RESEARCH ARTICLE

Ecological Solutions and Evidence | BRITISH ECOLOGICAL SOCIETY

Automating field-based floral surveys with machine learning

Nicholas Sookhan[1,2] (iD) | **Shane Sookhan**[3] | **Devlin Grewal**[2] | **J. Scott MacIvor**[1,2,3] (iD)

[1]Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario, Canada

[2]Department of Biological Sciences, University of Toronto Scarborough, Toronto, Ontario, Canada

[3]Department of Physical and Environmental Sciences, University of Toronto Scarborough, Toronto, Ontario, Canada

Correspondence
Nicholas Sookhan
Email: nicholas.sookhan@mail.utoronto.ca

Abstract

1. The abundance and diversity of flowering plant species are important indicators of pollinator habitat quality, but traditional field-based surveying techniques are time-intensive. Therefore, they are often biased due to under-sampling and are difficult to scale.

2. Aerial photography was collected across 10 sites located in and around Rouge National Urban Park, Toronto, Canada using a consumer-grade drone. A convolutional neural network (CNN) was trained to semantically segment, or identify and categorize, pixel clusters which represent flowers in the collected aerial imagery. Specifically, flowers of the dominant taxa found in the depauperate fall flowering plant community were surveyed. This included yellow flowering *Solidago* spp., white *Symphyotrichum ericoides/lanceolatum* and purple *Symphyotrichum novae-angliae*. The CNN was trained using $930\,m^2$ of manually annotated data, ~1% of the mapped landscape. The trained CNN was tested on 20% of the manually annotated data concealed during training. In addition, it was externally validated by comparing the predicted drone-derived floral abundance metrics (i.e. floral area (m^2) and the number of floral patches) to the field-based count of floral units estimated for $34\,4\,m^2$ plots.

3. The CNN returned accurate multiclassification when evaluated against the testing data. It obtained a precision score of 0.769, a recall of 0.849, and an F1 score of 0.807. The automated floral abundance counting yielded estimates that were strongly correlated with field-based manual counting. In addition, flower segmentation using the trained CNN was time-efficient. On average, it took roughly the same amount of time to segment the flowers occurring in an entire drone scene as it took to complete the abundance count of a single quadrat. However, the training process, particularly manual data annotation, was the most time-consuming component of the study.

4. *Practical implication*: Overall, the analysis provided valuable insights into automated flower classification and abundance estimation using drone imagery and machine learning. The results demonstrate that these tools can be used to provide accurate and scalable estimates of pollinator habitat quality. Further research should consider diverse wildflower systems to develop the generalizability of the methods.

1 | INTRODUCTION

In an era of global change, consistent, reproducible, and rigorous monitoring of landscapes are needed to document the impact of ecological pressures on biological diversity (Jetz et al., 2019). The automated processing of components of ecological data streams has increasingly become recognized as part of the solution to the data-deficiency crisis in ecology (Besson et al., 2022; Lürig et al., 2021; Tuia et al., 2022). An example of this deficiency is seen in the mapping of the abundance and diversity of flowering plant species, which are key indicators of habitat quality (Szigeti et al., 2016). Flying insects, such as pollinators that depend on flowering plants, operate within large local landscapes, and often forage in radii larger than 500 m (Kremen et al., 2007). Estimates of floral abundance and diversity are thus generated from quadrats and extrapolated to determine habitat quality (Kearns & Inouye, 1993). Unfortunately, traditional field-based methods used to assess pollinator habitat quality are time-intensive, especially across large spatial extents (Breeze et al., 2021). Thus, assessments of habitat quality are often biased due to under-sampling (Breeze et al., 2021; Szigeti et al., 2016).

In precision agriculture, drones have been paired with machine learning algorithms to automate the mapping of environmental indicators (Benos et al., 2021). Commercial drone platforms are used to map agricultural landscapes with multispectral and light detection and ranging (LiDAR) sensors (Tsouros et al., 2019). Aerial photography captured using low altitude drone flights is orthorectified into sub-meter resolution scenes of entire agricultural landscapes (Sa et al., 2018). For example, drone orthomosaics, which are orthorectified imagery datasets that are geometrically corrected for perspective and terrain effects, are used as the input to machine learning classifiers to automate crop classification and weed detection (Chen et al., 2019; Sa et al., 2018).

Artificial neural network (ANNs) techniques have caused a paradigm shift in automation due to their applicability to a wide breadth of tasks including classification problems (Samek et al., 2021). Convolutional neural networks (CNNs), a class of ANN used in image analysis, outperform other machine learning approaches applied to crop management (Benos et al., 2021). These algorithms generate a label for each pixel in input imagery, a process called semantic segmentation. To accomplish this, imagery data are resampled to different pixel densities to first identify and then label pixel clusters based on abstractions of pixel properties, with sequential downsampling layers used to extract fine to coarse-grained features and upsampling of the encoded feature space then used to predict the semantic segmentation. Accurate semantic segmentation, which refers to the categorization of pixel clusters into classes, is achieved via supervised learning using training data that is manually annotated with expert knowledge. During the cycles of training, known as training epochs, the algorithm iteratively adjusts the layer weights, which are the parameters within each layer of the neural network that determine the strength and direction of connections between neurons. This adjustment allows the feature extraction space to learn pixel properties that correspond to the user-provided target for semantic segmentation. To evaluate generalizability, ensuring that the model works well with new data, the trained classifier is then compared with the manual data annotation of novel testing data. In other words, the classifier is tested using data that it has had no prior exposure to. Deployed CNN classifiers are accurate, generalizable, and scalable when paired with drone-derived data products (Mittal et al., 2020).

In environmental management, research into the application of machine learning techniques to drone-derived data products has shown potential for increasing the efficiency and effectiveness of automated processing (Besson et al., 2022; Lürig et al., 2021). Commercial drone platforms which allow the collection of high-resolution multimodal imagery can be prohibitively expensive for applications in environmental management. However, machine learning techniques have been successfully paired with true colour red, green, and blue (RGB) imagery which can be obtained from more cost-effective consumer-grade drone platforms. This approach has been applied to automate the mapping of flowering plants (de Sá et al., 2018; Hicks et al., 2021; Hill et al., 2017). In de Sa et al. (2018), the authors used a random forest binary classifier paired with drone orthomosaic imagery to map the floral cover of *Acacia longifolia* (Andrews) Willd. (Fabaceae), a mass flowering shrub that is invasive in Portuguese dunes. It was found that automated drone-based monitoring was more efficient than traditional methods as it took a quarter of the time required to monitor flowering in the field. In another example, Hicks et al. (2021) sampled pollinator habitat quality using a CNN object detector. The CNN was trained to automate the counting of floral units in ground-level photography of $1 m^2$ field quadrats. It was determined that the automated assessment took 5.5 s per quadrat compared with 3.2 min for manual counting, a >30× improvement in time efficiency with minimal loss in quality for morphologically distinct taxa.

An automated drone-based biomonitoring program should provide accurate estimates of foraging resource availability that are scalable to spatial extents appropriate to flying insects. In this project, we train a CNN classifier to semantically segment flowers using aerial photography collected from a consumer-grade drone. Environmental management is historically underfunded, and therefore the exploration of consumer-grade drone platforms to investigate low-cost solutions is essential (Anderson & Gaston, 2013). We map semi-natural meadows located in Rouge National Urban Park (RNUP) Ontario, Canada. We focus on fall flowering plants, a low diversity system where the mass flowering of predominantly Asteraceae provides key floral resources for insects that are active late in the season (Ginsberg, 1983; Livingstone et al., 2020). These

include members of the yellow flowering *Solidago* L. (common: Goldenrod), purple *Symphyotrichum novae-angliae* (L.) G.L.Nesom (common: New England aster), as well as *Symphyotrichum ericodes* (L.) G.L.Nesom (common: White Heath Aster) and white *Symphyotrichum lanceolatum* (Willd.) G.L.Nesom (common: Lance-leaved aster).

