scientific reports

OPEN

Artificial intelligence correctly classifies developmental stages of monarch caterpillars enabling better conservation through the use of community science photographs

Naresh Neupane[1]✉, Rhea Goswami[1], Kyle Harrison[1], Karen Oberhauser[2], Leslie Ries[1] & Colin McCormick[1]

Rapid technological advances and growing participation from amateur naturalists have made countless images of insects in their natural habitats available on global web portals. Despite advances in automated species identification, traits like developmental stage or health remain underexplored or manually annotated, with limited focus on automating these features. As a proof-of-concept, we developed a computer vision model utilizing the YOLOv5 algorithm to accurately detect monarch butterfly caterpillars in photographs and classify them into their five developmental stages (instars). The training data were obtained from the iNaturalist portal, and the photographs were first classified and annotated by experts to allow supervised training of models. Our best trained model demonstrates excellent performance on object detection, achieving a mean average precision score of 95% across all five instars. In terms of classification, the YOLOv5l version yielded the best performance, reaching 87% instar classification accuracy for all classes in the test set. Our approach and model show promise in developing detection and classification models for developmental stages for insects, a resource that can be used for large-scale mechanistic studies. These photos hold valuable untapped information, and we've released our annotated collection as an open dataset to support replication and expansion of our methods.

Insects dominate the terrestrial fauna, and holometabolous insects, which undergo complete metamorphosis (e.g., beetles, moths and butterflies, flies, ants and bees, and some smaller taxa), represent 83% of all insects and ~50% of all known animal species. Lepidoptera alone (moths and butterflies) comprise 10% of all described animals[1]. As a whole, insects compose the bulk of terrestrial animal biomass, collectively outweighing all terrestrial vertebrates, including humans and livestock. Insects largely form the base of the animal food web and provide substantial ecosystem services, including nutrient cycling, pollination, and pest control[1]. Unfortunately, evidence is mounting that overall insect biomass has declined with estimates from 30 to 50% over the last 20–50 years and numerous studies indicate that the populations of many beneficial insects are likely to decline in the future[2,3]. A major research challenge is that multiple factors, such as changes land-use, climate, and agricultural practices, are difficult to tease apart, leading to a laundry list of stressors being referred to as "death by a thousand cuts"[4]. While the relative contribution of each stressor is likely to vary by region, to capture dynamics at the largest scales (even continental or global), data on the species of interest are required at those scales. The most widely available large-scale biodiversity datasets offer only occurrence data (species, location, date) of adults, and while this has helped researchers track range dynamics, they are not able to capture mechanistic processes at smaller scales[5]. For instance, the vast majority of large-scale Lepidoptera research is done using occurrences of adults, yet most individuals in nature are juveniles because they never emerge as adults, dying first from a variety of factors[6]. The juvenile stages of insects that undergo complete metamorphosis are egg, larvae, pupae, and adult (most other insects are hemimetabolous, with juveniles resembling adults). Of these earlier stages,

[1]Georgetown University, Washington, DC, USA. [2]University of Wisconsin, Madison, WI, USA. ✉email: nn343@georgetown.edu

larvae (caterpillars for moths and butterflies) are the most apparent with eggs and pupae notoriously difficult to find. Insect larvae go through a distinct set of developmental stages, called instars. Tracking dynamics at each earlier stage will open a window for more mechanistic understanding of how the environment impacts growth and development of this critical group of insects. Here, we test the ability of computer vision to identify butterfly larvae to specific instars.

For butterflies in North America, the most iconic species is the migratory monarch butterfly (*Danaus Plexippus)*. Monarchs have become a model system for understanding migratory insects in particular and butterflies in general. Globally, monarchs have a high public profile[7] and currently are undergoing yearly review for potential listing under the United States (US) Endangered Species Act (ESA)[8]. The migratory monarch butterfly has two fairly distinct migratory populations in North America (Florida also hosts a large sedentary population), which have been declining since at least the mid-1970s in the east[9] and since at least the mid-1990s in the west[10]. The best-known eastern migratory population covers the area east of the Rocky Mountains and migrates to central Mexican forests for overwintering. In the spring, adults that spent the winter in Mexico travel northward to the southern US, breed, and produce the next generation that moves further north, expanding throughout the Northeastern and Midwestern US and southeastern Canada where they produce 2–3 more generations[11]. In the fall, butterflies in the final generation of the year return to Mexico. The western migratory population inhabits the area west of the Rockies and overwinters in southern California along the Pacific coast. During the summer, they migrate up to the Rocky Mountain range.

The base data source for this research is iNaturalist (https://www.inaturalist.org), a photo upload platform which contains vast and growing numbers of timestamped, geolocated images; it surpassed 100 million observations in 2022[12]. Monarch caterpillar photos are the most common on iNaturalist compared to other Lepidoptera (personal observation) and now comprise a rich and quickly growing data set. While many photo contributors may not know the name of the specimens they have photographed, a widespread effort to automate species identification using computer vision and deep learning has been highly successful at providing a source of species classification[13]. This is complemented by community-sourced identification verification by expert naturalists[14]. These photos hold valuable information about species' ecology, development, and evolution, but the species are rarely classified or annotated for other observable characteristics (i.e., morphological or situational data). There are other platforms as well, including observation.org and flickr. These expanding resources make these photographs a still largely untapped resource for biodiversity and conservation research[15]. Indeed, the burgeoning field of "imageomics", especially when paired with computer vision and deep learning, has the potential to transform the way we study the vast diversity on our planet[16]. Each species requires a tailored spatial and temporal scope to effectively track their population dynamics. In the case of eastern migratory monarchs[11], understanding the dynamics of this single panmictic population[9] could be vastly improved with developmental data throughout their breeding range in eastern North America.

Our goal is to assess whether computer vision can classify monarch caterpillars by their developmental instar using photographs uploaded by citizen scientists, a new application in the emerging fields of imageomics as well as "phenometrics" (observable features of organisms)[15]. More specifically, we also aim to train a deep-learning model that can accurately (1) localize caterpillars in a photograph and (2) classify caterpillars into their respective developmental stages, from instar 1 to 5. This novel process could allow for the tracking of developmental stages and other observable information on a continental scale; an invaluable data set on insect biogeography providing insights into many dynamics, including adaptability to climate or land use change and information on impacts of conservation actions. For imageomics to be a promising application in large-scale phenometrics research, differentiation between relevant classes must be apparent in visual representations[17], which is possible for most caterpillars[18] including for monarchs (Fig. 1). Photos uploaded to web portals are often classified or identified to the species level, rarely categorized by the insect's specific developmental life stage or condition. When this information is available, for instance, from the "caterpillars of eastern North America project" on iNaturalist[19], where all annotations are manual, it only specifies that the photo is a caterpillar rather than an adult, not the specific instar stage. We do note that certain Lepidoptera species are far more apparent at the caterpillar stage, especially those that develop gregariously in webs and are pest species, such as the fall webwork (*Hyphantria cunea*) and the eastern tent caterpillar (*Malacosoma americanum*), or species where highly apparent caterpillars wander widely, such as the wooly bear (*Pyrrharctia isabella*). Yet for the vast majority of Lepidoptera, adults are the most apparent.

With the growing popularity of convolutional neural networks (CNNs), their application in insect detection and identification has notably expanded in recent years. While the use of computer vision for insect detection is particularly prevalent in agriculture for pest management and control[20], our focus is on conserving beneficial insects[3], specifically butterflies, that are not only vital to agriculture but also important from a broader conservation perspective.

