A machine learning pipeline for automated insect monitoring

Aditya Jain[*†] Mila - Quebec AI Institute

Fagner Cunha[*] Federal University of Amazonas

Michael Bunsen[*] Mila - Quebec AI Institute

Léonard Pasi[‡] EPFL

Anna Viklund[‡] Daresay

Maxim Larrivée Montreal Insectarium

David Rolnick McGill University Mila – Quebec AI Institute

Abstract

Climate change and other anthropogenic factors have led to a catastrophic decline in insects, endangering both biodiversity and the ecosystem services on which human society depends. Data on insect abundance, however, remains woefully inadequate. Camera traps, conventionally used for monitoring terrestrial vertebrates, are now being modified for insects, especially moths. We describe a complete, open-source machine learning-based software pipeline for automated monitoring of moths via camera traps, including object detection, moth/non-moth classification, fine-grained identification of moth species, and tracking individuals. We believe that our tools, which are already in use across three continents, represent the future of massively scalable data collection in entomology.

1 Introduction

The Earth is undergoing a sixth mass extinction event, where an *eighth of all species* may become extinct by 2100 [1–3]. Insects account for about half of all living species on earth and 40% of the animal biomass [4], but both the diversity and abundance of insects are undergoing a precipitous decline [5] as a result of several factors, in which climate change figures prominently. The "insect apocalypse" significantly increases the risk of breakdown of ecosystem functions on which human society depends [6]. Monitoring insects is therefore a crucial component of climate change adaptation.

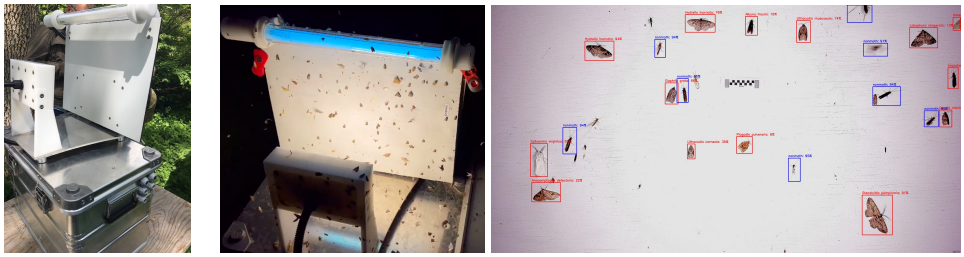
Traditional insect collection and identification by entomologists is hard to scale, due to the massive number of insect species and a lack of experts, with certain geographies and taxonomic groups especially poorly covered. The emergence of high-resolution cameras, low-cost sensors, and processing methods based on machine learning (ML) has the potential to fundamentally change insect monitoring methods [7]. Camera traps powered by computer vision models for terrestrial vertebrates monitoring are now commonplace [8], and specialized camera trap hardware for insect monitoring has begun to gain momentum [9–13]. A common group of focus for such studies has been moths[14–16], which serve vital ecological roles and represent a fifth of all insect species. Importantly, most moths can readily be attracted with UV light and are frequently visually distinguishable up to species or genus, making them ideal targets for camera traps. As hardware for moth-monitoring has grown more common, however, there is a need for scalable data processing techniques to match the influx of data. Prior methods have been greatly limited in the species and geography covered, as well as requiring extensive manual labelling for training the algorithms.

[*]Equal contribution.

[†]Corresponding author (aditya.jain@mila.quebec).

[‡]Work done while at Mila - Quebec AI Institute.

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2023.



(a) A moth camera trap (b) Trap in operation (c) Machine learning predictions on a raw image

Figure 1: Figure 1a and Figure 1b depict a moth camera trap used by our partners. Figure 1c shows our insect localization and species prediction on a raw image, with red boxes showing insects classified as moths (with fine-grained species predictions), and blue boxes showing non-moths.

This work describes a complete software and ML framework that transforms raw photos from insect camera traps into species-level moth data (see Fig. 1). Our system is motivated by the following goals: 1) Model predictions should be highly accurate across moth species, 2) the presence of non-moths must be accounted for, 3) the algorithms should work across different hardware setups (camera, lighting, etc.) and geographic regions, 4) extensive manual labelling of training images should not be required, 5) the pipeline should be fast to run, 6) the system should be user-friendly for non-ML experts, and 7) the framework should easily be extended to other insect groups. Our tools are currently supporting multiple deployments in Canada, the USA, the UK, Denmark, and Panama.

