PLOS ONE

RESEARCH ARTICLE

Data-centric AI approach for automated wildflower monitoring

Gerard Schouten[1,2]*, **Bas S. H. T. Michielsen**[1,3], **Barbara Gravendeel**[2,4]

1 School of ICT, Fontys University of Applied Sciences, Eindhoven, Netherlands, **2** Naturalis Biodiversity Center, Leiden, Netherlands, **3** Copernicus Institute of Sustainable Development, Utrecht University, Utrecht, Netherlands, **4** Radboud Institute for Biological and Environmental Sciences, Radboud University, Nijmegen, Netherlands

* g.schouten@fontys.nl

Check for updates

OPEN ACCESS

Citation: Schouten G, Michielsen BSHT, Gravendeel B (2024) Data-centric AI approach for automated wildflower monitoring. PLoS ONE 19(9): e0302958. https://doi.org/10.1371/journal.pone.0302958

Editor: Xiaoyong Sun, Shandong Agricultural University, CHINA

Received: April 16, 2024

Accepted: August 19, 2024

Published: September 9, 2024

Copyright: © 2024 Schouten et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The annotated EWD dataset is publicly available on the Harvard Dataverse Network instance for the Netherlands (also known as DataverseNL) at https://doi.org/10.34894/U4VQJ6. The Python code to pre-process the data, train and test the model as well as the inference script to run the model on unseen images is available at: https://github.com/bshtmichielsen/flowerpower.

Funding: The author(s) received no specific funding for this work.

Abstract

We present the Eindhoven Wildflower Dataset (EWD) as well as a PyTorch object detection model that is able to *classify* and *count* wildflowers. EWD, collected over two entire flowering seasons and expert annotated, contains 2,002 top-view images of flowering plants captured 'in the wild' in five different landscape types (roadsides, urban green spaces, cropland, weed-rich grassland, marshland). It holds a total of 65,571 annotations for 160 species belonging to 31 different families of flowering plants and serves as a reference dataset for automating wildflower monitoring and object detection in general. To ensure consistent annotations, we define species-specific floral count units and provide extensive annotation guidelines. With a 0.82 mAP (@IoU > 0.50) score the presented baseline model, trained on a balanced subset of EWD, is to the best of our knowledge superior in its class. Our approach empowers automated quantification of wildflower richness and abundance, which helps understanding and assessing natural capital, and encourages the development of standards for AI-based wildflower monitoring. The annotated EWD dataset and the code to train and run the baseline model are publicly available.

1. Introduction

Wildflowers, or angiosperms, are plants producing flowers that, once pollinated, develop into seed-bearing fruits. They are a crucial part of our ecosystem. In the history of life, they are relative newcomers. The first flowering plants emerged about 130 million years ago–approximately 70 million years after the first appearance of mammals–following a period of rapid evolution [1, 2]. At present, there are more than 300,000 species of flowering plants worldwide [3]; they outnumber all other plants combined and are the dominant vegetation on land [4]. Flowers exhibit a large diversity in color, shape and texture to attract pollinators. Besides playing crucial roles in natural ecosystems, plants provide a great variety of essential services to human life. Root systems of plants improve soil health, purify water, and prevent erosion. In addition to their aesthetic and cultural value, plants strongly regulate daily temperature variations in built environments [5] and provide building materials, food as well as a plethora of

Competing interests: The authors have declared that no competing interests exist.

medicines [6, 7]. Moreover, the photosynthesis process of plants mitigates climate change by capturing carbon dioxide.

An imperative condition for sustainable ecosystem services, as the ones mentioned above, is a diverse and resilient web-of-life [8]. Today, ecosystem services are degrading because of biodiversity loss [9]. To better understand ecosystems and maintain their services it is essential to have large-scale monitoring programs to assess biodiversity trends reliably and consistently over long periods of time [10–12].

Current practice for wildflower monitoring includes systematic surveys that rely on manual counts by professionals and nature enthusiasts complemented with data collection supported by citizen-science platforms. These routines have limitations hampering large-scale monitoring. Manual counting is arduous, error prone, and labor-intensive. Species identification–although supported nowadays by well-designed AI-enabled smartphone apps, such as Pl@ntNet [13] or PlantSnap [14]–is still challenging for the general public. The situation is further exacerbated by the fact that taxonomic knowledge is becoming increasingly scarce [15, 16]. Another drawback, especially for the more ad-hoc data collection in open citizen-science platforms, is biased data [17]. Orchids or other 'iconic' flowering plant families are overrepresented, while less sexy ones are underrepresented.

Given these issues, wildflower monitoring would be greatly helped by the development of AI solutions that can perform reliable automated wildflower counts from images of patches of land. The smartphone apps mentioned earlier cannot do this. They are driven by AI models that have been trained on photographic close-ups of flowers, and only solve a *classification* problem, implicitly assuming that each new input contains only one item. Classifying and localizing multiple wildflowers in overview images is a different, more complex challenge that requires an *object detection* approach [18]. Although visual object detection is advancing rapidly in other domains (e.g. autonomous driving and healthcare), there are to our knowledge currently next to no solutions in the domain of wildflower monitoring.

It is our aim to make a first step towards creating an AI solution that can carry out reliable automated counts from overview images containing multiple wildflowers of various species. This challenge involves two major elements: (i) the creation of an annotated dataset of such overview images, and (ii) the creation of an object detection model trained on the annotated dataset.

In this paper, we present:

1. The Eindhoven Wildflower Dataset (EWD). This is an expert-annotated reference dataset of over 2,000 bird's-eye view high-resolution wildflower images, captured in five different Dutch landscape types–roadsides, urban green spaces, extensively farmed cropland, weed-rich grassland, and marshland. We limited ourselves to herbs and collected images in two entire flowering seasons.

2. A state-of-the-art object detection model trained on EWD. Our model is based on off-the-shelf open-source software and serves as a baseline. It is intended as a stepping stone for new innovative modelling efforts in the domain of automated wildflower monitoring.

Many deep learning papers aim at improving predictive models or creating novel, more efficient algorithms. Contrary to this algorithm-centric perspective, we adopt a data-centric AI approach. This refers to a recent shift visible in the literature from focusing on the performance of models to the quality of the underlying data used to train and evaluate these models [19]. This is a fundamental change, where data is viewed as a profound asset, on par with model code. To obtain high-quality data for wildflower monitoring, we formulate explicit annotation guidelines [20] and introduce floral count units compatible with botanical

nomenclature. Moreover, before training the model we created a balanced subset of EWD and tiled the images to preserve resolution.

2. Background and related work

2.1 Datasets as benchmarks

The role of image-based reference datasets varies from supporting the famous general-purpose computer vision competitions, such as PASCAL VOC [21], COCO [22], ImageNet Large Scale Visual Recognition Challenge [23], and Open Images V4 [24] to benchmarks for domain-specific quests. For instance, in the medical domain open datasets with skin lesion images played a major role in creating models that can diagnose malicious melanomas [25–27] or, more recently, public CT and X-ray chest data significantly accelerated the development of clinical apps that are able to estimate COVID-19 infection status [28, 29]. Similarly, the Audi A2D2 dataset [30] pushes forward the research field of autonomous driving. Even synthetic data is utilized as reference data, e.g., for simulating driving scenarios with fast context changes, such as deteriorating weather conditions [31].

In short, the proliferation of annotated image-based public datasets in the scientific literature fast-tracks the creation of novel model architectures [32] and improvement of evaluation metrics, e.g. [21, 33]. It allows the scientific community to compare computer vision algorithms and pushes the field towards increasingly complex and challenging problems. Furthermore, they trigger more strict annotation protocols to raise data quality [20].

2.2 Public image-based flower datasets

Table 1 lists the most important public datasets with flower images taken 'in the wild' that target either the species identification or monitoring use case. The top six datasets in Table 1 include only close-up images of individual wildflowers. The Kaggle dataset with only five very distinctive wildflower species mainly serves an educational purpose. It contains low resolution images scraped from Flickr and Google Images. The Oxford 17 dataset [34] has also been acquired by searching the web; it contains 80 images for each of the 17 flowering plant species included in this study from the UK; Oxford 102 [35] holds, as the name indicates, 102 plant taxa from the UK represented by 40 to 258 images per class. It is more challenging due to small inter-class variances and large intra-class variances. The Jena 30 dataset [36] consists of

Table 1. Public flower datasets with images taken 'in the wild'.

Dataset (year)	# images | type (# ann.)	# species	Purpose	Computer vision task	Annotation guidelines
Kaggle Flower (—)	4,242 | RGB	5	Educational challenge	Classification	No
Oxford 17 (2006)	1,360 | RGB	17	Species identification	Classification	No
Oxford 102 (2008)	8,189 | RGB	102	Species identification	Classification	No
Jena 30 (2017)	1,479 | RGB	30	Species identification	Classification	No
PlantCLEF (2022)	2.9 M | RGB	80K	Species identification	Classification	No
HFD100 (2022)	10,700 | Hyperspectral	100	Species identification	Classification	No
Nectar (2019)	1997 | RGB (25K)	25	Wildflower monitoring, nectar index estimation	Object detection	No
EWD (ours) (2024)	**2,002 | RGB (65.6K)**	**160**	**Wildflower monitoring**	**Object detection**	**Yes**

https://doi.org/10.1371/journal.pone.0302958.t001

iPhone6 images of 30 wildflower species (11 to 70 images per class) and was collected during an entire flowering season in semi-arid grasslands around Jena in Germany. The PlantCLEF challenge [37] targets plant identification at a global scale. It is a fast-growing database currently holding 2.9 million images of about 80,000 species. It contains images of both flowers and leaves. The train data has two types of classification labels: either 'trusted' (validated) or 'web' (might suffer from species identification errors). The holdout test set consists of tens of thousands of pictures verified by experts. HDF100 extends flower identification into the hyperspectral domain [38].

