ORIGINAL RESEARCH

Automatic flower detection and phenology monitoring using time-lapse cameras and deep learning

Hjalte M. R. Mann[1,2], Alexandros Iosifidis[2], Jane U. Jepsen[3], Jeffrey M. Welker[4,5,6], Maarten J. J. E. Loonen[7] & Toke T. Høye[1]

[1]Department of Ecoscience and Arctic Research Center, Aarhus University, C.F. Møllers Allé 8, 8000, Aarhus C, Denmark

[2]Department of Electrical and Computer Engineering – Signal Processing and Machine Learning, Aarhus University, Finlandsgade 22, 8200, Aarhus N, Denmark

[3]Department of Arctic Ecology, Fram Centre, Norwegian Institute for Nature Research, Hjalmar Johansens gt.14, 9007, Tromsø, Norway

[4]Department of Ecology and Genetics, University of Oulu, Oulu, Finland

[5]University of the Arctic, Rovaniemi, Finland

[6]Department of Biological Sciences, University of Alaska, Anchorage, Alaska, USA

[7]Arctic Centre, University of Groningen, Groningen, the Netherlands

Abstract

The advancement of spring is a widespread biological response to climate change observed across taxa and biomes. However, the species level responses to warming are complex and the underlying mechanisms are difficult to disentangle. This is partly due to a lack of data, which are typically collected by direct observations, and thus very time-consuming to obtain. Data deficiency is especially pronounced in the Arctic where the warming is particularly severe. We present a method for automated monitoring of flowering phenology of specific plant species at very high temporal resolution through full growing seasons and across geographical regions. The method consists of image-based monitoring of field plots using near-surface time-lapse cameras and subsequent automated detection and counting of flowers in the images using a convolutional neural network. We demonstrate the feasibility of collecting flower phenology data using automatic time-lapse cameras and show that the temporal resolution of the results surpasses what can be collected by traditional observation methods. We focus on two Arctic species, the mountain avens *Dryas octopetala* and *Dryas integrifolia* in 20 image series from four sites. Our flower detection model proved capable of detecting flowers of the two species with a remarkable precision of 0.918 (adjusted to 0.966) and a recall of 0.907. Thus, the method can automatically quantify the seasonal dynamics of flower abundance at fine scale and return reliable estimates of traditional phenological variables such as the timing of onset, peak, and end of flowering. We describe the system and compare manual and automatic extraction of flowering phenology data from the images. Our method can be directly applied on sites containing mountain avens using our trained model, or the model could be fine-tuned to other species. We discuss the potential of automatic image-based monitoring of flower phenology and how the method can be improved and expanded for future studies.

Introduction

Climate change is influencing phenology across taxa and ecosystems and a trend towards advancement of spring is widely observed (Collins et al., 2021; Hällfors et al., 2020; Menzel et al., 2020; Parmesan, 2006). However, species vary in their responses to climate change and some species and populations elicit seemingly counterintuitive phenological responses (Cook et al., 2012; Prevéy et al., 2019; Rafferty et al., 2020; Thackeray et al., 2016). For instance,

species that are apparently non-respondent to warming can have opposite responses to fall/winter and spring warming, and observed trends in first flowering dates across species may thus depend on their relative sensitivity to these variables (Cook et al., 2012). Differential responses between interacting and interdependent species, for example flowers and their pollinators, to climate changes may disrupt their temporal synchrony (Visser & Both, 2005). To disentangle and quantify the intricate effects of climate change on phenology, we must study phenology at the species level. However, recording phenology is logistically challenging and is mostly done at a coarse temporal resolution with potentially poor change detection ability as a consequence (Arft et al., 1999).

Flower phenology is often described by a single variable related to the date of first flowering (Fox & Jönsson, 2019; Miller-Rushing et al., 2008; Prevéy et al., 2019). Such data may be too simplistic to capture how flowering seasons change in relation to climate change (Iler et al., 2013b; Inouye et al., 2019; Myers-Smith et al., 2020). For example, first, peak, and last flowering may not shift uniformly, and relying solely on the date of first flowering may overestimate shifts in peak flowering and fail to predict shifts in the last flowering, and hence fail to adequately detect responses to climate change (CaraDonna et al., 2014). Recording flower phenology across full growing seasons may remedy this, but this is rarely feasible within traditional sampling schemes relying on human observations (Koide et al., 2019; Prevéy et al., 2017; Rafferty et al., 2020). For manual sampling, data collection is limited by the frequency with which field sites and sampling plots can be visited. With the magnitude of reported trends in phenology at a few days per decade (CaraDonna et al., 2014; Menzel et al., 2020; Post et al., 2018; Thackeray et al., 2016), even weekly observations of individual plots introduce substantial uncertainty in the estimates of long-term trends. Further, manual sampling is prone to observer bias (Richardson et al., 2007). Recording species-specific flower phenology at sufficient temporal and spatial resolution, such that effects are not masked by sampling error or insufficient sampling frequency, requires efficient and preferably automated methods.

The Arctic is warming at three times the rate of the global average (AMAP, 2021), and responses of Arctic species to climatic changes can be particularly dramatic (Høye, Post, et al., 2007; Post et al., 2018). Yet, phenology data for the Arctic have been limited (Diepstraten et al., 2018; Metcalfe et al., 2018), though recent synthesis efforts have produced a broader foundation using traditional methods (Prevéy et al., 2021). Long-term monitoring data are immensely valuable for studying phenology dynamics, particularly for investigating the complex

responses to climate change (Iler et al., 2013b; Oberbauer et al., 2013). However, maintaining long-term monitoring schemes is challenging, especially in the Arctic (Post & Høye, 2013). Non-permanent funding, evolving methodologies, changing study priorities and observer error can have negative effects on data integrity and complicate analyses and interpretation of long-term datasets. Thus, methods that can facilitate reliable and efficient long-term data collection in remote and challenging locations are in high demand.

Image-based monitoring of phenology based on greenness analyses has been demonstrated both on the landscape level (Brown et al., 2016) as well as for specific species, such as *Dryas integrifolia* (Beamish et al., 2016). These studies have highlighted the effectiveness of image-based monitoring in reducing the load of fieldwork while increasing the resolution of the obtained data. However, image-based methods often still require substantial manual efforts in order to extract ecologically relevant data, which limits the scale at which they can be applied. In recent years, artificial intelligence systems, particularly deep learning, have undergone rapid advancement, but despite the great potential of especially image-based methods such as convolutional neural networks (CNNs), they are not yet well established within the field of ecology (Christin et al., 2019; Høye et al., 2021; Weinstein, 2018). Adapting and implementing such methods would facilitate a temporal and spatial upscaling of ecological data and allow ecologists to study ecological phenomena at the scale at which they occur (Estes et al., 2018; Pimm et al., 2015; Weinstein, 2018).