2 Machine learning pipeline

Our machine learning pipeline in a multi-stage process (Figure 2): 1) An object detector localizes all insects in the image (§2.2), 2) a binary image classifier distinguishes moths from other arthropods (§2.3), 3) the classified moths pass through a fine-grained moth species classifier to predict the species (§2.4), and 4) a tracking algorithm tracks the moths across frames to count the number of individuals (§2.5). Such a modular structure allows parallel development, improvement, and evaluation of each module. We now discuss each module in detail.

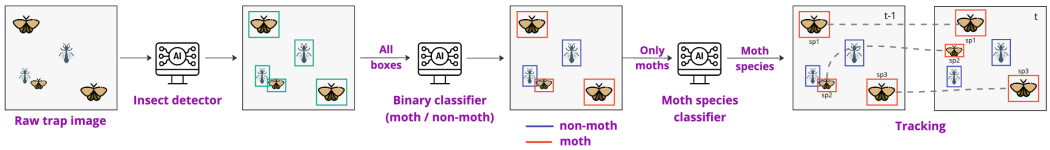


Figure 2: Machine learning workflow.

2.1 Training data for image classification

We need labelled data to train the binary and fine-grained species classifier. However, labelling data from the trap images directly is not scalable. The trap data likely would not include all rare species, labelling would be redone periodically with every change in hardware and new geographic location, and experts for species identification are scarce. We perform zero-shot transfer learning – training on a visually different dataset from the moth trap images. Specifically, we use annotated data indexed by the Global Biodiversity Information Facility (GBIF) [17], in particular from the citizen science platform iNaturalist [18]. Unlabeled moth trap data is provided by a network of partner ecologists.

Our pipeline for obtaining data from GBIF is as follows. Given a regional list of moth species, we compare it against the GBIF taxonomy backbone [19]. We remove duplicate entries, merge synonym names into accepted names, and investigate doubtful, fuzzy, and unmatched names to make a processed checklist. We use the unique taxon key for each species to fetch images and metadata (such as location and publishing institution) using the Darwin Core Archive (DwC-A) [20] file provided by GBIF. Appendix A.1 discusses additional steps we apply to clean the data.

2.2 Insect detection

Object detection, drawing bounding boxes around objects of interest and classifying them into different categories, is now a mature problem in computer vision [21]. Our setting involves a uniform pale background, fixed camera angle and distance from the screen, which makes it possible to obtain decent performance with classical image processing techniques, such as background subtraction and blob detection [14]. However, the accuracy of these algorithms remains limited due to additional challenges, such as the extreme diversity of insects attracted to the detector, cases of overlap between individual insects, inconsistent lighting, and the gradual buildup of dirt on the screen. Even foundation models for object detection, such as the Segment Anything Model (SAM) [22], have limitations on this relatively simple data, include failure to detect the smaller moths, cropping out legs and antennae, and low inference speed. Hence, we trained a custom model for this task.

The challenge is the lack of annotated data for the trap images to train an object detector. We use a weak annotation system leveraging SAM to generate a synthetic training dataset to minimize manual annotation. First, we use SAM to segment nearly 4k insect crops from 300 trap images from five trap deployments. Second, we review the crops to remove undesirable images, which gives us 2600 clean crops. Each crop review only takes a split second, much faster than drawing a bounding box. Third, the crops are randomly pasted on empty background images with simple augmentations (flips and rotations) to create a large simulated labelled dataset of 5k images. We train two versions of Faster R-CNN models [23] on this synthetic dataset: a slow model (ResNet-50-FPN backbone) and a fast model (MobileNetV3-Large-FPN backbone). The latter is 6 times faster than the former on a CPU while having a near-similar accuracy. Due to lack of ground truth evaluation, we visually analyze the performance of these models (App. 2.2).

2.3 Moth / non-moth classification

Since moths are not the only arthropods that appear on the camera trap screen, we train an image classifier to differentiate between moths and other insects. Using expert knowledge of the taxa likely to appear on the screen, we fetch 350,000 images from GBIF for each moth and non-moth category. The moth group consists of adult-stage images of moth species from multiple regions worldwide. The non-moth group comprises Trichoptera, Formicidae, Ichneumonidae, Diptera, Orthoptera, Hemiptera, Pholcidae, Araneae, Opiliones, Coleoptera, and Odonata. For this binary classification task, we train a ResNet-50 model [24], pre-trained on ImageNet-1K [25]. The model achieves an accuracy of 96.24% on a held-out set from GBIF data and 95.10% on 1000 expert-annotated insect crops from camera trap images.

