
IMAGO: ULTRACHEAP SCALING FOR FLORAL VISITATION MONITORING *

Caleb Sowers

Department of Entomology
Kansas State University
Manhattan, Kansas, USA

{Caleb Sowers}cdsowers@ksu.edu

ABSTRACT

Emerging evidence of widespread pollinator declines has reinvigorated scientific interest in the application of graph theory to mutualistic networks to help predict the asymptotic behavior of ecological systems. These theoretical advances have yielded few useful conservation products, primarily owing to the stifling cost of sampling pollination networks at scale. I introduce *Imago*, a software tool for automatically extracting plant-pollinator visitation networks from image and video. *Imago* employs a novel 'fat' detector architecture, where a general classification model is applied to each of an arbitrary number of partially-overlapping crops of a large parent photo. This technique subverts the need for an expensive bounding-box detection dataset, while also allowing a unique level of deployment-time accuracy vs. performance configurability. I validate the tool's performance against manual analysis for an empirical pollination network study, demonstrating that *Imago* can potentially lower per-hour sampling costs for pollination ecology by several orders of magnitude. Finally, I speculate on the possible impacts of ultracheap continuous sampling for network-theoretic ecology, particularly for enabling considerations of pollination dynamics across time and space.

Keywords ecology · applied computer vision · detection

1 Introduction

Animal pollination of flowering plants is a process critical for maintaining global food security and biodiversity. However, global pollinator populations are threatened by a synthesis of primarily anthropogenic threats, including indiscriminate pesticide application, habitat destruction, and climatic disruptions. These forces particularly endanger non-cosmopolitan native bees, whose relatively specialized requirements are especially difficult to fulfill relative to those of generalist bees with a worldwide distribution.

The cooperative relationship between plants and their pollinators presents mutual benefits for the fitness and biodiversity of both parties (cite). Diverse floral communities that exist across a wide swath of ecological niche space are needed to maintain the full diversity of pollinator species for a given ecosystem. Thus, plant-pollinator networks, which map the topology and 'strength' of the mutualistic relationships between an ecosystem's plant and pollinators, have emerged as an important tool for theoretically understanding the function of pollination for maintaining diversity and stability in biological systems.

Plant-pollinator networks (i.e. pollination networks) are a subset of bipartite graphs for which the two disjoint, independent sets U and V respectively represent the set of all flowering plants and pollinators within an ecosystem. Although it is popular to assume that pollination networks are square (ie. of equal cardinality) because of the relative ease of modeling balanced bipartite graphs, empirical plant-pollinator networks are often imbalanced, and composition of the sets U and V (hereafter plant P and animal A) can only be determined through field observation. Significant effort is required to determine the full set of linkages between plants and pollinators for a specific location, and many efforts are thought to suffer from undersampling [need to cite this]. Despite its expense, a fully-realized *topological*

**Citation: Authors. Title. Pages.... DOI:000000/11111.*

plant-pollinator network is only a first-order approximation of the interreal dynamics of a pollination system. Its most fundamental (and perhaps most spurious) implicit assumption is that all plant-pollination linkages equally influence the community's dynamics. Most theorists find this to be an unreasonable assumption. As such, some pollination networks additionally consider the relative intensity of interactions by weighting the edges of this graph to reflect the number of times a particular plant-animal interaction was observed. The best method by which to weight these *quantitative plant-pollinator networks* is hotly contested by theoretical ecologists, who have yet to reconcile basic theoretical components for weighting the marginal impact of an additional association. Perhaps the most commonly-used definition of interaction strength fails to distinguish between per-capita effects and interaction frequency, yielding the strange property that abundant species (typically generalists) provide higher per-capita effects than rare species. This assumption converts the bulk of theoretical and empirical evidence, and yet this contradiction was only remonstrated in 2019 [1].

Even if theorists agreed on a maximally informative quantitative weighting schema, empirical applications would still be fundamentally constrained by a dearth of empirical information regarding the quality of plant-pollinator associations (i.e. amount of pollen transferred & rewards conferred) and the persistent challenge of insufficient sampling. The significant challenges that afflict field ecologists in sampling topological plant-pollinator networks are amplified when dealing with quantitatively-weighted networks. Let us assume that a quantitative measure of linkage intensity will minimally comprise of the relative commonality of each linkage. Then, at minimum, a proper sample must now saturate the relative accumulation curve of each interaction. For any community of non-trivial species richness, any proper measure of linkage abundance which requires that the least common interaction type occur at least twice will require approximately an order of magnitude greater sampling effort than that required to sample a topological presence-absence network² This demands that we treat the temporal aspect of the network with greater precision: over what timespan do we define our network? With sufficient sampling intensity, bi-weekly samples may be sufficient to encounter at least one example of each plant-pollinator interaction, or at least capture a large fraction of the set of all possible linkages. However, this temporal resolution may be wholly insufficient for generating approximately unbiased estimates of the relative contribution of a linkage to a community of linkages, regardless of the precise mathematical form that this contribution may take. Until excellent empirical quantitative networks can be cheaply sampled, network-theoretic pollination ecology may be stuck as a cacophony of chattering theorists, with nary a testable hypothesis in sight.

This challenge is presented at a particularly critical juncture. The planet is experiencing a catastrophic collapse in biodiversity, and yet is it estimated that 80% of insect biodiversity remains unknown to science[2][3]. Archaic practices in biodiversity monitoring will be wholly insufficient for cataloguing the world's biological complexity³, much less comprehending its dynamics to such a degree that science will be able to develop efficient, effective targeted interventions in service of biodiversity maximization. The few extant attempts to unravel the temporal dynamics of pollination networks are often marred by significant compromises, presumably to dampen the burden of repeated sampling. While limiting a study to a single site[5], waiting more than a month between samples[6], or surveying phenologically-compressed, species-poor arctic systems[7] may not compromise the integrity of any single inquiry, such concessions *do* inevitably constrain these studies' contributions towards a generalized understanding of the spatio-temporal structure of pollination.