Our first objective was to train a generalized CNN algorithm (see Figure 1a) for multiclassification of the focal flowering plant taxa using true colour RGB imagery collected with a consumer-grade drone. The floral classification maps predicted by the classifier can then be used to generate accurate drone-derived estimates of pollinator habitat quality across habitat patches that vary in environmental context (e.g. the community composition of plants which compose the background in which segmented flowers occur). Our second objective was to explore the impact of flight altitude on the accuracy of CNN classifier predictions. Flight altitudes are 7, 15, and 30m above-ground level (AGL), which are operationally defined as low, mid, and high altitude, respectively. Compared with high altitudes, low-altitude flights sample a small amount of area relative to the local landscape in which pollinators operate (Table S1). However, spatial resolution is high at low altitude which may allow the discrimination of features not resolvable at higher altitudes. Our final objective was to compare the floral abundance estimates derived from field-based quadrat sampling estimates (i.e. the count of floral units) to automated drone predictions (i.e. floral area (m^2) and the number of floral patches; Figure 1b).

2 | MATERIALS AND METHODS

2.1 | Site description

The Rouge National Urban Park (RNUP) is located within the Greater Toronto Area, which is Canada's largest urban agglomeration. The focal flowering plant taxa are historically abundant within the region, but occupancy and abundance vary substantially among sites (Livingstone et al., 2020). In total, eight 2500m^2 sites within the RNUP were sampled (Figure S1 and Table S2). These sites have been previously surveyed to study the drivers and impacts of plant invasion at the park (Livingstone et al., 2020). An additional field site at the RNUP (A: 43°50′21.96″ N, 79°12′13.87″ W) and another at the University of Toronto Scarborough (CF: 43°47′34.76″ N, 79°11′16.75″ W) were established to measure field-based quadrat sampling estimates for external validation.

2.2 | Drone mapping

The drone image acquisition programme was completed during 2021 between September 5 and September 13 using a consumer-grade drone (DJI Phantom 4 Pro V2, DJI, Shenzhen, China). The DJI Phantom 4 Pro is equipped with a 1-inch complementary metal-oxide semiconductor (CMOS) true colour RGB sensor with a resolution of 20.7 megapixels (5472×3648 pixels) at a focal length of 8.8mm. Imagery was collected using a gridded flight pattern consisting of a series of parallel flight lines over the mapped area with high overlap (forward=70% and side=80%) between sites to account for field-of-view limitations associated with the consumer-grade sensor used here (FOV: 84° 8.8mm/24mm [35mm format equivalent]). Further, we used short (<20min) programmed missions at three altitudes (7, 15, and 30m AGL) and at a maximum speed of 3m/s (Table S1). The three altitudes were selected through qualitative field testing to maximize the resolution of imagery that could be captured per flight. Flights were kept short to account for the draw on battery life associated with flight stabilization in variable wind conditions. Flight speed was limited by the shutter speed of the consumer-grade sensor (8–1/2000s) to reduce the effect of 'motion-blur' in the collected imagery. Each site was mapped with a

FIGURE 1 Flowchart of methods used in the study to classify flowers in drone photography. (a) The convolutional neural network was trained using the training and validation datasets; (b) This was followed with evaluation using the testing dataset as well as external validation using field-based quadrat estimates.
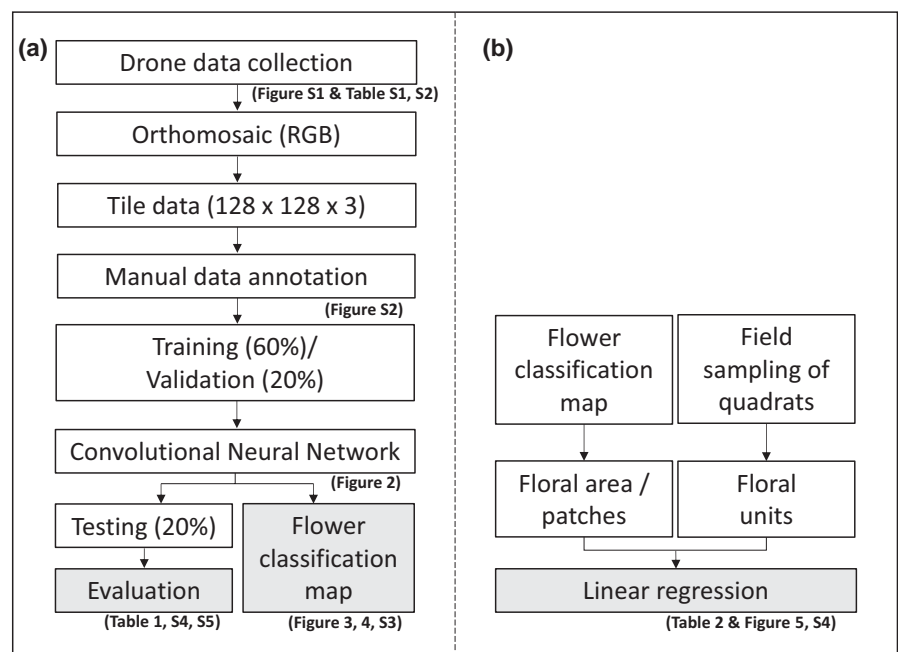
flight programmed using at least one of the three altitudes (Table S1). According to the specifications of the sensor, imagery collected at the 7m, 15m, and 30m flight altitudes has ground sampling distances (GSD) of 0.19cm/pixel, 0.41cm/pixel, and 0.82cm/pixel, respectively. At some sites, multiple programmed flights were flown to either increase spatial coverage at a single flight altitude and/or to map the location at multiple altitudes. Orthomosaics constructed with imagery collected at 7m flight altitude only covered 800 m^2 which is a low spatial coverage relative to the 2500m^2 site areas. Therefore, to achieve higher spatial coverage, two flight missions were flown at sites mapped at the 7m flight altitude. This doubled flight time and resulted in at most 1600m^2 of spatial coverage (Table S1). At 15m, up to 3400m^2 could be mapped with a single flight mission which resulted in complete site coverage. The 30m flight altitude mapped up to 12,400m^2, or in other words almost five times the site area in a single flight. All missions were flown between 11:45 and 14:45h and were programmed using Litchi (www.flylitchi.com), a low-cost drone mission planning software. Imagery was orthorectified using Agisoft Metashape Professional (version 1.8.3). The orthorectification process is seen as a crucial step in increasing imagery fidelity since it can help remedy distortions caused by topographical relief and lens distortion (Park et al., 2022).

2.3 | Floral surveying

Field surveys were completed to validate the drone-derived metrics of floral abundance using field-derived metrics. Thirty-four 4m^2 quadrats were established across the two field sites established for this study (CF=9, A=25). The floral units of each plant species were counted within quadrats from September 11 to September 13. Floral units were defined as single flowers or inflorescences (umbels for Apiaceae and flower heads for Asteraceae). At the remaining sites, visual assessments were completed to determine the flowering plant species present. Visual assessments are comparable to formal quadrat surveys when assessing plant species occupancy (Morrison, 2016).

Vegetative surveys previously conducted at the RNUP field sites have grouped *Solidago* spp. (Livingstone et al., 2020). Thus, during floral surveys and following further analysis, *Solidago* species were grouped at the genus level (i.e. *Solidago* spp.). An additional 10 flowering plant taxa were identified to genus or species (Table S3). Quadrats were dominated by four showy mass flowering Asteraceae taxa: *Solidago* spp., *Symphyotrichum ericoides*, *Symphyotrichum lanceolatum*, and *Symphyotrichum novae-angliae*. These taxa were found to be both locally abundant and common during visual assessments of sites. Crucially, each of these plant genera and species occurs at sizes and colour ranges that are captured within the specifications of the consumer-grade sensor used here when following the collection method described here.

2.4 | Ground truthing

Drone orthomosaics for all altitudes were split into tiles (128 pixels×128 pixels×3 channels), 1% of which were randomly sampled.

To produce the ground truth dataset, flowering parts of plant taxa in the randomly sampled tiles were segmented and labelled using QGIS (https://www.qgis.org). Flowering parts that belonged to *Symphyotrichum ericoides* and *Symphyotrichum lanceolatum* were grouped together because they could not be distinguished from each other at altitudes above 7m.

Following manual data annotation, data from all altitudes were split into training, validation, and testing sets with a 3:1:1 ratio (3563 training, 1192 validation, and 1197 testing tiles). A stratified approach was used where tiles from each orthomosaic were split independently and then collapsed into training, validation, and testing sets. Orthomosaics from all altitudes were used together for training as opposed to creating a series of altitude-specific CNNs to provide a more accurate comparison of model performance at different altitudes. There was a class imbalance in the ground truth data such that *Solidago* spp. was found in 2863 image tiles while *Symphyotrichum ericoides/lanceolatum* and *Symphyotrichum novae-angliae* occurred in 815 and 352, respectively.