Our aim for this work was to investigate the feasibility of automating the monitoring of species-specific flower phenology at a large spatial scale and high temporal resolution using near-surface time-lapse cameras. Further, we explored the potential for using deep learning for the extraction of phenology data from the image series by automated detection and counting of flowers. We tested the method on two plant species, *Dryas octopetala* and *D. integrifolia* (Murray, 1997). The species are among the most common in tundra plant phenology monitoring programs (Welker et al., 1997) and automated methods would therefore be particularly relevant for them. Previous studies showed promising results for the detection of *Dryas* flowers (Ärje et al., 2019; Tran et al., 2018), although individual separation of flowers was poor, which is a requirement for accurate estimation of flower phenology based on automatic flower counts. We expect that a standardized and automated monitoring workflow will facilitate a better understanding of flower phenology and the effects of climate change through large and accurate datasets. Further, image-based monitoring has great potential for expansion and could be leveraged, for

example, in studies on plant reproductive phenology and interactions with flower-visiting insects. Image series can hold a great deal of information and as methodologies for image analysis evolve, existing image datasets can be re-analyzed and new results can be derived.

Materials and Methods

Study sites and study species

For this study, we collected 20 images time-lapse image series from three different locations. From one location images were collected over two seasons. We hereafter refer to each location/year combination as a site. Five image series were collected from each of the following sites (sampling years given in parentheses): Narsarsuaq, South Greenland (2018 and 2019), Thule, North-West Greenland (2018), and Ny-Ålesund, Svalbard (2019). The cameras were permanently positioned above cushions of *D. integrifolia* (Narsarsuaq, Thule) or *D. octopetala* (Ny-Ålesund) before the start of the flowering season and recorded images throughout the season.

The two species of *Dryas* are widespread, perennial, cushion-forming evergreen dwarf-shrubs and are native to arctic and alpine regions of Europe, Asia, and North America with a geographically separated distribution except for a likely hybrid zone in Northeast Greenland (Høye, Ellebjerg, & Philipp, 2007; Philipp & Siegismund, 2003). The insect-pollinated flowers have white petals and a yellow center and are held erect above the cushion. Flowers of *D. octopetala* are characterized by having eight petals, while *D. integrifolia* have up to 11. Since the two species are very similar in appearance and each of our sites only contains one species, we treated the two species as one.

Image collection

We used consumer-grade wildlife cameras (Moultrie Wingscapes TimelapseCam Pro, Moultrie Products, Birmingham, AL, USA) positioned above cushions of *Dryas*. The cameras were mounted facing toward the ground on custom-made metal mounting frames (Fig. 1). Images were recorded at the highest possible resolution (20 MP). The cameras allow for manual focusing at a short range (down to 15 cm) and are equipped with a LED flash for photography in low lighting. We powered the cameras by either solar power using a central solar panel, a 12 V 100 Ah lead battery, and a power distributor for a clutch of six cameras, or by lithium AA batteries. On AA batteries and with a time-lapse interval of 1 min, a camera can run continuously for more than 2 weeks and record 24 000 images at 20 MP, which typically fits on a 128 GB

SD memory card. Decreasing the time-lapse interval reduces power consumption and increases runtime. On a continuous power source, such as solar power, the runtime of the cameras is restricted purely by the memory. We successfully ran the cameras with 512 GB SD memory cards allowing for storage of almost 100 000 images and a constant runtime of around 70 days with 1-min imaging frequency.

Image preprocessing

While the original time-lapse frequency was higher to allow for potential detection of flower visitors at a later stage, here we subsampled the time-lapse image series to 1-h intervals. Maintaining a relatively high temporal resolution of 1 h ensures that any diurnal variation in the images is captured. As some cameras ran for an extending period before and after the flowering season, we truncated the tails of each series to a maximum of flowering season length ± 50%. These series are hereafter referred to as the 1-h series. All series spanned full growing seasons, meaning the image collection was initiated before the first flowering and terminated after the last flowering, except for three series: NARS 2019 D and E failed to capture the onset of flowering and had one and 22 flowers in bloom, respectively, in the first image of each series. THUL 2018 A failed to capture the end of the flowering and had one flower in bloom in the last image of the series. Two series had missing data spanning more than a full day within the flowering season: THUL 2018 E: 5 days (DOY 190–194); NYAA 2019 D: 1 day (DOY 194). The cameras collected images for a total of 1274 days.

To create a set of images on which to test the manual and automatic derivation of phenology variables, we identified the images containing the first and the last flower, respectively, for each series and randomly sampled a set of images from each series within these limits. The images from the time-lapse series vary in lighting conditions throughout the day and the flowers can change appearance across the day and the flowering season. Further, the background varies between the different sites and throughout the season. In order to train a flower detection model with high generalization ability, the training data should cover the full variation in the appearance of the flowers. As manual flower annotations were to be used for both manual derivation of phenology variables as well as for training an automatic flower detection model, we ensured that this variation was captured in the annotated images by sampling images in batches across all series and sequentially annotating them. Each batch contained 25 randomly sampled images from each 1-h series. If <25 images were left for a given image series, sampling from that series was skipped. Images that did

Figure 1. A cluster of six time-lapse cameras mounted above cushions of *Dryas integrifolia* and a solar panel in Narsarsuaq, Greenland, 2019. Photo: HMRM.

not contain flowers were removed in the annotation process (72 images did not contain flowers and were removed before training and testing the automatic flower detection model). The sampling approach ensures a fair evaluation of the automatic detection method in the sense that the annotator cannot use information from the previous or following images which is also the case for the detection model. However, in cases where the manual approach for quantifying phenology from images is adopted, leveraging information from flanking images is likely to ease the annotation workflow and reduce the risk of noise in the annotations. In total, eight batches, amounting to 3771 images were sampled and annotated for flowers. These images are hereafter referred to as the sampled series. The number of images per day depends on the length of the flowering season for a given series. The sampled series contained a minimum of 4.6 images on average per day (NARS 2018 C) and maximum of 18.8 images on average per day (THUL 2018 A). The sampling approach ensures a dataset that is balanced in regards to the total number of images sampled from each series, which is important for a fair assessment of the automatic flower detection model.

Manually deriving flower phenology from time-lapse images

We focused on three key phenological variables; onset, peak, and end of flowering, defined as the times when 10%, 50%, and 90% of the flowers in bloom had been recorded, respectively. To manually derive these, we annotated all flowers in bloom (a total of 33 548) in the sampled image series using the rectangular annotation tool in the *VIA* VGG annotation software (Dutta &

Zisserman, 2019). Flowers were defined as in bloom from when the yellow reproductive organs were visible in the opening bud until the moment the flower had lost >50% of its petals or when the petals had changed color to yellow/brown. Flowers that crossed the edge of the image frame or that were covered by other flowers were only annotated when >40% of the flower was visible. This decision was taken by considering the standard practice in object detection methods, where performance is derived based on the overlap, calculated as the intersection over union (IoU), between bounding boxes from manual annotations and detections. Overlap above a certain threshold is considered a correct detection. Standard threshold choices for generic object recognition are IoU-50%, IoU-70%, and IoU-90%. We calculated the cumulative sum of annotated flowers across the season for each of the 20 series and extracted the time points for the observation closest to each of the phenology variables (10%, 50%, and 90%). Owing to the high temporal resolution of the data, there was no need for linear interpolation between observations.