Among the most severe manifestations of the global biodiversity information gap is in the field of bee conservation, where an emerging chorus of dire findings (albeit, contested findings⁴) suggest that either global bee diversity is in precipitous decline, or our global capacity for monitoring bee diversity and population dynamics has treacherously declined in recent decades[9]. Any decline in insect abundance or diversity carries the potential for first-order knock-on effects to pollination services, and thus pollination-dependent plant fitness. However, the nature and magnitude of such effects are largely unknown, except for that limited knowledge gleaned from small-scale pollination-exclusion experiments and empirically unverified simulation work. Ultimately, at the present moment, ecology has no means of

²I assume that at least one of the involved species (probably the pollinator) is Poisson-distributed with low λ [$\lambda < 1$]. Although species may not be Poisson-distributed over the domain of the season, this is a reasonable assumption for any given sampling day for which we know that it is biologically plausible that a given pollinator may occur. Provided these relevant conditions (phenology, environment, obligate partner presence, etc.), we assume that occurrences of rare pollinators within a sample day are approximately independent, the rate of their occurrence is approximately constant (notwithstanding weather changes throughout the day), and that two occurrences will not occur simultaneously for $t \ll 1sec$.

³Biodiversity monitoring capacity is especially limited in the Earth's tropics, which are home to a robust majority of the world's insect biodiversity[4].

⁴It is worth noting that evidence for 'global' insect abundance and diversity declines is, at best, ambiguous. The evidence is particularly weak in the United States, where *Crossley et al.* found in 2020 that there exists no evidence for such declines in the US[8]. Importantly, this survey assessed a broad variety of taxa across >5,300 medium- to long-term time series collected at US Long Term Ecological Research sites. More than half of the time series were collected at sites that are significantly disturbed by human activities. The methodology employed in these long-term surveys are unusually robust, and this finding should enjoy an elevated epistemological status.

predicting the asymptotic behavior of plant-pollinator systems[10]. Network approaches to pollination ecology are the means by which the field can gain such a capability. Spatio-temporal evolution is the next frontier of network science: future work will treat pollination networks as dynamical systems whose linkages and weights are best represented as time series[11]. Given the gravity of the current moment, this capability is needed as quickly as possible.

The quandry presented by potential global insect decline, the accompanying problem of data poverty, and the resultant urgency and present-day infeasibility of predictive mutualistic network theory make apparent the pressing need for rapid innovation in biodiversity and pollination monitoring technologies. I identify the primary constraint in contemporary monitoring techniques as being its absolute dependency on human observation.

1.1 Contemporary best practices in visitation network sampling

A plant-pollinator visitation network consists of the set of all observed interactions between plants and plausible pollinators, where an interaction is defined as physical contact between the pollinator and a reproductive structure of a blooming flower. Although the purpose and design of the studies that involve these networks vary, their fundamental requirements are constant. Visitation records minimally require plant and pollinator taxonomic identities and conventionally include observation time and place.⁵

Because fine-scale identification from momentary observation is impractical, observers attempt to non-destructively net and euthanize the pollinator for post-hoc identification. If this netting is unsuccessful, observers record an approximate 'gross identification' of the pollinator. To ensure standardized sampling times, observers are expected to pause their sampling timers while netting and recording an observation. Observers monitor a small ($3m^2$) defined area for a particular amount of time (ex. 6 minutes) before rotating to another plot. During a typical field season, this lab may deploy 8 persons for 10 weeks of fieldwork at an estimated cost of \$68,841⁶. The lab typically maintains 14 field sites, each of which is sampled for 90 minutes per week, for a total of 1,260 minutes (21 hours) of sample time per week for a total of 210 hours of sampling time at an estimated cost of \$328/hr. Because most expenses are tied to labor, fixed costs are relatively low and the cost of additional sampling time scales approximately linearly⁷. An additional 100 hours of sampling time costs approximately \$30,000⁸.

Pollination monitoring using today's methodologies is far too expensive to achieve the spatiotemporal scale and resolution required for the next generation of network-theoretical pollination ecology. Although fixed costs will persist for any methodology, their share of a project's budget approaches zero at the limit of sampling intensity, whereas marginal costs approaches 100% of project budget. Consequentially, the most effective means to reduce the cost of generating large-scale visitation networks is to minimize the direct and indirect cost of labor via automation.

The objective of this study was to radically reduce the cost of scaling plant-pollinator visitation network studies by developing a sampling strategy that minimizes the marginal cost of additional sampling time. In service of this objective, I aimed to develop a tool to accurately extract plants, pollinators, and visitations from image datasets at a minimal marginal cost. In this paper, I describe this tool and empirically assess its ability to extract a plant-pollinator visitation network from photos via a field study. Finally, I evaluate the potential costs and benefits of employing image-based sampling in place of or alongside conventional methods, and postulate novel macro-scale monitoring capabilities enabled by *Imago*.

2 Materials & Methods

2.1 Dataset acquisition

Floral diversity, density, and visitations were sampled across 9 remnant tallgrass prairie sites in Eastern Kansas during summer 2021. For each site, 3 parallel 50m transects were defined and circumscribed by a square 50m plot boundary. Care was taken to ensure that plots were established along relatively flat saddles at least 100m from nearby groves and lakes. A floral survey was completed along each transect at the beginning of pollinator sampling for each sample day. Taxonomic identities and flowering head count were recorded for all anthesis-stage flowering plants within 0.5m of

⁵I employ the standard sampling practices of the University of Washington's Brosi Lab, as employed at the Rocky Mountain Biological Laboratory, as a representative example of best practices in visitation network sampling.

⁶Housing:\$30800, Lab space:\$4685, Station fees:\$13356, Airfare: \$4000, Wages:\$16000. Estimates are based on 2022 RMBL fee schedule, and assume 6 employees on lab payroll paid \$400 weekly.

⁷This cost estimate is for a plant-pollinator network with only 'gross ID' field identifications, and does not include the cost of obtaining species-level identifications for the netted pollinator specimens.

⁸See Appendix for full simulated cost comparison.

each transect centerline. An additional systemic walk was performed across the extent of each plot to identify and record any flowering species not found along a transect. Both components were performed by a single observer.