Multiple studies have shown the success of CNNs in identifying adult butterflies[21]. For example, [22–24]reported models that were more than 85% accurate in identifying certain butterfly species. [25]used the GoogLeNet pretrained CNN architecture and achieved an accuracy of up to 97% in identifying and classifying four species of butterflies. [26]obtained a detection accuracy of up to 98% for a set of 10 species. From the image detection point of view, most research has focused on adult butterflies but not on more fine-scaled developmental stages, such as instars. The success of CNNs in identifying adult butterflies is partly due to their larger size and often charismatic wings; while many other insects are smaller, or have large variations of size within a species, and in those cases, detection accuracies often fall[27–29]. In our case, the body sizes of instars vary significantly, with the first instar (L1), ranging in length from 2 to 6 mm, being much more difficult to localize in photographs compared to the 5th (L5) instar (25–45 mm) (Fig. 1). Our method is comparable to the approach used for locating and identifying ladybird beetles and identifying life history timing of fall webworm, using a two-step automated detector in iNaturalist photographs, which resulted in 92% accuracy by combining image processing and deep

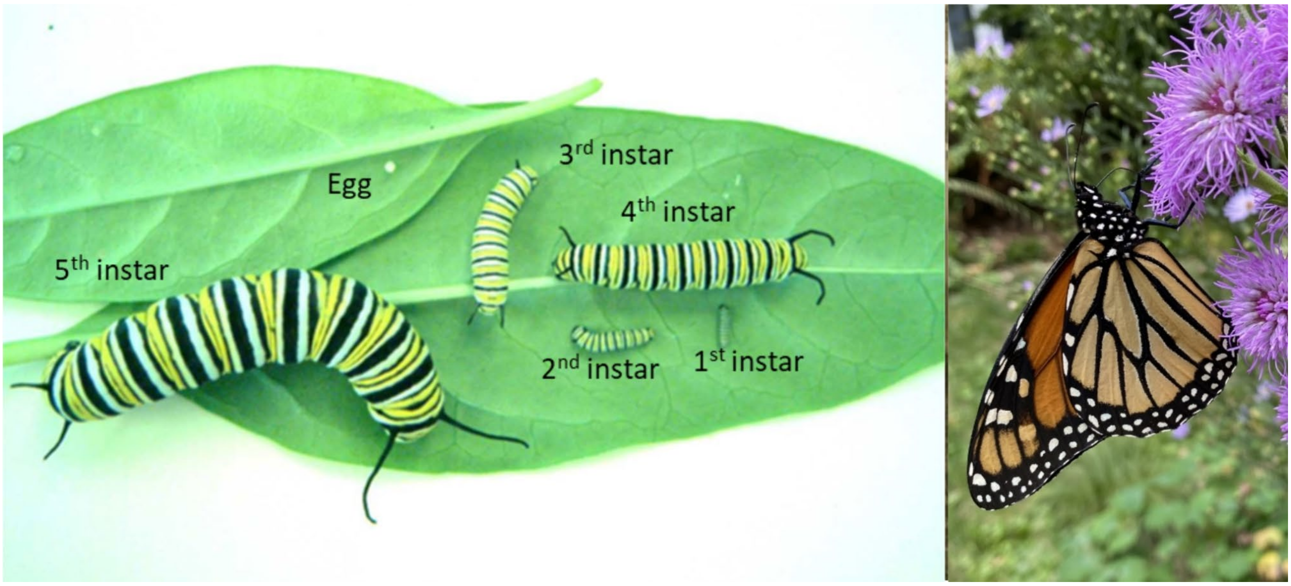

Fig. 1. Monarch butterflies (adult shown to right) proceed through five developmental instars as juveniles. In the field, size is a useful factor in instar-identification, combined with other morphological differences that can be easily learned. However, since photograph sources almost never have scale bars and background features that could be used for scale, such as host plant leaves, vary a great deal in size, only morphological features can be used to distinguish caterpillar stages using field photographs. Photo credit: Karen Oberhauser (printed with permission).

learning techniques[30,31]. Note that this process was trained on adult insects and not their earlier developmental stages. To determine the tractability of creating a computer vision pipeline for locating butterfly caterpillars and identifying them to instar, we constructed a dataset of citizen scientist photographs of monarch caterpillars at all instar stages from iNaturalist and manually annotated both the location and instar stage in each photo to create a training dataset (Fig. 2).

Results

To evaluate the performance of the trained models in detecting and localizing caterpillars, we used the mean average precision (mAP) metric, the most common single metric used in modern object detection and classification[32]. Because this metric depends on the choice of intersection-over-union (IOU) threshold chosen for a positive localization (detection), we use two versions of it: the value with the IOU threshold set to 0.50 (mAP50) and the average value across a range of IOU thresholds from 0.50 to 0.95 (mAP50-95). Overall, all four model sizes demonstrated high accuracy in detecting and localizing caterpillars (e.g., mAP50 ~ 0.95 and 0.94 in validation and test sets, respectively, see Table 1). The highest values of mAP50 for the hold-out test set (the most unbiased measure of model performance) were achieved by the medium and large models, while the highest values of mAP50-95 were achieved by the large and extra-large models. Balancing resources and performance, we conclude that the large model has the best overall performance. These findings align with previous research by[33], emphasizing the relatively higher performance of the YOLOv5 large version at a reasonable speed and computational cost.

Because training neural networks is a stochastic process, it is important to select a sufficient number of training epochs to fully optimize the model and assess whether it is overfitting, meaning it performs well on the training data but less effectively on a hold-out test set[34]. At the end of training, all four model sizes were used to test the model on the test dataset. Here we present results of the loss value function from the YOLOv5 large model (Fig. 3). This showed stable mAP50 and mAP50-95 values for the validation set, indicating that more training epochs are not likely to improve the models' performance. For example, in all loss graphs, during training, the gradient descent initially increases rapidly, then gradually slows down, stabilizing after 200 epochs. Also, both mAP50 and mAP50-95 of the best trained model evaluated on the holdout test set are very similar to the corresponding values evaluated on the validation data set. This indicates that there was no model overfitting. Despite this high performance, all of the models faced some challenges in accurately identifying L2 larvae, often confusing them with L3s. The model that we selected here, YOLOv5l, demonstrated up to 95% accuracy in classifying L3 images and 94% accuracy in identifying L5 images in the validation set (Fig. 4a). In general, the prediction accuracy for all classes was above 0.87. However, the model encountered challenges when classifying L2, where 11% of the images were erroneously predicted, confusing L2 with L3 larvae.

Comparatively, when tested on the test sets (Fig. 4b), there were slight reductions in the model accuracies. The prediction accuracy ranged from 81% for L4 to 94% for L5. The model exhibited some misclassifications. In 13% of the cases, it mistakenly labeled L1 as L2, and similarly, it inaccurately identified L2 as L3. Additionally, the model returned some false positives (incorrect prediction of instars in background areas), although this

