that combining consistent conditions and similar interval lengths between acquisition days is a challenging task.

While previous research, such as the work of Bender et al. (2020), focuses on collecting data from many different data sources (e.g., imagery, climate, and soil data), monitoring a large number of
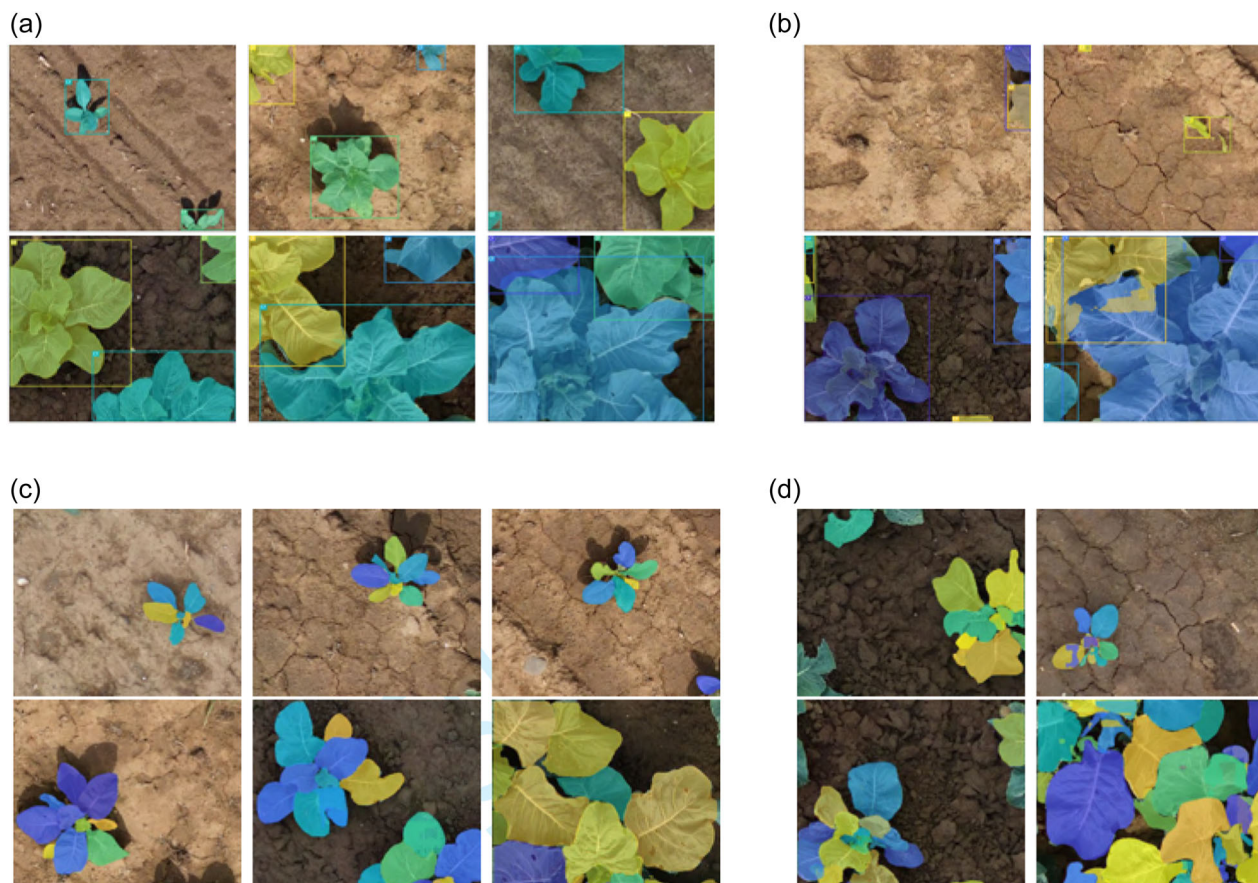
**FIGURE 16** Plant and leaf instance segmentation results. The different colors indicate the different instances. In the visualization of the leaf instance segmentation results, we concentrate on the visualization of the masks for clarity and omit the bounding boxes. The examples shown in (a) and (c) show accurate results and those in (b) and (d) show improvable results. (a) Accurate plant instance segmentation results, (b) improvable plant instance segmentation results, (c) accurate leaf instance segmentation results, and (d) improvable leaf instance segmentation results.