A hybrid methodology of image-based random walks and video-based static monitoring was employed to jointly maximize the sampling intensity and methodological standardization possible for a two-person sampling team. A strictly standardized plot-based sampling schema as described above was determined to be impractical for a small field team. Tallgrass prairie floral assemblages are relatively sparse, so I surmised that a 1-2 person sampling team would be able to recover significantly more of the plant-pollinator links present for any sample day through selective allocation of sampling effort on the basis of floral density⁹. 1-3 observers performed a random walk around the 50m square plot for a cumulative total of at least 3hrs per sample. When a pollinator was seen visiting an active inflorescence, observers attempted to capture a photo¹⁰ of the interaction and then net the pollinator. In order to evaluate the potential of Imago as a tool for enhancing studies which require a higher degree of standardization, as well as its potential for passive sampling, I implement video-based sampling. For each sample, at least 12 mins of video was recorded for each flower species. A video camera¹¹ was positioned on a tripod at a minimum of 2ft from each floral specimen, and was aimed such that the target inflorescence was centered in frame. All photo and video data were offloaded from cameras into unique site-observer folders following each sample.

Before beginning each sample's visitation monitoring, a weather station¹² was erected adjacent to plot boundaries. All camera internal clocks were synchronized with that of the weather station. Temperature, relative humidity, photosynthetically active insolation, and wind speed/direction were logged at 1 minute intervals for the entire duration of visitation sampling. Each of 9 sites was visited approximately 3 times by across the length of the summer season. By season's end, 178 hours of video and 17,771 images collected across 106 hours of random walks combined for a total of 284 sample hours. Following the conclusion of the field season, all data were assessed for plant/insectoid presence and insectoid-inflorescence visitations using a bespoke image processing tool I term *Imago*.

2.2 Processing pipeline

2.2.1 Design considerations

Imago is a pipeline¹³ of sequential image-processing algorithms, some of which are derived or co-opted from existing open-source software. The tool was designed to be a flexible, general tool for zero-shot automatic visitation extraction. Contemporary high-performance object detection algorithms are black boxes which present few opportunities for understanding or addressing their own limitations short of retraining. Today's popular open-source object detection models reliably locate easily-resolved objects belonging to a modest set of relatively homogeneous classes. While computationally cheap detection of common objects is an appropriate objective for most current applications, the resultant compromises implicitly limit the utility of even finetuned object detection models to problems of small-object detection. Even the coarse detection of insects and plants within the same photo is challenging. Automated approaches must contend with extreme scale differences between plants and their pollinators, the tendency for pollinators to be partially or completely obscured, and the unusual incidence of blurriness due to pollinator movement and depth-of-field effects. The nature of these challenges is dependent upon camera type, photographer skill, and image capture methodology. As such, a visitation extraction algorithm should be highly configurable: the assumptions and quantitative heuristics employed by the pipeline should be maximally specifiable by the specialist user in order to best fit their application's needs. Most importantly, this user should be able to arbitrarily tune the accuracy vs. compute-cost trade-off: the ability of the tool to extract visitations from a complex scene should be bounded by computational resources rather than an algorithm's architectural compromises. One-size-fits-all image detection algorithms fulfill few to none of these design requirements.

Today's two-stage and multi-stage object detection algorithms propose and refine image regions for classification with remarkable efficiency. However, the architectural marriage of lossy region proposition and localization with region classification has yielded monolithic tools which cannot easily take advantage of additional compute to improve detection completeness and which require expensive localized (i.e. bounding box) training data. This precludes the application of object detection techniques to domains where 1) large labeled detection datasets are unavailable 2) the

⁹The likelihood of ultra-low floral density in any given patch is possibly greater for systems with long-summers, where floral phenologies overlap relatively less than in ex. high-alpine systems. Given that per-flower visitation activity is strongly positively driven by floral density for lower densities[12], permissive sampling methodologies can be especially valuable for revealing more topological links per sampling hour in long-summer systems such as the tallgrass prairie.

¹⁰All photos were captured with either a Canon 250D with EF-S 55-250m zoom lens or a Sony A6000 with E 55-210m zoom lens.

¹¹All videos were captured with a Canon 250D with EF-S 55-250m zoom lens recording in 4k resolution at 30fps.

¹²Spectrum WatchDog 2900ET

¹³See Figure 1.

graphical properties of target images diverge from those for which existing architectures are optimized. Although general taxonomic classifiers can be trained from publicly-available datasets, general specimen-detectors cannot, and today's detectors systemically under-detect small objects relative to large ones, rendering these techniques an awkward fit for plant-pollinator detection. *Imago* reformulates the multi-stage detector by de-optimizing region proposal and performing region refinement *after* classification. This 'dumb detector' architecture substantially reduces the costs of applying automation to complex computer vision problems in niche domains which lack the economies of scale to support the development of highly optimized, bespoke detection tools. I demonstrate that a robust classifier, when applied to many indiscriminately-selected image regions, can function as a 'fat detector' *without* expensive localized datasets. This approach produces many overlapping detections, which can then be refined to reduce redundancy using the some of the same region-refinement techniques employed by multi-state object detectors. By 'punting' region proposal efficiency, *Imago* is able to couple detection accuracy with additional compute resources, at the expense of computational complexity and baseline localization accuracy. At a high level, the tool slices input image data into sub-images, assesses each for plant or insectoid presence, and then assesses the relative proximity of detected specimens to determine the plant-pollinator visitations within each image. *Imago* can be conceptualized as consisting of five linked stages: pre-processing, slicing, specimen detection, post-processing, and visitation detection.

Parameter values which were held constant between both trial configurations are provided below (those which vary are provided in Results).

2.2.2 Pre-processing

The preprocessing stage prepares raw photos and videos for further processing by converting videos into image sequences and associating all images with appropriate metadata. Videos are converted into images by frame sampling. The tool supports sampling at any frequency i frames per second, although constraints on high-speed storage space and processing power may influence parameter choice (I chose to sample $i = 2$ fps). If weather data is provided, *Imago* creates a database associating each image and sampled frame with spatiotemporal metadata and weather conditions from the nearest minute.