Automatically deriving flower phenology from time-lapse images

For automatically deriving the phenology variables from the images, we trained a CNN to detect the flowers in the images. There are many deep learning networks for object detection publicly available, with You Only Look Once (YOLO) and the Faster Region-based Convolutional Neural Network (Faster R-CNN) (Redmon et al., 2016; Ren et al., 2017) being among the most well-known. Generally, YOLO has the advantage of fast processing times, while Faster R-CNN obtains higher

detection accuracy (Redmon & Farhadi, 2018). As we prioritized accuracy rather than the speed we implemented the Mask R-CNN (He et al., 2017), an extension of the Faster R-CNN, for the detection of flowers in images. We used an open-source implementation of Mask R-CNN based on Tensorflow (Abadi et al., 2016) and Keras (Chollet, 2015) and written in Python. The code is publicly available on GitHub (https://github.com/matterport/Mask_RCNN) (Abdulla, 2017). The Matterport Mask R-CNN repository provides pre-trained weights for the COCO dataset (328 k images and 2.5 million labeled instances) (T. Y. Lin et al., 2015). Before training, we tested the influence of adjusting key hyperparameters on the detection accuracy of the Mask R-CNN model (Fig. S1) and found a model with standard settings but with an extended image augmentation pipeline performed the best, and we used this approach for our model. We trained our model by fine-tuning the pre-trained weights on our dataset of annotated flowers, a process referred to as transfer learning (Girshick et al., 2014). Transfer learning leverages large datasets for training the model to detect and discriminate features in images, and the model is then subsequently adapted to a specific use case by fine-tuning on a smaller dataset. The training was continued until overfitting, detected by a rise in the validation error, and early stopping was used, that is, the model that was output at the epoch step returning the lowest validation error was chosen as the final model. This approach ensures that the model with the best ability to generalize is chosen. We tested the influence of image resolution on detection accuracy and inference speed when running detections (Fig. S2) and found that downscaling the images to a resolution of 760×427 pixels from the original 6080×3420 pixels gave the best detection accuracy while also vastly increasing inference speed compared to higher resolution images. We used this image resolution for flower detection with our model.

Training and testing the flower detection model

Training a neural network to detect a specific type of object requires a set of training and validation data, specifically a set of images in which the objects have been annotated. The model learns the features of the object of interest from the training data and repeatedly measures its prediction performance on the validation data. The performance achieved on the validation set is used to decide when to stop the training process. Additionally, a test set, an independent set of annotated images to evaluate the generalization ability of the deep learning model (i.e., its ability to perform well on images that are not

included in the training process) is required. We used the set of manually annotated images from the 20 sampled series to train and test our flower detection model. To ensure independency between images used for training, validation, and testing, all images from a single camera were used in only one of these sets. Three image series from each of the four sites were used for training, one from each site for validation, and one from each site for testing, yielding a total of 12 training series, four validation series, and four test series. The decision on which specific series from a site would be assigned to training, validation, and testing was based on minimizing the deviation from a 5:2:3 split in terms of number of individual flower annotations. The split was chosen to ensure sufficient data in all three sets. The final training set contained 18 008 annotations in 2183 images from 12 image series, the validation set contained 6468 annotations in 752 images from four image series (NARS 2018 E, THUL 2018 C, NARS 2019 A, NYAA 2019 D), and the test set contained 9073 annotated flowers in 764 images from four image series (THUL 2018 A, NARS 2018 D, NARS 2019 B, and NYAA 2019 C). Including images from all four sites in the training maximizes the degree of variation in the training data. Similarly, including image series from each site in the test set ensures that the generalization capability of the model is tested.

To evaluate the accuracy of our flower detection model, we compared the model detections in the test set of images with the manual annotations in the same images. We present the precision (ratio of the number of true positives to the total number of predictions), recall (ratio of the number of true positives to the actual number of objects), and $F1$ score (weighted average of precision and recall) using an IoU threshold value of 0.5 to determine satisfactory overlap between bounding boxes of the manual flower annotations and the detections. Thus, the manual annotations represent a ground truth on which the automatic detections are evaluated. Note that any discrepancies between the detections and ground truth are always considered an error of the CNN and not an error of the manual annotations. Collectively, the method of evaluating the precision of the flower detection model is therefore conservative. For practical reasons, the shift between flowering stages was considered instant during the annotation process. However, as it is in reality gradual, we hypothesized that the false-positive predictions produced by the flower detection model could be biased toward flowers in stages before and after what we considered a flower in bloom. Further, only flowers where more than 40% of the flower was visible were annotated, but the flower detection model may be capable of recognizing flowers even if they are only partly visible, which would also be counted as false positives. To investigate these
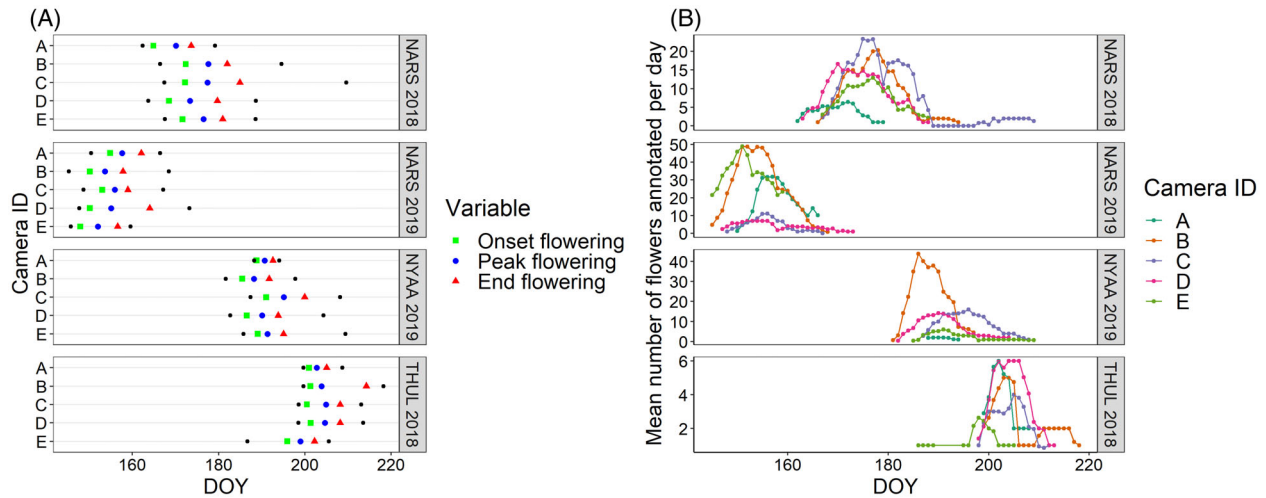
Figure 2. (A) The phenology variables extracted from the flower detections for each of the 20 image series. Black dots show time for the first and last flower in the series. (B) Mean number of flowers annotated per day through the season for each image series.

issues further, we quantified what proportion of the false positives (i.e., bounding boxes from the automatic flower detections that did not overlap with manual annotations) were actually flowers in any stage or size. We obtained an adjusted precision score for the model by including these as correct positives. False-positive detections that could not be recognized as a flower with certainty were not included and neither were single petals lost from flowers.