The bottom two rows in Table 1 refer to datasets containing top-view images with tagged bounding boxes (BB) around wildflowers. Both support the monitoring use case, i.e., the use of object detection models that can *classify* and *localize*. The Nectar dataset [39] consists of nearly 2,000 Canon Powershot G10 (14.7 Mpixels) images of flowering plant species collected in one flowering season in one habitat type, viz. weed-rich grasslands in the UK. All 25 species are easy to outline and have predominantly simple inflorescences, i.e., solitary flowers or flower heads. The aim of the Hicks et al. study [39] was to estimate nectar sugar mass as an indicator for effectiveness of nature-inclusive farming. In other words, these authors used wildflower monitoring as a means to an end. In our study, wildflower monitoring with computer vision is the main topic. Although the EWD dataset has a comparable number of Canon EOS Mark IV (30.4 Mpixel) images, it holds a significantly larger number of flowering plant species. The EWD images were collected over two entire flowering seasons in five different habitat types around the city of Eindhoven in the Netherlands. Even more importantly, what makes EWD novel and unique, is that the annotated images adhere to strict guidelines and that well-defined floral count units are used, also for more complex inflorescences. In short, EWD complies with the data-centric AI principle to boost data quality.

2.3 Challenges and solutions for automated wildflower monitoring

Automated wildflower monitoring has its own unique computer vision challenges. Backgrounds can be very cluttered, and flowers 'in the wild' are often occluded or damaged. Moreover, flowers develop from bud to blooming stage and subsequently wither or transform into fruits, i.e., they are not stable in size, shape and color but change during their life cycle. And above all, wildflower identification is a large vocabulary problem with few majority classes and many minority classes [40]. This requires a classifier that can handle class imbalance and that supports fine-grained class differentiation, i.e. small inter-class variances (related flowering species might be quite similar to one another) as well as large intra-class variances (individuals of the same species may vary considerably in their morphology). Fig 1 illustrates some of these challenges with EWD images.

2.3.1 Classification. Let us first focus on the 'easier' *classification* problem. Over the years various computer vision algorithms have been proposed to deal with this challenge. A straightforward method to identify wildflowers in images is explicitly coding morphological features (color, texture, shape) with handcrafted image processing filters [41, 42]. More elaborate approaches compute local image descriptors [43], such as *scale-invariant feature transform* [44] or *histogram of gradients* [45], followed by a machine learning classifier. The descriptor or early learn-from-data-implementations resulted in significantly higher precision and recall rates compared to the algorithms without machine learning [35, 36]. An extensive survey of wildflower identification with traditional image processing as well as the descriptor-classifier approach is given by Wäldchen et al. [46].

Since the landmark paper of Krizhevsky et al. [47]–in which an 8-layer convolutional neural network (CNN) was presented that won the ILSVRC challenge–and the subsequent designs of

Fig 1. Computer vision challenges for top-view wildflower images taken 'in the wild'. The panels show from top left to bottom right: a) *Cardamine pratensis* is difficult to detect on a cluttered background with a touch of frost, b) overlapping umbels of *Anthriscus sylvestris* are hard to count, c) wildflowers change over time, *Taraxacum officinale* in blooming (white circle) and fruiting stage (yellow circle), and d) *Daucus carota* in blooming (white circle) and fruiting stage (yellow circle).

https://doi.org/10.1371/journal.pone.0302958.g001

even more powerful CNN architectures [48–50], wildflower identification (along with many other computer vision problems) entered the realm of deep learning. Training a CNN with generalization capabilities requires a large, labelled dataset with intra-class diversity. A popular and efficient deep learning approach is transfer learning [51], in which a large convolutional neural network with pretrained weights is used as a visual encoder followed by a shallow classifier network that is trained to discriminate (or learn) several domain-specific object categories, such as flowering plant species. Currently, deep learning is state-of-the-art for flower identification. Using this technique, Wäldchen et al. [52] report accuracy levels of approximately 95% for the Oxford 102 benchmark dataset (containing 102 species, see Table 1), thereby outperforming the best descriptor-classifier result. Recent work of Fei et al. [53] shows that for the industrial case of quality grading systems for fresh-cut flowers, four channel attention-based CNNs (RGB and depth) reach accuracy levels beyond 99%.

2.3.2 Object detection. Rather than classifying flowers (from isolated or cropped-out close-ups), we aim to apply an *object detection* algorithm that can reliably identify and count wildflowers from overview images. Object detection seeks to locate and recognize object instances (from a large number of predefined classes) in natural images. For recent surveys of the developments in object detection, see [18, 54]. Adopting object detection for wildflower monitoring adds a further challenge to the ones mentioned above. Whereas a 'simple' classifier answers the question "*What object is in the image*?", object detection models answer the question "*What objects are where in the image*?". Answering the latter question is compatible with the monitoring use case, i.e., identifying wildflower instances and counting them.

Object detection models are trained with images annotated with tagged bounding boxes (BBs) around objects of interest. The annotations can be formalized as set $\{(b_i^g, c_i^g)\}$, where b_i^g

denotes the ground-truth BB of object *i* in the image, and c_i^g the ground-truth class label. In inference mode, an object detection model takes an image as input and the standard output is a set of object proposals, each with an 'objectness' score. This results in the set $\{(b_j, c_j, p_j)\}$, where b_j denotes the predicted BB of object *j*, c_j the predicted category, and p_j the corresponding confidence score (indicating the likelihood of membership to a set of objects *vs*. background).

The metric to evaluate object detection models is mean averaged precision (mAP). It is a figure of merit between 0 and 1 expressing the precision of an object detection model (a high score refers to a more accurate model). It is computed from the test set by applying the following steps.

1. *Determine true positive (TP) and false positive (FP) object proposals*. When a prediction for a certain annotated object (b_i^g, c_i^g) satisfies the following three conditions: i) the predicted and ground-truth BBs, b_j and b_i^g respectively, have a sufficient degree of overlap, ii) the predicted label c_j is correct, and iii) the confidence score p_j is higher than a preset threshold, the prediction is considered as a TP. Otherwise, it is considered as a FP. The 'sufficient degree' of overlap condition is operationalized with the intersection over union (IoU) areas for the predicted and ground-truth BB, i.e., $(b_j \cap b_i^g)/(b_j \cup b_i^g)$. An IoU > 0.5 is generally considered a good prediction if precise localization is not a priority [21]. In the research presented in this paper we follow this common practice.

2. *Compute average precision (AP) per class*. Based on the TPs and FPs, precision-recall pairs can be constructed by varying the confidence score threshold. From the obtained precision-recall pairs (allowing precision to be regarded as a downwards-sloping function of recall), the average precision (AP) per class can be derived from the area under the curve [23].

3. *Average the AP scores*. Finally, mAP is calculated by averaging the AP scores across classes.

In practice, an object detection model will now and then predict multiple BBs that sufficiently overlap and have matching class labels with above-threshold confidence scores. In this case the prediction with the highest IoU is the TP, and duplicates are designated as FPs.

There are two categories of object detection models: Two-stage region-based CNNs (RCNNs) [55–57], and one-stage detectors such as single shot detectors (SSDs) and the YOLO (You Only Look Once) algorithm family [58–60]. RCNNs are in general more accurate, whereas one-stage models are known for their near real-time performance in inference mode [54]. Given the goal of this study, speed is of minor importance, so we focus on the RCNN models. In RCNN models, first of all category-independent region proposals are generated from an image with a so-called Region Proposal Network, next CNN-like features are extracted from these regions, and then category-specific classifiers are used to determine the category labels of the proposals.

So far, object detection has been explored sparsely for automatic wildflower monitoring. Ärje et al. [61] applied it to record plant phenology from time-lapse images; they targeted only one flowering plant species on Greenland: *Dryas integrifolia*. Hicks et al. [39] used it to derive a nectar sugar mass index from detected floral resources. And Gallman et al. [62] adopted this AI technology along with drone-based image acquisition and the construction of georeferenced orthomosaics for detecting wildflowers in mountainous areas. The proposed solutions of Hicks et al. [39] and Gallman et al. [62] can identify and count 25 common wildflowers. Note that the Gallman et al. [62] dataset is not available, so it is not included in Table 1.

3. Eindhoven wildflower dataset

3.1 Collecting the EWD images

All 2,002 EWD images encompass untrampled soil area of approximately $1m^2$ and were taken (near-)vertically downward from a height in the range of 1.5–1.9m. Most images (91%) were collected near and in the city of Eindhoven, the Netherlands. This region is located in the northeastern corner of the Kempens Plateau. The soil primarily consists of cover sand deposited during glacial periods and is intersected by several streams, the Dommel being the most significant. The area has a temperate oceanic climate. The center of the city of Eindhoven is an urban heat island, characterized by higher temperatures compared to the surrounding rural areas, particularly in the summer.

Data collection is limited to non-woody herbaceous flowering plants. Maximum allowed vegetation height was 1m. Images were acquired with a 24-105mm/f4 lens mounted to a full-frame SLR camera (Canon Mark IV, 30.4 Mpixel). Exposure strategy: Fix ISO-setting at 800, maintain a shutter time faster than 1/50s to avoid motion blur caused by camera movement, and maximize aperture setting to create a large depth of focus.

EWD images were collected over two flowering seasons under various weather conditions in five different landscape types: i) roadsides (including dikes, bike trails, and railway borders; 832 images), ii) urban green spaces (parks, playgrounds, allotments; 237 images), iii) extensively farmed croplands (53 images), iv) weed-rich grasslands (457 images), v) marshland (including pools, fens, canals, and ditches; 423 images). The specific locations were selected not so much to create a dataset that is representative for the areas' ecologies, but instead to mitigate the long-tailed species distribution one normally finds in biological data. In other words, using our knowledge of local landscapes, we aimed to include sufficient variation in the dataset, so that the model would have enough examples of a wide range of wildflowers to train on.