2.5 | Convolutional neural network (CNN)

A CNN was trained for patch-wise semantic segmentation of the flower taxa using the machine learning platform Tensorflow v2.10 with the Keras v2.10 interface in Python v3.9.13 (Abadi et al., 2016). 'Patch-wise' refers to the process of dividing the input image into smaller and manageable sections or patches and then performing segmentation on each patch individually rather than on the entire image at once. The CNN was constructed using the U-Net architecture (Ronneberger et al., 2015) (details in Figure 2). RESNET-50, a CNN designed for image classification, was used as the encoder. A key strength of image classifiers is that the knowledge gained from being pre-trained on several hundreds of thousands of images can be transferred to other CNNs to improve the efficiency of model fitting, particularly when small datasets are used in training (Bosilj et al., 2020). This process is referred to as transfer learning and has previously been demonstrated to improve the performance of crop plant semantic segmentation in low-altitude drone data products (Bosilj et al., 2020). In this case, the implemented RESNET-50 encoder was pre-trained on the ImageNet 2012 dataset (1000 classes and 1.28 million training images; He et al., 2015). The output semantic segmentation had four classes: background (i.e. vegetation, bare soil, and other features in the landscape), *Solidago* spp., *Symphyotrichum ericoides/lanceolatum* and *Symphyotrichum novae-angliae*, respectively.

The CNN was trained and deployed using a NVIDIA GeForce RTX 3060. The training and validation ground truth datasets were used to train the CNN. The CNN had a total of 32,561,694 parameters, 12,467,096 of which were trainable. The classifier was trained for up to 300 epochs with a training and validation batch size of 32. An early-stopping criterion was implemented whereby training was halted if validation loss dropped by less than 0.0001 across epochs (patience=20). The number of epochs, training and validation batch size and validation loss threshold were chosen through experimental
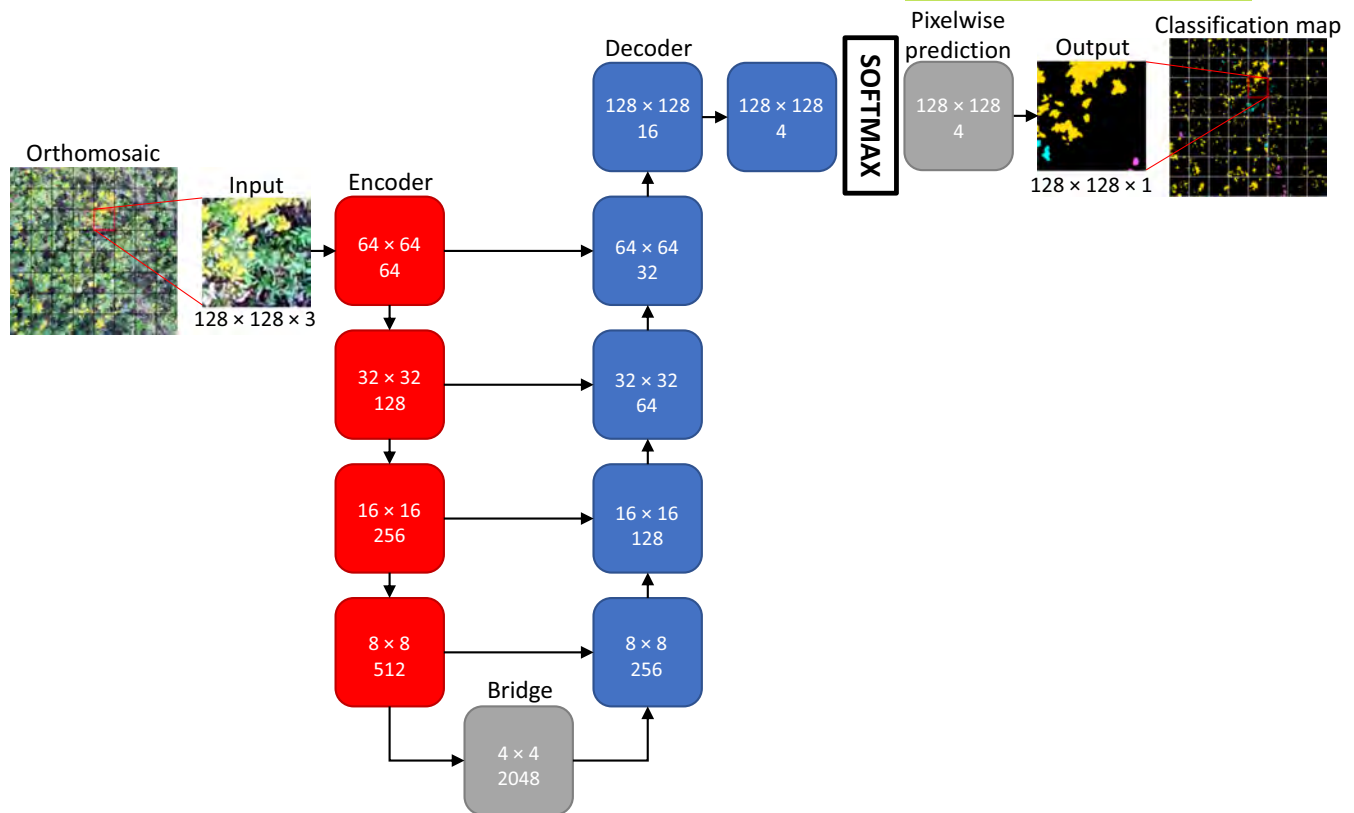
FIGURE 2 Patch-wise semantic segmentation approach using U-Net architecture. U-Net architecture is comprised of five blocks: Input, encoder, bridge, decoder, and output. Drone orthomosaics are split into overlapping tiles (128 pixels × 128 pixels × 3 channels). Input tiles are then encoded using the RESNET-50 architecture. Input resolution is reduced by half and feature channel depth is doubled during each step of the contracting path. The bridge block is an encoded representation of the input tile with low resolution and high feature channel depth. The bridge block is fed to the decoder which doubles input resolution and halves feature channel depth during each step of the expansive path. During decoding, feature maps from the corresponding resolution in the contracting path are concatenated to the expansive path using skip connections. This improves the incorporation of coarse- to fine-grained context information. Softmax activation is applied to the decoded image which produces a pixel-wise prediction where feature channel depth is equal to the number of classes (128 × 128 × 4). Argmax is applied pixel-wise to output the class with the highest predicted probability per pixel (128 × 128 × 1). Non-overlapping sections of semantic segmentation tiles are joined to produce the overall flower classification map.

testing to prevent overfitting, maximize model efficiency and ensure stability on the hardware used. The categorical focal Jaccard loss (i.e. the sum of categorical focal loss and Jaccard loss) was used as the loss function.

Data were pre-processed using min–max scaling. In addition, input tiles were colour augmented and rotated during training and validation (Figure S2) to reduce model overfitting (Shorten & Khoshgoftaar, 2019). Colour augmentations were implemented using the 'imgaug' package in Python (Jung et al., 2020). Brightness, saturation, and temperature augmentation were applied in random order to simulate variation in lighting conditions.

Following training, the CNN was deployed to classify flowers in imagery using a patch-wise semantic segmentation approach. To do this, were split into overlapping tiles (128 × 128 × 3). Tiles overlapped by half of their length and width (i.e. 64 pixels). Semantic segmentation was then completed on the overlapping tiles using the trained classifier. The centre of each predicted tile (64 × 64 × 1) was extracted. Then, the cropped tiles were stitched together to construct the overall floral classification map. This approach was used

to account for decline in quality of classification towards the edges of input imagery that is typically observed in semantic segmentation (Liu et al., 2018).

2.6 | Model evaluation

The testing data, which were novel to the trained classifier, were compared with the floral classification maps to evaluate the quality of classification. The intersection over union (IoU), precision, recall, and F1 scores were used. Ranging from 0 to 1, the IoU score measures the overlap between the ground truth and prediction pixels divided by the union of ground truth and prediction pixels. The higher the value, the more similar the prediction is to the ground truth. The precision, recall, and F1 scores were calculated from the confusion matrix comparing the testing data to the predicted floral classification maps. Precision is the proportion of pixels that were labelled positive that are true-positive where precision score = true-positive/(true-positive + false-positive).