Applying the model to the test set and comparing it to the manual annotations in the same images accurately measures its ability to detect flowers. However, in order to test the feasibility of automatically deriving the phenology variables from image series without preprocessing, we also ran the flower detection model on the 1-h image series and compared the derived phenology variables to the sampled series.

Results

Manually derived flowering phenology

The manual flower annotations gave detailed information on the phenology of the flowers in each plot. With several images annotated for flowers per day for each series, we obtain flower counts across the season in very high temporal resolution. In Figure 2, we present the phenology variables obtained for each series as well as the distribution of flowers through the season represented as the mean number of flowers annotated per day (number of flowers annotated per image through the season is given in Fig. S3).

Automatic flower detection

Our flower detection model reached a precision of 0.918 and a recall of 0.907. Thus, 8.2% (N = 738) of the detected flowers were false positives and 9.3% (N = 843) of the 9073 annotated flowers in the test set were not detected (i.e., false negatives). The performance corresponds to an $F1$ score of 0.912. However, we found that 58.8% (N = 434) of the 738 false positives produced by the model actually constituted flowers, but that these were either categorized in a stage other than bloom or were <40% visible and hence not considered valid observations during annotation. Including these as correct detections increased the precision of the flower detection model to 0.966 ($F1$ = 0.936). Figure 3 shows example images from each of the four test series with the corresponding results obtained with the flower detection model.

Automatically-derived flowering phenology

The automatic flower detections on the sampled images as well as on the 1-h image series from the test set closely resembled the seasonal flowering dynamics derived by the manual annotations (Fig. 4A presents the mean number of flowers per day while the number of flowers per image is given in Fig. S4). Consequently, the cumulative sum of flower counts followed very similar trajectories for the annotations of the sampled series, the detections on the sampled series, and the detections on the 1-h series for the four test series (Fig. 4B). The derived onset, peak, and end of flowering showed much more variation between
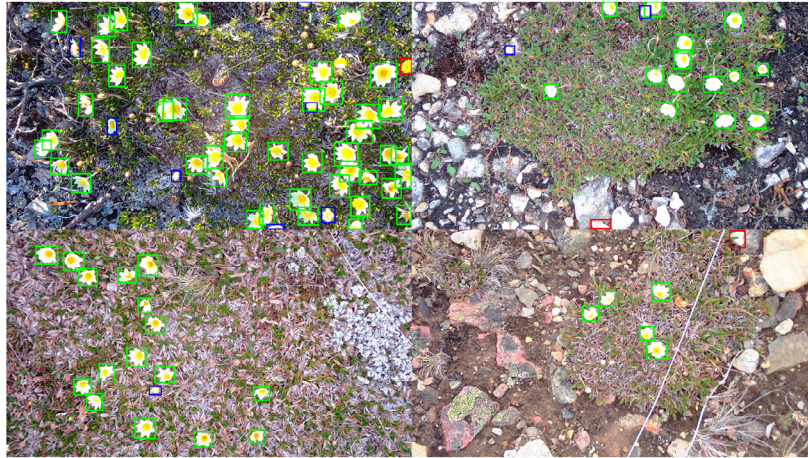
Figure 3. Example image from each of the four test series with bounding boxes for positives (correct flower detections) as green rectangles, false negatives (flowers that were not detected) as blue rectangles, and false positives (incorrect flower detections) as red rectangles. Top left: NARS 2019 B, top right: NYAA 2019 C, bottom left: NARS 2018 D, bottom right: THUL 2018 A.
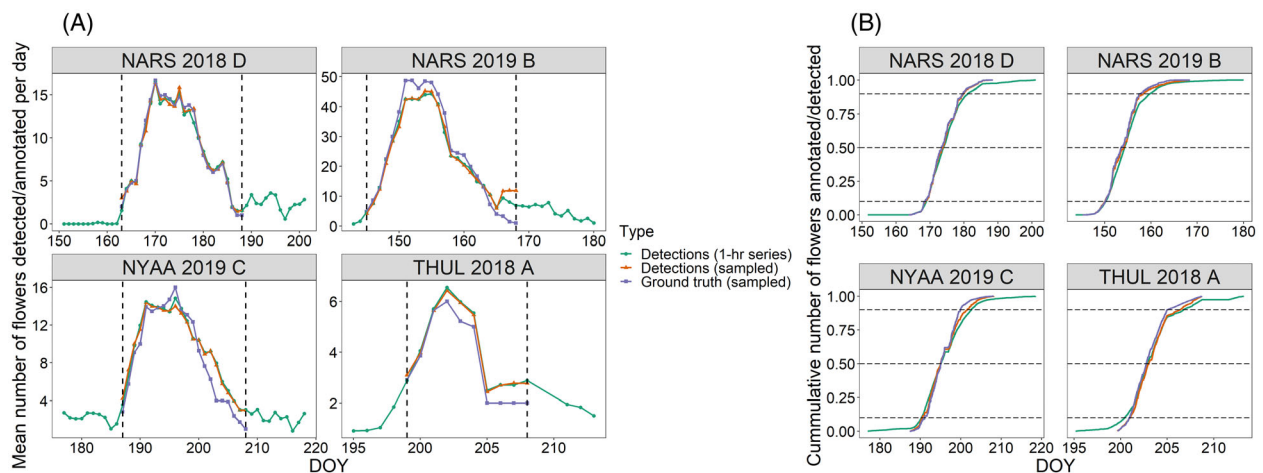

Figure 4. (A) Mean number of flowers annotated (sampled images) and detected (sampled series and 1-h series) per day through the season for each of the four image series in the test set. Hatched lines show the times for the first and last flower in the series. (B) Cumulative sum of the number of flowers annotated (sampled images) and detected (sampled series and 1-h series) through the season for each of the four image series. Hatched lines mark 10%, 50%, and 90%.

the four test series than between detections and ground truth observations. The deviation between the ground truth onset of flowering and those derived from the automatic detections on the sampled series was 5.0 ± 6.7 h (mean ± SD). For peak flowering, the deviation was 8.0 ± 4.4 h, and for end flowering the deviation was 21.5 ± 18.0 h. The deviation between ground truth values and those derived from the automatic detections on the full series was 11.0 ± 7.1, 8.0 ± 7.5, and 46.5 ± 13.0 h for onset, peak, and end flowering, respectively.