2.2.3 Slicing

Images are then sliced at h scales of values S_{1-h} , where each square slice of scale s_i overlaps neighboring slices of the same scale by a ratio ϕ . Slicing is performed using a modified version of the Python package *SAHI*[13] which downscales and passes slices directly to the remainder of the pipeline.

2.2.4 Specimen detection

Each slice is passed to a general specimen identification model¹⁴, which returns a taxonomic label and confidence level for all identifications above specified minimum confidence threshold c_g . If the detected specimen is within the *Bombus sp.* genus, the slice is then passed to a bumblebee-specific model¹⁵ for a second identification. If this second identification differs from that produced by the general model and is produced with a confidence score greater than threshold c_b , the slice's taxonomic identification is overwritten. Both models are quantized according to the TensorflowLite specification, and thus run on CPU only (along with the rest of the pipeline). If *Imago* is being used outside of North America, where one is likely to encounter bumblebee species which are unidentifiable by the model, then the pipeline can be configured to only use the general identification model. I specify confidence thresholds of $c_g = 0.7$ and $c_b = 0.8$.

2.2.5 Post-processing

For any single image, increasing the number of unique image slices queried by identification models via multiscale slicing monotonically increases the proportion of ground-truth specimens which are detected by the pipeline. However, some specimens may be identifiable at multiple scales, causing the systemic overcounting of specimens. The ratio of detected specimens to ground-truth specimens will increase as slicing is performed at more scales or with greater overlaps. Thus, post-inference duplicate detection filtering is needed to minimize the recall vs. precision trade-off.

Duplicate detections are filtered per-species via a custom implementation of non-maximum suppression[cite: gil werman], a common method among object detectors for selecting one box from many overlapping boxes. I assume that the highest-confidence detection for a species is functionally analogous to the highest-confidence bounding box

¹⁴This general identification model was produced by iNaturalist and was originally utilized in the popular identification app *Seek*[14].

¹⁵This model is a quantized version of the BeeMachine identification model[15].

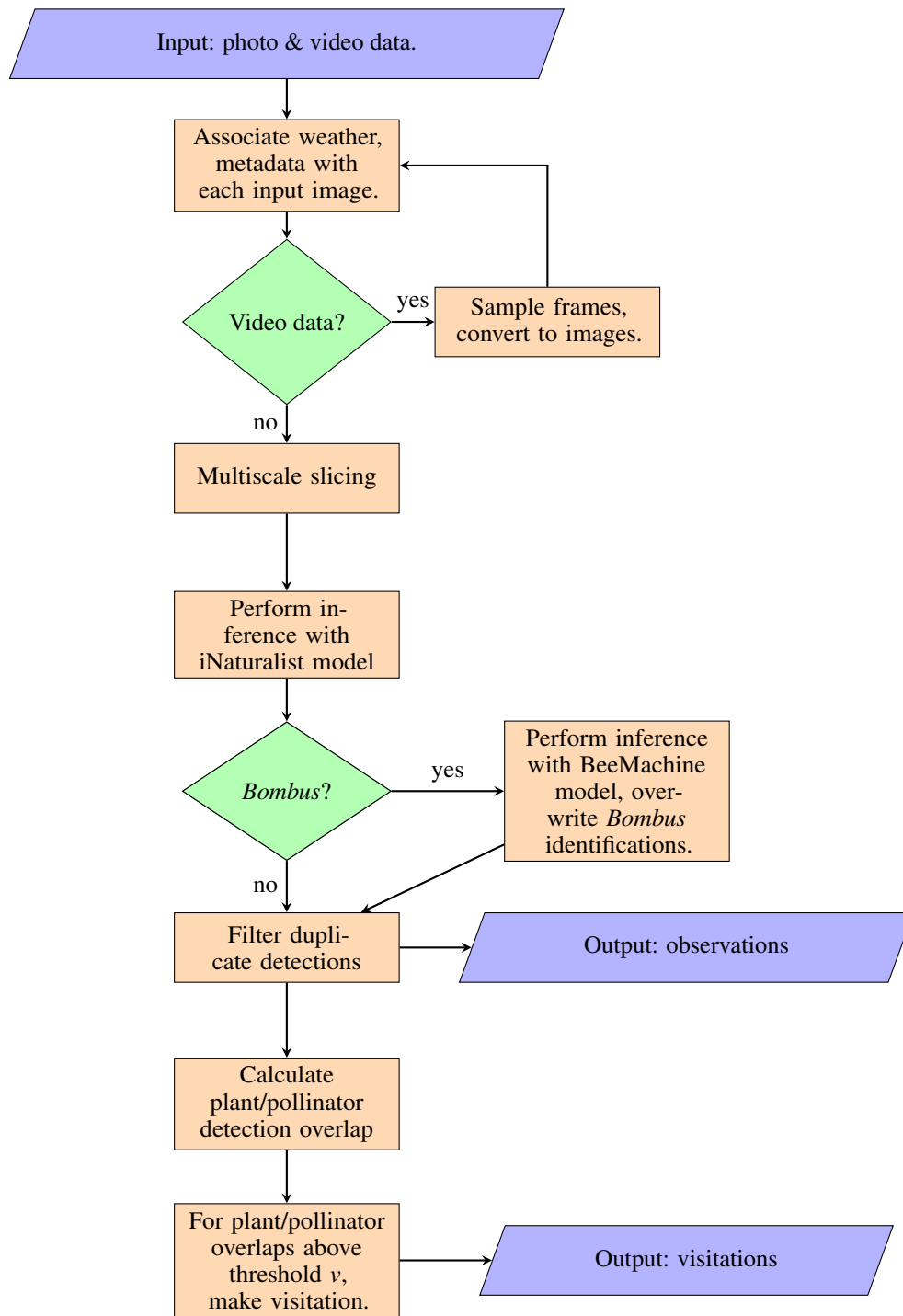


Figure 1: *Imago* extracts detects plants and pollinators by slicing input images, classifying slices, and post-processing positive slices to remove duplicate detections caused by multi-scale inference. This creates a dataset containing all the plants and pollinators detected in the input image dataset. This observation dataset can then be mined to find plant-pollinator visitations in the input dataset. For any image, plant-pollinator visitations are extracted by assessing the overlap of all slices containing a plant or pollinator. If any plant and pollinator overlap to a degree greater than overlap threshold v , then a plant-pollinator visitation event is recorded in the outputted visitation dataset. Several parameters can be tuned within the pipeline to optimize its behavior according to particular datasets or analytic goals.