different instances that are specifically needed for training DL methods has not been considered sufficiently. In our work, we provide a large image number of different instances but we lack additional data, such as soil and climate data. Climate and soil characteristics are important factors to determine or predict plant growth. ML methods only learn features that are present in the training data. Thus, if external factors such as climate and soil change, the growth of plants is influenced as well. The lack of this information can cause ML models to be prone to error in their results when applied to new data. Therefore, we suggest improving the data acquisition by capturing additional soil and climate data. Another suggestion for improvement of data acquisition is the field design. To avoid systematic effects in the data, reference plots and other selected areas like defoliation plots should be distributed equally within the field, because it is difficult at the beginning of data collection to predict how much which plants (crops and weeds) will grow and how the location will affect growth and thus, the acquired data.

Regarding our baseline experiment considering instance segmentation, we observe that the application shows satisfactory results on

our data. To improve the results, future studies could integrate prior knowledge about the shape and structure of plants and leaves. Weyler et al. (2022) show an example of how to improve the results for plant and leaf segmentation by developing a combined approach of neural networks and clustering to simultaneously determine leaf and plant instances. However, to the best of our knowledge, the approach is not used for plants as large and highly overlapping as in our provided later developmental stages (see Figure 16c,d).

Another way to improve the result of instance segmentation is to vary the field of view per image. In further experiments, we observe that when applying a modified field of view, plants are segmented with only a few errors. For these experiments, we reduce the threshold by up to 20% depending on the extension of the field of view. The masks and bounding boxes of the predictions match the plants. However, the size of the objects differs from the training data due to the change in scale. This causes the model to have less confidence in its predictions, even if they are correct. Even with a changed image size, cauliflower plants can be easily distinguished from weeds and each other. As it brings greater variability to the data, adding images with a larger field of view to the training set could lead

to further improvements in the results. However, this would require labeling more data.

We recommend our dataset for further methodological developments or as evaluation dataset for existing approaches as used in the work of Günder et al. (2022).

# 7 | CONCLUSION AND FUTURE DIRECTIONS

In this *article*, we have introduced the GrowliFlower dataset, which is a georeferenced, image-based UAV time-series dataset of two monitored cauliflower fields during their entire growth period. The proposed dataset was described, and we discussed the data collection process, which may helpful for other similar data collection procedures. The proposed dataset comprises weekly RGB and multispectral UAV orthophotos and image time-series of individual plants reflecting weekly plant growth. In subset of the proposed dataset, in-situ reference measurements, for example, plant size, are also available, and another subset provides pre- and post-defoliation images to demonstrate the relationship the interior and exterior of the cauliflower plant. The proposed dataset also contains annotations with segmented plant and leaf instances, as well as annotations on stems. The data are available at http://rs.ipb.uni-bonn.de/data/. The proposed dataset is intended to promote the use and evaluation of ML methods and foster close collaboration between different disciplines, for example, agricultural sciences, remote sensing, and ML. We have also presented baseline results of two applications of the proposed dataset using the Mask R-CNN model, that is, plant instance segmentation and leaf instance segmentation tasks. In addition, we expect that the findings and descriptions presented in this paper will help realize effective data collection processes that are transferred to other areas.

## DATA AVAILABILITY STATEMENT
The dataset presented in this study will be freely available once the paper is published. We will be happy to provide the first version to the reviewers if needed.

## ORCID
*Jana Kierdorf* http://orcid.org/0000-0003-1145-1555
*Ribana Roscher* http://orcid.org/0000-0003-0094-6210

## REFERENCES
Ahmadi, A., Halstead, M. & McCool, C. (2021) Virtual temporal samples for recurrent neural networks: applied to semantic segmentation in agriculture. In: *DAGM German Conference on Pattern Recognition*. Springer, pp. 574–588.

Ali, I., Greifeneder, F., Stamenkovic, J., Neumann, M. & Notarnicola, C. (2015) Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. *Remote Sensing*, 7(12), 16398–16421.

Arad, B., Balendonck, J., Barth, R., Ben-Shahar, O., Edan, Y., Hellström, T., Hemming, J., Kurtser, P., Ringdahl, O., Tielen, T. et al. (2020) Development of a sweet pepper harvesting robot. *Journal of Field Robotics*, 37(6), 1027–1039.