Discussion

In this study, we demonstrate how time-lapse cameras can be used for detailed monitoring of flowering phenology of specific plant species across the length of growing seasons. Image-based monitoring with time-lapse cameras allows for automatic data collection at a much-needed increased temporal resolution compared to traditional methods. Depending on the time-lapse frequency of the cameras and assuming a constant power source and sufficient memory, the cameras can run through full growing

seasons, only requiring attendance at setup and data retrieval. This makes the image-based methodology optimal for monitoring at logistically challenging sites, especially for long-term monitoring, as equipment can be continually used across multiple seasons. Across the 20 cameras, we collected images for a total of 1274 days. We obtained detailed information on flower phenology in 20 sample plots at four different sites from manual flower annotations of a randomly sampled subset of images. For large-scale monitoring covering many sites, the production of manual flower annotations becomes unfeasible. Our dedicated deep learning model resolves the challenge of extracting relevant data from the image series and can compute very accurate flowering phenology curves. Collectively, the presented method is an automated pipeline for detailed, species-specific recording of flowering phenology across seasons.

Phenology variables are traditionally derived by interpolation between flower counts from, typically, weekly observations of permanent plots (Iler et al., 2013a), which introduces considerable uncertainty to the estimates. We derived the onset and end of flowering based on 10% and 90%, respectively, of the flower counts instead of, for example, the timings of the first and last single flower observation. This avoids excessive emphasis on extreme outliers and reduces sensitivity to flower abundance in the plot. The high temporal resolution of our image-based method means that we can identify these variables based on much more well-defined distributions of observations. Therefore, it is reasonable to assume that the true date of events can be more accurately estimated by high-frequency time-lapse image series than by traditional direct observation methods. Such increased accuracy will facilitate an improved understanding of the relationship between phenology and the drivers of phenological change and potentially the role of phenology in biotic interactions.

CNNs are most often trained and evaluated on large and diverse standard datasets such as the COCO dataset and the ImageNet dataset (3.2 million images) containing many everyday objects such as people, electronic consumer products, food products, animals, vehicles, etc. (Deng et al., 2009; He et al., 2017; Lin et al., 2015; Redmon & Farhadi, 2018). Such benchmarks demonstrate the capabilities of the CNNs but are perhaps less relevant for specific *in situ* applications in ecology. Here, we trained and applied the CNN on images collected in the field to test the feasibility of automatic monitoring of flower phenology. We used all images from a single camera for either training, validation, or testing, ensuring that the ability of the CNN to generalize and detect flowers in images with unique backgrounds was tested. This approach increases the risk of detection errors compared

to using images from the same camera for both training and testing, but the ability to generalize is relevant in the context of large-scale ecological monitoring. Despite the difficult setting, the model detected flowers with high accuracy and the detections represented the true flowering phenology closely. Further, we were able to accurately estimate traditional key phenological variables on the basis of the cumulative sum of flower detections. This, together with the fact that the image series were collected with consumer-grade time-lapse cameras, makes the method relevant and accessible for many research programs.We treat the transitions between flower stages (e.g., from bud to flower or from flower to senesced flower) as abrupt, while they are in reality gradual. The high temporal resolution of the image series exacerbates the risk of errors during this transition phase. We note that similar problems may occur for both manual flower annotations and automatic detections as well as for manual observation in the field. Naturally, basing the flower count on images makes it possible to confirm results at any time by inspecting the images. On a related note, flowers that were partly occluded and where <40% of a flower was visible were not annotated. However, the model proved capable of detecting many of these flowers, which were the cause of a large proportion of the false positives. We count these as errors, but in fact, it shows that the model is able to generalize and should be considered a quality. These issues should be taken into account when assessing the precision of flower detection models.

We treat the transitions between flower stages (e.g., from bud to flower or from flower to senesced flower) as abrupt, while they are in reality gradual. The high temporal resolution of the image series exacerbates the risk of errors during this transition phase. We note that similar problems may occur for both manual flower annotations and automatic detections as well as for manual observation in the field. Naturally, basing the flower count on images makes it possible to confirm results at any time by inspecting the images. On a related note, flowers that were partly occluded and where <40% of a flower was visible were not annotated. However, the model proved capable of detecting many of these flowers, which were the cause of a large proportion of the false positives. We count these as errors, but in fact, it shows that the model is able to generalize and should be considered a quality. These issues should be taken into account when assessing the precision of flower detection models.

Our method can still be further improved. Visually inspecting the false-positive detections revealed that the model had a tendency to falsely detect stones as flowers, especially when they were overexposed. In some cases, a single stone was the cause of a high proportion of the false positives within a series. For example, 78% of the

false positives in the sampled THUL 2018 A series were repeated detections of a single stone, which due to over-exposure might resemble petals of the flowers. Cameras with greater flexibility in the settings, better optical properties, or a broader range of backgrounds in the training data could alleviate this problem, which may also be less pronounced for flowers that are not white. For our study, we used a deep learning model based on the Faster R-CNN framework. This framework has proven capable of a wide variety of detection problems (Jiang & Learned-Miller, 2017; Lin & Chen, 2018; Wan & Goudos, 2020) and is thus a good candidate for novel applications. However, we note that object detection in computer vision is an active research topic and new models may achieve higher performance.

If a constant level of false-positive detections from the detection model is assumed, the accuracy of the estimation of onset, peak, and end of flowering will be sensitive to the number of flowers within a plot. As the ratio between correct positives and false positives will improve with an increasing number of flowers, so will the precision of the detection model. Further, false positives affect the cumulative sum of positives and thereby the accuracy of the estimated phenology variables. For example, a long tail of false positives before the actual flowering starts will advance the estimated onset of flowering, while a long tail of false positives after flowering has ended will delay the estimated end of flowering. Finally, the estimate of phenology variables is sensitive to the number of flowers on which it is based, as few flowers may not represent the true distribution of flowering in the population well. This goes for traditionally collected phenology data as well as for our automated method. Mounting the cameras higher above the ground would allow for covering a larger area and could ensure a minimum number of flowers, but would reduce image resolution and possibly affect detection accuracy negatively. It is also possible to couple near-surface cameras with drone surveys for phenological context (Beamish et al., 2020). However, here we have shown that even with tails at a maximum of 50% $\pm$ the length of the flowering season and for plots with few as well as many flowers, the method produces accurate estimates of the phenology metrics.