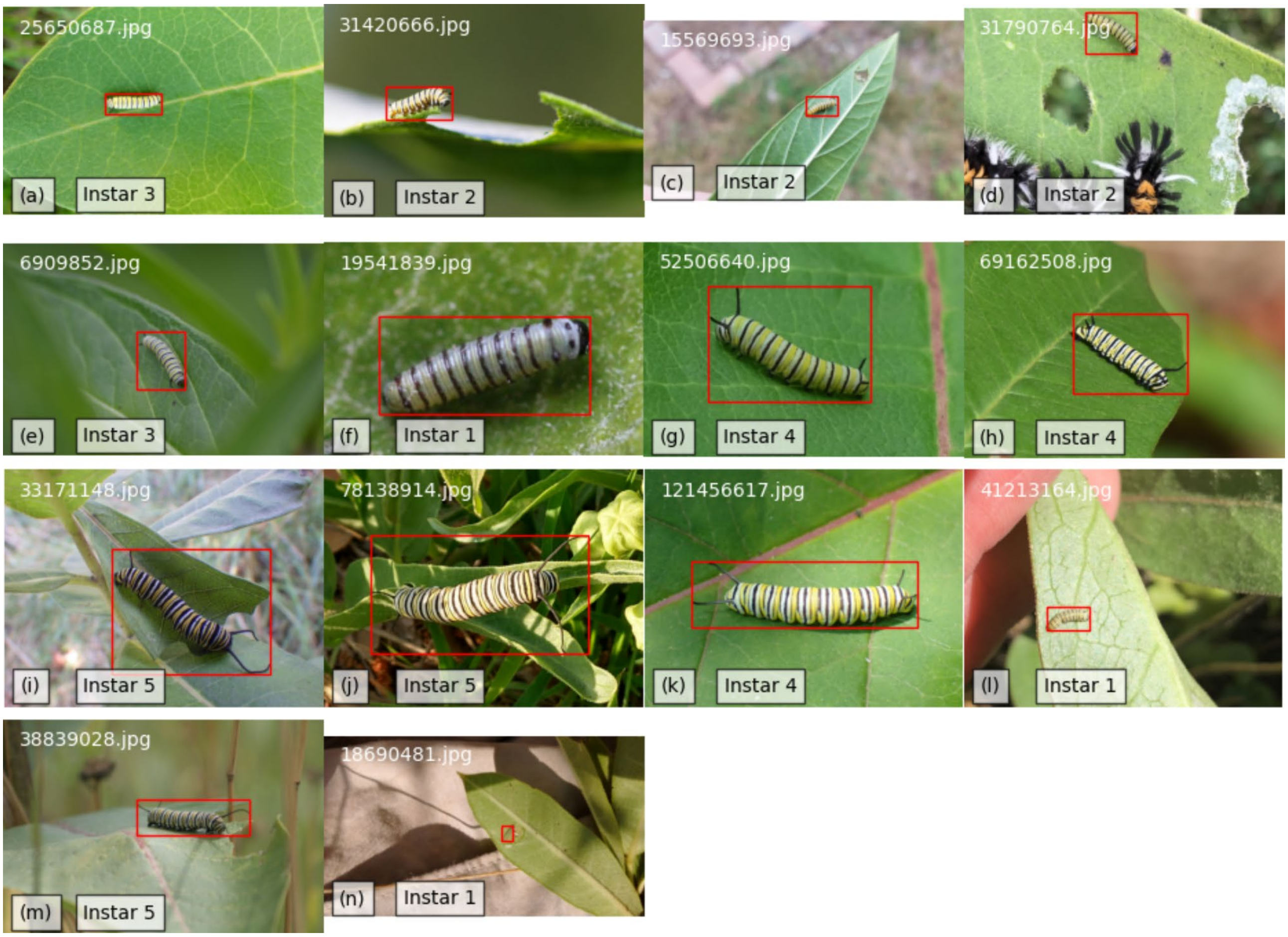

Fig. 2. Monarch caterpillars progress through five developmental stages (instars), depicted in panels (**a**–**p**). These images are representative samples of 1 photos used in the test set. Our team of experts annotated each image, providing both the caterpillar's location (bounding box) and its respective instar assignment. Photo credit: (**a**) 25,650,687 © Norman Murray, all rights reserved, (**b**) 31,420,666 © Rodrigo Solis Sosa, some rights reserved (CC-BY-NC), (**c**) 15,569,693 © Kim Smith, some rights reserved (CC-BY-NC-ND), (**d**) 31,790,764 © Emma Horrigan, all rights reserved, (**e**) 6,909,852 © Mark Kluge, some rights reserved (CC-BY-NC), (**f**) 19,541,839 © Even Dankowicz, some rights reserved (CC-BY), (**g**) 52,506,640 © Lyell Slade, lls, some rights reserved(CC-BY-NC), (**h**) 69,162,508 © Michael (Mike) Ostrowski, some rights reserved (CC-BY-SA), (**i**) 33,171,148 © David Weisenbeck, some rights reserved (CC-BY), (**j**) 78,138,914 © Ian Shelburne, all rights reserved, (**k**) 121,456,617 © Chris Buelow, some rights reserved (CC-BY-NC), (**l**) 41,213,164 © Lauren J. Simpson, some rights reserved (CC-BY-NC), (**m**) 38,839,028 © Meghan Pierce, some rights reserved (CC-BY-NC), (**n**) 18,690,481 © Royce J. Bitzer (iowabiologist), some rights reserved (CC-BY-NC). Each number corresponds to an iNaturalist observation ID, accessible via the base URL https://www.inaturalist.org/observations/followed by the respective observation number. The photos are printed with permission from the photographer.

	mAP50 (valid)	mAP50-95 (valid)	mAP50 (test)	mAP50-95 (test)
Small	0.921	0.812	0.925	0.814
Medium	0.943	0.842	0.947	0.853
Large	0.951	0.862	0.947	0.855
Extra-large	0.945	0.864	0.939	0.861

Table 1. Best trained model performance (mean average precision, mAP) on the validation and hold-out test sets.

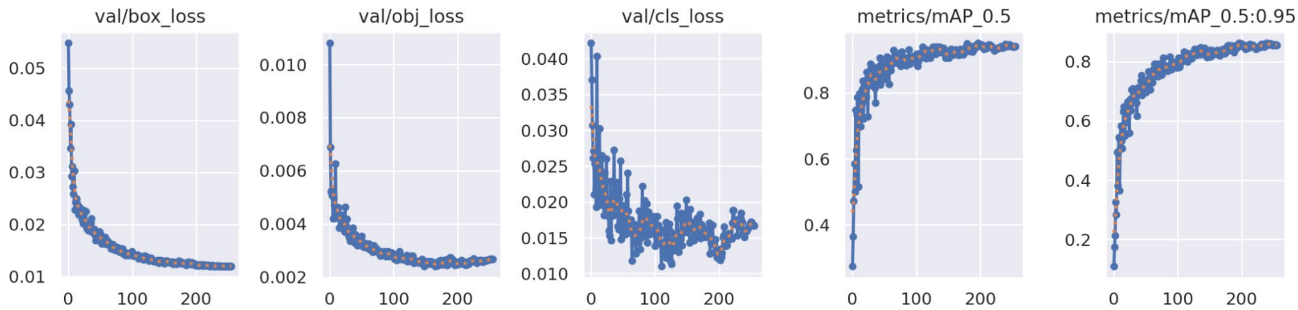

Fig. 3. Evolution of the box loss, object loss, class loss, mAP50, and mAP50-95 for the large model on the validation set, during the 256 epochs of training. Box loss represents the mean of IoU loss, objectness loss represents the mean of the object detection loss, and classification loss represents the mean of classification loss in the validation dataset. In all three metrics of losses, the loss curves become stable, with small fluctuations after about 200 epochs. This means, after about 200 epochs the model is fully trained, indicating that further epochs are not likely to improve performance. Based on this, we trained the model with 256 epochs.

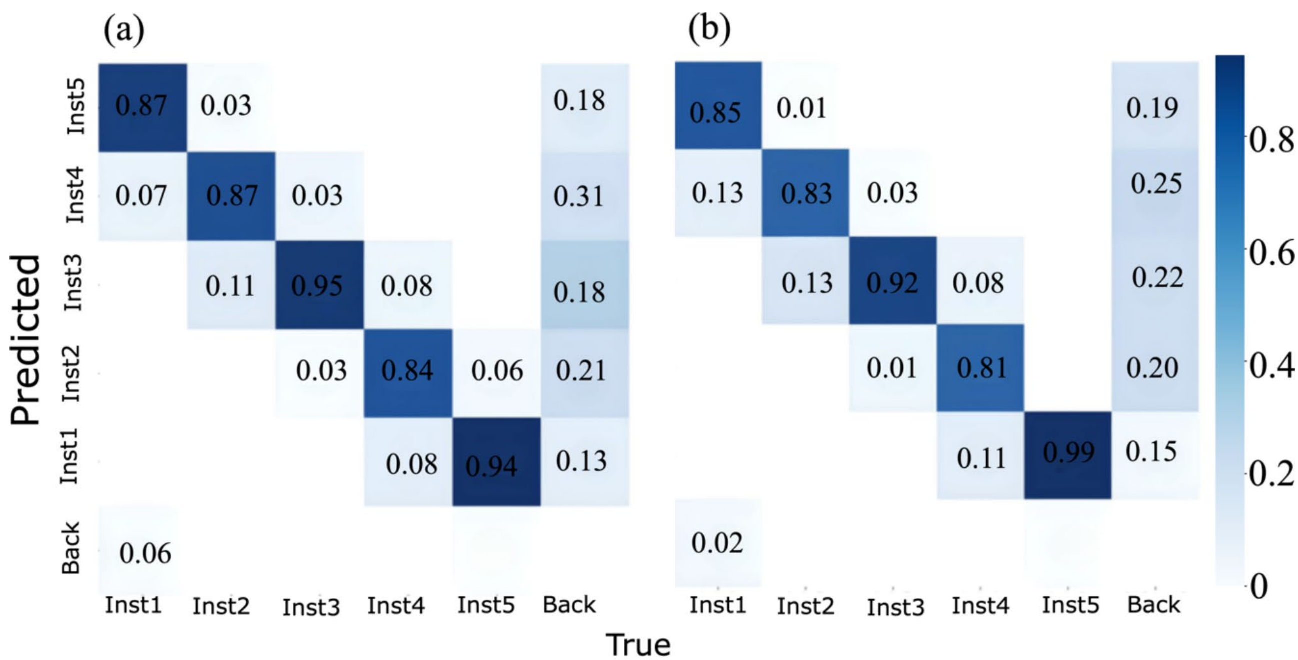

Fig. 4. The confusion matrix displays the classification accuracy for the fully trained large model, on (**a**) the validation set (left) and (**b**) the hold-out test set (right). The validation set comprised 383 (18%) images, while the holdout test set contained 213 (10%) images of instars from categories 1–5.Monarch caterpillars (larvae) go through five developmental stages, known as instar 1 to instar 5. The size ranges for each stage are as follows: 1st instar (2–6 mm), 2nd instar (6–9 mm), 3rd instar (10–14 mm), 4th instar (13–25 mm), and 5th instar (25–45 mm) (for details, see Supplementary A[51]). In instances where the predicted label matches the true label, the value along the diagonal becomes 1.

varies by the IOU threshold chosen. At a confidence threshold of 0.25, more than 20% of the photos had the true background mistakenly identified as containing L2, L3, and L4 (Fig. 4b). This could be improved with better image quality and more images for training[35].