2.4 Fine-grained moth species classification

The most challenging algorithm in our pipeline is for species-level identification of moths, representing a fine-grained classification task in computer vision. As there are almost 160,000 known moth species in the world [5], many with very limited data, we train separate models on regional species lists. However, the task is still quite challenging, as a regional list typically contains several thousand species, and closely related species may have only subtle visual differences. Another issue is that the number of training examples per species follows a long-tail distribution, i.e., some species have many images, while most have only a few (see App. A.3). This can bias the model towards the majority species, even if the rare species may be of particular ecological interest.

We train separate models on regional species lists using the standard ResNet-50 architecture [24]. Models are trained using AdamW optimizer [26], cosine decay learning rate schedule with linear warm up [27], and label smoothing [28]. To minimize the performance degradation caused by the distribution shift between GBIF training data and the data from the moth traps, we apply a set of strong data augmentation operations: random crop, random horizontal flip, RandAugment [29], and a mixed-resolution augmentation that simulates the relatively low resolution of the cropped images from the moth traps. More details on the hyper-parameters are in App. A.4. To mitigate the impact of majority classes on the model, we limit the number of training examples per species to 1,000. On a species list from the region of Quebec and Vermont, USA, our model achieves an accuracy of 86.14% on the GBIF held-out test set, and 77.81% on a small expert-labelled moth trap test set. We provide additional results in App. A.5.

2.5 Tracking individual moths

Since the final goal is to count the number of individuals for each moth species in a night's data, tracking becomes an essential part of the system. For memory reasons, the camera trap does not continuously collect images, taking pictures at a fixed interval or when activated by insect motion. This means that movements between frames can be relatively large and "jerky".

We approach object tracking by noting that instances of the same moth in consecutive frames will likely be close to each other, similar in size, and similar in the feature space of the species classifier. Formally, we calculate the cost of the assignment between any two moth crops in two consecutive images as a weighted combination of four factors: 1) Intersection over union, 2) ratio of crop sizes, 3) distance between box centres, and 4) similarity in classification model feature space. The lower the cost, the more likely the match. We use linear sum assignment [30] for optimal matching of the moth crops, with unmatched individuals indicating that a moth has either appeared for the first time, or has left the image. Due to a lack of ground truth data for tracking, our evaluation for this module is currently qualitative.

3 Pathway to impact

In order for ML tools to be usable for ecologists without extensive ML experience, we have developed a tool, the AMI Data Companion, an open-source software package (link omitted for anonymity) designed to assist researchers with these steps. (AMI stands for **A**utomated **M**onitoring of **I**nsects.) Data from field deployments of camera systems can be reviewed before processing takes place, for example, to confirm that the devices ran as scheduled. Images can be processed on demand or placed into a fault-tolerant queue for long-running operations. The choice of model is configurable for each operation (e.g. object detection, tracking, species classification). The software provides a graphical interface that can be run on all major desktop operating systems, as well as a command-line interface that can be run on multiple server nodes in parallel. We are also developing an online version of the application (in progress). The AMI Web Platform (Figure 3) is designed to address challenges in taxonomic alignment, access to compute infrastructure, the curation of training data, and the need to collaborate with experts from around the world.

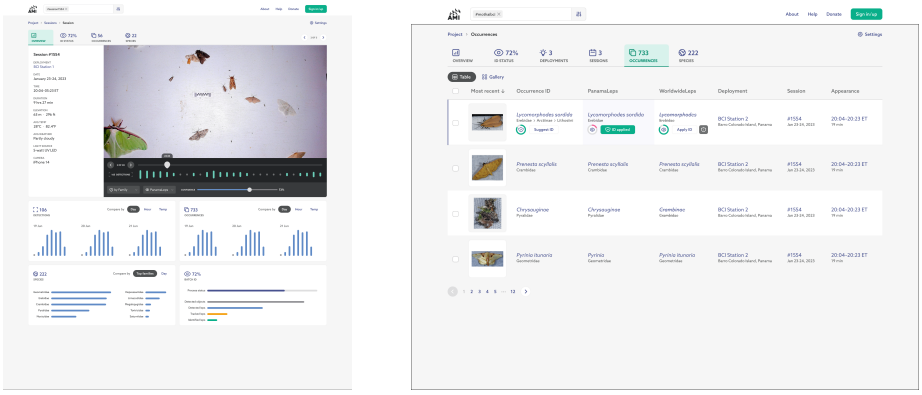


Figure 3: A preview of the AMI Web Platform in development.