Bender, A., Whelan, B. & Sukkarieh, S. (2020) A high-resolution, multimodal data set for agricultural robotics: a ladybird's-eye view of brassica. *Journal of Field Robotics*, 37(1), 73–96.

Blok, P.M., van Henten, E.J., van Evert, F.K. & Kootstra, G. (2021) Image-based size estimation of broccoli heads under varying degrees of occlusion. *Biosystems Engineering*, 208, 213–233.

Chaparro, D., Piles, M., Vall-Llossera, M., Camps, A., Konings, A.G. & Entekhabi, D. (2018) L-band vegetation optical depth seasonal metrics for crop yield assessment. *Remote Sensing of Environment*, 212, 249–259.

Chebrolu, N., Lottes, P., Schaefer, A., Winterhalter, W., Burgard, W. & Stachniss, C. (2017) Agricultural robot dataset for plant classification, localization and mapping on sugar beet fields. *The International Journal of Robotics Research*, 36(10), 1045–1052.

Chi, M., Plaza, A., Benediktsson, J.A., Sun, Z., Shen, J. & Zhu, Y. (2016) Big data for remote sensing: challenges and opportunities. *Proceedings of the IEEE*, 104(11), 2207–2219.

Cordts, M., Omran, M., Ramos, S., Scharwächter, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S. & Schiele, B. (2015) The cityscapes dataset. In: *CVPR workshop on the future of datasets in vision*, Vol. 2.

Debolini, M., Schoorl, J.M., Temme, A., Galli, M. & Bonari, E. (2015) Changes in agricultural land use affecting future soil redistribution patterns: a case study in Southern Tuscany (Italy). *Land Degradation & Development*, 26(6), 574–586.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. & Fei-Fei, L. (2009) ImageNet: a large-scale hierarchical image database. In: *2009 IEEE conference on computer vision and pattern recognition*. IEEE, pp. 248–255.

Drees, L., Junker-Frohn, L.V., Kierdorf, J. & Roscher, R. (2021) Temporal prediction and evaluation of brassica growth in the field using conditional generative adversarial networks. *Computers and Electronics in Agriculture*, 190, 106415.

Feller, C., Bleiholder, H., Buhr, L., Hack, H., Hess, M., Klose, R., Meier, U., Stauss, R., Boom, T. & Weber, E. (1995) Phenological growth stages of vegetable crops. ii. Fruit vegetables and pulses. Coding and description according to the extended BBCH scale with illustrations. *Nachrichtenblatt des Deutschen Pflanzenschutzdienstes*, 47(9), 217–232.

Förster, A., Behley, J., Behmann, J. & Roscher, R. (2019) Hyperspectral plant disease forecasting using generative adversarial networks. In: *2019 IEEE international geoscience and remote sensing symposium (IGARSS 2019)*. IEEE, pp. 1793–1796.

Grenzdörffer, G. (2019) Automatic generation of geometric parameters of individual cauliflower plants for rapid phenotyping using drone images. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, XLII-2/W13, 329–335.

Günder, M., Ispizua Yamati, F.R., Kierdorf, J., Roscher, R., Mahlein, A.-K. & Bauckhage, C. (2022) Agricultural plant cataloging and establishment of a data framework from UAV-based crop images by computer vision. *GigaScience*, 11, giac054.

Halstead, M., Denman, S., Fookes, C. & McCool, C. (2020) Fruit detection in the wild: the impact of varying conditions and cultivar. In: *2020 digital image computing: techniques and applications (DICTA)*. IEEE, pp. 1–8.

He, K., Gkioxari, G., Dollár, P. & Girshick, R. (2017) Mask R-CNN. In: *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969.

Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A.A. (2017) Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125–1134.

Kierdorf, J., Weber, I., Kicherer, A., Zabawa, L., Drees, L. & Roscher, R. (2022) Behind the leaves: estimation of occluded grapevine berries with conditional generative adversarial networks. *Frontiers in Artificial Intelligence*, 5, 830026. https://doi.org/10.3389/frai.2022.830026

Kierdorf, J., Zabawa, L., Lucks, L., Klingbeil, L., Kuhlmann, H. et al. (2019) Detection and counting of wheat ears by means of ground-based image acquisition. *Bornimer Agrartechnische Berichte*, 102, 158–167.