We also evaluated performance using the F1 score, which is the harmonic mean of precision and recall, and serves as a comprehensive index for evaluating model performance across different confidence thresholds[36]. Upon analyzing the F1 curves, we observe that L1 and L5 exhibit the highest F1 value, indicating strong model performance in accurately classifying these instar stages (Fig. 5). Conversely, L2 and L3 demonstrate relatively lower F1 evaluations, suggesting more difficulty in identifying these developmental stages. The combined performance across all instar classes, YOLOv5l achieves F1 score of 0.9 at a confidence level of 0.81. A smoother curve indicates higher prediction confidence and a lower occurrence of False Positives (FP) and False Negatives (FN). Specifically, the curves exhibit smoothness for L1 and L5, while they show more fluctuations for L2 and L3. The model maintains a high F1 score across a broad range of confidence thresholds for L1 and L5. Therefore,

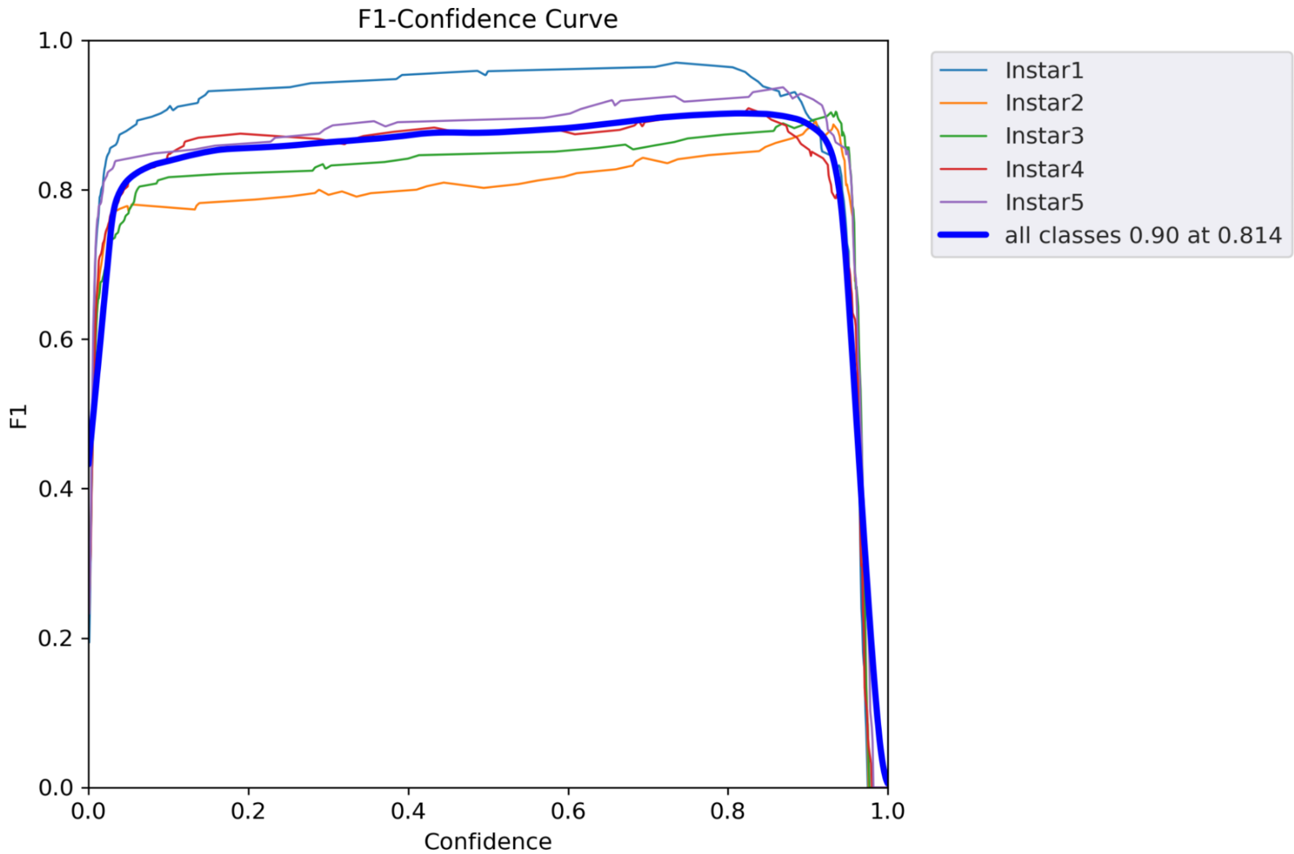

Fig. 5. F1 score curve of a YOLOV5 large weight model. F1 score is a measure of model's accuracy and it is the harmonic mean of precision and recall. Higher F1 score signifies superior performance, with the optimal threshold for the model prediction identified where the F1 score peaks. Notably, this YOLOv5 large model averaged for all classes achieves significantly high confidence (0.9), while concurrently optimizing the F1 (0.81) score.

it is evident that the models encountered challenges in accurately classifying L2 and L3 larvae, but demonstrated better performance with increased confidence and precision when classifying L1 and L5.

Next, we used the trained YOLOv5l model to perform inference (detection and localization) of instars on each of these images in the test set (Figs. 2 and 6). For 158 (74%) of these images, the model detected a single instar of the correct class; for 13 (6%) of these images, the model detected a single instar of the wrong class; and for 42 (20%) of these images, the model identified multiple instars, despite each image containing only a single instar. The confidence scores of the model were relatively high when it detected the correct class (including multiple detections where the most confident one was correct), with a mean of 0.949 (SD = 0.047). When the model detected the incorrect class, the confidence scores were lower, yet still high with a mean of 0.911 (SD = 0.061).

One notable challenge of YOLOv5l was ensuring precise instar localization accuracy in the photograph. The model struggles more with images featuring very small instars or those that occupy a minimal area in the photograph, because lower resolution images often present challenges in object (adult mosquito) detection and classification, as highlighted by[28,37]. To explore this, we first calculated the number of pixels within the manually labeled instar bounding box for each image in our hold-out test set, as a way to quantify how "big" or "small" the instar is within an image. For the first two sets of images (a single detection) we calculated the IOU of the manual bounding box and the model prediction, and plotted the result against the number of pixels within the manual bounding box (see Fig. 7). From this it is clear that lower localization accuracy (smaller IoU) is associated with smaller instars (fewer pixels in the manually annotated bounding box), demonstrating that the model does indeed perform less effectively with smaller objects.

Discussion

Deep learning techniques have found a wide range of applications in recent years, with several recent studies using such techniques to identify adult butterfly species[22,23,25,38]. Also, advancements have been made in detecting and classifying insect instar stages, particularly for agricultural pests[20]. Deep learning techniques have been successfully applied to classify bagworm instars, providing high accuracy and valuable data for pest management[39,40]. Similarly, YOLO-based deep learning frameworks have been used to detect and classify whitefly life stages in soybean crops, demonstrating the potential of these methods for pest control[41]. However, to date, we are unaware of any studies that have used deep learning for identifying and classifying butterfly