In contrast, recall is the proportion of true-positive pixels that were labelled positive, where recall score = true-positive/(true-positive + false-negative). The precision and recall scores range from 0 to 1 and emphasize different aspects of classification. For a drone orthomosaic, the precision score will be high if the flowers segmented and labelled by the classifier correspond to flowers that occur in the mapped landscape, even if a large proportion of flowers were not identified by the classifier. In contrast, the recall score will be high if most of the flowers that occurred in the landscape were correctly segmented and labelled by the classifier, even if a large proportion of background were also misclassified as flowers. Finally, the $F1$ score is the harmonic mean of precision and recall.

Evaluation metrics were calculated for each flowering plant taxa independently. Ground truth data were grouped by flight altitude to explore the relationship between flight altitude and the quality of classification.

2.7 | External validation

For each taxon, the drone-derived floral abundance metrics were calculated with the quadrats and compared with the field-based counts. The metrics were calculated using floral classification maps generated from data collected at either the 15 m or 30 m flight altitudes. Drone-derived floral abundance metrics include the floral area (m^2) and the count of predicted floral patches. At the plot level, the field-based count of flowers was regressed on the drone-derived metrics using linear regression. For each flight altitude, a separate linear regression model was fit for each flower taxa. The R^2 was used to assess the strength of each relationship. The linear regression models were fitted using R v4.2.2 (R Core Team, 2022).

3 | RESULTS

In total, it took 4 h and 59 min to complete all drone flights and an additional 9 h and 56 min to construct the corresponding drone orthomosaics which mapped 88,200 m^2 across the RNUP and the University of Toronto Scarborough field site (Table S2). The construction of the manually annotated dataset was the most time-consuming component of the analysis pipeline taking ~124 h, or 1.25 min per tile. This was followed by model training a total of 58 epochs which took a total time of 55 min. Finally, prediction of the flower classification maps took a total of 3 h and 23 min across the 12 orthomosaics, an average of 17 min per orthomosaic. Overall, all steps took a total of 143 h and 13 min.

3.1 | Flight altitude

At the 7 m flight altitude (GSD = 0.19 cm/pixel), flower morphology was discernable and the white flowering *Symphyotrichum* species, *S.*

ericoides and *S. lanceolatum*, could be identified from each other, and from other white flowering plant species present (Figure 3). Spatial resolution decreased with flight altitude which made the determination of morphological features used to distinguish these species more difficult. At the 15 m (GSD = 0.41 cm/pixel) and 30 m flight altitude (GSD = 0.82 cm/pixel), taxa were identifiable by spectral characteristics and shape, but other morphological features were generally hard to distinguish (15 m: Figure S3; 30 m: Figure 4). Therefore, it was not possible to discriminate between the two white *Symphyotrichum* species. *Solidago* and *Symphyotrichum novae-angliae* could be reliably identified in aerial photography collected at any flight altitude due to a combination of a unique floral hue and morphology within the context of the landscape.

3.2 | Model evaluation

The confusion matrix used to calculate the model evaluation metrics is reported in Table S4. It was found that precision, recall, and the $F1$ scores were high (Table 1). The classifier was most accurate at segmenting and labelling floral units belonging to *Solidago* ($F1 = 0.82$) this was followed by *Symphyotrichum novae-angliae* ($F1 = 0.79$) and *Symphyotrichum ericoides/lanceolatum* ($F1 = 0.70$; Table 1). There were no clear trends in classifier performance between flight altitudes (see Table S5 for a breakdown by flight altitude).

The three focal taxa were generally not confused in classification and other flowering plants were not falsely classified as any of the three focal taxa (Table S4). Background pixels were generally not misclassified as flowers vice-versa, except for *Symphyotrichum ericoides/lanceolatum* which had relatively low precision and recall, respectively.

3.3 | External validation

It took a total of 8 h and 40 min to complete quadrat sampling of the 34 plots or 15.30 min per plot. The drone-derived metrics of floral abundance were strongly correlated with the counts of floral units measured in the field using quadrat sampling at both the 15 m and the 30 m flight altitudes (15 m: Figure 5, 30 m: Figure S4). However, the relationships were weaker when fitted with the 30 m data than with the 15 m data where the R^2 was up to 0.26 lower (Table 2).

4 | DISCUSSION

4.1 | Flight altitude

The spatial resolution of an orthomosaic was negatively correlated with drone flight altitude during the capture of aerial photography, with the drop-off in resolution related to the pixel resolution, field-of-view, and focal length range of the consumer-grade sensor. Despite this, multiclassification performed well across the three
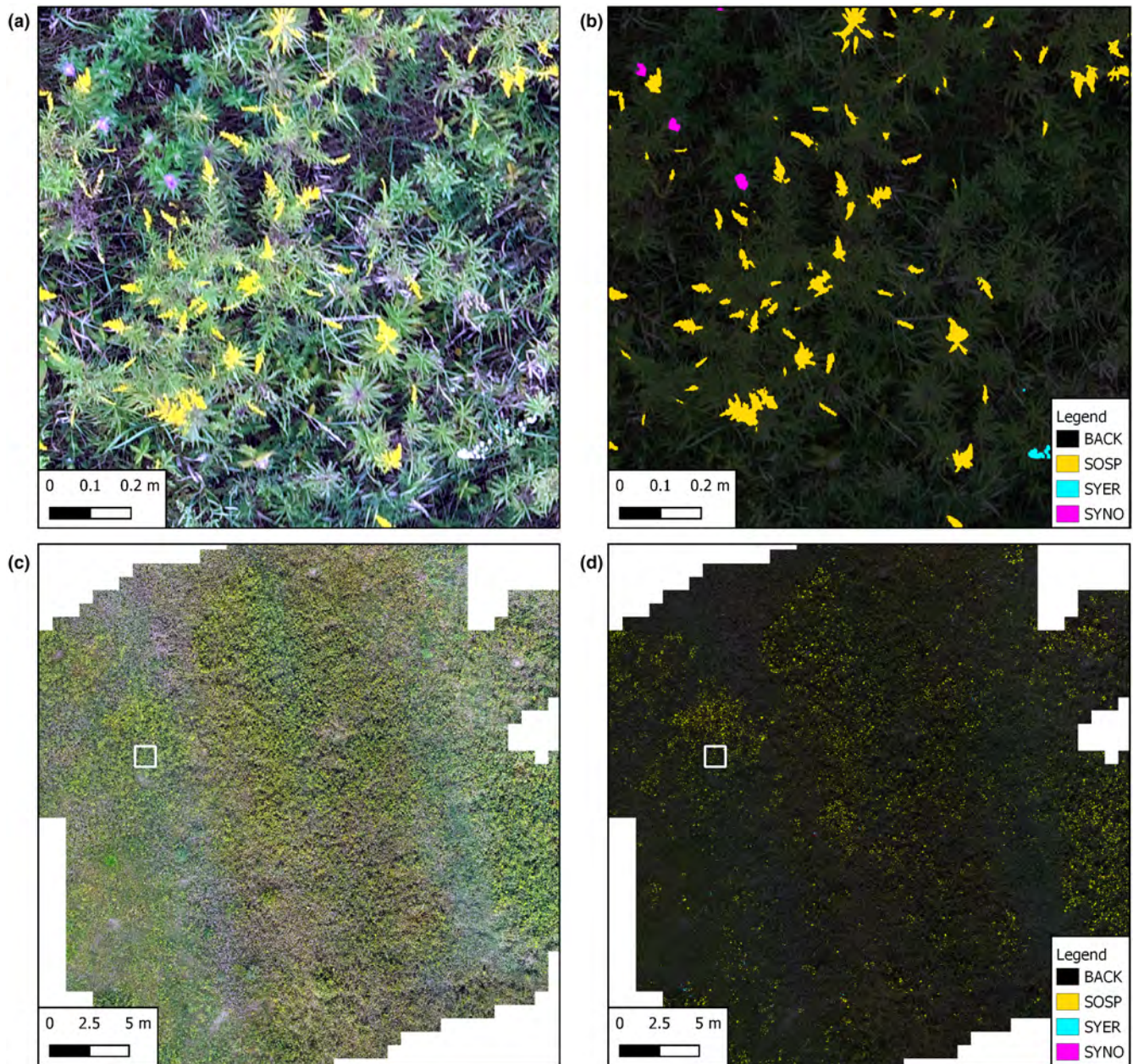
FIGURE 3 Example of drone orthomosaic and flower classification map from imagery collected at 7 m altitude. (a) Close-range view of drone orthomosaic; (b) Close-range view of classification map; (c) Long-range-view of drone orthomosaic, the white box denotes area depicted in close-range view; (d) Long-range view of classification map.