proposal in that it is the window which best frames the object of interest¹⁶. For every species detected within an image, the IoU (intersection over union) is calculated between every detection slice d and the set of all detection slices a . All isolated detections d for which $IoU_{d,a} = 0$ are added to a 'keep' list L_s . All other d are assigned to non-overlapping stacks s , such that between-stack overlap $IoU_{s1,s2,s\dots} = 0$, while within-stack overlap $IoU_{s1_1,s1_2,s1\dots} < 0$. For every s , the highest-confidence detection d_{max} is added to L_s . Then, the IoU between d_{max} and all other d for the same s is calculated. All d that overlap with d_{max} by more than the minimum threshold m are added to L_s . For both trials, $m = 0.01$.

2.2.6 Visitation detection

Insectoid visitations to plants are determined by comparing the relative positions of each clade's detections within each photo. The smallest pixel distance between the bounding boxes containing each unique plant-pollinator pair is calculated. If this distance is smaller than the minimum visitation overlap distance v , then a visitation is recorded. v is specified assuming a 6000 x 4000 input photo resolution, although this value and the calculated smallest pixel distance are both linearly scaled according to each image's input resolution. The sensitivity for which an insectoid is determined to be visiting an adjacent or overlapping plant can be modulated by adjusting v . Because this technique cannot distinguish between insectoid visitations to plants' reproductive vs. non-reproductive components, these visitation interactions are distinguished from pollination interactions, which are a subset of the former. Visitations more closely approximate pollination interactions if image data are collected in such a way that is triggered by manual observation of a likely pollination interaction.

2.2.7 Experimental regimes

Specimen detections and visitations were manually validated for two configurations of *Imago* across two randomly-selected 180-image subsets. Two experimental regimes were specified: configuration *a* simulates a 'lean' pipeline application, for which extremely poor quality images are manually excluded and processing power is at a premium. These conditions reflect those which would be encountered in ex. a smartphone application, wherein users select particular photos for processing. In contrast, *b* is a 'fat' configuration which simulates a scientific application of *Imago* for which detection completeness is weighed much more heavily than computational efficiency, and for which human inputs are minimized. In this scenario, images are *not* pre-filtered (this would be impractical for large, messy field image datasets), but a much larger proportion of region proposal space is sampled.

Specifically, for configuration *a*, input images are sliced at 3 scales, 28 extremely low-quality images are manually excluded, and a minimum distance visitation criterion $v = 500$ px was specified. However, blurry images where individual insects or flowers can be discerned were preserved, even if these objects are unlikely to be precisely identifiable to even an ideal taxonomist. Configuration *b* is more computationally expensive, slicing at $h = 14$ scales. Additionally, a more lenient minimum distance visitation criterion $v = 1000$ px was enforced. Metrics for experiment *b* are calculated across the entire 180-image dataset: no manual pruning of poor-quality images is performed.

2.2.8 Validation metrics

Recall measures the ability of *Imago* to detect and identify to species every target present for an image, and is generically calculated by $(n_{predicted}/n_{actual})$. For pollinators, recall is measured as $n_{polls_{detected}}/n_{polls_{actual}}$. However, plant recall is defined as $n_{plants_{species_{detected}}}/n_{plants_{species_{actual}}}$, to reflect *Imago*'s inability to discern plant individuals from tightly-framed inflorescence visitation images (a consequence of the extreme scale difference between flowering plants and their insectoid visitors). Qualitative precision measures the likelihood that an image for which a positive prediction is produced actually contains a positive sample. For pollinator detections, quantitative precision measures the likelihood that the correct number of pollinator individuals is detected, given that at least one is detected for that image. In the case of plant detections, quantitative precision measures that the correct number of *species* was detected, given that at least one plant was detected. Finally, quantitative visitation precision measures the likelihood that the correct number of visitations were extracted from an image, given that *Imago* detected at least one plant-pollinator interaction for that image. The average time to process each photo (standardized at 6000px x 4000px) was recorded. All processing was performed with a Intel 10700F processor. Because the pipeline is very I/O intensive, all images were stored on a low-latency solid-state drive (SSD).

¹⁶I believe this assumption to be sound, except for the purpose of reducing qualitative false positives (erroneous detection of species not present within a photo). The rate of qualitative false positives necessarily increases when region proposals are generated for a greater number of scales h with higher overlap ϕ . This issue could be addressed by increasing the minimum classification thresholds c_g and c_b for larger samples of region proposal space. This proposal is motivated by the observation that the likelihood that the classifier evaluated a slice containing a well-framed view of each specimen within a photo increases monotonically with the number of unique slices. However, I do not examine the form of $c_{g,b}(h, \phi)$ here.