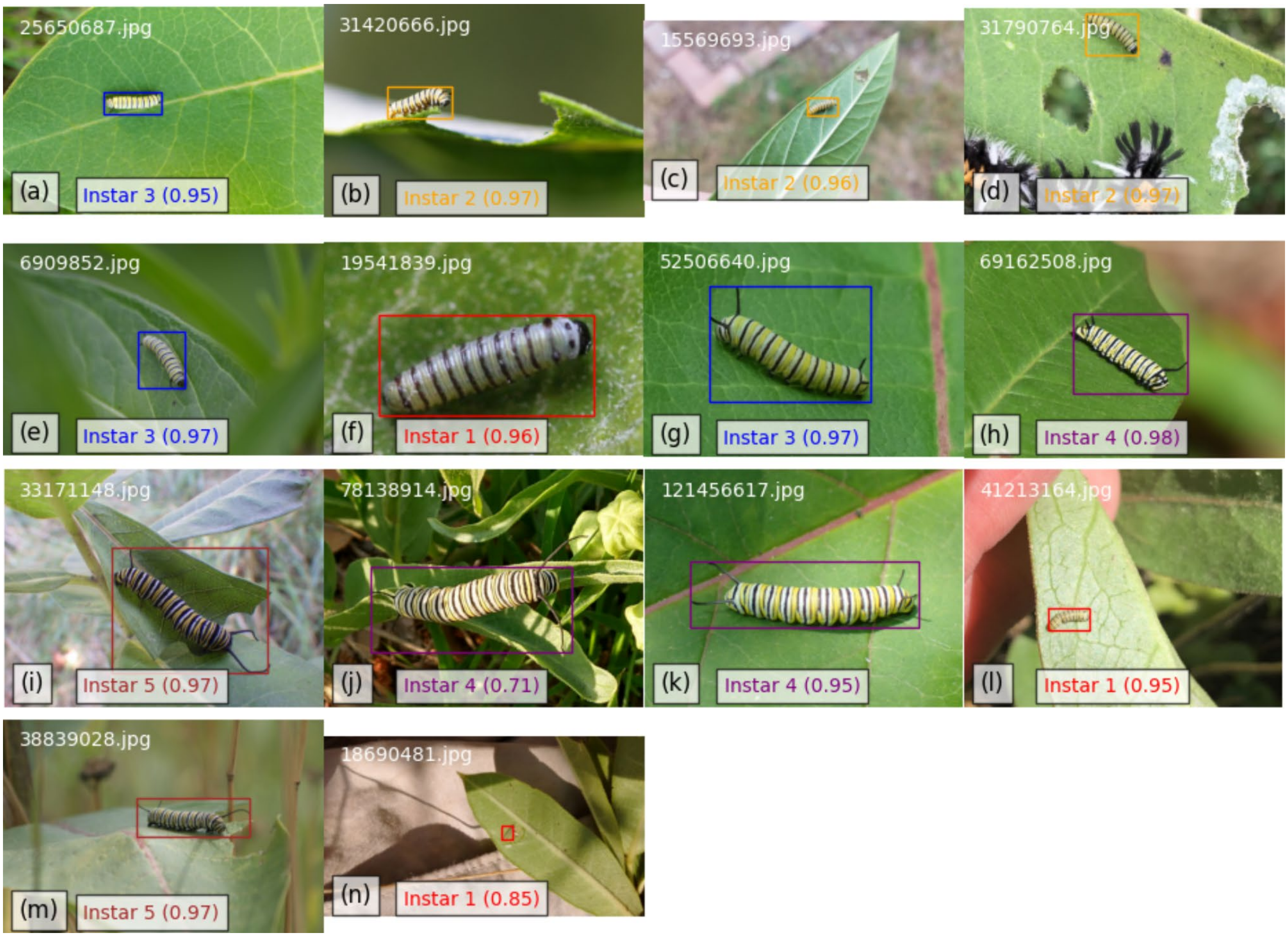

Fig. 6. YOLOV5l model detections and classifications of monarch instars (**a**–**p**). The true expert classified labels are displayed in Fig. 2. Predicted instar category and the confidence values are as indicated. Photo credit: (**a**) 25,650,687 © Norman Murray, all rights reserved, (**b**) 31,420,666 © Rodrigo Solis Sosa, some rights reserved (CC-BY-NC), (**c**) 15,569,693 © Kim Smith, some rights reserved (CC-BY-NC-ND), (**d**) 31,790,764 © Emma Horrigan, all rights reserved, (**e**) 6,909,852 © Mark Kluge, some rights reserved (CC-BY-NC), (**f**) 19,541,839 © Even Dankowicz, some rights reserved (CC-BY), (**g**) 52,506,640 © Lyell Slade, lls, some rights reserved(CC-BY-NC), (**h**) 69,162,508 © Michael (Mike) Ostrowski, some rights reserved (CC-BY-SA), (i) 33,171,148 © David Weisenbeck, some rights reserved (CC-BY), (**j**) 78,138,914 © Ian Shelburne, all rights reserved, (**k**) 121,456,617 © Chris Buelow, some rights reserved (CC-BY-NC), (**l**) 41,213,164 © Lauren J. Simpson, some rights reserved (CC-BY-NC), (**m**) 38,839,028 © Meghan Pierce, some rights reserved (CC-BY-NC), (**n**) 18,690,481 © Royce J. Bitzer (iowabiologist), some rights reserved (CC-BY-NC). Each number corresponds to an iNaturalist observation ID, accessible via the base URL https://www.inaturalist.org/observations/ followed by the respective observation number. The photos are printed with permission from the photographer.

caterpillar instars, particularly those of monarchs. Indeed, a few applications we have seen are for a select number of pests[42,43], including aphid nymphs - but it is worth noting that these do not undergo complete metamorphosis, so the juveniles look very similar to the adults[44]. To address the gap in application of these tools, we developed a deep learning model based on the YOLOv5 framework which is capable of localizing and categorizing monarch instars into their five developmental stages, L1-5. Monarch butterflies are one of the most widely studied and popular insect species. Unfortunately, their population is declining[9]. Categorizing their larval developmental stages allows a more mechanistic data set for range-wide analysis of the factors influencing population dynamics[45]. Our study could easily be adapted to identify and classify the instars of other butterflies; even though species show great morphological differences, monarchs are typical of the other species in its subtribe (Danainae) in terms of the progressive changes in instar appearance, and this is true of many other groups[18].

The training photographs used in this study were annotated and classified by coauthors Oberhauser, Ries, and Neupane, (see Methods and the rules outlined in Appendix A). Oberhauser is one of the leading experts on monarchs and runs the monarch larva monitoring program (mlmp.org) where hundreds of volunteers have been trained to identify monarch caterpillars in the field to instar[46]. Ries and Neupane have also published widely on the monarch[47–49]. It can be difficult to accurately classify caterpillars' developmental stages, even for experts,

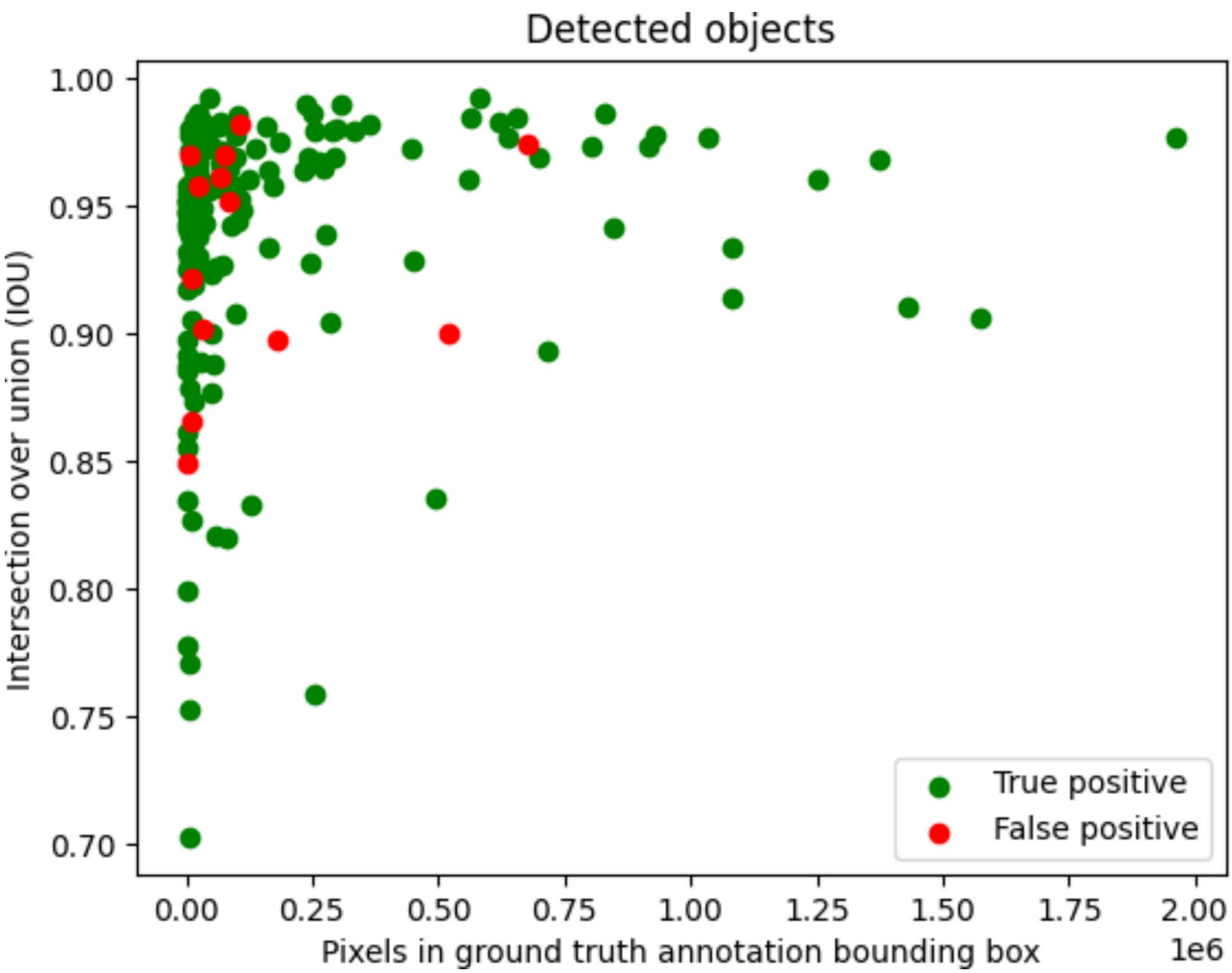

Fig. 7. IOU of the manual bounding box and the model-predicted bounding box, as a function of the number of pixels in the manual bounding box. Images are taken from the hold-out test set (see text).

without a clear photograph that displays key morphological features. While our model may improve through iterative training and potentially identify features missed by human annotators[50,51], there is also a risk that errors in the manual annotations could be propagated into the model and affect its accuracy. This highlights the importance of carefully checking the annotations, adding more training photos when possible, and iteratively improving the model to minimize potential bias or inaccuracies. Our results are comparable to studies using deep learning for mosquito larvae (e.g., up to 97%, as in[52,54] and locusts (e.g., up to 96%)[55].