tested flight altitudes. Semi-natural meadows are complex environments where plant structural and spectral diversity is high. Complex backgrounds pose a challenge to accurate automated segmentation and labelling (Sa et al., 2018). Although this is the case, focal taxa which included yellow and white flowering parts, were not confused with senescing vegetative or reproductive plant matter. In addition, the increased spatial resolution had a trade-off with spatial coverage of mapping. At the low (7 m) flight altitude, it would be necessary to complete multiple flights to achieve full spatial coverage of the sampled sites (2500 m^2), while in contrast, at the mid (15 m) and high (30 m) altitude, landscape-level coverage was achieved with a single flight. The spatial extent of the sampled sites was designed

to test questions relating to plant ecology (Livingstone et al., 2020) and does not translate to meaningful coverage of the local landscape within which pollinators operate (Kremen et al., 2004).

At first glance, it might be inferred that high-flight altitude is all-together superior to low altitude when capturing aerial photography for the estimation of pollinator habitat quality. However, moderate performance loss was observed in the external validation procedure which compared mid and high-altitude drone surveying in terms of correlation to field counts. Furthermore, it is important to note that the flowering plant species pool was depauperate during the time of year that the region was sampled. In the fall, late flowering members of the Asteraceae family produce
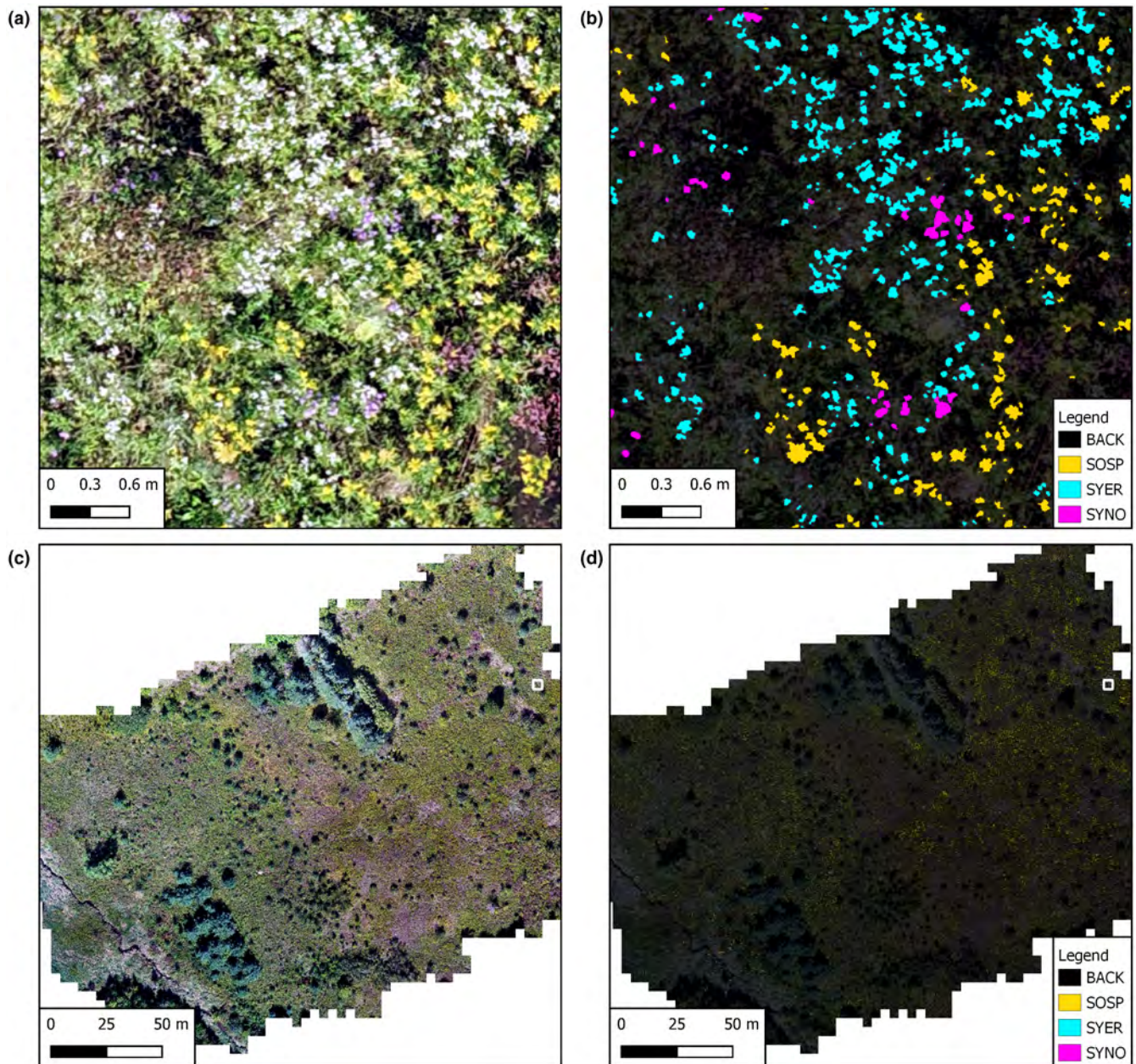
FIGURE 4 Example of drone orthomosaic and flower classification map from imagery collected at 30 m altitude. (a) Close-range view of drone orthomosaic; (b) Close-range view of classification map; (c) Long-range view of drone orthomosaic, the white box denotes area depicted in close-range view; (d) Long-range view of classification map.

TABLE 1 Metrics values evaluating CNN classifier performance.

Metric	Overall	SOSP	SYER	SYNO
Jaccard	0.677	0.697	0.541	0.646
Precision	0.769	0.776	0.709	0.800
Recall	0.849	0.872	0.696	0.770
F1	0.807	0.821	0.702	0.785

Note: The column 'Overall' represents the weighted metric values among classified taxa. The other columns are the classified taxa where 'SOSP' is *Solidago* spp., 'SYER' is *Symphyotrichum ericoides/lanceolatum* and 'SYNO' is *Symphyotrichum novae-angliae*.