3 Results

<i>Regime</i>	<i>Task</i>	<i>Counts</i>		<i>Recall</i>	<i>Precision</i>		
		<i>n images</i>	<i>predicted/actual</i>		<i>Qual.</i>	<i>Quant.</i>	
config <i>a</i> <i>h</i> = 3 selective <i>v</i> = 500px <i>t</i> = 3.3s/image	Visitations (all)	152	34/67	0.51	0.97	0.55	
	Detections (all)	<i>Polls</i>	152	45/84	0.54	1	0.43
		<i>Plants</i>		110/162	0.68	0.99	0.87
	Detections (well-resolved)	<i>Polls</i>	119	39/60	0.65	0.98	0.45
<i>Plants</i>			72/106	0.68	0.99	0.84	
Detections (blurry)	<i>Polls</i>	33	2/7	0.29	1	0.21	
	<i>Plants</i>		20/30	0.67	0.97	0.91	
config <i>b</i> <i>h</i> = 14 non-selective <i>v</i> = 1000px <i>t</i> = 70.1s/image	Visitations (all)	180	188/119	0.71	0.99	0.54	
	Detections (all)	<i>Polls</i>	180	122/125	0.82	1	0.57
		<i>Plants</i>		196/176	0.87	0.99	0.85
	Detections (well-resolved)	<i>Polls</i>	149	105/109	0.82	1	0.6
<i>Plants</i>			162/145	0.92	0.99	0.83	
Detections (blurry/very blurry)	<i>Polls</i>	31	16/14	0.92	1	0.4	
	<i>Plants</i>		26/25	0.88	0.96	0.92	

Table 1: Comparison of detector performance for ‘lean’ (*a*) and ‘fat’ (*b*) configurations. Bolded text indicates the better value for each metric. Ties are additionally italicized.

Imago capably extracts plants, insectoids, and their visitations from image data. Task performance is strongly dependent upon pipeline configuration. The ‘fat’ configuration *b* exhibits much stronger recall for both visitations (71% vs. 51%) and detections (*all pollinators*: 82% vs. 54%; *all plants*: 87% vs. 68%). Pollinator recall especially benefits from additional slicing scales, and a majority of pollinator detections occurred a smaller scale ($\leq 1400px$, see Figure 2). This advantage persists despite configuration *b* facing a relatively more difficult task in processing images which were not pre-filtered on the basis of quality.

Increased recall does not come at the expense of precision. A very high degree of qualitative precision ($> 96\%$) persists across all categories for both configurations: nearly all images for which a positive sample was predicted indeed contained a positive sample. Additionally, both configurations generally predict the correct number of plant species within an image ($> 83\%$).

However, both configurations struggle to correctly predict the correct number of insectoid individuals present within an image. The slim configuration *a* exhibits better quantitative pollinator precision than is the fat configuration *b*, but both

Visitation recall performance depends upon the leniency of the minimum distance visitation criterion *v* in addition to plant/pollinator recall. I find that visitation recall for configuration *b* almost exactly matches what would be expected if this metric was purely a function of pollinator and plant recall ($0.82 * 0.87 = 0.7134 \approx 0.71$), whereas configuration *a* exhibits *better* visitation recall than would be predicted by this proposition ($0.54 * 0.68 = 0.3672 < 0.51$). This result is likely an artefact of poor plant localization for configuration *b* due to NMS being an inappropriate region refinement algorithm for high-*h* (see *Discussion*).

3.1 Inferred network validation

4 Discussion

Imago produces lossy, conservative approximations of ground truth for measures of both plant and pollinator presence and interactions. Imago is heavily biased Type-II errors of omission: while an image for which Imago reports a detection or visitation almost always contains one, the tool frequently fails to record interactions which would have been discernible by a human observer. This characteristic was an expected outcome of Imago’s architecture, which was designed in a way that maximizes the predictability of the trade-offs made at each stage in the pipeline. minimizes false-positives by using a highly generalized classifier, which is inherently more robust against false positives caused by erroneous classification of out-of-sample objects. Visitations benefit from another layer of architectural robustness: a false positive requires the overlapping detection of both a plant and a pollinator, meaning that the likelihood of an