Kusumam, K., Krajník, T., Pearson, S., Duckett, T. & Cielniak, G. (2017) 3D-vision based detection, localization, and sizing of broccoli heads in the field. *Journal of Field Robotics*, 34(8), 1505–1518.

Lary, D.J., Alavi, A.H., Gandomi, A.H. & Walker, A.L. (2016) Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7(1), 3–10.

Lillesand, T., Kiefer, R.W. & Chipman, J. (2015) *Remote sensing and image interpretation*, 7th ed. Hoboken, New Jersey: John Wiley & Sons.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. & Zitnick, C.L. (2014) Microsoft COCO: common objects in context. In: *European conference on computer vision*. Springer, pp. 740–755.

Lottes, P., Behley, J., Milioto, A. & Stachniss, C. (2018) Fully convolutional networks with sequential information for robust crop and weed detection in precision farming. *IEEE Robotics and Automation Letters*, 3(4), 2870–2877.

Lottes, P., Hörferlin, M., Sander, S. & Stachniss, C. (2017) Effective vision-based classification for separating sugar beets and weeds for precision farming. *Journal of Field Robotics*, 34(6), 1160–1178.

Milioto, A., Lottes, P. & Stachniss, C. (2018) Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs. In: *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, pp. 2229–2235.

Minervini, M., Fischbach, A., Scharr, H., & Tsaftaris, S.A. (2016). Finely-grained annotated datasets for image-based plant phenotyping. *Pattern Recognition Letters*, 81, 80–89.

Mosleh, M.K., Hassan, Q.K. & Chowdhury, E.H. (2015) Application of remote sensors in mapping rice area and forecasting its production: a review. *Sensors*, 15(1), 769–791.

Muresan, H., & Oltean, M. (2018) Fruit recognition from images using deep learning. *Acta Universitatis Sapientiae, Informatica*, 10, 26–42.

Nock, C.A., Vogt, R.J. & Beisner, B.E. (2016) Functional traits. In: *eLS*. Chichester: John Wiley & Sons, Ltd.

Persia, M., Barca, E., Greco, R., Marzulli, M.I. & Tartarino, P. (2020) Archival aerial images georeferencing: a geostatistically-based approach for improving orthophoto accuracy with minimal number of ground control points. *Remote Sensing*, 12(14), 2232.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N. et al. (2019) Deep learning and process understanding for data-driven earth system science. *Nature*, 566(7743), 195–204.

Ren, M. & Zemel, R.S. (2017) End-to-end instance segmentation with recurrent attention. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6656–6664.

Romera-Paredes, B. & Torr, P.H.S. (2016) Recurrent instance segmentation. In: *European conference on computer vision*. Springer, pp. 312–329.

Roscher, R., Volpi, M., Mallet, C., Drees, L. & Wegner, J.D. (2020) SemCity Toulouse: a benchmark for building instance segmentation in satellite images. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-5-2020, 109–116.

Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T. & McCool, C. (2016) DeepFruits: a fruit detection system using deep neural networks. *Sensors*, 16(8), 1222.

Scharr, H., Minervini, M., Fischbach, A. & Tsaftaris, S.A. (2014) Annotated image datasets of rosette plants. In: *European Conference on Computer Vision*. Zürich: Suisse, pp. 6–12.

Scharr, H., Minervini, M., French, A.P., Klukas, C., Kramer, D.M., Liu, X., Luengo, I., Pape, J.-M., Polder, G., Vukadinovic, D. et al. (2016) Leaf segmentation in plant phenotyping: a collation study. *Machine Vision and Applications*, 27(4), 585–606.

Verger, A., Baret, F. & Weiss, M. (2014) Near real-time vegetation monitoring at global scale. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(8), 3473–3481.

Verrelst, J., Malenovský, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J.-P., Lewis, P., North, P. & Moreno, J. (2019) Quantifying vegetation biophysical variables from imaging spectroscopy data: a review on retrieval methods. *Surveys in Geophysics*, 40(3), 589–629.

Ward, D. & Moghadam, P. (2018) Synthetic Arabidopsis dataset. In: *CSIRO. Data collection*.

Weiss, M., Jacob, F. & Duveiller, G. (2020) Remote sensing for agricultural applications: a meta-review. *Remote Sensing of Environment*, 236, 111402.