large floral displays in southern Ontario. This is especially true of members of the *Solidago* and *Symphyotrichum* genera in RNUP. Due to the simplicity of the floral landscape, it was possible to sort flowers into taxonomic units. During manual data annotation, the coarse-grained features that could be resolved in mid and high-flight altitude drone orthomosaics correlated well with floral morphological traits that were diagnostic of the focal taxa. This would not be the case in a high-diversity wildflower system, where high-resolution taxonomic classification would require a complex suite of floral traits that are not resolved at the mid- and high-flight altitude.
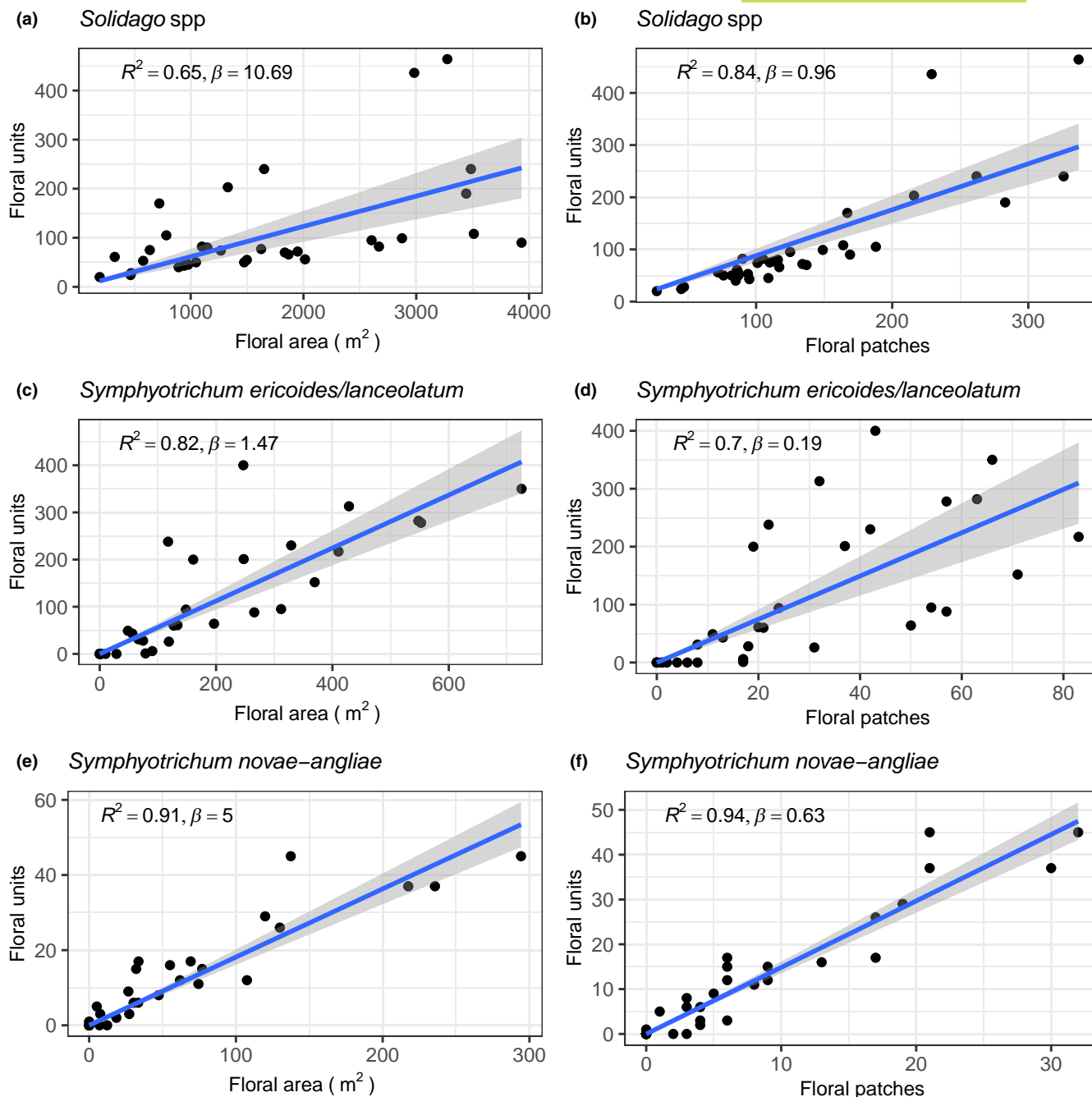
FIGURE 5 Comparison between floral counts using quadrat sampling (*y*-axis) and the drone-derived floral metrics predicted by the CNN classifier (*x*-axis). The drone-derived floral metrics were calculated using drone orthomosaics collected at 15 m altitude. Each data point corresponds to the estimates from a sampled quadrat. 'R^2' is goodness-of-fit of the linear regression measured as the adjusted R^2 and 'β' is the effect size relating the drone-derived metric to the quadrat sampling metric. (a) Floral unit count using quadrat sampling regressed on drone-derived floral area for *Solidago* species; (b) Floral unit count regressed on drone-derived floral patch count for *Solidago* species; (c) Floral unit count regressed on floral area for *Symphyotrichum ericoides/lanceolatum*; (d) Floral unit count regressed on floral patch count for *Symphyotrichum ericoides/lanceolatum*; (e) Floral unit count regressed on floral area for *Symphyotrichum novae-angliae*; (f) Floral unit count regressed on floral patch count for *Symphyotrichum novae-angliae*.

In practice, the CNN classifier was able to distinguish between *Symphyotrichum ericoides/lanceolatum* and *Daucus carota*. These were the only two taxa common in the landscape that have white flowers. The feature extraction space of the trained CNN classifier discriminated taxa using coarse differences in the pixel colour and shape of proposed segments. Floral morphology of the Apiaceae family, of which *Daucus carota* is a member, vary substantially from Asteraceae, for example in the arrangement and shape of inflorescences. In contrast, it was not possible to discriminate between *Symphyotrichum ericodes* and *Symphyotrichum lanceolatum*. The

TABLE 2 Goodness-of-fit (R^2) of the linear regression models where the floral counts using quadrat sampling were regressed on the drone-derived floral metrics predicted by the CNN classifier.

Taxa	Drone	Altitude	
		15 m	30 m
SOSP	Floral patches	0.84	0.74
	Floral area (m^2)	0.65	0.64
SYER	Floral patches	0.70	0.63
	Floral area (m^2)	0.82	0.71
SYNO	Floral patches	0.94	0.73
	Floral area (m^2)	0.91	0.65

Note: For each altitude, the R^2 model is listed for the regression model fitted to each flower taxa. 'SOSP' is *Solidago* spp., 'SYER' is *Symphyotrichum ericoides/lanceolatum* and 'SYNO' is *Symphyotrichum novae-angliae*.

diagnostic floral traits required to distinguish these species, such as variation in the number of ray florets and the arrangement of inflorescences, are not resolved at the spatial resolution provided by mid to high orthomosaics.

When applied to systems with a speciose flowering plant community, classifiers trained and deployed on high-altitude true colour RGB drone orthomosaics will likely, at best, perform multiclassification of flowers based on spectra instead of taxonomic classes. Unfortunately, floral visual cues used by pollinators are not strongly correlated with pollen and nectar reward quality and quantity (Ortiz et al., 2020). Thus, an intuitive recommendation is that high-altitude true colour RGB drone-derived automated floral mapping is limited to the abundance and distribution of conspicuous flowers and therefore has potential in the biomonitoring of invasive species on the landscape level (de Sá et al., 2018; Hill et al., 2017). In contrast, classifiers trained and deployed on low-altitude data can be used in the mapping of high-resolution taxonomic units that are more appropriate for the estimation of pollinator habitat quality.

4.2 | Drone-derived metrics as indicators of pollinator habitat quality

The drone-derived metrics included floral area which is an important visual and olfactory cue to pollinators (Chittka & Raine, 2006). Foraging insects utilize multiple sensor modalities when locating and learning efficient search paths to floral patches that occur within the landscape (Sprayberry, 2018). The size of a flowering patch is correlated with the size of the visual and olfactory stimuli that it presents, which in turn moderates forager efficacy (Sprayberry, 2018). In other words, in landscapes that feature high floral areas, pollinators will spend less time searching and learn efficient routes to available foraging resources. Larger flowering patches are associated with heightened pollinator activity (for example higher visitation rates: Blaauw & Isaacs, 2014; Dauber et al., 2010), and therefore can

provide insight into locations where pollinators concentrate foraging effort (Baldock et al., 2019).

The count of floral units is surrogate for floral resource availability in a habitat patch (Kearns & Inouye, 1993). Foraging resource availability can be measured with a tabulated species list of estimated per-floral unit foraging reward. Given a habitat patch, foraging resource availability is measured as the sum of reward estimates weighted by the floral unit abundance of species in that habitat patch (Hicks et al., 2016). As accurate abundance data are necessary, quadrat and transect sampling are the preferred methods of field collection (Szigeti et al., 2016). Unfortunately, these methods are expensive and time-consuming; for Hegland et al. (2010), it took a single surveyor 8 h and 40 min to complete the collection of data from 34 4 m^2 quadrats. In their meta-analysis of pollination studies that estimated floral resource availability, Szigeti et al. (2016) found that studies generated estimates from an average of 0.69% sampling coverage. While Szigeti et al. (2016) was unable to provide a recommendation for adequate sampling coverage, this value is clearly low. This is because flowering plants tend to occur in a clustered spatial dispersion pattern (Hatfield & LeBuhn, 2007) which increases the sampling effort required for an unbiased estimate of density used in extrapolation to the landscape (Aberdeen, 1958).