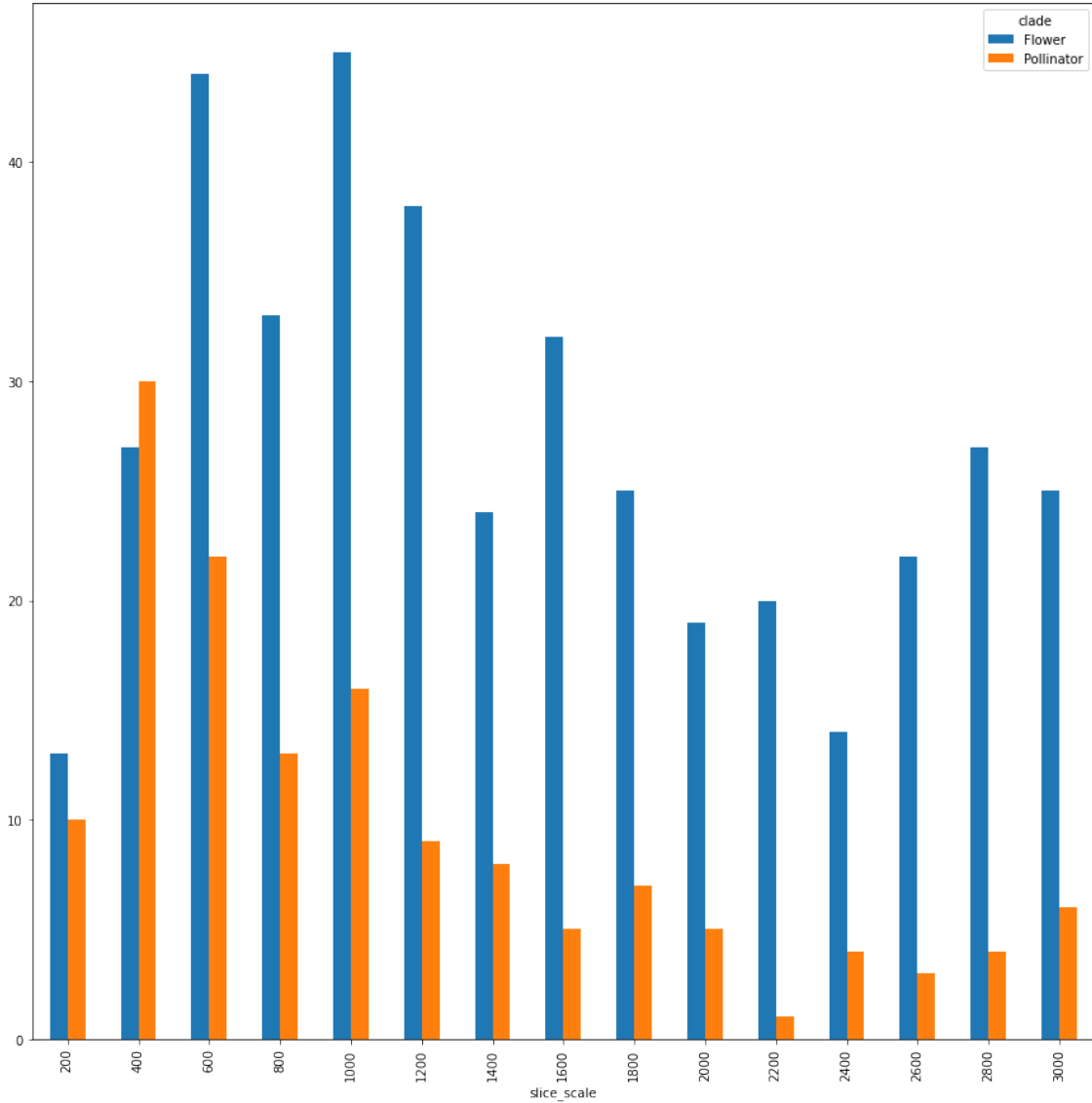


Figure 2: The scales at which plants and pollinators are resolved are bimodally distributed. Plants are most commonly identifiable around a peak of 1000px and a smaller peak around 2800px. A much greater proportion of pollinator detections are at smaller scales. Most detections are clustered around a peak of 400px, while a smaller number are resolved at a scale of 3000px. Expanding the domain of slicing scales would likely yield additional detections of both plants and pollinators. These metrics were obtained using the 180-image subset used for the $h = 14$ slicing regime described in Table 1.

erroneous positive is a compounded function of the (already-low) likelihood of at least one erroneous false-positive plant or pollinator detection in the same region as another false or true positive detection.

Imago provides multiple 'knobs' with which one can fine-tune the tool's detection and filtering processes to produce a more idealized outcome. The number of slicing scales and the amount of slice overlap can both be modulated to increase the likelihood that the classifying model gets a 'good look' at each plant or pollinator in the dataset. Such decisions will be left to the discretion of the user and their computational resources, as increases to these parameters will increase the time complexity of the algorithm by approximately $O(n)$ and $O(n^2)$, respectively.

Some researchers may care about the number of pollinators present in each photo. Quantitative precision could be improved by

The ability of the tool to detect pollinators is strongly dependent upon the quality of the input image. The tool particularly struggles to recover pollinators from blurry images. Although I hypothesize that the error distribution is similar for Imago-aided and manual network sampling, comparing the quantitative rates of errors across methodologies is challenging. Importantly, it is unclear how the recall performance of *Imago* compares to mainstream techniques for sampling visitation network, as these techniques produce unquantifiable uncertainties. Manual observational sampling can only quantify the missed interactions that were seen and leave no persistent record which can be reviewed *post-hoc* to quantify sampling uncertainty or to backfill missing observations. The same basic factors constrain the performance of both an unaided human observer and one sampling with the aid of Imago. A pollinator is more likely to be missed if it is, for example, small, or inconspicuously colored, or it moves in an unpredictable way. It is thus likely that samples produced by Imago are biased in a similar way to those produced by manual sampling. However, image-based sampling enables the researcher to measure and retroactively improve the quality of their samples. Autonomous imaging techniques, in particular, could potentially provide an excellent baseline for ascertaining sampling bias for small or elusive pollinators. Video and timelapse data provide a persistent ground-truth for pollinator activity which is completely independent of human observers: no human must notice a pollinator for it to be included in this data. One could study this record to estimate the rate at which humans miss the presence of inconspicuous pollinators and perhaps apply corrections to records generated by photo or manual observational methods.

The tool consistently missed some flowers (ex. *Erigeron speciosus*) whose species cannot be reliably determined from their inflorescence alone. However, this issue is partially mitigated for studies which pair visitation samples with floral surveys, as a complex of morphologically similar taxa will often only be represented by a single species for a single sample unit. In these circumstances, the tool still generally returns a cluster of medium-confidence detections of similar flowers (ex., a cluster of 4 *Erigeron* species, $p < 0.6$), but none of these detections will surpass the threshold for inclusion. A lower threshold for inclusion could be specified, and species-level identifications could then be backfilled. A method could be devised to automatically detect such clusters of multiple taxonomically-similar flower identifications. The user could either be prompted to clarify these identifications, or the program could itself reference the floral species survey to determine the correct species for an ambiguous detection. This capability would improve Imago's recall performance for both plant detections and visitations.

The relatively enhanced visitation recall of the slim configuration *a* is likely a product of larger post-refinement plant bboxes. Almost all images w/ pollinator also contain a proximate plant, since that is the condition for taking photo. Thus, if appropriate bboxes are conserved for both plant and a visiting pollinator, then the likelihood of a visitation should be nearly independent of v . Indeed, a stricter visitation condition v can only result in *poorer* visitation recall, implying that v cannot explain the better-than expected performance of *a*. Thus, the relatively poorer $recall_{visitation}(recall_{detection_{pollinator}}, recall_{detection_{plant}})$ performance of *b* must hail from poorer localization of either plants or their visitors. Because moving from scenario *a* to *b* improves pollinator recall ($0.54 \rightarrow 0.82$) 38% more than it does plant recall ($0.68 \rightarrow 0.87$), and because pollinators are much more likely to be detected at small scales than are plants (see Figure 2), Therefore, I conclude that post-refinement plant detections are probably more poorly localized for configuration *b*. This conclusion is supported by mechanistic analysis. Given that plants are much larger than pollinators, and many more plant detections are at large scales, it should be much more likely that all plant detections in a photo overlap than pollinators. The region refinement algorithm NMS only guarantees that 1) all non-overlapping detections and 2) the highest-confidence detection within each stack are preserved. Thus, pollinator detections are less likely to be mistakenly discarded on the basis of region refinement. It is unclear that plant duplicate detections should be filtered, given that the purpose of plant detection is region localization rather than individualization. If plant filtering is desirable, then NMS is probably a poor choice. Plant region refinement should instead prioritize the construction of tight region proposals, likely via biased conservation of dense, small detection slices for stacks with at least one very-high confidence detection.