Weyler, J., Magistri, F., Seitz, P., Behley, J., & Stachniss, C. (2022). In-field phenotyping based on crop leaf and plant instance segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2725–2734.

Weyler, J., Milioto, A., Falck, T., Behley, J. & Stachniss, C. (2021) Joint plant instance detection and leaf count estimation for in-field plant phenotyping. *IEEE Robotics and Automation Letters*, 6(2), 3599–3606.

Zabawa, L., Kicherer, A., Klingbeil, L., Milioto, A., Topfer, R., Kuhlmann, H. & Roscher, R. (2019) Detection of single grapevine berries in images using fully convolutional neural networks. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*.
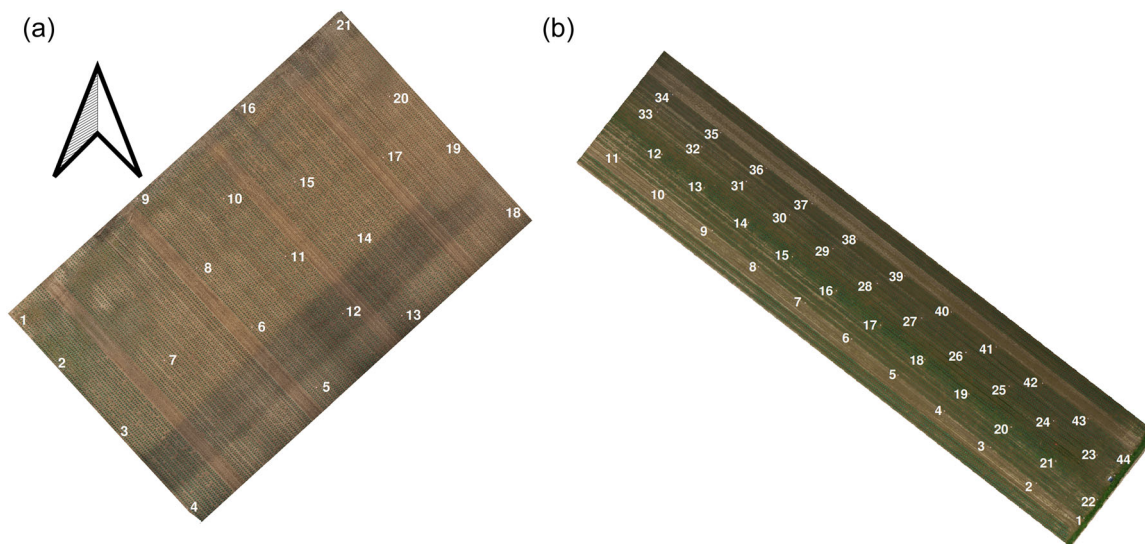
## APPENDIX A: APPENDICES

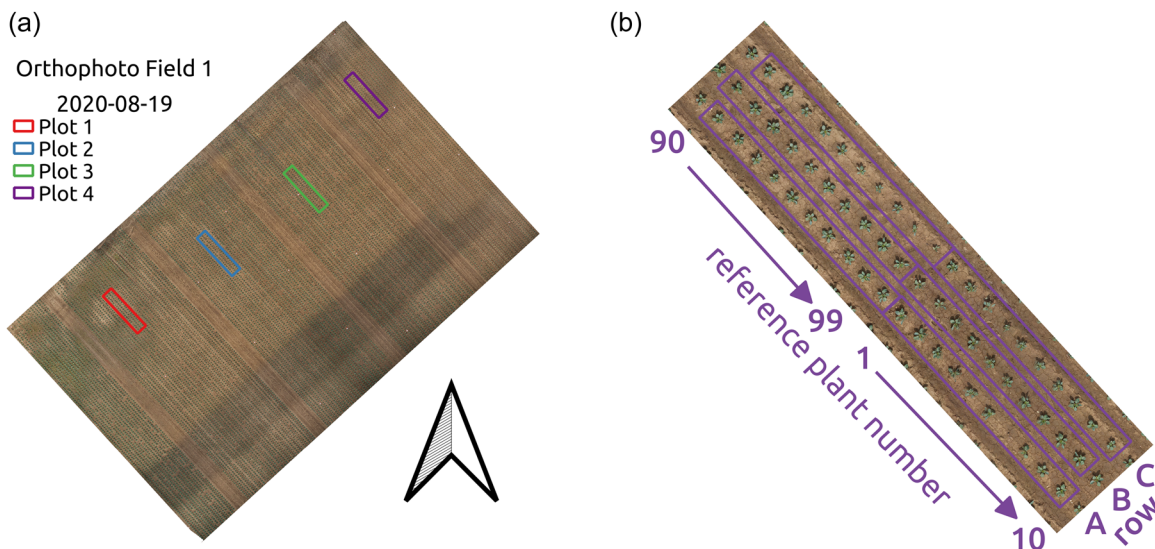**FIGURE A1** Location of ground control points in fields 1 and 2.