These tools have enabled our work to used across a range of partner ecology organizations, and we are actively scaling up deployments worldwide. It is our hope that the data our methods provide to entomologists will help better inform land use decisions and policy making for climate change adaptation and conservation.

Acknowledgement

We sincerely thank our partners and their team without whom this project could not have been conceived and deployed. Here is a non-exhaustive list of key people: Kent McFarland[1], David Roy[2], Tom August[2], Alba Gomez Segura[2], Marc Bélisle[3], Toke Thomas Høye[4] and Kim Bjerge[4]. Thanks to Michelle Lin[5] and Yuyan Chen[5] for their feedback on the manuscript.

References

[1] Gerardo Ceballos, Paul R Ehrlich, Anthony D Barnosky, Andrés García, Robert M Pringle, and Todd M Palmer. Accelerated modern human–induced species losses: Entering the sixth mass extinction. *Science advances*, 1(5):e1400253, 2015.

[2] Gerardo Ceballos, Paul R Ehrlich, and Rodolfo Dirzo. Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the national academy of sciences*, 114(30):E6089–E6096, 2017.

[3] Gerardo Ceballos, Paul R Ehrlich, and Peter H Raven. Vertebrates on the brink as indicators of biological annihilation and the sixth mass extinction. *Proceedings of the National Academy of Sciences*, 117(24):13596–13602, 2020.

[4] Nigel E Stork et al. How many species of insects and other terrestrial arthropods are there on earth. *Annual review of entomology*, 63(1):31–45, 2018.

[5] David L Wagner. Insect declines in the anthropocene. *Annual review of entomology*, 65:457–480, 2020.

[6] Sandra Díaz, Josef Settele, Eduardo S Brondízio, Hien T Ngo, John Agard, Almut Arneth, Patricia Balvanera, Kate A Brauman, Stuart HM Butchart, Kai MA Chan, et al. Pervasive human-driven decline of life on earth points to the need for transformative change. *Science*, 366(6471):eaax3100, 2019.

[7] Roel van Klink, Tom August, Yves Bas, Paul Bodesheim, Aletta Bonn, Frode Fossøy, Toke T Høye, Eelke Jongejans, Myles HM Menz, Andreia Miraldo, et al. Emerging technologies revolutionise insect ecology and monitoring. *Trends in ecology & evolution*, 2022.

[8] Ruth Y Oliver, Fabiola Iannarilli, Jorge Ahumada, Eric Fegraus, Nicole Flores, Roland Kays, Tanya Birch, Ajay Ranipeta, Matthew S Rogan, Yanina V Sica, et al. Camera trapping expands the view into global biodiversity and its change. *Philosophical Transactions of the Royal Society B*, 378(1881):20220232, 2023.

[9] Kim Bjerge, Hjalte MR Mann, and Toke Thomas Høye. Real-time insect tracking and monitoring with computer vision and deep learning. *Remote Sensing in Ecology and Conservation*, 8(3):315–327, 2022.

[10] Kim Bjerge, Jamie Alison, Mads Dyrmann, Carsten Eie Frigaard, Hjalte MR Mann, and Toke Thomas Høye. Accurate detection and identification of insects from camera trap images with deep learning. *PLOS Sustainability and Transformation*, 2(3):e0000051, 2023.

[11] Jozsef Suto. Codling moth monitoring with camera-equipped automated traps: A review. *Agriculture*, 12(10):1721, 2022.

[12] Quentin Geissmann, Paul K Abram, Di Wu, Cara H Haney, and Juli Carrillo. Sticky pi is a high-frequency smart trap that enables the study of insect circadian activity under natural conditions. *PLoS Biology*, 20(7):e3001689, 2022.

[1]Vermont Center for Ecostudies
[2]UK Centre for Ecology & Hydrology
[3]Université de Sherbrooke
[4]Aarhus University
[5]Mila - Quebec AI Institute

[13] Jamie Alison, Jake M Alexander, Nathan Diaz Zeugin, Yoko L Dupont, Evelin Iseli, Hjalte MR Mann, and Toke T Høye. Moths complement bumblebee pollination of red clover: a case for day-and-night insect surveillance. *Biology Letters*, 18(7):20220187, 2022.