One blind spot: species identification accuracy. - iNaturalist model: is there literature on its accuracy? - The IDs are much stronger than what is possible for a human observer, who can only identify to a gross ID in the field. - IDs

produced by the model can always be reviewed post-hoc, unlike field IDs. - They could capture every pollinator, but at obvious additional expense and latency.

- Estimated cost. - Approximate time investment (labeling training inference cleaning of results). - Compare field-deployable options for pollination monitoring. - Manual monitoring - Active image-based sampling (what I did)
- Benefits: - Generally higher-quality image and video data than that what would be produced by place-and-forget cameras left in the field. - For photo-sampling: if observer is unable to net specimen, then they will at least have picture. This resolves a sampling inconsistency wherein observers with taxonomic expertise can infer approximate identifications for missed pollinators, while less experienced observers may fail to identify or provide erroneous classification of the same specimens. - Video-based monitoring allows for the quantification of Type II error rate, as video data can be analyzed post-hoc to measure visitations missed by *Imago*. While motion-triggered camera traps tremendously reduce storage requirements per sample hour, sampling error cannot be quantified unless paired with manual or trigger-free sampling. *Imago* allows for a flexible data retention policy. - Challenges: - Monitoring via citizen science (iNat) - Could compare my sample (from field) to that uploaded to iNat in the region over the same time period. - Passive image-based monitoring. - Benefits: - Could capture several magnitudes of order more interaction data. - Fill in sample gaps (weather, changing conditions over season, very short-lived species) - Challenges: - Technical

5 Future Work

- **Future goal: develop device for unattended active plant-pollinator monitoring.** - Revisit this goal: may not be necessary if 'dumb' passive image-based sampling works well enough. - Model-assisted labeling. - Potential to substantially speedup the biggest burden in this methodology.

6 Conclusion

- The full pipeline is published as a docker image. - How to adapt the pipeline for other studies / ecosystems? - Brief details, indicate that more extensive documentation will be available in a GitHub repo. - *Imago* should be particularly useful for those applications which are relatively insensitive to Type II errors and minor taxonomical errors, but which would benefit greatly from increased temporal and spatial scale.

Acknowledgments

References

- [1] Fernanda S. Valdovinos. Mutualistic networks: moving closer to a predictive theory. 22(9):1517–1534. [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ele.13279](https://onlinelibrary.wiley.com/doi/pdf/10.1111/ele.13279).
- [2] Nigel E. Stork. How many species of insects and other terrestrial arthropods are there on earth? 63:31–45.
- [3] Showket A. Dar, Mohmmad Javed Ansari, Yahya Al Naggar, Shafia Hassan, Syed Nighat, Syed Burjes Zehra, Rizwan Rashid, Mudasar Hassan, and Barkat Hussain. *Causes and Reasons of Insect Decline and the Way Forward*. IntechOpen. Publication Title: Global Decline of Insects.
- [4] Carlos García-Robledo, Erin K. Kuprewicz, Christina S. Baer, Elizabeth Clifton, Georgia G. Hernández, and David L. Wagner. The erwin equation of biodiversity: From little steps to quantum leaps in the discovery of tropical insect diversity. 52(4):590–597. [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/btp.12811](https://onlinelibrary.wiley.com/doi/pdf/10.1111/btp.12811).
- [5] Niels Kaiser. Functional integrity of plant-pollinator communities in restored habitats in mauritius.
- [6] Alicia M. Basilio, Diego Medan, Juan P. Torretta, and Norberto J. Bartoloni. A year-long plant-pollinator network. 31(8):975–983. [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1442-9993.2006.01666.x](https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1442-9993.2006.01666.x).
- [7] Jens M. Olesen, Jordi Bascompte, Heidi Elberling, and Pedro Jordano. Temporal dynamics in a pollination network. 89(6):1573–1582. [_eprint: https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1890/07-0451.1](https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1890/07-0451.1).
- [8] Michael S. Crossley, Amanda R. Meier, Emily M. Baldwin, Lauren L. Berry, Leah C. Crenshaw, Glen L. Hartman, Doris Lagos-Kutz, David H. Nichols, Krishna Patel, Sofia Varriano, William E. Snyder, and Matthew D. Moran. No net insect abundance and diversity declines across US long term ecological research sites. 4(10):1368–1376.
- [9] Eduardo E. Zattara and Marcelo A. Aizen. Worldwide occurrence records suggest a global decline in bee species richness. 4(1):114–123. Publisher: Elsevier.

- [10] Eva Delmas, Mathilde Besson, Marie-Hélène Brice, Laura A. Burkle, Giulio V. Dalla Riva, Marie-Josée Fortin, Dominique Gravel, Paulo R Guimarães, David Hembry, Erica Newman, Jens M. Olesen, Mathias M. Pires, Justin D. Yeakel, and Timothée Poisot. Analyzing ecological networks of species interactions.
- [11] Jordi Bascompte and Pedro Jordano. *Mutualistic Networks*. Princeton University Press. Google-Books-ID: 52SYDwAAQBAJ.
- [12] Carla J. Essenberg. Explaining variation in the effect of floral density on pollinator visitation. 180(2):153–166. Publisher: The University of Chicago Press.
- [13] Fatih Cagatay Akyon, Sinan Onur Altinuc, and Alptekin Temizel. Slicing aided hyper inference and fine-tuning for small object detection. Publication Title: arXiv original-date: 2021-01-30T12:54:53Z.
- [14] Seek app version 2.x. original-date: 2018-09-27T22:48:59Z.
- [15] Brian J. Spiesman, Claudio Gratton, Richard G. Hatfield, William H. Hsu, Sarina Jepsen, Brian McCornack, Krushi Patel, and Guanghui Wang. Assessing the potential for deep learning and computer vision to identify bumble bee species from images. 11(1):7580.