During the field work a log with a species list per image was recorded. Wildflowers were identified in the field using the 24^{th} edition of the Heukels' Flora [63] supported by the ObsIdentify app [64]; the field worker (first author) has over 35 years of experience with botanizing. As said, the sampling strategy used can be summarized as a wisely chosen orchestration that ensures data collection with a fair number of wildflower species as well as a reasonably balanced number of instances per species. Contrary to [39] we did not avoid human elements in the images, such as pavement, litter, sewer gullies, etc. The abovementioned measures guarantee a rich reference dataset for developing AI models that support real-life wildflower monitoring.

To prevent disclosure of locations of rare and red list wildflowers GPS coordinates (latitude and longitude) are removed from the metadata of the collected images. No permits were required to collect the field data, all visited terrains (roadsides, urban parks, cropland, grassland, marshland) were open to the public and freely accessible.

3.2 Inflorescences and floral count units

All flowers are part of an inflorescence. Inflorescences vary in type among different plant families. An inflorescence type encodes both the branching pattern and the order in which its flowers open [65]. In this paper we adopt the formal, internationally agreed botanical inflorescence types [66, 67]. Some are relatively simple and form easy-to-outline structures (solitary flowers, spikes, and capitula or flower heads), others are more complex (dichasium, raceme, verticillaster, cyme, umbel, corymb, thyrse, panicle). For most wildflowers, inflorescences are excellent proxies for counting individual plants, as indicated by the EWD examples in Fig 2. By adopting inflorescences as floral count units (FCUs) we embody the building plan of wildflowers into

Fig 2. Examples of floral count units (FCUs) that match standard botanical inflorescence types. From top left to bottom right: a) FCU of *Papaver rhoeas* is a solitary flower, b) FCU of *Dactylorhiza praetermissa* is a spike, c) FCU of *Cirsium dissectum* is a capitulum (flower head), d) FCU of *Pedicularis palustris* is a raceme, e) FCU of *Salvia pratensis* (white circle) is a verticillaster, FCU of *Lotus corniculatus* (yellow circle) is a cyme, and f) FCU of *Cicuta virosa* is a compound umbel (typically containing 10 umbellets, each with up to 100 flowers). In each image only five FCUs are marked per species.

https://doi.org/10.1371/journal.pone.0302958.g002

the annotations. However, for a few species the standard inflorescences are impractical or ambiguous as floral count units in top-view relevées. For instance, it might be difficult or even impossible to outline the entire inflorescence, e.g., because the number of flowers varies too much (see Fig 3A), or it contains multiple flowers that typically bloom one at a time (see Fig 3B). In these circumstances we typically annotate single flowers. In case of nested inflorescences, the chosen FCU depends on how dense the secondary inflorescences are packed. As an example, for *Tanacetum vulgare* (corymb of capitula) the largest inflorescence type (corymb) is taken as FCU (see Fig 3C). One the other hand, for *Centaurea cyanus* (panicle of capitula) it is more convenient to take the individual flower heads as FCU (see Fig 3D). An important constraint for defining FCUs is to be consistent for look-alike species. For example, since the inflorescence and the FCU of *Bellis perennis* is a capitulum, a flower head should also be used for other EWD wildflowers that have white ray florets and yellowish disc florets, like *Leucanthemum vulgare*, *Matricaria chamomilla*, and *Erigeron annuus*. The complete list of FCUs that we

Fig 3. Examples of FCUs that do not match standard botanical inflorescence types. From top left to bottom right: a) The inflorescence of *Caltha palustris* is a cyme holding a variable number of 1 to 7 flowers. Since it is often impossible to outline these cymes (dashed grey circles are indicative) we therefore propose to use single flowers (white circles) as FCU. b) The stalked cyme of *Convolvulus arvensis* supports 1 to 2 flowers that bloom one after the other. White circles mark flowers that currently (t_0) bloom; the adjacent grey dashed circles indicate possible future or past positions (t^*) where flowers of the same cyme will appear (or disappeared). In this case we also propose to use single flowers as FCU. Right panel: c) The inflorescence of *Tanacetum vulgare* is a corymb of capitula. For this nested inflorescence it is convenient to use the flat-topped corymbs as FCU (large white circles) and not the individual flower heads (small grey dashed circles). d) Just the opposite: the inflorescence of *Centaurea cyanus* (white circle) and *Matricaria chamomilla* (yellow circle) is a panicle of capitula. It both cases it is impossible to outline the panicle (dashed grey circles are indicative) and we propose to use the individual flower heads (white circles) as FCU. In each image only five FCUs are marked per species.

https://doi.org/10.1371/journal.pone.0302958.g003

propose for each EWD species can be found in S1 Appendix. For 1 out of 5 flowering plant species the defined FCU does not have a one-to-one correspondence to its botanical inflorescence type.

Furthermore, each species-specific FCU is supplemented with two conversion factors obtained from manual inspection of digitized herbarium sheets collected in the Netherlands (near the city of Eindhoven) and registered as a collection item at the Naturalis Biodiversity Center. The first conversion factor represents the average number of flowers per FCU, and the second conversion factor represents the average number of FCUs per plant. These mappings extend FCU counts, as predicted by an object detection model, to other use cases such the assessment of nectar index, pollen density, winter food supply for birds, etc. and support demographic studies of plant populations.

In summary, what we register in each relevée are FCUs, defined for every species based on mostly practical considerations of countability. The FCUs can subsequently be converted into number of flowers or number of plants, using conversion tables.

3.3 Adding annotations to EWD

Adding annotations to real-world images is labor intensive and by nature a subjective process [68]. We took several precautions to make this process as objective as possible. First, EWD

images were annotated at species level by a plant expert making use of the field notes and respecting the species-specific FCUs (usually inflorescences) as listed in S1 Appendix. A few species that are visually indistinguishable in the field are merged: i) *Ranunculus acris* and *Ranunculus repens* are annotated as *Buttercup* (aggregate)*, ii) *Matricaria chamomilla* and *Matricaria maritima* are annotated as *Chamomile* (aggregate)*, and iii) unidentified yellow composites with only ray florets are marked as *Yellow Composite* (aggregate)*. Secondly, a 4-eyes principle was applied: Every annotation is double-checked.

To further formalize the annotations, we drafted a set of both generic and domain-specific guidelines. These rules make the annotation process exhaustive and consistent, hence ensuring high data quality. The generic guidelines comprise the following commonly accepted best practices for object detection:

1. *Draw a tight bounding box (BB) around an object of interest*. Objects of interest must be (reasonably) sharp.
2. *Annotate every object of interest*. Objects of interest that are present in images but not annotated 'confuse' AI models (they are considered background).
3. *Extend BBs of occluded objects*. Sometimes an object is partially covered by another object in an image. If that is the case, ensure that the occluded object is annotated with a BB as if it were in full view.
4. *Annotate cut-off objects at image borders if at least half of an object of interest is visible*.

Besides generic ones, domain-specific annotation guidelines are needed to cope with the difficulties shown in Fig 1. EWD is challenging since multiple species exhibit large visual similarities and because the set of images covers a variety of flower stages–roughly: in bud, blooming, fruiting stage–drastically impacting the appearance of the wildflowers. Therefore, we propose the following three additional rules:

5. *Use floral count units as entities for annotation*. For most flowering plant species FCUs map one-to-one to botanically defined inflorescences. See also S1 Appendix.
6. *Only annotate wildflowers in blooming stage*. As indicated above, wildflowers undergo changes in size, shape, and color throughout their life cycle. They develop from bud to blooming stage and eventually wither or transform into fruits. We only annotate wildflowers, or more precisely FCUs (rule#5), in blooming stage and consider two scenarios. First, for FCUs corresponding to solitary or single flowers, we annotate when the flowers meet the following two constraints, associated with the start and end of the blooming stage: i) the flowers are open AND at least half of the petals are unfolded and fresh, and ii) the colors of the flower are not yet fading AND more than half of the petals are still present. Second, for all other FCUs that map on inflorescences consisting of multiple flowers (spikes, capitula, racemes, umbels, etc.), we i) start annotating when at least one of the flowers belonging to the FCU is open and fresh, and ii) stop annotating when colors of the FCU fade OR blooming flowers become a minority within the FCU (i.e. most of them are already withered).
7. *Only annotate floral count units that are larger than 10mm* containing flowers that grow in easy-to-distinguish clusters. For solitary flowers 'size' refers to the dorsal-ventral axis. For other inflorescences it refers to the length of the largest axis. Motivation of this rule is that–given the striking visual similarities between closely related plant species–enough details of the FCU must be visible for proper identification by AI.