[14] Kim Bjerge, Jakob Bonde Nielsen, Martin Videbæk Sepstrup, Flemming Helsing-Nielsen, and Toke Thomas Høye. An automated light trap to monitor moths (lepidoptera) using computer vision-based tracking and deep learning. *Sensors*, 21(2):343, 2021.

[15] Dimitri Korsch, Paul Bodesheim, and Joachim Denzler. Deep learning pipeline for automated visual moth monitoring: Insect localization and species classification. 2021.

[16] Jonas Mielke Möglich, Patrick Lampe, Mario Fickus, Sohaib Younis, Jannis Gottwald, Thomas Nauss, Roland Brandl, Martin Brändle, Nicolas Friess, Bernd Freisleben, et al. Towards reliable estimates of abundance trends using automated non-lethal moth traps. *Insect Conservation and Diversity*, 2023.

[17] GBIF.org. GBIF Home Page. 2023. Available from: `https://www.gbif.org` 28 September 2023.

[18] iNaturalist. Available from https://www.inaturalist.org. Accessed 2023-09-28.

[19] GBIF Secretariat. GBIF Backbone Taxonomy. Checklist dataset https://doi.org/10.15468/39omei accessed via GBIF.org on 2023-09-28.

[20] GBIF (2021) Darwin Core Archives – How-to Guide, version 2.2. Copenhagen: GBIF Secretariat. https://ipt.gbif.org/manual/en/ipt/2.5/dwca-guide.

[21] Li Liu, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu, and Matti Pietikäinen. Deep learning for generic object detection: A survey. *International journal of computer vision*, 128:261–318, 2020.

[22] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. *arXiv preprint arXiv:2304.02643*, 2023.

[23] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.

[24] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[25] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115:211–252, 2015.

[26] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

[27] Tong He, Zhi Zhang, Hang Zhang, Zhongyue Zhang, Junyuan Xie, and Mu Li. Bag of tricks for image classification with convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 558–567, 2019.

[28] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

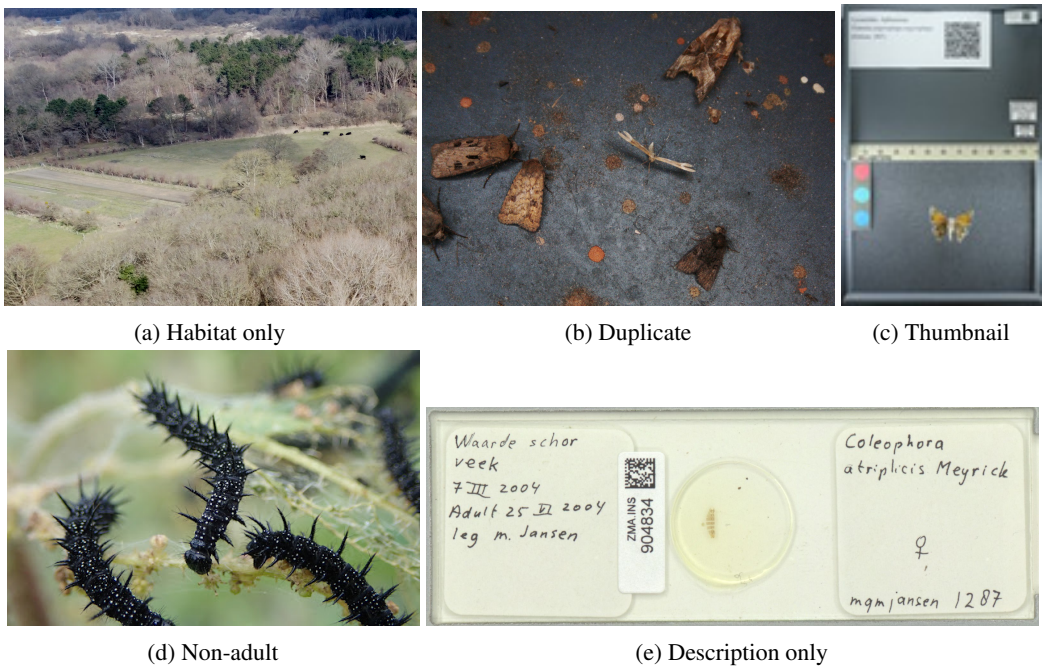
[29] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 702–703, 2020.

[30] David F Crouse. On implementing 2d rectangular assignment algorithms. *IEEE Transactions on Aerospace and Electronic Systems*, 52(4):1679–1696, 2016.