The machine learning techniques utilized in this study demonstrate the potential of automating the landscape-level measurement of floral resource availability. This is because the estimated linear models found strong correlations between the drone-derived metrics and the field-based counts of floral units. Although this was the case, it would be preferable for the CNN classifier to complete object detection of floral units, such that floral units are counted and tabulated by species from the provided drone orthomosaics. The CNN object detector deployed by Hicks et al. (2021) accomplished this task using ground-level photos of 1 m^2 field quadrats. The trained classifier was not only able to annotate floral units accurately, but also label annotations to high taxonomic resolution, typically to species. Automated counts were multiplied by corresponding species-specific pollen and nectar measurements to estimate quadrat-level pollen and nectar availability. In this study, the features of floral units belonging to Asteraceae observed were resolved in low-altitude aerial photography. Thus, although spatial coverage is low, it should be possible to replicate the method presented by Hicks et al. (2021) using low-altitude orthomosaics. In this case, multiple drone flights can be strategically placed. This will allow accurate classification and adequate sampling coverage while reducing the time and cost of estimating pollinator habitat quality.

5 | CONCLUSIONS

The results demonstrate that CNN classifiers can be deployed to accurately map wildflowers using consumer-grade drone imagery. Here, a CNN was trained to perform on imagery collected across meadow

habitat patches that varied in environmental conditions, specifically vegetative invasion intensity. Only moderate performance loss was observed with increasing flight altitude. Nevertheless, it was found that the automated drone-based floral monitoring correlated strongly with quadrat-based floral counts measured in the field. Crucially, this highlights that the approach can be used for the rapid assessment of floral resources. Standard approaches to floral resource estimation are time-consuming, especially across large spatial extents (Szigeti et al., 2016). Hence, by training with imagery collected at several flight altitudes, an altitude-agnostic CNN can be deployed to adjust automated mapping to data collection needs (e.g. consideration of the flight altitude-related trade-off between accuracy and sample coverage).

The bottleneck of this approach is the construction of a manually annotated dataset. This is a time-consuming but necessary process as high-quality ground truth data are required for training. However, as shown in this study, the return on initial investment will be high as a deployed CNN classifier can provide low-cost and rapid monitoring of floral resources. This is important in applications where high spatio-temporal resolution is needed, such as in long-term biomonitoring programmes or collaborations over large study areas containing common habitat types (Besson et al., 2022; Breeze et al., 2021; Szigeti et al., 2016).

AUTHOR CONTRIBUTIONS

Nicholas Sookhan, Shane Sookhan, and J. Scott MacIvor designed the project; Nicholas Sookhan, Shane Sookhan, and Devlin Grewal collected data and completed manual data annotation; Nicholas Sookhan and Shane Sookhan analysed the data; all authors contributed to the writing and editing of the manuscript drafts and final version.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/2688-8319.12393.

DATA AVAILABILITY STATEMENT

The Python code and the data on the flowering plant taxa described from field measurements and derived from drone imagery are published on Zenodo: https://doi.org/10.5281/zenodo.13834747 (Sookhan, 2024). The drone orthomosaics, predicted floral classification maps, and the trained Tensorflow model (HDF5 format) are available from the Dryad repository: https://doi.org/10.5061/dryad.nvx0k6f1t (Sookhan et al., 2024).

ORCID

Nicholas Sookhan �iD https://orcid.org/0009-0008-0279-1375
J. Scott MacIvor �iD https://orcid.org/0000-0002-2443-8192

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., … Zheng, X. (2016). TensorFlow: Large-scale machine learning on heterogeneous distributed systems. ArXiv:1603.04467 [Cs]. http://arxiv.org/abs/1603.04467

Aberdeen, J. (1958). The effect of quadrat size, plant size, and plant distribution on frequency estimates in plant ecology. *Australian Journal of Botany*, 6(1), 47–58.

Anderson, K., & Gaston, K. J. (2013). Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Frontiers in Ecology and the Environment*, 11(3), 138–146. https://doi.org/10.1890/120150

Baldock, K. C. R., Goddard, M. A., Hicks, D. M., Kunin, W. E., Mitschunas, N., Morse, H., Osgathorpe, L. M., Potts, S. G., Robertson, K. M., Scott, A. V., Staniczenko, P. P. A., Stone, G. N., Vaughan, I. P., & Memmott, J. (2019). A systems approach reveals urban pollinator hotspots and conservation opportunities. *Nature Ecology & Evolution*, 3(3), 363–373. https://doi.org/10.1038/s41559-018-0769-y

Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. *Sensors*, 21(11), 3758. https://doi.org/10.3390/s21113758

Besson, M., Alison, J., Bjerge, K., Gorochowski, T. E., Høye, T. T., Jucker, T., Mann, H. M. R., & Clements, C. F. (2022). Towards the fully automated monitoring of ecological communities. *Ecology Letters*, 25, 2753–2775. https://doi.org/10.1111/ele.14123

Blaauw, B. R., & Isaacs, R. (2014). Larger patches of diverse floral resources increase insect pollinator density, diversity, and their pollination of native wildflowers. *Basic and Applied Ecology*, 15(8), 701–711. https://doi.org/10.1016/j.baae.2014.10.001

Bosilj, P., Aptoula, E., Duckett, T., & Cielniak, G. (2020). Transfer learning between crop types for semantic segmentation of crops versus weeds in precision agriculture. *Journal of Field Robotics*, 37(1), 7–19. https://doi.org/10.1002/rob.21869

Breeze, T. D., Bailey, A. P., Balcombe, K. G., Brereton, T., Comont, R., Edwards, M., Garratt, M. P., Harvey, M., Hawes, C., Isaac, N., Jitlal, M., Jones, C. M., Kunin, W. E., Lee, P., Morris, R. K. A., Musgrove, A., O'Connor, R. S., Peyton, J., Potts, S. G., … Carvell, C. (2021). Pollinator monitoring more than pays for itself. *Journal of Applied Ecology*, 58(1), 44–57. https://doi.org/10.1111/1365-2664.13755

Chen, Y., Lee, W. S., Gan, H., Peres, N., Fraisse, C., Zhang, Y., & He, Y. (2019). Strawberry yield prediction based on a deep neural network using high-resolution aerial orthoimages. *Remote Sensing*, 11(13), 1584. https://doi.org/10.3390/rs11131584

Chittka, L., & Raine, N. E. (2006). Recognition of flowers by pollinators. *Current Opinion in Plant Biology*, 9(4), 428–435. https://doi.org/10.1016/j.pbi.2006.05.002

Dauber, J., Biesmeijer, J. C., Gabriel, D., Kunin, W. E., Lamborn, E., Meyer, B., Nielsen, A., Potts, S. G., Roberts, S. P. M., Sõber, V., Settele, J., Steffan-Dewenter, I., Stout, J. C., Teder, T., Tscheulin, T., Vivarelli, D., & Petanidou, T. (2010). Effects of patch size and density on flower visitation and seed set of wild plants: A pan-European approach. *Journal of Ecology*, *98*(1), 188–196. https://doi.org/10.1111/j.1365-2745.2009.01590.x

de Sá, N. C., Castro, P., Carvalho, S., Marchante, E., López-Núñez, F. A., & Marchante, H. (2018). Mapping the flowering of an invasive plant using unmanned aerial vehicles: Is there potential for biocontrol monitoring? *Frontiers in Plant Science*, *9*, 293. https://doi.org/10.3389/fpls.2018.00293

Ginsberg, H. S. (1983). Foraging ecology of bees in an old field. *Ecology*, *64*(1), 165–175. https://doi.org/10.2307/1937338

Hatfield, R. G., & LeBuhn, G. (2007). Patch and landscape factors shape community assemblage of bumble bees, Bombus spp. (Hymenoptera: Apidae), in montane meadows. *Biological Conservation*, *139*(1), 150–158. https://doi.org/10.1016/j.biocon.2007.06.019

He, K., Zhang, X., Ren, S., & Sun, J. (2015, December 10). Deep residual learning for image recognition. *ArXiv.Org*. https://arxiv.org/abs/1512.03385v1

Hegland, S. J., Dunne, J., Nielsen, A., & Memmott, J. (2010). How to monitor ecological communities cost-efficiently: The example of plant–pollinator networks. *Biological Conservation*, *143*(9), 2092–2101. https://doi.org/10.1016/j.biocon.2010.05.018