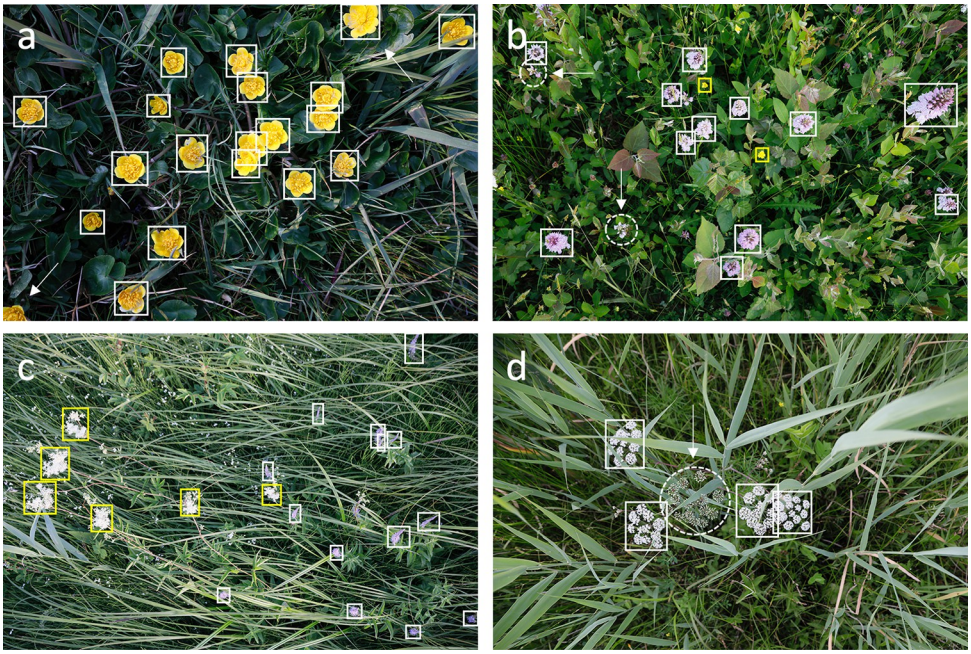

Fig 4. Annotated relevées compliant with the above guidelines. Note the tight and sometimes overlapping BBs (rule#1, rule#3). From top left to bottom right: a) *Caltha palustris*, the arrows point at cut-off flowers, they are only annotated when at least half of the FCU is visible (rule#4), b) *Dactylorhiza maculata* (white BBs), the arrows and dashed circles mark spikes that contain no open flowers (rule#6), and *Ranunculus repens* (yellow BBs) annotated as *Buttercup**, c) *Veronica longifolia* (white BBs) and *Filipendula ulmaria* (yellow BBs), the ubiquitous tiny 2-3mm *Galium palustre* 'speckles' are not annotated (rule#7), d) *Cicuta virosa*, the arrow and dashed circle mark an umbel that is still in bud and hence not annotated (rule #6).

https://doi.org/10.1371/journal.pone.0302958.g004

Fig 4 illustrates how these seven rules work out in practice. Strictly applying the above rules enable the AI to unlearn unwanted patterns, such as objects that are either out of scope (e.g., flowers that are in bud or withered) or of low quality (e.g., unidentified flowers that appear as unsharp colored blobs); by doing so they are treated as background. Annotations are added with the open-source tool labelImg and are stored per image in PASCAL VOC format [21]. After the annotation process EWD consists of 2,002 {<datetime-habitat>.jpg, <datetime-habitat>.xml} file pairs. The image file (*.jpg) and annotation file (*.xml) of each pair are given the same unique name that encodes date and time of the image exposure followed by one of the five landscape types.

Summarizing, EWD holds 65,571 annotations for 160 species and exhibits a long-tailed species distribution (which is typical for biological field data); 19 wildflowers have 1,000 or more annotations per species, 22 wildflowers have annotation counts in the range [500, 999], 61 wildflowers have annotation counts in the range [100, 499], and 58 species have fewer than 100 annotations. The most common species in EWD is *Buttercup* (aggregate)* with 4,196 annotations, followed by *Berteroa incana* with 3,177 annotations. The number of annotations per image varies from 0 (some images have no annotations, e.g., because the depicted wildflowers are not in blooming stage) to a maximum 704 with an average of 32.75 annotations per image. The number of annotated species per image varies between 0 and 7, with a median of 1. Table 2 presents the EWD key figures per month.

Per species statistics of EWD, along with the defined FCU, the average number of individual flowers per FCU and the average number of FCUs per flowering plant, are given in S1 Appendix.

Table 2. EWD key figures per month.

Blooming period	# images	# images per landscape type {R, U, C, G, M}	# species	# annotations
Year 2021	**994**	{**419**, **122**, **42**, **220**, **191**}		**38,515**
April	75	{18, 14, —, 38, 5}	10	3,103
May	296	{69, 31, 12, 123, 61}	46	14,297
June	205	{108, 18, 25, 13, 41}	51	11,585
July	215	{63, 40, 5, 27, 80}	56	4,745
August	146	{125, 19, —, —, 2}	45	3,421
September	57	{36, —, —, 19, 2}	20	1,364
Year 2022	**1,008**	{**413**, **115**, **11**, **237**, **232**}		**27,056**
March	68	{9, 9, —, —, 50}	10	1,765
April	122	{19, 34, —, 66, 3}	16	3,983
May	209	{99, 13, —, 63, 34}	54	6,661
June	298	{102, 49, 11, 51, 85}	75	7,817
July	211	{111, 10, —, 48, 42}	60	4,451
August	80	{59, —, —, 5, 16}	47	1,767
September	20	{14, —, —, 4, 2}	16	612

The landscape types are abbreviated with R (roadside), U (urban green), C (cropland), G (grassland), and M (marsh).

https://doi.org/10.1371/journal.pone.0302958.t002

4. Data-centric AI demonstrator

To demonstrate the potential use of EWD, we trained an object detection model based on industry standard tools. More specifically, we used the open-source PyTorch framework and selected a state-of-the-art F-RCNN model [69], as provided in the TorchVision library.

This end-to-end deep learning object detection model consists of a visual encoder (ResNet50) with pretrained weights followed by a classification and localization head that jointly learn in a multi-task setting [57, 70]. The choice for transfer learning with standard tools was motivated by the fact that we do not necessarily aim to train the best model possible but rather seek to explore the impact of improving the quality of the data on the model's ability to estimate species richness and abundance. Thus, we opted for creating a baseline model with a relatively simple and fast training procedure and put more effort into preprocessing and balancing the data.

4.1. Preprocessing the data

Given that the EWD images have a resolution of 6,720x4,480 pixels and that the pretrained model has a maximum input size that is significantly lower (1,333 pixels on either axis), the default approach of automatically resizing the image to match the maximum input size would drastically reduce the resolution of the EWD data. This might inhibit the model's ability to correctly identify detected wildflowers. Our solution to this problem was to use a mosaic of the original EWD images by cutting them into tiles that are (roughly) equal to the maximum input size, thus eliminating or strongly reducing the degradation of image resolution. However, as shown in Fig 5, a simple cutting approach would inadvertently cut through annotated FCUs, which would make them less suitable as training data. Therefore, we devised a cutting approach that is aware of the presence of annotations in the image and attempts to find a less destructive way to cut the image into a mosaic. This produces a trade-off situation between perfectly matching the maximum input size of the F-RCNN model on the one hand and cutting through none of the annotations on the other hand. Our approach uses a 20% margin on

Fig 5. Smart tiling approach for the annotated EWD images. For the mosaic of the 30.4 Mpixel EWD images a setting was chosen that results in 15 (5×3) tiles. Starting with equally sized tiles (blue dashed lines) margin modifications of +/- 20% are allowed to minimize object cuts. In this example, the number of object cuts is reduced from nine (blue BBs) to four (white BBs). To prevent double counts, only the largest part of cut-off objects (yellow BBs)–in accordance with rule #4 of Section 3.3 –is added to the set of tiled annotations. White and yellow BBs are drawn a bit larger for cosmetic reasons.

https://doi.org/10.1371/journal.pone.0302958.g005

the model's maximum input size and then finds the cut lines that are least destructive within that margin.

4.2 Balancing the data

The next issue to address is the highly imbalanced nature of the EWD, having some wildflowers being overly abundant and many others being rather rare. Using a long-tailed dataset such as the EWD as training data is likely to create bias in the model because of the classifier having insufficient training opportunity on sparsely represented classes and will probably favor over-represented species [71]. Again, rather than manipulating the training process [72], e.g., by using over/under-sampling techniques, or even adopting complex learning solutions such as adversarial neural networks that can potentially correct for bias to some degree [73, 74], we chose a solution focused on the input data.

The solution we propose is to train with a balanced subset of EWD, i.e. use a fixed predefined number of annotations per species for training, evaluating, and testing. To create this subset, we went through the tiled images of EWD, selecting a combination of tiles that produces the required numbers. A similar approach was adopted by Koch et al. [75] for bird images extracted from a Norwegian citizen science platform. For our object detection case this is not a straightforward process as, the selecting of an image tile containing annotations for one specific species may collaterally include annotations of other species as well. Although it is theoretically possible to calculate all tile permutations of possible solutions that bring the exact number of required annotations per species maximizing the total number of species included, this is in practice not feasible due to its extremely high computational load. Hence, we followed a near-ideal heuristic approach. This approach guarantees that the preset number of required annotations per species are selected but as a drawback may drop a flowering plant species

when a solution is not found within a certain number of attempts. Following this approach, we generated a stratified partition for 49 flowering plant species with the following fixed number of annotations per species: 250 for training, 50 for evaluation, and 50 for testing. Note that these numbers are not hard-coded but parameterized, see S2 Train & Test Notebook.

When dividing the tiles across training, evaluation, and testing sets, we did not use any constraint on what original image the tiles were taken from. This means that tiles taken from the same image could be assigned to different sets. For example, four out of the six tiles taken from one image might be part of the training set, one of the evaluation set and one of the test set. We adopted this approach to maximize the number of tiles used for training the model.

4.3 Training the model

We used a relatively simple and fast training process with standard data augmentation practices and a transfer learning approach for the 49 object classes (flowering plant species) and one additional catch-all background category. Training happened in train-and-evaluate loops, called epochs. Instead of fixing the number of epochs, after every epoch the evaluation metric 'mAP @IoU > 0.5' was compared to the average of this metric for the last five epochs, and if the value for the current epoch was lower than this average, the training process was stopped. In practice, the backpropagation process needed 13 to 15 epochs to converge, see S2 Train & Test Notebook.