Accurately identifying caterpillar developmental stages provides a novel data set for understanding their ecological role and supporting conservation efforts by making it possible to examine phenotypical differences in different environments. The YOLO framework efficiently and accurately located caterpillars in photographs and identified their developmental stages (Figs. 2 and 6). This technology is particularly valuable for platforms like iNaturalist, where large datasets can be quickly processed with annotations that provide valuable ecological and evolutionary data at the individual scale. and compile. While, other platforms provide photographs of caterpillars; (for instance observations.org hosts thousands of butterfly photos, including caterpillar photos) there is currently no way to filter by life stage (observation.org); indeed, this is also true of most iNaturalist Lepidoptera photos. However, adding a step where the algorithm classifies each photo to life stage would be a trivial task for a computer vision algorithm[56]. Implementing models like this is crucial to advance large-scale mechanistic research, a substantial recent advancement[57]. These types of processes could be developed into pipelines that provide important information about developmental stages, their condition and, combined with appropriate environmental data, identify key stressors. Such stage-specific data may be especially useful for conservation biologists, because they can increase the accuracy of population viability models - which are often used in species management. Such data is also critical for research on insectivores who are directly impacted by the dynamics of herbivorous insects. Ornithologists in particular are interested in caterpillar dynamics, because caterpillars are the primary food source for insectivorous birds[58] and several studies have recently identified insect declines as a possible reason for bird declines[59]. In light of this, further mechanistic research remains feasible even by being able to group larvae by their adjacent developmental stages (for example, 2/3 or 4/5).

Among the models evaluated (YOLOv5 small, medium, large, and extra-large), YOLOv5 large emerged as the top performer in terms of mAP (Table 1). The model's average prediction accuracy was above 0.94 for moderate localization accuracy (mAP50) and above 0.86 for high localization accuracy (mAP50-95). This is in agreement with the findings of[60] who reported comparable performance for various YOLO versions for small insect and pest detection. The mAP scores for identification of various instar categories displayed differences (above 92% for all instar categories, Fig. 5). It is possible that the model performance could be improved if more training data were included in the future. The model achieved a high accuracy rate, correctly identifying and classifying 85% of the images (42 out of 50). To visually demonstrate the predictions of the models, we show 14 images from the holdout test set (Fig. 6), which displays the predictions made by the trained YOLOv5l model, with the associated confidence value of the prediction.

Another notable strength of the trained YOLOv5l model is its ability to confidently detect instars in many of the images, with 100% certainty (not shown), while in many other cases, the certainties were above 95% (for example, in 10 out of 14 photos in Figs. 2 and 6). However, there were certain instances where the model made incorrect classifications. L2/3 were mistakenly classified as L3/2 (not shown), and L4/5 was misclassified as L5/4 (Fig. 6j). Those errors occurred with relatively low confidence values, reaching up to 0.4 in most cases and up to 0.71 in one instance, suggesting that the model's performance could benefit from a larger training dataset. We observed that the model tends to have lower accuracy in localizing instars in images with smaller sizes (fewer pixels). This raises the question of whether the model performs better when dealing with higher resolution images (Fig. 7). Notably, accuracy at later instars likely has more opportunity to provide more informative data. Not only do L4 and L5 take up more space in an image compared to L1-L3 and thus provide more clarity to annotate the photos. Further, for biological reasons, annotations of younger instars are likely not as meaningful as for L5s; monarch caterpillars only spend 2–3 days in each of the earlier instars (assuming a consistent, comfortable temperature throughout), where most errors occur (L1-L4), whereas they spend 6 days (40% of the full caterpillar lifespan) as an L5[61].

During this process, it took each of our expert annotators approximately 4–5 h to identify, classify, and annotate 200 photos. Images of L1-3 were generally smaller than L4-5, making them more difficult to locate in the images, and distinguishing between L4 and L5 was challenging even for experts. In such a situation, a precise computer vision model for classification would be highly valuable, especially since an iterative cycle of modeling and annotation of more photos will continue to increase automated annotation accuracy. Indeed, the machine algorithm may even find features that human annotators have not discovered to help distinguish between instars. This means that the identification and classification rules we adhered to (see Appendix A) can be improved by future research.

Our results showed that identification of caterpillars to their developmental instars can provide a potentially transformative source of data for entomological research on moths and caterpillars. These are among the most highly photographed insects in biodiversity photo upload platforms such as iNaturalist[62]. This exponentially growing dataset provides a flow of data that far outpaces our ability to annotate photos manually, and the development of a pipeline to locate and identify caterpillars to their developmental stage would be of great value to entomologists studying insect ecology and conservation. Further, this first step opens the possibility of further classifying images for other ecologically relevant information such as color morphs (indicative of thermal adaptation), host plant or co-located species (indicative of species interactions) and potentially many other factors, so has potentially unlocking a treasure trove of ecological and evolutionary data[63].

Methods

For this work we selected a one-stage deep learning object identification framework, You-Only-Look-Once (YOLO). YOLO is a popular object detection algorithm that has been widely used for detecting ("localizing") objects in images and videos[64–67]. YOLO is computationally more efficient and has relatively higher detection accuracy compared to other CNN-based models[64,65,68]. In particular, it has been used successfully in detecting small objects, such as insects, which are challenging to detect with other methods[64,69,70].

To effectively apply computer vision techniques like YOLO for object detection, it is essential to use accurately labeled and classified photographs for model training. Incorrectly annotated photos may lead to a decrease in model accuracy. Here, we relied on our expert knowledge of monarchs to identify and classify caterpillar photographs by instar, which were then used as the training data (details in the Discussion section), (specifically co-authors Oberhauser, Ries, and Neupane; details in the Discussion section). Oberhauser in particular has trained volunteers to identify monarch caterpillars to instars in the field since 1999[36].

The photos used in this study were obtained from the citizen science-based program iNaturalist. We downloaded 2,562 photographs taken within the United States and tagged by citizen scientists as monarch caterpillars (Fig. 2). Most photos contained only one caterpillar, but some featured multiple caterpillars, often in multiple developmental stages. The caterpillars were mostly present in natural settings such as plant leaves, though some were photographed in artificially reared laboratory backgrounds. Image quality and resolution varied depending upon the camera used and exposure details. To ensure the accuracy of the labeling of the photos, we relied on our expert team members and followed a standard protocol of categorization (detailed in Supplementary A) to manually categorize instar photos into five categories, L1-L5[72]. Images with caterpillars of multiple species were discarded. Photographs were manually annotated using an online web portal https://www.makesense.ai, enclosing identified caterpillars within tight rectangular bounding boxes and recording the instar (as displayed in Fig. 2). From these annotated photographs, 425 were randomly selected from each category (L1-L5) to create a balanced training set. Of the resulting 2,125 photographs, 1,529 (72%) were used for training, 383 (18%) were used for validation, and 213 (10%) were held out for final testing. All photos used in this study have copyrights that allow sharing for non-commercial purposes and upon request we provide links to all those photographs for further testing and analysis (contact corresponding author).

Model framework

We selected YOLO version 5 for our study. YOLOv5 is a deep learning model which has been steadily improved in recent years[36,64,65]. It has become increasingly popular due to its user-friendliness and its capability to detect small objects accurately and efficiently[73,74]. This makes YOLOv5 particularly suitable for the detection of larvae in their early developmental stages (L1 and L2) which are generally much smaller in the photo frame compared to larger instars.

The YOLOv5 model structure is composed of five components: Input, Backbone, Neck, Head and Prediction. The "Input" component is where the images are supplied as input. The "Backbone" component is a convolutional neural network which generates various image features at different spatial resolutions. The "Neck" component combines the image features and forwards them to the "Head" component. The Head component consumes the features to generate the predicted bounding boxes and the predicted categories. Finally, the "Prediction" component combines the predicted bounding boxes and the predicted categories and outputs the result[75,76].

Our model architecture uses an input image size of 640 × 640 pixels and three color (red, green, and blue) channels, which is resized into four images of size 320 × 320 pixels of the three-color channels. This feeds into a convolutional layer followed by a batch normalization layer and a concatenation layer. The final output includes information about 5 object types, the likelihood of each type, and the position of each object in the image. In other words, the output gives the predicted bounding box and classification categories. For more information on the model architecture, please refer to[77].