**FIGURE A2** Visual overview of (a) reference plots for in-situ measurements in field 1 and (b) the respective design of reference plot 4 (including reference plants and ordering of reference plant numbers). The plot design is valid for all reference plots in field 1.

**TABLE A1** Monitored abiotic and biotic stresses.

| Abbreviation | Meaning | Abbreviation | Meaning |
|---|---|---|---|
| P | Plant | L | Leaf/leaves |
| nP | No plant | wL | Without leaves |
| Pl | Plant lying down | oL | Old leaves |
| wP | Whole plant | yL | Yellowish leaves |
| 2P | Two plants | rL | Reddish leaves |
| bb | Blind bud | pgL | Pale green leaves |
| pd | Planted too deep | pygL | Pale yellowish green leaves |
| A | Aphids present | sg | Stunted growth with many shoots |
| C | Coal fleas present | dT | Damage to leaves caused by tractor |
| F | Flies present | | |
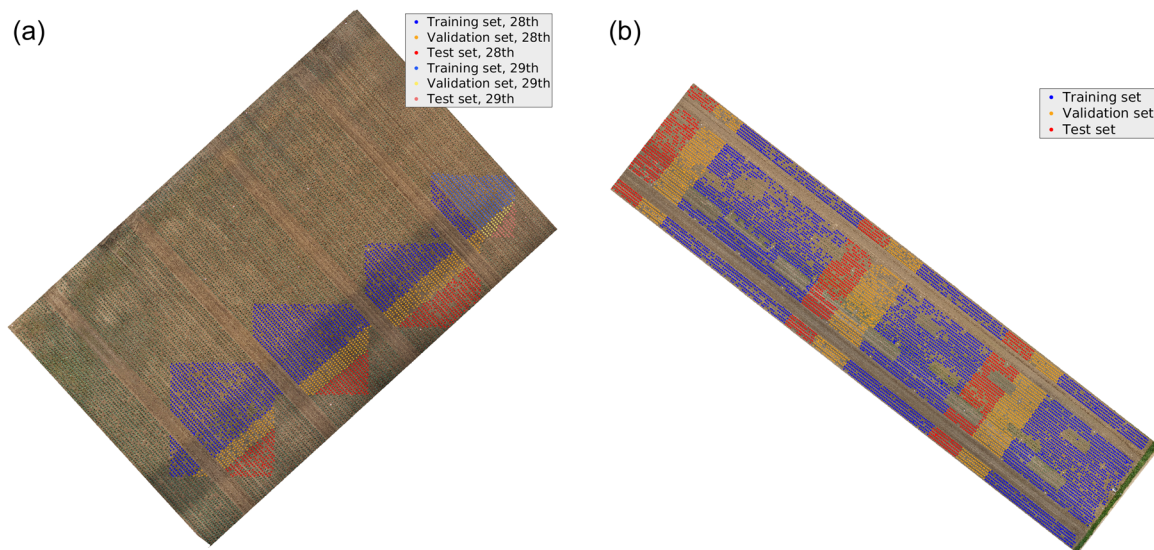