4.4 Performance of the model

The optimized model after training is able to generate predictions consisting of BB proposals with species labels and confidence scores. The overall model performance (Section 2.4) on the holdout test set has a mAP of 0.82 (@IoU > 0.5).

The AP results for each of the flowering plant species in the test set are shown in Fig 6. For the first 48 species the AP score gradually decreases from 0.99 for *Centaurea cyanus* to 0.45 for *Ranunculus aquatilis*. The rightmost species *Anthriscus sylvestris* is difficult to detect and shows a sharp drop to an AP of 0.19. The AP score is weakly correlated with the complexity of the species FCU. Most high score species have simple FCUs (solitary flowers, spikes, capitula), whereas the majority of wildflowers with low scores have complex FCUs (like racemes, cymes, corymbs and umbels).

More insight into counts and miscounts of the model is provided by Fig 7, which shows the confusion matrix for the test set with detections that have a confidence score > 0.5. The rows

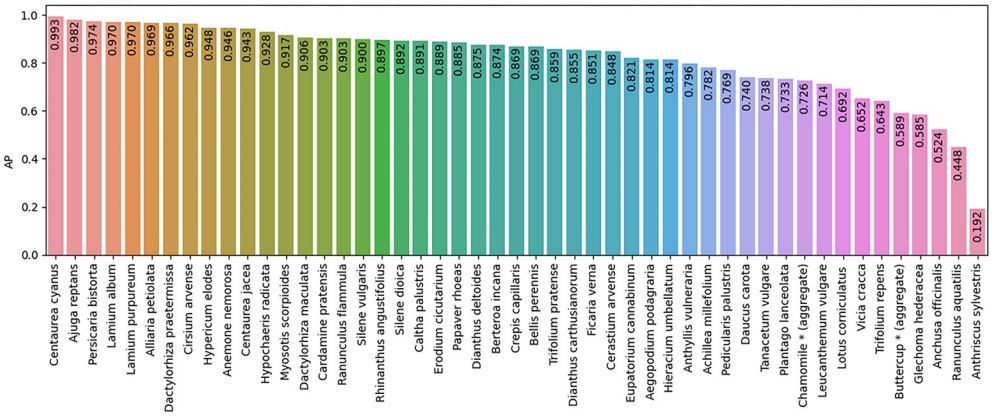

Fig 6. Average precision (AP) for each the 49 flowering plants.

https://doi.org/10.1371/journal.pone.0302958.g006

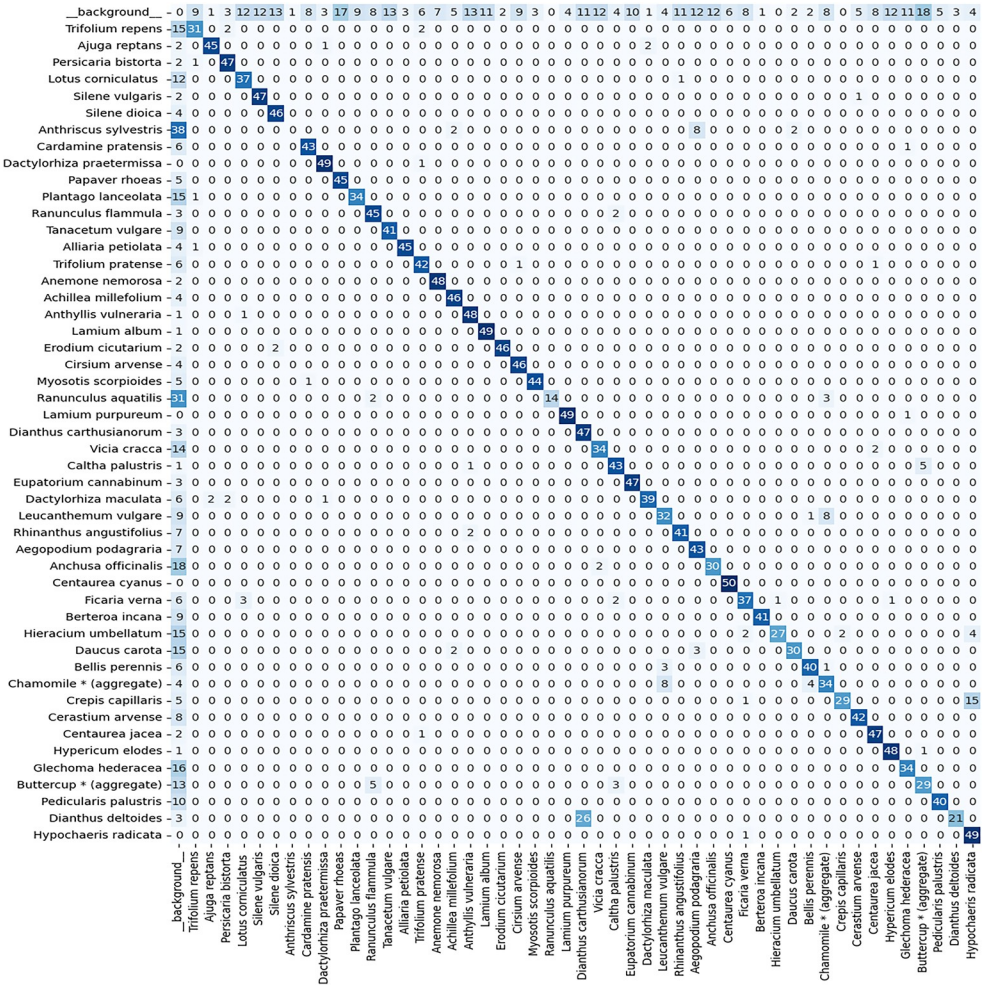

Fig 7. Confusion matrix for the holdout test set. The rows hold true wildflower annotations, the columns contain detections generated by the baseline model. The row labelled _background_ refers to spurious detections (false positives); the column labelled _background_ holds undetected wildflowers (misses). The confidence score threshold used to construct this confusion matrix is set at 0.5.

https://doi.org/10.1371/journal.pone.0302958.g007

of the matrix hold the actual annotations (they add up to 50 for each species); the columns indicate predictions. A perfect object detection model results in a confusion matrix with no off-diagonal counts.

Note that the diagonal is by far the most conspicuous pattern in the confusion matrix, indicating that the model generalizes well and is able to come up with reasonable wildflower counts retrieved from unseen test images. Disregarding the spurious detections, the overall percent correct score is 79.2%, with chance level performance at 1/49 = 2.0%. Although this percent correct score is high, there are a few other noticeable confusion patterns. First of all, the row labelled as _background_ shows that there are quite some spurious wildflower detections in the background parts of the image, and secondly, the column labelled as _background_ demonstrates that a fair number of annotated wildflowers still go undetected. Species that have particularly low diagonal scores are *Anthriscus sylvestris* (also lowest AP value), *Ranunculus aquatilis*, and *Dianthus deltoides*. Many *Anthriscus sylvestris* annotations (38 out of 50) and *Ranunculus aquatilis* annotations (31 out of 50) go undetected because they have low confidence scores (below the threshold) and are therefore treated by the model as

background. *Dianthes deltoides* counts suffer less from the abovementioned 'disappearance-in-the-background' problem (3 out of 50) but are mainly confused with *Dianthes carthusianorum* (26 out of 50), a species from the same genus that is morphologically similar. On the other end of the spectrum, species that can be counted very well by the model are the eye-catching sky-blue *Centaurea cyanus* (50 diagonal hits; also highest AP value), *Dactylorhiza praetermissa*, *Lamium album*, *Lamium purpureum* and *Hypochaeris radicata* (49 diagonal hits). Furthermore, a notorious confusion is found between the two look-alike yellow composite species *Hypochaeris radicata* and *Crepis capillaris*.

4.5 Inference with the model

Whereas training, evaluation and testing is about *creating* an optimized model and probing its performance, inference is about *applying* a trained model to generate predictions for new, unseen data which is not labelled. Only if the input used for inference is of the same nature and quality as the input used for training, evaluation and testing, it can be assumed that the model generalizes well [76].

In our case, this means that top-view images of 1 m^2-sized relevées, with a resolution comparable to EWD images, should be used in order to obtain reliable wildflowers counts. Moreover, since our baseline model is trained and tested with tiles, these unseen images should also be divided in a 5x3 mosaic of tiles. Because these unseen images have no annotations (there is no à priori information in inference mode) we partition the image in equally-sized tiles, which are offered one-by-one to the baseline model, see S3 Inference Notebook.

5. Discussion

This paper presents two data products: i) a reference dataset for wildflower monitoring, and ii) an object detection model that can identify and count flowering plants from top-view images.

The EWD dataset holds 2,002 images with 65.6K expert annotations for 160 species, belonging to 31 different families and is collected in two entire flowering seasons (begin of March till end of September). We have substantially extended FCU definitions for species with more complex (nested) inflorescences and our annotation process is supported by strict guidelines. This makes EWD a reference dataset that can be used for benchmarking computer vision algorithms. Moreover, we have added complementary botanical information to convert detected FCUs to individual flower or whole plant counts, thereby supporting other use cases related to wildflower monitoring.

Our baseline model–trained on a balanced subset of EWD–enables automated wildflower monitoring for 49 species with a mAP of 0.82. This high score makes the knowledge encapsulated by model actionable, i.e., it can be applied in practice for automatic large-scale wildflower monitoring. Other F-RCNN object detection models for wildflower monitoring, i.e., [39, 62], having 25K and 10K annotations respectively, collected wildflower data in a limited period within one flowering season (May till August) and their models are able to identify and count only 25 flowering plant species, half the number of species that our model can handle. In contrast to our study, they did not follow a strict annotation protocol and focused primarily on the model, not on the data perse. Despite being trained with fewer species, the model of Hicks et al. [39] achieved a similar FCU error detection rate of 1–2 in 10 (precision 86%, recall 88%), without reporting the standard mAP metric. The Gallman et al. study [62] did not explicitly use FCUs and reported a mAP score of 0.48.