Image-based setups coupled with computer vision techniques are increasingly being suggested for monitoring flowers, especially in agricultural settings (Jiang et al., 2020; Palacios et al., 2020; Wang et al., 2021). Such methods can facilitate automated and efficient assessment of phenology and crop yield forecasting, but often requires human camera operators of custom-made mobile imaging platforms. Here, we present a method for efficiently obtaining flower phenology data in a logistically challenging natural setting with permanently installed cameras. To explore the current and future effects of climate change, standardized long-term image-based monitoring schemes should be implemented and maintained to ensure rigorous and efficient data collection along with the continuous development of methods to handle and analyze the output of these schemes. Importantly, publicly available high-quality and preferably annotated datasets are very relevant for developing methods with high-performance and broad applicability. In parallel, the focus should be on continuous digitization and analysis of herbarium specimens to create a historical context for evaluating newly collected data. Computer vision techniques have proven valuable for the analysis of herbarium specimens (Pearson et al., 2020; Schuettpelz et al., 2017) and the availability of large herbarium datasets will be increasingly relevant as these techniques are refined. We have focused our work on two species of *Dryas*, as these are commonly included in the long-term monitoring of flower phenology, but the overall methodology is generic and we note that it would be suitable for many other plant species. For the Arctic in particular, *Silene acaulis*, *Saxifraga oppositifolia*, *Rubus chamaemorus*, and other species with clearly defined flowers positioned close to the ground are relevant candidate species. The automatic detection and counting of other species require a fine-tuning of the model to the species of interest using annotated images.

In the present case, there were no flowering plant species visually similar to the Dryas species within the sample plots. In other settings, this might be the case. When applying a pretrained model as the one presented here in a new setting, the accuracy of the model detections should be tested. In the case of our model, if other visually similar species are detected as *Dryas*, fine-tuning the model with images containing other flower species in the background included in the training data could help alleviate the issue. Alternatively, as deep learning methods have proven capable of accurate species classification despite low variation between classes, expanding the training data to include annotations of other species could facilitate multiclass detection and allow for comparative studies between co-occurring species (Nilsback & Zisserman, 2008; Spiesman et al., 2021). In any case, the presented method can serve as a guideline for developing and testing solutions for automatic flower phenology monitoring in other settings.

We emphasize that there are many benefits to image-based monitoring of phenology even without the deep learning processing pipeline in place. Camera-based monitoring elicits minimal disturbance to permanent plots and image series constitute an archive of the phenology of the captured species and contain additional relevant information that can be extracted from the images, either

manually or automatically. For instance, our method could be expanded to include the detection of flower visitors to count and classify pollination events through the season. This would facilitate monitoring of interactions between flowers and their visitors in very high temporal and spatial resolution. To discern the effects of climate change on the phenology and abundance of flowers and how changes may affect interactions within the community, we need such species-level knowledge (Prevéy et al., 2019; Tang et al., 2016). With our method, flower visits could even be pinpointed to the individual flower with manual or automatic tracking of single flowers through the season and these data could be coupled to the reproductive success of the flowers.

In conclusion, we have presented an automated pipeline for monitoring flower phenology of specific species at high temporal resolution and across regions. We have shown how state-of-the-art computer vision and deep learning methods can be applied to images collected *in situ* and used to extract ecologically relevant data. The methodology is easily expandable to new sites and optimal for long-term monitoring of plots. Archived images can ensure reproducibility of results and can be re-analyzed as new questions arise or new methods are developed. The system facilitates cost-efficient monitoring of vegetation plots at unprecedented temporal resolution across full growing seasons and our results demonstrate the great potential of automatic image-based long-term monitoring of flower phenology. The presented flower detection model and associated code is available at https://github.com/TECOLOGYxyz/AutomaticFlowerPhenology.

Acknowledgments

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J. et al. (2016) TensorFlow: a system for large-scale machine learning. In: *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16)*, pp. 265–283. https://doi.org/10.1016/0076-6879(83)01039-3

Abdulla, W. (2017) Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. *GitHub Repository*. Github. Available from: https://github.com/matterport/Mask_RCNN [Accessed 17th October 2019].

AMAP. (2021) Arctic climate change update 2021: key trends and impacts. In: *Summary for policy-makers*. Tromsø, Norway: Arctic Monitoring and Assessment Programme (AMAP), p. 16.

Arft, A.M., Walker, M.D., Gurevitch, J., Alatalo, J.M., Bret-Harte, M.S., Dale, M. et al. (1999) Responses of tundra plants to experimental warming: meta-analysis of the international tundra experiment. *Ecological Monographs*, **69** (4), 491–511. https://doi.org/10.1890/0012-9615(1999)069 [0491:ROTPTE]2.0.CO;2

Ärje, J., Milioris, D., Tran, D.T., Jepsen, J.U., Raitoharju, J., Gabbouj, M. et al. (2019) 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*.

Beamish, A.L., Nijland, W., Edwards, M., Coops, N.C. & Henry, G.H.R. (2016) Phenology and vegetation change measurements from true colour digital photography in high Arctic tundra. *Arctic Science*, **2**(2), 33–49. https://doi.org/10.1139/as-2014-0003

Beamish, A.L., Raynolds, M.K., Epstein, H., Frost, G.V., Macander, M.J., Bergstedt, H. et al. (2020) Recent trends and remaining challenges for optical remote sensing of Arctic tundra vegetation: a review and outlook. *Remote Sensing of Environment*, **246**, 111872. https://doi.org/10.1016/j.rse.2020.111872

Brown, T.B., Hultine, K.R., Steltzer, H., Denny, E.G., Denslow, M.W., Granados, J. et al. (2016) Using phenocams to monitor our changing earth: toward a global phenocam network. *Frontiers in Ecology and the Environment*, **14**(2), 84–93. https://doi.org/10.1002/fee.1222

CaraDonna, P.J., Iler, A.M. & Inouye, D.W. (2014) Shifts in flowering phenology reshape a subalpine plant community. *Proceedings of the National Academy of Sciences of the United States of America*, **111**(13), E9993. https://doi.org/10.1073/pnas.1815637115

Chollet, F. (2015) Keras. *GitHub Repository*. Available from: https://github.com/fchollet/keras [Accessed 1st November 2021].