**FIGURE A3** Separation of plants fields 1 and 2 in GrowliFlowerT in training (blue), validation (yellow), and testing (red) sets. For field 1, the two planting days are separated using dark colors for July 28, 2020 and light colors for July 29, 2020.
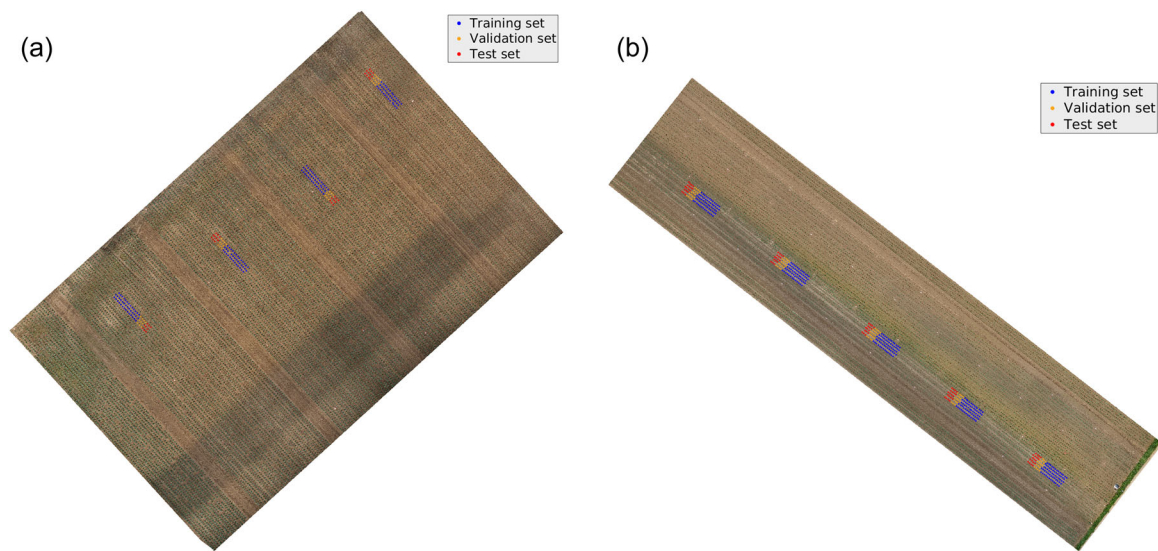
**FIGURE A4** Separation of reference plants in fields 1 and 2 in GrowliFlowerR in training (blue), validation (yellow), and testing (red) sets.
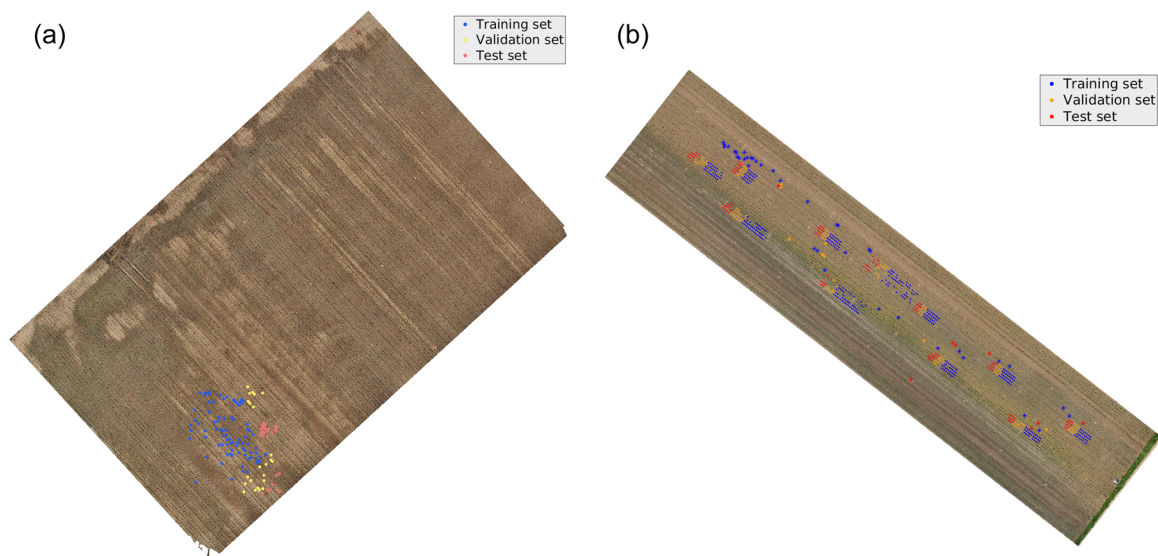


**FIGURE A5** Separation of defoliated plants in fields 1 and 2 in GrowliFlowerD in training (blue), validation (yellow), and testing (red) sets.