5.1. Quality of annotations

The first data product, EWD, is an expert-annotated dataset. Platforms like Amazon's Mechanical Turk have made it possible to outsource annotation work as a crowdsourced task,

at relatively low cost. Although several methods–either ex-ante or ex-post–exist to correct for unreliable labels of crowd workers [77–80], this remains a sub-optimal approach [81]. High-quality consistent expert annotations along with a well-defined data acquisition protocol are paramount for a data-centric AI approach. They turn datasets into reference datasets. By providing a solid ground-truth, reference datasets allow for benchmark studies, varying from comparing different computer vision algorithms, developing metrics, to evaluating innovative techniques such as semi-supervised learning [82, 83] that have the potential to alleviate the burden of adding huge numbers of manual annotations.

Although the quality of the EWD annotations is even further enhanced by formulating explicit guidelines (Section 3.3), some rules are hard to objectify. Especially rule#1 could be made more explicit by exploring the use of state-of-the-art no-reference image quality metrics for sharpness [84].

5.2 Model limitations and possible improvements

The second data product is our wildflower object detection model. It builds upon a F-RCNN backbone; an end-to-end deep learning architecture that is a proven off-the-shelf solution. The model demonstrates the potential use of the EWD dataset. It can be adapted easily to accommodate more wildflowers by reducing the required number of annotations per species (currently set at 250 for training, 50 for evaluation, and 50 for testing). The main motivation of the modelling choices made is preserving data quality. A critical reflection on the modelling choices is given below.

Although our preprocessing solution–partitioning the image and feeding high resolution tiles to the F-RCNN–avoids the tenacious 'small object detection' problem [85], it creates a minor artifact as well. In particular, the smaller parts of cut-off objects (for which the annotations are removed, see Fig 5) generate some spurious detections, even though the model is trained with annotations that contain at least half of the FCU (see Section 3.3; rule #4). Another artifact arises from balancing the data. Because of this, the model is trained on a subset of flowering plant species and might generate false hits in those few test image tiles that happen to hold look-alike species on which it was not trained. Both imperfections account for the relative high number of hits in the _background_ row of Fig 7.

For several reasons, we used transfer learning instead of 'training-from-scratch'. First of all, bypassing the vast number of weights that belong to visual encoders in the training process keeps the model optimizing process fast and flexible. In addition, it strongly reduces the carbon footprint of training the AI [86]. Finally, recent literature [87] indicates that using pre-trained networks in many practical situations leads to performance gains. Hence, we stick to the transfer learning paradigm, and explore in our future work: i) cost-function engineering tailored to long-tailed distributions, such as proposed by Tan et al. [88], and ii) more fundamental 'beyond vision' solutions as put forward below.

The most important shortcoming of object detection models is that proposed bounding boxes with a class and score are considered as independent entities and that predictions are only based on local visual content. They have no mechanism to explicitly relate localized predictions in the same image to one another. In the physical world, visual objects usually coexist with other related objects (a fortiori for wildflowers), and human perception very much relies on this fact [89]. Henceforth, an intelligent agent or voting procedure that considers neighboring wildflowers might boost model performance significantly. To give a trivial example: the best species predictor for a non-canonical view of an occluded flower that is hard to recognize is probably a look-alike species (same color) in the proximity. So, picking matches from neighboring species labels for predictions with a low confidence score might be a promising direction to explore.

Another limitation is that our wildflower monitoring framework is agnostic with respect to flowering time and landscape type. Ecologists implicitly include this type of information while doing fieldwork, i.e., they implicitly rule out or confirm–based on date and habitat–'suggestions' created by their visual system. Modifying confidence scores of wildflowers by explicitly encoding this type of domain knowledge is envisioned as another promising future direction.

5.3 Ongoing work and outlook

Current work focuses on extending the 'AI demonstrator' (Section 4) with: i) a user interface supporting interaction with different stakeholders [90], and ii) a mobile sensor with guidance for correct image acquisition and streaming cloud storage. These mobile sensors could be drones for professional use or smartphones for the general public.

In future work the presented method for automated wildflower monitoring will be validated in the field and extended to other flowering plant species. It is a valuable tool that can be used for ecological and environmental research, e.g., to investigate whether wildflower occurrence and richness can be used as a proxy for mowing regimes, nitrogen depletion or climate change. Policy makers might use the results of automated wildflower monitoring for spatial planning activities and biodiversity conservation [91]. In the long run, we envisage that AI-enabled monitoring solutions will be tied to more coarse-grained ecological indicators [92].

Finally, we hope to motivate and engage AI researchers to use the EWD reference dataset and create improved solutions for automatic wildflower monitoring.

6. Conclusions

Humans have a knotty relation with nature. On the one hand, using technology we put a tangible mark on our planet causing climate change and an unprecedented biodiversity loss. On the other hand, our technology has the potential to shape transitions towards a more sustainable world. This study belongs to the latter category. It aims at showcasing automated wildflower monitoring using AI. Wildflower diversity is a profound part of our natural capital.

More specifically, this paper demonstrates how the quality of automated wildflower monitoring can be enhanced by adopting a data-centric AI approach. Distinct aspects of our work that treat data as a first-class citizen are: i) a protocol for collecting relevées (Section 3.1), ii) the introduction of FCUs for a wide variety of flowering plant species (Section 3.2), iii) expert annotations according to strict guidelines (Section 3.3), iv) preprocessing that preserves image resolution (Section 4.1), and v) training the F-RCNN-ResNet50 object detection model with a balanced subset in order to obtain unbiased predictions (Section 4.2).

Our baseline model can recognize and count 49 flowering plant species with a mAP of 0.82, thereby outperforming contemporary automated wildflower monitoring solutions. We hope to inspire and encourage the computer vision research community to create improved AI-enabled solutions for wildflower monitoring using EWD as a reference dataset.

Supporting information

S1 Appendix. EWD species characteristics and statistics.
(DOCX)

S1 File. Train & test notebook: HTML printout of the Python code to train and test the model.
(HTML)

S2 File. Inference notebook: HTML printout of the Python code to do inference with the model.
(HTML)

Author Contributions

Conceptualization: Gerard Schouten.

Data curation: Gerard Schouten, Barbara Gravendeel.

Formal analysis: Bas S. H. T. Michielsen.

Investigation: Gerard Schouten.

Software: Bas S. H. T. Michielsen.

Visualization: Gerard Schouten, Bas S. H. T. Michielsen.

Writing – original draft: Gerard Schouten.

Writing – review & editing: Bas S. H. T. Michielsen, Barbara Gravendeel.

References

1. Niklas KJ. The evolutionary biology of plants. Chicago: University of Chicago Press; 1997.
2. Crane PR, Friis EM, Pedersen KR. The origin and early diversification of angiosperms. In: Gee H, editors. Shaking the Tree. Readings from Nature in the History of Life. Chicago: University of Chicago Press; 2000. pp.233–250.
3. Ollerton J, Winfree R, Tarrant S. How many flowering plants are pollinated by animals? Oikos. 2011 Feb 21; 120(3):321–326. https://doi.org/10.1111/j.1600-0706.2010.18644.x
4. Crepet WL, Niklas KJ. Darwin's second "abominable mystery": Why are there so many angiosperm species? Am J Bot. 2009 Jan 1; 96(1):366–381. https://doi.org/10.3732/ajb.0800126 PMID: 21628194
5. Gravendeel B. Towards climate adaptive urban jungles. Nijmegen: Radboud University; 2022.
6. Houghton PJ. The role of plants in traditional medicine and current therapy. J Altern Complement Med. 1995; 1(2):131–143. https://doi.org/10.1089/acm.1995.1.131 PMID: 9395610
7. Fowler MW. Plants, medicines and man. J Sci Food Agric. 2006 Aug 9; 86(12):1797–1804. https://doi.org/10.1002/jsfa.2598
8. Mace GM, Norris K, Fitter AH. Biodiversity and ecosystem services: a multilayered relationship. Trends Ecol Evol. 2011 Sep 22; 27(1):19–26. https://doi.org/10.1016/j.tree.2011.08.006 PMID: 21943703
9. Brondizio ES, Settele J, Díaz S, Ngo HT. Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Bonn: IPBES secretariat; 2019. https://doi.org/10.5281/zenodo.3831673
10. Sutherland WJ, Pullin AS, Dolman PM, Knight TM. The need for evidence-based conservation. Trends Ecol Evol. 2004 Jun; 19(6):305–308. https://doi.org/10.1016/j.tree.2004.03.018 PMID: 16701275
11. Diaz S, Lavorel F, de Bello F, Quétier F, Grigulis K, Robson TM. Incorporating plant functional diversity effects in ecosystem service assessments. Proceedings of the National Academy of Sciences; 2007 Dec 26; 104(52):20684–20689. https://doi.org/10.1073/pnas.0704716104 PMID: 18093933
12. Klein, DJ, McKown MW, Tershy BR. Deep learning for large scale biodiversity monitoring. Bloomberg Data for Good Exchange Conference; 2015 Sep 28; New York.
13. Pl@ntNet: Identify, explore and share your observations of wild plants [internet]. 2024 [Cited 2024 Feb 15]. Available from: https://identify.plantnet.org
14. PlantSnap: A plant expert at your fingertips [Internet]. 2024 [Cited 2024 Feb 15]. Available from: https://www.plantsnap.com
15. Gaston KJ O'Neill MA. Automated species identification: why not? Philos Trans R Soc Lond B Biol Sci. 2004 Apr 29; 359(1444):655–667. https://doi.org/10.1098/rstb.2003.1442
16. Pilgrim SE, Cullen LC, Smith DJ, Pretty J. Ecological knowledge is lost in wealthier communities and countries. Environ Sci Technol. 2008 Jan 8; 42(4):1004–1009. https://doi.org/10.1021/es070837v PMID: 18351064