Christin, S., Hervet, É. & Lecomte, N. (2019) Applications for deep learning in ecology. *Methods in Ecology and Evolution*, **10**(10), 1632–1644. https://doi.org/10.1111/2041-210x.13256

Collins, C.G., Elmendorf, S.C., Hollister, R.D., Henry, G.H.R., Clark, K., Bjorkman, A.D. et al. (2021) Experimental warming differentially affects vegetative and reproductive phenology of tundra plants. *Nature Communications*, **12**(1), 1–12. https://doi.org/10.1038/s41467-021-23841-2

Cook, B.I., Wolkovich, E.M. & Parmesan, C. (2012) Divergent responses to spring and winter warming drive community level flowering trends. *Proceedings of the National Academy*

of Sciences of the United States of America, **109**(23), 9000–9005. https://doi.org/10.1073/pnas.1118364109

Deng, J., Dong, W., Socher, R., Li, L., 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. https://doi.org/10.1109/CVPR.2009.5206848

Diepstraten, R.A.E., Jessen, T.D., Fauvelle, C.M.D. & Musiani, M.M. (2018) Does climate change and plant phenology research neglect the Arctic tundra? *Ecosphere*, **9**(9), e02362. https://doi.org/10.1002/ecs2.2362

Dutta, A. & Zisserman, A. (2019) The VIA annotation software for images, audio and video. In: *MM 2019 - Proceedings of the 27th ACM International Conference on Multimedia*, pp. 2276–2279. https://doi.org/10.1145/3343031.3350535

Estes, L., Elsen, P.R., Treuer, T., Ahmed, L., Caylor, K., Chang, J. et al. (2018) The spatial and temporal domains of modern ecology. *Nature Ecology and Evolution*, **2**(5), 819–826. https://doi.org/10.1038/s41559-018-0524-4

Fox, N. & Jönsson, A.M. (2019) Climate effects on the onset of flowering in the United Kingdom. *Environmental Sciences Europe*, **31**(89). https://doi.org/10.1186/s12302-019-0271-4

Girshick, R., Donahue, J., Darrell, T. & Malik, J. (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. *ArXiv*, 1311.2524v5. https://doi.org/10.1109/CVPR.2014.81

Hällfors, M.H., Antão, L.H., Itter, M., Lehikoinen, A. & Lindholm, T. (2020) Shifts in timing and duration of breeding for 73 boreal bird species over four decades. *Proceedings of the National Academy of Sciences of the United States of America*, **117**(31). https://doi.org/10.1073/pnas.1913579117

He, K., Gkioxari, G., Dollar, P. & Girshick, R. (2017) Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, 2980–2988. https://doi.org/10.1109/ICCV.2017.322

Høye, T.T., Ärje, J., Bjerge, K., Hansen, O.L.P., Iosifidis, A., Leese, F. et al. (2021) Deep learning and computer vision will transform entomology. *Proceedings of the National Academy of Sciences of the United States of America*, **118**(2), e2002545117. https://doi.org/10.1073/pnas.2002545117

Høye, T.T., Ellebjerg, S.M. & Philipp, M. (2007) The impact of climate on flowering in the high Arctic-the case of dryas in a hybrid zone. *Arctic, Antarctic, and Alpine Research*, **39**(3), 412–421. https://doi.org/10.1657/1523-0430(06-018)[HOYE]2.0.CO;2

Høye, T.T., Post, E., Meltofte, H., Schmidt, N.M. & Forchhammer, M.C. (2007) Rapid advancement of spring in the high Arctic. *Current Biology*, **17**(12), 449–451. https://doi.org/10.1016/j.cub.2007.04.047

Iler, A.M., Høye, T.T., Inouye, D.W. & Schmidt, N.M. (2013a) Long-term trends mask variation in the direction and magnitude of short-term phenological shifts. *American Journal of Botany*. https://doi.org/10.3732/ajb.1200490

Iler, A.M., Høye, T.T., Inouye, D.W. & Schmidt, N.M. (2013b) Nonlinear flowering responses to climate: are species approaching their limits of phenological change? *Philosophical Transactions of the Royal Society B: Biological Sciences*, **368**(1624). https://doi.org/10.1098/rstb.2012.0489

Inouye, B.D., Ehrlén, J. & Underwood, N. (2019) Phenology as a process rather than an event: from individual reaction norms to community metrics. *Ecological Monographs*, **89**(2), 1–15. https://doi.org/10.1002/ecm.1352

Jiang, H. & Learned-Miller, E. (2017) Face detection with the faster R-CNN. *Proceedings - 12th IEEE International Conference on Automatic Face and Gesture Recognition*, 650–657. https://doi.org/10.1109/FG.2017.82

Jiang, Y., Li, C., Xu, R., Sun, S., Robertson, J.S. & Paterson, A.H. (2020) DeepFlower: a deep learning-based approach to characterize flowering patterns of cotton plants in the field. *Plant Methods*, **16**(1), 1–17. https://doi.org/10.1186/s13007-020-00698-y

Koide, D., Ide, R. & Oguma, H. (2019) Detection of autumn leaf phenology and color brightness from repeat photography: accurate, robust, and sensitive indexes and modeling under unstable field observations. *Ecological Indicators*, **106**, 105482. https://doi.org/10.1016/j.ecolind.2019.105482

Lin, P. & Chen, Y. (2018) Detection of strawberry flowers in outdoor field by deep neural network. *2018 3rd IEEE International Conference on Image, Vision and Computing*, 482–486. https://doi.org/10.1109/ICIVC.2018.8492793

Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D. et al. (2015) Microsoft COCO: common objects in context. *ArXiv*, 1405.0312. https://doi.org/10.1007/978-3-319-10602-1_48

Menzel, A., Yuan, Y., Matiu, M., Sparks, T., Scheifinger, H., Gehrig, R. et al. (2020) Climate change fingerprints in recent European plant phenology. *Global Change Biology*, **26**, 2599–2612. https://doi.org/10.1111/gcb.15000

Metcalfe, D.B., Hermans, T.D.G., Ahlstrand, J., Becker, M., Berggren, M., Björk, R.G. et al. (2018) Patchy field sampling biases understanding of climate change impacts across the Arctic. *Nature Ecology & Evolution*, **2**, 1443–1448. https://doi.org/10.1038/s41559-018-0612-5

Miller-Rushing, A.J., Inouye, D.W. & Primack, R.B. (2008) How well do first flowering dates measure plant responses to climate change? The effects of population size and sampling frequency. *Journal of Ecology*, **96**(6), 1289–1296. https://doi.org/10.1111/j.1365-2745.2008.01436.x

Murray, D.F. (1997) Systematics of the ITEX species. *Global Change Biology*, **3**(S1), 10–19. https://doi.org/10.1111/j.1365-2486.1997.gcb131.x

Myers-Smith, I.H., Kerby, J.T., Phoenix, G.K., Bjerke, J.W., Epstein, H.E., Assmann, J.J. et al. (2020) Complexity revealed in the greening of the Arctic. *Nature Climate Change*, **10**, 106–117. https://doi.org/10.1038/s41558-019-0688-1

Nilsback, M.E. & Zisserman, A. (2008) Automated flower classification over a large number of classes. *6th Indian Conference on Computer Vision, Graphics and Image Processing*, 722–729. https://doi.org/10.1109/ICVGIP.2008.47

Oberbauer, S.F., Elmendorf, S.C., Troxler, T.G., Hollister, R.D., Rocha, A.V., Bret-Harte, M.S. et al. (2013) Phenological response of tundra plants to background climate variation tested using the international tundra experiment. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **368**(1624), 20120481. https://doi.org/10.1098/rstb.2012.0481

Palacios, F., Bueno, G., Salido, J., Diago, M.P., Hernández, I. & Tardaguila, J. (2020) Automated grapevine flower detection and quantification method based on computer vision and deep learning from on-the-go imaging using a mobile sensing platform under field conditions. *Computers and Electronics in Agriculture*, **178**(June). https://doi.org/10.1016/j.compag.2020.105796