A Appendix

A.1 Examples of images removed during dataset cleaning

After fetching images and metadata, we clean the data by removing the following images: duplicates, thumbnails, non-adult images (based on metadata or using a life-stage classifier when metadata is not available), and images from datasets that we have identified as containing primarily descriptive information rather than specimen pictures. The Figure 4 shows some examples of images we removed during our dataset cleaning procedure.



(a) Habitat only (b) Duplicate (c) Thumbnail

(d) Non-adult (e) Description only

Figure 4: Examples of images that are removed during the dataset cleaning procedure. Some images are used as placeholders and do not contain any animals. (a) is an extreme case used by hundreds of thousands of occurrences. In some cases, the same picture has more than one species, and each individual is counted as a single occurrence, with the same image being referenced by all of them (b). Some occurrences have more than one image, and some of them are small images (thumbnails) (c). The scope of our models is only the adult individuals; non-adult pictures (d) should be removed. Finally, some pictures do not have any specimens, only descriptions (e).

A.2 Insect detector analysis

We initially thought blob detection should have good performance for this problem setting, given its relative simplicity. However, it fails when there is a lot of insect activity on the screen or inconsistent lightning. We cherry-pick images with a clean background and low density of insects to generate annotations using blob detection automatically. We then train a Faster R-CNN with a ResNet-50-FPN backbone (blobdata-ResNet50) on this data. In addition to being slow, this model is prone to two types of errors: missed detections (i.e. false negatives), especially on smaller moths, and double detections (i.e. bounding boxes that group multiple moths). Additionally, the model occasionally made multiple predictions on the same large moth.

We attribute the above issues to a need for diversity in training data. Hence, as discussed in subsection 2.2, we train two Faster R-CNN models on synthetic data generated using SAM: one with a heavier ResNet-50-FPN backbone (syntheticdata-ResNet50) and one with a lighter MobileNetV3-Large-FPN backbone (syntheticdata-MobileNetV3). Figure 5 shows a representative sample of the differences using visual inspection.