17. Schermer M, Hogeweg L. Supporting citizen scientists with automatic species identification using deep learning image recognition models. Biodivers Inf Sci Stand. 2018 May 17; e25268. https://doi.org/10.3897/biss.2.25268

18. Zou Z, Shi Z, Guo Y, Ye J. Object detection in 20 years: A survey. Proceedings of the IEEE. 2023 Jan 27; 111(3):257–276. https://doi.org/10.1109/JPROC.2023.3238524

19. Whang SE, Roh Y, Song H, Lee JG. Data collection and quality challenges in deep learning: A data-centric AI perspective. VLDB J. 2023 Jan 3; 32:791–813. https://doi.org/10.1007/s00778-022-00775-9

20. Rädsch T, Reinke A, Weru V, Tizabi MD, Schreck N, Kavur AE, et al. Labelling instructions matter in biomedical image analysis. Nat Mach Intell. 2023 Mar 2; 5:273–283. https://doi.org/10.1038/s42256-023-00625-5

21. Everingham M, van Gool L, Williams CKI, Winn J, Zisserman A. The Pascal Visual Object Classes (VOC) challenge. Int J Comput Vis. 2009 Sep 9; 88(2):303–338. https://doi.org/10.1007/s11263-009-0275-4

22. Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. Microsoft COCO: Common objects in context. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. Computer Vision–ECCV 2014, Lecture Notes in Computer Science. 2014 Sep 6–12; Zurich, Switzerland. 8693:740–755. https://doi.org/10.1007/978-3-319-10602-1_48

23. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet Large Scale Visual Recognition Challenge. Int J Comput Vis. 2015 Apr 11; 115(3):211–252. https://doi.org/10.1007/s11263-015-0816-y

24. Kuznetsova A, Rom H, Alldrin N, Uijlings J, Krasin I, Pont-Tuset J, et al. The open images dataset V4. Int J Comput Vis. 2020 Mar 13; 128(7):1956–1981. https://doi.org/10.1007/s11263-020-01316-z

25. Codella NCF, Nguyen Q-B, Pankanti S, Gutman DA, Helba B, Halpern AC, et al. Deep learning ensembles for melanoma recognition in dermoscopy images. IBM J Res Dev. 2017 Sep 8; 61(4/5):1–15. https://doi.org/10.1147/JRD.2017.2708299

26. Raumanns R, Schouten G, Joosten M, Pluim JPW, Cheplygina V. ENHANCE (ENriching Health data by ANnotations of Crowd and Experts): A case study for skin lesion classification. J Mach Learn Biomed Imaging. 2021 Dec 31; 1:1–26. https://doi.org/10.59275/j.melba.2021-geb9

27. Wen D, Khan SM, Xu AJ, Ibrahim H, Smith L, Caballero J, et al. Characteristics of publicly available skin cancer image datasets: A systematic review. Lancet Digit Health. 2021 Nov 9; 4(1):e64–e74. https://doi.org/10.1016/S2589-7500(21)00252-1 PMID: 34772649

28. Zaffino P, Marzullo A, Moccia S, Calimeri F, De Momi E, Bertucci B, et al. An open-source COVID-19 CT dataset with automatic lung tissue classification for radiomics. Bioengineering. 2021 Feb 16; 8(2):26. https://doi.org/10.3390/bioengineering8020026 PMID: 33669235

29. Ghose P, Alavi M, Tabassum M, Ashraf U, Biswas M, Mahbub K, et al. Detecting COVID-19 infection status from chest X-ray and CT scan via single transfer learning-driven approach. Front Genet. 2022 Sep 21; 13. https://doi.org/10.3389/fgene.2022.980338 PMID: 36212141

30. Geyer J, Kassahun Y, Mahmudi M, Ricou X, Durgesh R, Chung AS, et al. A2D2: Audi autonomous driving dataset. arXiv:2004.06320v1 [Preprint]. 2020 [cited 2024 Feb 19]. Available from: https://arxiv.org/abs/2004.06320

31. Sun T, Segu M, Postels J, Wang Y, van Gool L, Schiele B, et al. SHIFT: A synthetic driving dataset for continuous multi-task domain adaptation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022 Jun 18–24; New Orleans, USA. pp.21371–21382. https://doi.org/10.1109/CVPR52688.2022.02068

32. Huston P, Edge VL, Bernier E. Reaping the benefits of open data in public health. Can Commun Dis Rep. 2019 Oct 3; 45(11):252–256. https://doi.org/10.14745/ccdr.v45i10a01 PMID: 31647060

33. Maier-Hein L, Reinke A, Godau P, Tizabi MD, Buettner F, Christodoulou E, et al. Metrics reloaded: Pitfalls and recommendations for image analysis validation. arXiv:2206.01653v7 [Preprint]. 2023 [cited 2024 Feb 19]. Available from: https://arxiv.org/abs/2206.01653

34. Nilsback M-E, Zisserman A. A visual vocabulary for flower classification. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). 2006 Jun 17–22; New York, USA. 2:1447–1454. https://doi.org/10.1109/CVPR.2006.42

35. Nilsback M-E, Zisserman A. Automated flower classification over a large number of classes. Proceedings of the 6th Indian Conference on Computer Vision, Graphics & Image Processing. 2008 Dec 16–19; Bhubaneswar, India. pp.722–729. https://doi.org/10.1109/ICVGIP.2008.47

36. Seeland M, Rzanny M, Alaqraa N, Wäldchen J, Mäder P. Plant species classification using flower images–A comparative study of local feature representations. PLoS One. 2017 Feb 24; 12(2):e0170629. https://doi.org/10.1371/journal.pone.0170629 PMID: 28234999

37. Krishna NH, Rakesh M, Ram Kaushik R. Plant species identification using transfer learning–PlantCLEF 2020. CLEF 2020 Working Notes. 2020 Sep 22–25; Thessaloniki, Greece.

38. Zheng Y, Zhang T, & Fu Y. A large-scale hyperspectral dataset for flower classification. Knowl Based Syst. 2022 Jan 25; 236:107647. https://doi.org/10.1016/j.knosys.2021.107647

39. Hicks D, Baude M, Kratz C, Ouvrard P, Stone G. Deep learning object detection to estimate the nectar sugar mass of flowering vegetation. Ecol Solut Evid. 2021 Sep 22; 2(3):e12099. https://doi.org/10.1002/2688-8319.12099

40. Liu Z, Miao Z, Zhan X, Wang J, Gong B, Yu SX. Large-scale long-tailed recognition in an open world. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019 Jun 15–20; Long Beach, USA. pp.2532–2541. https://doi.org/10.1109/CVPR.2019.00264

41. Cho S-Y, Lim P-T (2006). A novel virus infection clustering for flower images identification. Proceedings of the 18th International Conference on Pattern Recognition (ICPR). 2006 Aug 20–24; Hong Kong, China. pp.1038–1041. https://doi.org/10.1109/ICPR.2006.144

42. Hong S-W, Choi L. Automatic recognition of flowers through color and edge-based contour detection. Proceedings of the 3rd International Conference on Image Processing Theory, Tools and Applications (IPTA). 2012 Oct 15–18; Istanbul, Turkey. pp.141–146. https://doi.org/10.1109/IPTA.2012.6469535

43. Tuytelaars T, Mikolajczyk K. Local invariant feature detectors: A survey. Found Trends in Comput Graph Vis. 2008 Jun 15; 3(3):177–280. https://doi.org/10.1561/0600000017

44. Lowe DG. Distinctive image features from scale-invariant keypoints. Int J Comput Vis. 2004 Nov; 60 (2):91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

45. Dalal N, Triggs B. Histograms of oriented gradients for human detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). 2005 Jun 20–25; San Diego, USA. 1:886–893. https://doi.org/10.1109/CVPR.2005.177

46. Wäldchen J, Mäder P. Plant species identification using computer vision techniques: A systematic literature review. Arch Comput Methods Eng. 2017 Jan 7; 25(2):507–543. https://doi.org/10.1007/s11831-016-9206-z PMID: 29962832

47. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS). 2012 Dec 3–6; Lake Tahoe, USA. 1:1097–1105.

48. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE Conference of Computer Vision and Pattern Recognition (CVPR). 2016 Jun 27–30; Las Vegas, USA. pp.770–778. https://doi.org/10.1109/CVPR.2016.90

49. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556v6 [Preprint]. 2015 [cited 2024 Feb 20]. Available from: https://arxiv.org/abs/1409.1556

50. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. Proceedings of the IEEE Conference of Computer Vision and Pattern Recognition (CVPR). 2016 Jun 27–30; Las Vegas, USA. pp.2818–2826. https://doi.org/10.1109/CVPR.2016.308

51. Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, et al. A comprehensive survey on transfer learning. Proceedings of the IEEE. 2020 Jul 7; 109(1):43–76. https://doi.org/10.1109/JPROC.2020.3004555

52. Wäldchen J, Rzanny M, Seeland M, Mäder P. Automated plant species identification—Trends and future directions. PLoS Comput Biol. 2018 Apr 5; 14(4):e1005993. https://doi.org/10.1371/journal.pcbi.1005993 PMID: 29621236

53. Fei Y, Li Z, Zhu T, Ni C. A lightweight attention-based convolutional neural networks for fresh-cut flower classification. IEEE Access. 2023 Feb 13; 11:17283–17293. https://doi.org/10.1109/ACCESS.2023.3244386.