Parmesan, C. (2006) Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution, and Systematics*, **37**(1), 637–669. https://doi.org/10.1146/annurev.ecolsys.37.091305.110100

Pearson, K.D., Nelson, G., Aronson, M.F.J., Bonnet, P., Brenskelle, L., Davis, C.C. et al. (2020) Machine learning using digitized herbarium specimens to advance phenological research. *Bioscience*, **70**(7), 610–620. https://doi.org/10.1093/biosci/biaa044

Philipp, M. & Siegismund, H.R. (2003) What can morphology and isozymes tell us about the history of the *Dryas integrifolia-octopetala* complex? *Molecular Ecology*, **12**(8), 2231–2242. https://doi.org/10.1046/j.1365-294X.2003.01875.x

Pimm, S.L., Alibhai, S., Bergl, R., Dehgan, A., Giri, C., Jewell, Z. et al. (2015) Emerging technologies to conserve biodiversity. *Trends in Ecology and Evolution*, **30**(11), 685–696. https://doi.org/10.1016/j.tree.2015.08.008

Post, E. & Høye, T.T. (2013) Advancing the long view of ecological change in tundra systems. introduction. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **368**, 20120477. https://doi.org/10.1098/rstb.2012.0477

Post, E., Steinman, B.A. & Mann, M.E. (2018) Acceleration of phenological advance and warming with latitude over the past century. *Scientific Reports*, **8**, 3927. https://doi.org/10.1038/s41598-018-22258-0

Prevéy, J.S., Elmendorf, S., Bjorkman, A., Alatalo, J., Ashton, I., Assmann, J. et al. (2021) The tundra phenology database: more than two decades of tundra phenology responses to climate change. *Arctic Science* (in Press). https://doi.org/10.1139/as-2020-0041

Prevéy, J.S., Rixen, C., Rüger, N., Høye, T.T., Bjorkman, A.D., Myers-Smith, I.H. et al. (2019) Warming shortens flowering seasons of tundra plant communities. *Nature Ecology & Evolution*, **3**, 45–52. https://doi.org/10.1038/s41559-018-0745-6

Prevéy, J.S., Vellend, M., Rüger, N., Hollister, R.D., Bjorkman, A.D., Myers-Smith, I.H. et al. (2017) Greater temperature sensitivity of plant phenology at colder sites: implications for convergence across northern latitudes. *Global Change Biology*, **23**, 2660–2671. https://doi.org/10.1111/gcb.13619

Rafferty, N.E., Diez, J.M. & Bertelsen, C.D. (2020) Changing climate drives divergent and nonlinear shifts in flowering phenology across elevations. *Current Biology*, **30**, 432–441. https://doi.org/10.1016/j.cub.2019.11.071

Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. (2016) You only look once: unified, real-time object detection. *ArXiv*, 1506.02640v5. https://doi.org/10.1109/CVPR.2016.91

Redmon, J. & Farhadi, A. (2018) YOLOv3: an incremental improvement. *ArXiv*, 1804.02767v1. https://doi.org/10.1109/CVPR.2017.690

Ren, S., He, K., Girshick, R. & Sun, J. (2017) Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Richardson, A.D., Jenkins, J.P., Braswell, B.H., Hollinger, D.Y., Ollinger, S.V. & Smith, M.L. (2007) Use of digital webcam images to track spring green-up in a deciduous broadleaf forest. *Oecologia*, **152**(2), 323–334. https://doi.org/10.1007/s00442-006-0657-z

Schuettpelz, E., Frandsen, P.B., Dikow, R.B., Brown, A., Orli, S., Peters, M. et al. (2017) Applications of deep convolutional neural networks to digitized natural history collections. *Biodiversity Data Journal*, **5**. https://doi.org/10.3897/BDJ.5.e21139

Spiesman, B.J., Gratton, C., Hatfield, R.G., Hsu, W.H., Jepsen, S., McCornack, B. et al. (2021) Assessing the potential for deep learning and computer vision to identify bumble bee species from images. *Scientific Reports*, **11**(1), 1–10. https://doi.org/10.1038/s41598-021-87210-1

Tang, J., Körner, C., Muraoka, H., Piao, S., Shen, M., Thackeray, S.J. et al. (2016) Emerging opportunities and challenges in phenology: a review. *Ecosphere*, **7**(8), e1436.

Thackeray, S.J., Henrys, P.A., Hemming, D., Bell, J.R., Botham, M.S., Burthe, S. et al. (2016) Phenological sensitivity to climate across taxa and trophic levels. *Nature*, **535**(7611), 241–245. https://doi.org/10.1038/nature18608

Tran, D.T., Høye, T.T., Gabbouj, M. & Iosifidis, A. (2018) Automatic flower and visitor detection system. *European Signal Processing Conference*, 2018-Septe, 405–409. doi:10.23919/EUSIPCO.2018.8553494

Visser, M.E. & Both, C. (2005) Shifts in phenology due to global climate change: the need for a yardstick. *Proceedings of the Royal Society B: Biological Sciences*, **272**(1581), 2561–2569. https://doi.org/10.1098/rspb.2005.3356

Wan, S. & Goudos, S. (2020) Faster R-CNN for multi-class fruit detection using a robotic vision system. *Computer Networks*, **168**, 107036. https://doi.org/10.1016/j.comnet.2019.107036

Wang, X.(.A.)., Tang, J. & Whitty, M. (2021) DeepPhenology: estimation of apple flower phenology distributions based on deep learning. *Computers and Electronics in Agriculture*, **185** (March), 106123. https://doi.org/10.1016/j.compag.2021. 106123

Weinstein, B.G. (2018) A computer vision for animal ecology. *Journal of Animal Ecology*, **87**(3), 533–545. https://doi.org/ 10.1111/1365-2656.12780

Welker, J.M., Molau, U., Parsons, A.N., Robinson, C.H. & Wookey, P.A. (1997) Responses of *Dryas octopetala* to ITEX environmental manipulations: a synthesis with circumpolar comparisons. *Global Change Biology*, **3**(S1), 61–73. https:// doi.org/10.1111/j.1365-2486.1997.gcb143.x

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. Flower detection performance on the test set for the standard model (1) and four modified models. Q: Extended augmentation, R: Extended augmentation and increased number of regions of interest (ROIs), S: increased number of ROIs but without extended augmentation, T: ResNet50 as backbone instead of ResNet101. Model Q returns the highest F1 score and was used as the flower detection model.

Figure S2. Left: The time required to process one image at different levels of downscaling of the original image dimensions (x and y).

Figure S3. Number of flowers manually annotated per image through the season for each of 20 image series.

Figure S4. Number of flowers annotated (sampled images) and detected (sampled series and 1-h series) per image through the season for each of the four image series in the four image series in the test set.