Hicks, D., Baude, M., Kratz, C., Ouvrard, P., & Stone, G. (2021). Deep learning object detection to estimate the nectar sugar mass of flowering vegetation. *Ecological Solutions and Evidence*, *2*(3), e12099. https://doi.org/10.1002/2688-8319.12099

Hicks, D. M., Ouvrard, P., Baldock, K. C. R., Baude, M., Goddard, M. A., Kunin, W. E., Mitschunas, N., Memmott, J., Morse, H., Nikolitsi, M., Osgathorpe, L. M., Potts, S. G., Robertson, K. M., Scott, A. V., Sinclair, F., Westbury, D. B., & Stone, G. N. (2016). Food for pollinators: Quantifying the nectar and pollen resources of urban flower meadows. *PLoS One*, *11*(6), e0158117. https://doi.org/10.1371/journal.pone.0158117

Hill, D. J., Tarasoff, C., Whitworth, G. E., Baron, J., Bradshaw, J. L., & Church, J. S. (2017). Utility of unmanned aerial vehicles for mapping invasive plant species: A case study on yellow flag iris (*Iris pseudocorus* L.). *International Journal of Remote Sensing*, *38*(8–10), 2083–2105. https://doi.org/10.1080/01431161.2016.1264030

Jetz, W., McGeoch, M. A., Guralnick, R., Ferrier, S., Beck, J., Costello, M. J., Fernandez, M., Geller, G. N., Keil, P., Merow, C., Meyer, C., Muller-Karger, F. E., Pereira, H. M., Regan, E. C., Schmeller, D. S., & Turak, E. (2019). Essential biodiversity variables for mapping and monitoring species populations. *Nature Ecology & Evolution*, *3*(4), Article 4. https://doi.org/10.1038/s41559-019-0826-1

Jung, A. B., Wada, K., Crall, J., Tanaka, S., Graving, J., Reinders, C., Yadav, S., Banerjee, J., Vecsei, G., Kraft, A., Rui, Z., Borovec, J., Vallentin, C., Zhydenko, S., Pfeiffer, K., Cook, B., Fernández, I., De Rainville, F.-M., Weng, C.-H., ... Laporte, M. (2020). Imgaug. https://github.com/aleju/imgaug

Kearns, C. A., & Inouye, D. W. (1993). *Techniques for pollination biologists*. University Press of Colorado.

Kremen, C., Williams, N. M., Aizen, M. A., Gemmill-Herren, B., LeBuhn, G., Minckley, R., Packer, L., Potts, S. G., Roulston, T., Steffan-Dewenter, I., Vázquez, D. P., Winfree, R., Adams, L., Crone, E. E., Greenleaf, S. S., Keitt, T. H., Klein, A.-M., Regetz, J., & Ricketts, T. H. (2007). Pollination and other ecosystem services produced by mobile organisms: A conceptual framework for the effects of land-use change. *Ecology Letters*, *10*(4), 299–314. https://doi.org/10.1111/j.1461-0248.2007.01018.x

Kremen, C., Williams, N. M., Bugg, R. L., Fay, J. P., & Thorp, R. W. (2004). The area requirements of an ecosystem service: Crop pollination by native bee communities in California. *Ecology Letters*, *7*(11), 1109–1119. https://doi.org/10.1111/j.1461-0248.2004.00662.x

Liu, Y., Ren, Q., Geng, J., Ding, M., & Li, J. (2018). Efficient patch-wise semantic segmentation for large-scale remote sensing images. *Sensors*, *18*(10), Article 10. https://doi.org/10.3390/s18103232

Livingstone, S. W., Isaac, M. E., & Cadotte, M. W. (2020). Invasive dominance and resident diversity: Unpacking the impact of plant invasion on biodiversity and ecosystem function. *Ecological Monographs*, *90*(4), e01425. https://doi.org/10.1002/ecm.1425

Lürig, M. D., Donoughe, S., Svensson, E. I., Porto, A., & Tsuboi, M. (2021). Computer vision, machine learning, and the promise of phenomics in ecology and evolutionary biology. *Frontiers in Ecology and Evolution*, *9*, 642774. https://www.frontiersin.org/articles/10.3389/fevo.2021.642774

Mittal, P., Singh, R., & Sharma, A. (2020). Deep learning-based object detection in low-altitude UAV datasets: A survey. *Image and Vision Computing*, *104*, 104046. https://doi.org/10.1016/j.imavis.2020.104046

Morrison, L. W. (2016). Observer error in vegetation surveys: A review. *Journal of Plant Ecology*, *9*(4), 367–379. https://doi.org/10.1093/jpe/rtv077

Ortiz, P. L., Fernández-Díaz, P., Pareja, D., Escudero, M., & Arista, M. (2020). Do visual traits honestly signal floral rewards at community level? *Functional Ecology*, *35*(2), 369–383. https://doi.org/10.1111/1365-2435.13709

Park, J., Cho, Y. K., & Kim, S. (2022). Deep learning-based UAV image segmentation and inpainting for generating vehicle-free orthomosaic. *International Journal of Applied Earth Observation and Geoinformation*, *115*, 103111. https://doi.org/10.1016/j.jag.2022.103111

R Core Team. (2022). *R (4.2.2): A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation (arXiv:1505.04597). *arXiv*. http://arxiv.org/abs/1505.04597

Sa, I., Popović, M., Khanna, R., Chen, Z., Lottes, P., Liebisch, F., Nieto, J., Stachniss, C., Walter, A., & Siegwart, R. (2018). WeedMap: A large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming. *Remote Sensing*, *10*(9), Article 9. https://doi.org/10.3390/rs10091423

Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Müller, K.-R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, *109*(3), 247–278. https://doi.org/10.1109/JPROC.2021.3060483

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*(1), 1–48.

Sookhan, N., Sookhan, S., Grewal, D., & MacIvor, J. S. (2024). Data from: Automating field based floral surveys with machine learning [dataset]. *Dryad*. https://doi.org/10.5061/dryad.nvx0k6f1t

Sookhan, N. (2024). nsookhan/floral_segmentation: Code for: Automating field based floral surveys with machine learning (V1.0). *Zenodo*. https://doi.org/10.5281/zenodo.13834747

Sprayberry, J. D. H. (2018). The prevalence of olfactory- versus visual-signal encounter by searching bumblebees. *Scientific Reports*, *8*(1), Article 1. https://doi.org/10.1038/s41598-018-32897-y

Szigeti, V., Kőrösi, Á., Harnos, A., Nagy, J., & Kis, J. (2016). Measuring floral resource availability for insect pollinators in temperate grasslands—A review. *Ecological Entomology*, *41*(3), 231–240. https://doi.org/10.1111/een.12298

Tsouros, D. C., Bibi, S., & Sarigiannidis, P. G. (2019). A review on UAV-based applications for precision agriculture. *Information*, *10*(11), 349.

Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., Mathis, A., Mathis, M. W., van Langevelde, F., Burghardt, T., Kays, R., Klinck, H., Wikelski, M., Couzin, I. D., van Horn, G., Crofoot, M. C., Stewart, C. V., & Berger-Wolf, T. (2022). Perspectives in machine learning for wildlife conservation. *Nature Communications*, *13*(1), Article 1. https://doi.org/10.1038/s41467-022-27980-y

SUPPORTING INFORMATION

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

Table S1. Flight altitude versus spatial coverage (m^2), flight speed (m/s) and ground sampling distance (GSD) (i.e. spatial resolution).

Table S2. Site coordinates, drone mapping and orthomosaic parameters, and ground truth breakdown by site.

Table S3. Forbs in flower that were observed during September ground truthing of sites at Rouge National Urban Park.

Table S4. A confusion matrix comparing the trained CNN classifier to the ground truth testing dataset.

Table S5. Metric values evaluating CNN classifier performance by flight altitude and classified taxa.

Figure S1. Locations of meadow field sites at Rouge National Urban Park. The size of red bounding boxes depicts the area sampled.

Figure S2. Data augmentation pipeline applied to input images ($128 \times 128 \times 3$) during training and validation of classifier.

Figure S3. Example of drone orthomosaic and flower classification map from imagery collected at 15m altitude.

Figure S4. Comparison between floral counts using quadrat sampling (*y*-axis) and the drone-derived floral metrics predicted by the CNN classifier (*x*-axis).