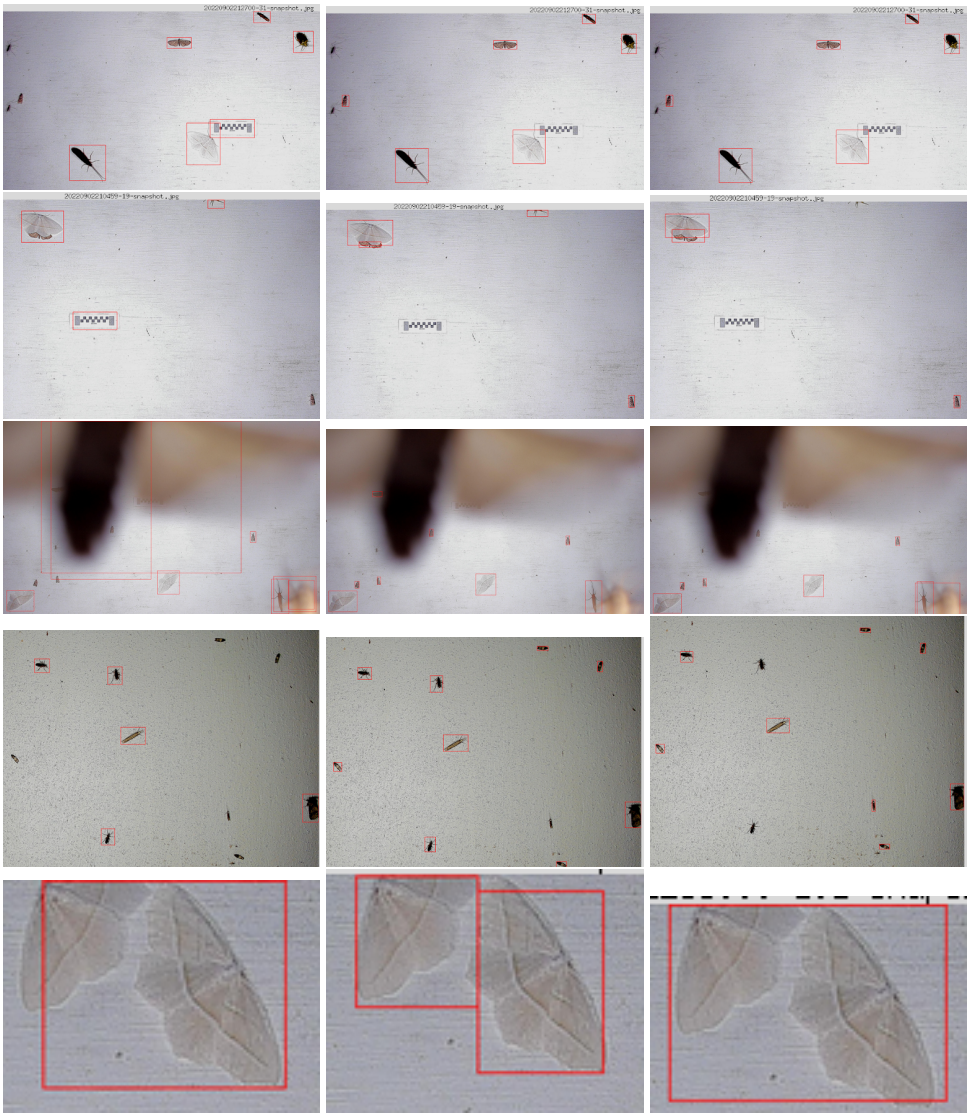


Figure 5: Left column: blobdata-ResNet50; middle column: syntheticdata-ResNet50; right column: syntheticdata-MobileNetV3. As seen in the first column, the model trained on a small amount of labelled data has many false positives and fails to detect insects close to each other. While the models trained on large amounts of synthetic data (last two columns) overcomes those challenges. We finally use the model with the MobileNetV3-Large-FPN backbone, as it is six times faster than its counterpart and similar in accuracy.

A.3 Number of images per species distribution

The number of images per species on GBIF for the Lepidoptera order follows a long-tail distribution, i.e., some species have many photos, while most have only a few, as shown in the Figure 6.

A.4 Species classifier hyper-parameters

Our classification algorithm uses the standard ResNet-50 [24] architecture, which is initialized with weights pre-trained on ImageNet-1K [25]. We use a low input resolution of 128 x 128, which is roughly the mean size of the cropped images. We found that this resolution produces better accuracy on moth trap images. The detailed training hyper-parameters are provided in Table 1.

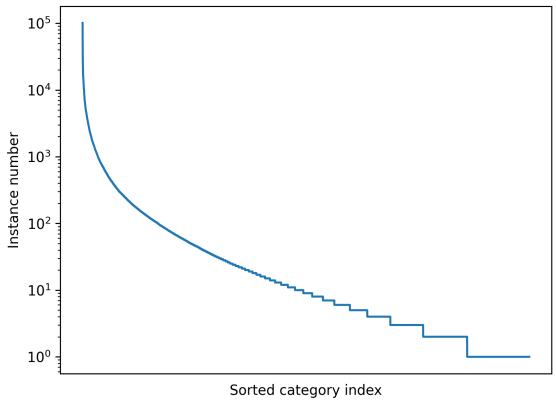


Figure 6: Number of images per species available in GBIF for the order Lepidoptera. The picture includes data from occurrences labelled as 'adult' from approximately 37,000 species. Categories are sorted by the number of images.

Table 1: Training hyper-parameters for fine-grained species classification task.

Hyper-parameters	Config
input resolution	128 x 128
optimizer	AdamW
learning rate	0.001
LR schedule	cosine
warmup epochs	2
training epochs	30
weight decay	1e-5
RandAug	N=2, M=9
label smooth	0.1

A.5 Classifier validation results

We provide validation results in Table 2 for two regional lists: Quebec-Vermont and UK-Denmark. However, it is important to note that the expert-labelled moth trap test set has only 1000 examples, with only 338 crops labelled at the species level, limiting the conclusions regarding our methods' generalisation. Obtaining more labelled data is one of our next steps.

Table 2: Test results for Quebec-Vermont and UK-Denmark regional lists. GBIF test sets are held-out from the GBIF training images. The expert-labelled moth trap test set for Quebec-Vermont contains 1000 examples, with 338 crops labelled at the species level. We predict at a higher taxonomic level (genus and family) by summing up the confidence of predictions within each higher taxon.

Dataset	Test set	Accuracy		
		Species	Genus	Family
Quebec-Vermont	GBIF	86.26%	90.20%	94.77%
	Moth trap	77.58%	77.00%	89.61%
UK-Denmark	GBIF	88.77%	92.43%	96.21%