YOLOv5 has been shown to be superior to its predecessors in terms of image processing, frames per second (FPS), and its mean average precision (mAP). For example, v5 can reach an FPS speed of up to 140, while v4 can only reach 50 (Yan et al., 2021). While this high frame rate is desirable in some settings (such as live video feeds) it was not crucial in the current work. There are multiple versions of YOLOv5 available, including small (YOLOv5s), medium (YOLOv5m), large (YOLOv5l), and extra-large (YOLOv5x). These versions have an increasing number of model parameters from small to extra-large, offering an increasing ability to model complex image data at the cost of progressively longer training times[75,76]. In this study, all four of these model versions were tested.

Modeling platform

The model code was written in the Python programming language (v3) using the yolov5 library[78] and the Pytorch deep learning framework. All code was run on Google Colab, a cloud service platform commonly used for machine learning research, with data stored on Google Drive. NVIDIA A100 GPUs with CUDA 12.0 were used for the calculations. We trained the model for 256 epochs with a batch size of 32, automatic training data augmentation, and default hyperparameters. The average epoch times were 5, 8, 11, and 17 s for small, medium, large, and extra-large models. After training was complete, the model weights that performed the best on predicting the validation set were evaluated on the hold-out test set, using automated test-time augmentation[79].

Data availability

All the image annotation data link is made publicly available for research purpose. The python code is attached on separate files. For further needs contact corresponding author: nn343@georgetown.edu.

Received: 8 July 2024; Accepted: 31 October 2024
Published online: 07 November 2024

References

1. Eggleton, P. The state of the world's insects. *Annu. Rev. Environ. Resour.* **45** (1), 61–82 (2020).
2. Van Klink, R. et al. Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. *Science*. **368** (6489), 417–420 (2020).
3. Neupane, N., Larsen, E. A. & Ries, L. Ecological forecasts of insect range dynamics: a broad range of taxa include winners and losers under future climate. *Curr. Opin. Insect Sci.*, p.101159. (2024).
4. Wagner, D. L., Grames, E. M., Forister, M. L., Berenbaum, M. R. & Stopak, D. Insect decline in the Anthropocene: Death by a thousand cuts. *Proceedings of the National Academy of Sciences*, *118*(2), p.e2023989118. (2021).
5. Buckley, L. B. et al. Can mechanism inform species' distribution models? *Ecol. Lett.* **13** (8), 1041–1054 (2010).
6. McMahon, D. P. & Hayward, A. Why grow up? A perspective on insect strategies to avoid metamorphosis. *Ecol. Entomol.* **41** (5), 505–515 (2016).
7. Gustafsson, K. M., Wolf, S. A. & Agrawal, A. A. Science-Policy-Practice interfaces: Emergent knowledge and monarch butterfly conservation. *Environ. Policy Gov.* **27** (6), 521–533 (2017).
8. International Union for Conservation of Nature. Migratory monarch butterfly now Endangered—IUCN RedList. (2022). https://www.iucn.org/press-release/202207/migratory-monarch-butterfly-now-endangered-iucn-red-list
9. Shirey, V. & Ries, L. Population dynamics and drivers of the eastern monarch (Danaus plexippus) across its full annual cycle: a cross-scale synthesis of a model migratory species. *Curr. Opin. Insect Sci.*, p.101132. (2023).
10. Schultz, C. B., Brown, L. M., Pelton, E. & Crone, E. E. Citizen science monitoring demonstrates dramatic declines of monarch butterflies in western North America. *Biol. Conserv.* **214**, 343–346 (2017).
11. Brower, L. P. Understanding and misunderstanding the migration of the monarch butterfly (Nymphalidae) in North America: 1857–1995. (1995).
12. Hartmann, M. C. et al. A text and image analysis workflow using citizen science data to extract relevant social media records: Combining red kite observations from Flickr, eBird and iNaturalist. *Ecological Informatics*, *71*, p.101782. (2022).
13. Amarathunga, D. C., Grundy, J., Parry, H. & Dorin, A. Methods of insect image capture and classification: A Systematic literature review. *Smart Agricultural Technology*, *1*, p.100023. (2021).
14. Waagmeester, A., Mietchen, D., Leachman, S. & Groom, Q. Using crowd-curation to improve taxon annotations on the Wikimedia infrastructure. *Biodivers. Inform. Sci. Stand.* (2019).
15. Lürig, M. D., Donoughe, S., Svensson, E. I., Porto, A. & Tsuboi, M. Computer vision, machine learning, and the promise of phenomics in ecology and evolutionary biology. *Front. Ecol. Evol.* **9**, 642774 (2021).

16. Archana, R. & Jeevaraj, P. E. Deep learning models for digital image processing: a review. *Artificial Intelligence Review*, *57*(1), p.11. (2024).
17. Kühl, H. S. & Burghardt, T. Animal biometrics: quantifying and detecting phenotypic appearance. *Trends Ecol. Evol.* **28**, 432–441. 10.1016/ j.tree.2013.02.013 (2013).
18. Wagner, D. L. *Caterpillars of Eastern North America: A Guide to Identification and Natural History* (Princeton University Press, 2010).
19. Di Cecco, G. J. & Hurlbert, A. H. Caterpillar patterns in space and time: insights from and contrasts between two citizen science datasets. In Caterpillars in the Middle: Tritrophic Interactions in a Changing World (541–556). Cham: Springer International Publishing. (2022).
20. Høye, T. T. et al. Deep learning and computer vision will transform entomology. *Proceedings of the National Academy of Sciences*, *118*(2), p.e2002545117 (2021).
21. Yasmin, R., Das, A., Rozario, L. J. & Islam, M. E. Butterfly Detection and Classification Techniques: A Review. *Intelligent Systems with Applications*, p.200214. (2023).
22. Almryad, A. S. & Kutucu, H. Automatic identification for field butterflies by convolutional neural networks. *Eng. Sci. Technol. Int. J.* **23** (1), 189–195 (2020).
23. Theivaprakasham, H. Identification of Indian butterflies using deep convolutional neural network. *J. Asia. Pac. Entomol.* **24** (1), 329–340 (2021).
24. Rajeena, P. P. et al. M. and A Novel Method for the Classification of Butterfly Species Using Pre-Trained CNN Models. *Electronics*, *11*(13), p.2016. (2022).
25. Arzar, N. N. K. et al. June. Butterfly species identification using convolutional neural network (CNN). In *2019 IEEE international conference on automatic control and intelligent systems (I2CACIS)* (pp. 221–224). IEEE. (2019).
26. Fauzi, F., Permanasari, A. E. & Setiawan, N. A. July. Butterfly Image Classification Using Convolutional Neural Network (CNN). In *2021 3rd International Conference on Electronics Representation and Algorithm (ICERA)* (pp. 66–70). IEEE. (2021).
27. Karar, M. E., Alsunaydi, F., Albusaymi, S. & Alotaibi, S. A new mobile application of agricultural pests recognition using deep learning in cloud computing system. *Alexandria Eng. J.* **60** (5), 4423–4432 (2021).
28. Kittichai, V. et al. Deep learning approaches for challenging species and gender identification of mosquito vectors. *Scientific reports*, *11*(1), pp.1–14. (2021).
29. Legaspi, K. R. B., Sison, N. W. S. & Villaverde, J. F. March. Detection and Classification of Whiteflies and Fruit Flies Using YOLO. In *2021 13th International Conference on Computer and Automation Engineering (ICCAE)* (pp. 1–4). IEEE. (2021).
30. Venegas, P. et al. Automatic ladybird beetle detection using deep-learning models. *Plos One*. **16** (6), e0253027 (2021).
31. Black, E. N., Blair, J. D., Van der Burg, K. R. & Marshall, K. E. Crowd-sourced observations of a polyphagous moth reveal evidence of allochronic speciation varying along a latitudinal gradient. *Plos One*. **18** (7), e0288415 (2023).
32. Padilla, R., Passos, W. L., Dias, T. L., Netto, S. L. & Da Silva, E. A. A comparative analysis of object detection metrics with a companion open-source toolkit. *Electronics*, *10*(3), p.279. (2021).
33. Hu, Z. et al. Research on Identification Technology of Field Pests with Protective Color Characteristics. *Applied Sciences*, *12*(8), p.3810. (2022).
34. Bejani, M. M. & Ghatee, M. A systematic review on overfitting control in shallow and deep neural networks. *Artif. Intell. Rev.* **54** (8), 6391–6438 (2021).
35. Jiang, P., Chen, Y., Liu, B., He, D. & Liang, C. Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks. *IEEE Access.* **7**, 59069–59080 (2019).
36. Jocher, G. Yolov5. Git Code. Available online: (2021). https://github.com/ultralytics/yolov5 (accessed on December 2022).
37. Koziarski, M. & Cyganek, B. Impact of low resolution on image recognition with deep neural networks: an experimental study. *Int. J. Appl. Math. Comput. Sci.* **28** (4), 735–744 (2018).
38. Xin, D., Chen, Y. W. & Li, J. Fine-grained butterfly classification in ecological images using squeeze-and-excitation and spatial attention modules. *Applied Sciences*, *10*(5), p.1681. (2020).
39. Johari, S. N. A. M. et al. Identification of bagworm (Metisa plana) instar stages using hyperspectral imaging and machine learning techniques. Computers and Electronics in Agriculture, 194, p.106739. (2022).
40. Johari, S. N. A. M. et al. Automatic classification of bagworm, Metisa plana (Walker) instar stages using a transfer learning-based framework. Agriculture, 13(2), p.442. (2023).
41. de Castro Pereira, R., Hirose, E., de Carvalho, O. L. F., da Costa, R. M. & Borges, D. L. Detection and classification of whiteflies and development stages on soybean leaves images using an improved deep learning strategy. *Computers and Electronics in Agriculture*, *199*, p.107132. (2022).
42. Kiobia, D. O. et al. A review of successes and impeding challenges of IoT-based insect pest detection systems for estimating agroecosystem health and productivity of cotton. Sensors, 23(8), p.4127. (2023).
43. Li, W. et al. Classification and detection of insects from field images using deep learning for smart pest management: A systematic review. *Ecological Informatics*, *66*, p.101460. (2021).
44. Chen, J. et al. Automatic segmentation and counting of aphid nymphs on leaves using convolutional neural networks. *Agronomy*, *8*(8), p.129. (2018).
45. Oberhauser, K. et al. A trans-national monarch butterfly population model and implications for regional conservation priorities. *Ecol. Entomol.* **42** (1), 51–60 (2017).
46. Ries, L. & Oberhauser, K. A citizen army for science: quantifying the contributions of citizen scientists to our understanding of monarch butterfly biology. *BioScience*. **65** (4), 419–430 (2015).
47. Neupane, N., Zipkin, E. F., Saunders, S. P. & Ries, L. Grappling with uncertainty in ecological projections: a case study using the migratory monarch butterfly. *Ecosphere*. **13** (1), e03874 (2022).
48. Zylstra, E. R. et al. Changes in climate drive recent monarch butterfly dynamics. *Nat. Ecol. Evol.* **5** (10), 1441–1452 (2021).
49. Saunders, S. P. et al. Multiscale seasonal factors drive the size of winter monarch colonies. Proceedings of the National Academy of Sciences, 116(17), pp.8609–8614. (2019).
50. Van Dyck, L. E., Kwitt, R., Denzler, S. J. & Gruber, W. R. Comparing object recognition in humans and deep convolutional neural networks—an eye tracking study. *Frontiers in Neuroscience*, *15*, p.750639. (2021).
51. Liu, G., Zhang, J., Chan, A. B. & Hsiao, J. H. Human attention-guided explainable artificial intelligence for computer vision models. *Neural Netw.*, p.106392. (2024).
52. Ortiz, A. S., Miyatake, M. N., Tünnermann, H., Teramoto, T. & Shouno, H. Mosquito larva classification based on a convolution neural network. In Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA) (pp. 320–325). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp). (2018).
53. Fuad, M. A. M. et al. Training of convolutional neural network using transfer learning for Aedes Aegypti larvae. *TELKOMNIKA (Telecommunication Comput. Electron. Control).* **16** (4), 1894–1900 (2018).
54. Asmai, S., Zukhairin, M. N. D. M., Jaya, A., Rahman, A. F. N. A. & Abas, Z. Mosquito larvae detection using deep learning. *Int. J. Innovative Technol. Exploring Eng. (IJITEE)*. **8** (12), 804–809 (2019).
55. Shuhan, L. U. & Ye, S. J. Using an image segmentation and support vector machine method for identifying two Locust species and instars. *J. Integr. Agric.* **19** (5), 1301–1313 (2020).

56. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. & Torralba, A. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2921–2929). (2016).
57. Heffernan, J. B. et al. Macrosystems ecology: understanding ecological patterns and processes at continental scales. *Front. Ecol. Environ.* **12** (1), 5–14 (2014).
58. Hinks, A. E. et al. Scale-dependent phenological synchrony between songbirds and their caterpillar food source. *Am. Nat.* **186** (1), 84–97 (2015).
59. Tallamy, D. W. & Shriver, W. G. Are declines in insects and insectivorous birds related? *Condor*. **123** (1), duaa059 (2021).
60. Cheng, Z. et al. A Lightweight Crop Pest Detection Method Based on Convolutional Neural Networks. *Applied Sciences*, *12*(15), p.7378. (2022).
61. Zalucki, M. P. Temperature and rate of development in Danaus plexippus L. and D. Chrysippus L.(Lepidoptera: Nymphalidae). *Australian J. Entomol.* **21** (4), 241–246 (1982).
62. Van Horn, G. et al. The inaturalist species classification and detection dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8769–8778). (2018).
63. Aristeidou, M. et al. How do young community and citizen science volunteers support scientific research on biodiversity? The case of inaturalist. *Diversity*, *13*(7), p.318. (2021).
64. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779–788). (2016).
65. Redmon, J. & Farhadi, A. YOLO9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7263–7271). (2017).
66. Sultana, F., Sufian, A. & Dutta, P. A review of object detection models based on convolutional neural network. *Intelligent computing: image processing-based applications*, pp.1–16. (2020).
67. Zhiqiang, W. & Jun, L. July. A review of object detection based on convolutional neural network. In *2017 36th Chinese control conference (CCC)* (pp. 11104–11109). IEEE. (2017).
68. Du, J. April. Understanding of object detection based on CNN family and YOLO. In *Journal of Physics: Conference Series* (Vol. 1004, No. 1, p. 012029). IOP Publishing. (2018).
69. Zhong, Y., Gao, J., Lei, Q. & Zhou, Y. A vision-based counting and recognition system for flying insects in intelligent agriculture. *Sensors*, *18*(5), p.1489. (2018).
70. Liu, J. & Wang, X. Tomato diseases and pests detection based on improved Yolo V3 convolutional neural network. *Front. Plant Sci.* **11**, 898 (2020).
71. Oberhauser, K. S. Overview: monarchs and people: Model programs for citizen science, education, and conservation in (eds Oberhauser, K. S., Nail, K. R. & Altizer, S. M.) Monarchs in a Changing World: Biology and Conservation of an Iconic Insect. Ithaca, USA: Cornell University Press. (2015).
72. Oberhauser, K. & Kuda, K. *A Field Guide to Monarch Caterpillars* (Danaus plexippus), 1997).
73. Jiang, P., Ergu, D., Liu, F., Cai, Y. & Ma, B. A review of Yolo algorithm developments. *Procedia Comput. Sci.* **199**, 1066–1073 (2022).
74. Yun, W., Kumar, J. P., Lee, S., Kim, D. S. & Cho, B. K. Deep learning-based system development for black pine bast scale detection. *Sci. Rep*. *12*(1), p.606. (2022).
75. Chen, Z. et al. Plant Disease Recognition Model Based on Improved YOLOv5. *Agronomy*, *12*(2), p.365. (2022).
76. Francies, M. L., Ata, M. M. & Mohamed, M. A. A robust multiclass 3D object recognition based on modern YOLO deep learning algorithms. *Concurrency Computation Pract. Experience*. **34** (1), e6517 (2022).
77. Yan, B., Fan, P., Lei, X., Liu, Z. & Yang, F. A real-time apple targets detection method for picking robot based on improved YOLOv5. *Remote Sensing*, *13*(9), p.1619. (2021).
78. Jocher, G. et al. ultralytics/yolov5: v4. 0-nn. SiLU () activations, Weights & Biases logging, PyTorch Hub integration. *Zenodo*. (2021).
79. Moshkov, N., Mathe, B., Kertesz-Farkas, A., Hollandi, R. & Horvath, P. Test-time augmentation for deep learning-based cell segmentation on microscopy images. *Sci. Rep*. *10*(1), p.5068. (2020).

Acknowledgements
This work was supported by the National Science Foundation #DEB-2017791 and Georgetown Environment Initiative/ Georgetown University. We acknowledge the amateur naturalist for the larval photo from iNaturalist.org and thank the platform for making it available for research.

Author contributions
N. N., K. O., L. R, C. M. C.: Conceptualization, Analysis, Writing; R. G., K. H., N. N., C. M. C. Data preparation and coding. N. N., K. O., L. R. Annotation and identification. All authors contributed to the paper and approved the submitted version.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-78509-w.

Correspondence and requests for materials should be addressed to N.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.