54. Liu L, Ouyang W, Wang X, Fieguth P, Chen J, Liu X, et al. Deep learning for generic object detection: A survey. Int J Comput Vis. 2019 Oct 31; 128(2):261–318. https://doi.org/10.1007/s11263-019-01247-4

55. Girshick R, Donahue J, Darell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2014 Jun 23–28; Columbus, USA. pp.580–587. https://doi.org/10.1109/CVPR.2014.81

56. Girshick R. Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2015 Dec 7–13; Santiago, Chile. pp.1440–1448. https://doi.org/10.1109/ICCV.2015.169

57. Ren S, He K, Ross G, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2016 Jun 6; 39(6):1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031 PMID: 27295650

58. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016 Jun 27–30; Las Vegas, USA. pp.779–788. https://doi.org/10.1109/CVPR.2016.91

59. Bochkovskiy AW, Wang C-Y, Liao H-YM. YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934v1 [Preprint]. 2020 [cited 2024 Feb 20]. Available from: https://arxiv.org/abs/2004.10934

60. Xue S, Li Z, Wu R, Zhu T, Yuan Y, Ni C. Few-shot learning for small impurities in tobacco stems with improved YOLOv7. IEEE Access. 2023 May 10. 11:48136–48144. https://doi.org/10.1109/ACCESS.2023.3275023

61. Ärje J, Milioris D, Tran DT, Jespen JU, Raitoharju J, Gabbouj M, et al. Automatic flower detection and classification system using a light-weight convolutional neural network. EUSIPCO Workshop on Signal Processing, Computer Vision and Deep Learning for Autonomous Systems. 2019 Sep 6; Coruña, Spain.

62. Gallmann J, Schüpbach B, Jacot K, Albrecht M, Winizki J, Kirchgessner, et al. Flower mapping in grasslands with drones and deep learning. Front Plant Sci. 2022 Feb 9; 12. https://doi.org/10.3389/fpls.2021.774965 PMID: 35222449

63. Duistermaat L (2020). Heukels' Flora van Nederland. Groningen; Noordhoff Uitgevers bv; 2020.

64. ObsIdentify: Recognize nature in one click! [Internet]. 2024 [Cited 2024 Feb 20]. Available from: https://observation.org/apps/obsidentify/

65. Benlloch R, Berbel A, Serrano-Mislata, A, Madueño F. Floral initiation and inflorescence architecture: A comparative view. Ann Bot. 2007 Aug 16; 100(3):659–676. https://doi.org/10.1093/aob/mcm146

66. Rickett HW. The classification of inflorescences. Bot Rev. 1944 Mar; 10:187–231. https://doi.org/10.1007/BF02861094

67. Weberling F. Morphology of flowers and inflorescences. Cambridge: Cambridge University Press; 1989.

68. Nowak S, Rüger S. How reliable are annotations via crowdsourcing: A study about inter-annotator agreement for multilabel image annotation. Proceedings of the International Conference on Multimedia Information Retrieval. 2020 Mar 29; Philadelphia, USA. pp.557–566. https://doi.org/10.1145/1743384.1743478

69. Li Y, Xie S, Chen X, Dollar P, He K, Girshick R. Benchmarking detection transfer learning with vision transformers. arXiv:2111.11429v1 [Preprint]. 2021 [cited 2024 Feb 21]. Available from: https://arxiv.org/abs/2111.11429

70. Zhang Y, Yang Q. A survey on multi-task learning. IEEE Transactions on Knowledge and Data Engineering. 2021 Mar 31; 34(12):5586–5609. https://doi.org/10.1109/TKDE.2021.3070203

71. Li Y, Wang T, Kang B, Tang S, Wang C, Li J. Overcoming classifier imbalance for long-tail object detection with balanced group softmax. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020 Jun 13–19; Seattle, USA. pp.10991–11000. https://doi.org/10.1109/CVPR42600.2020.01100

72. Lavoie M-A, Waslander SL. Class instance balanced learning for long-tailed classification. Proceedings of the 20th Conference on Robots and Vision (CRV). 2023 Jun 6–8; Montreal, Canada. pp.121–128. https://doi.org/10.1109/CRV60082.2023.00023

73. Kim B, Kim H, Kim K, Kim S, Kim J. Learning not to learn: Training deep neural networks with biased data. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019 Jun 15–20; Long Beach, USA. pp.9004–9012. https://doi.org/10.1109/CVPR.2019.00922

74. Abbasi-Sureshjani S, Raumanns R, Michels BEJ, Schouten G, Cheplygina V. Risk of Training Diagnostic Algorithms on Data with Demographic Bias. In: Cardoso J, van Nguyen H, Heller N, Abreu PH, Isgum I, Silva W, et al, editors. Interpretable and Annotation-Efficient Learning for Medical Image Computing, Lecture Notes in Computer Science. 2020 Oct 4–8; Lima, Peru. 12446:183–192. https://doi.org/10.1007/978-3-030-61166-8_20

75. Koch W, Hogeweg L, Nilsen EB, O'Hara RB, Finstad AG. Recognizability bias in citizen science photographs. R Soc Open Sci. 2023 Feb 1; 10(2). https://doi.org/10.1098/rsos.221063 PMID: 36756065

76. Bengio Y, Lecun Y, Hinton G. Deep learning for AI (Turning lecture). Commun ACM. 2021 Jun 21; 64(7):58–65. https://doi.org/10.1145/3448250

77. Raykar VC, Yu S, Zhao LH, Valadez Hermosillo G, Florin C, Bogoni L, et al. Learning from crowds. J Mach Learn Res. 2010 Oct 4; 11:1297–1322.

78. Yan Y, Rosales R, Fung G, Schmidt M, Hermosillo G, Bogini L, et al. Modeling annotator expertise: Learning when everybody knows a bit of something. Proceedings of the 13th International Conference on Artificial Intelligence and Statistics. 2010; Chia (Sardinia), Italy. 9:932–939.

79. Hu Q, He Q, Huang H, Chiew K, Liu Z. Learning from crowds under experts' supervision. In: Tseng VS, Ho TB, Zhou ZH, Chen ALP, Kao HY, editors. Advances in Knowledge Discovery and Data Mining, Lecture Notes in Computer Science. 2014 May 13–16; Tainan, Taiwan. 8443:200–211. https://doi.org/10.1007/978-3-319-06608-0_17

80. Giuffrida MV, Chen F, Scharr H, Tsaftaris SA. Citizen crowds and experts: observer variability in image-based plant phenotyping. Plant Methods. 2018 Jan 29; 14(12). https://doi.org/10.1186/s13007-018-0278-7 PMID: 29449872

81. Penna ND, Reid MD. Crowd & prejudice: An impossibility theorem for crowd labelling without a gold standard. arXiv:1204.3511v1 [Preprint]. 2021 [cited 2024 Feb 21]. Available from: https://arxiv.org/abs/1204.3511

82. Tarvainen A, Valpola H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. arXiv:1703.01780v6 [Preprint]. 2018 [cited 2024 Feb 21]. Available from: https://arxiv.org/abs/1703.01780

83. Ouali Y, Hudelot C, Tami M. An overview of deep semi-supervised learning. arXiv:2006.05278v2 [Preprint]. 2020 [cited 2024 Feb 21]. Available from: https://arxiv.org/abs/2006.05278

84. Rajevenceltha J, Gaidhane VH. A no-reference image quality assessment model based on neighborhood component analysis and Gaussian process. J Vis Commun Image Represent. 2024 Feb; 98:104041 https://doi.org/10.1016/j.jvcir.2023.104041

85. Tong K, Wu Y, Zhou F. Recent advances in small object detection based on deep learning: A review. Image Vis Comput. 2020 May; 97:103910. https://doi.org/10.1016/j.imavis.2020.103910

86. Verdecchia R, Sallou J, Cruz L. A Systematic Review of Green AI. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2023 Jun 5; 13(4):e1507 https://doi.org/10.1002/widm.1507

87. Balaiah T, Jeyadoss TJT, Thirumurugan SS, Ravi RC. A deep learning framework for automated transfer learning of neural networks. Proceedings of the 11th International Conference on Advanced Computing (ICoAC). 2019 Dec 18–20; Chennai, India. pp.428–432. https://doi.org/10.1109/ICoAC48765.2019.246880

88. Tan J, Wang C, Li B, Li Q, Ouyang W, Yin C, et al. Equalization loss for long-tailed object recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020 Jun 13–19; Seattle, USA. pp.11659–11668. https://doi.org/10.1109/CVPR42600.2020.01168

89. Bar M. Visual objects in context. Nat Rev Neurosci. 2004 Aug 1; 5(8):617–629. https://doi.org/10.1038/nrn1476 PMID: 15263892

90. Heck PM, Schouten G. Defining quality requirements for a trustworthy AI wildflower monitoring platform. Proceedings of the 2nd International Conference on AI Engineering–Software Engineering for AI (CAIN). 2023 May 15–16; Melbourne, Australia. pp.119–126. https://doi.org/10.1109/CAIN58948.2023.00029

91. Hellwig N, Schubert LF, Kirmer A, Tischew S, Dieker P. Effects of wildflower strips, landscape structure and agricultural practices on wild bee assemblages–A matter of data resolution and spatial scale? Agric Ecosyst Environ. 2022 Mar 1; 326:107764. https://doi.org/10.1016/j.agee.2021.107764

92. Kleijn D, Biesmeijer KJC, Klaassen RHG, Oerlemans N, Raemakers I, Schepers J, et al. Integrating biodiversity conservation in wider landscape management: Necessity, implementation and evaluation. In: Bohan DA, Vanbergen AJ, editors. Advances in Ecological Research. Academic Press; 2020. 63:127–159. https://doi.org/10.1016/bs.aecr.2020.08.004