

RESEARCH

Open Access



Arthropod Taxonomy Orders Object Detection in ArTaxOr dataset using YOLOX

Fatma M. A. Mazen^{1*}

*Correspondence:
fma04@fayoum.edu.eg

¹ Electronics and Communication
Department, Faculty
of Engineering, Fayoum
University, Fayoum, Egypt

Abstract

The detection and classification of insect species represent challenging computer vision tasks that have significant applications in zoology and agriculture. Fortunately, biologists and taxonomists have developed a systematic approach to organizing organisms, which results in a hierarchical classification system. Insect classification employs a hierarchical structure that includes object detection at the order level, family classification, and species classification. However, the conventional insect identification method is time-consuming and requires the expertise of highly skilled taxonomists to identify insects accurately based on morphological characteristics. This paper presents a pioneering study on the automatic detection and classification of Arthropod Taxonomy Orders using an enhanced variant of the You Only Look Once (YOLOX) framework along with the Arthropod Taxonomy Orders Object Detection (ArTaxOr) Dataset. The proposed ArTaxOr dataset encompasses diverse arthropod species such as insects, spiders, crustaceans, centipedes, millipedes, and isopods. Moreover, some images within this dataset depict multiple species with varying sizes, shapes, and colors. Accordingly, all images are resized to 640 × 640 to ensure compliance with the requisite input image size for YOLOX. Further, mosaic augmentation is employed to enhance the model's accuracy in recognizing small objects. In spite of natural complexity in a majority of dataset images, the proposed YOLOX-based model attained superior mean average precision. The outcomes of this study could act as a standard against which forthcoming research in this domain could be compared or judged.

Keywords: Arthropods, Mosaic augmentation, YOLOX, Insect classification, Roboflow, Object detection

Introduction

Insects exhibit remarkable levels of diversity, abundance, distribution, and adaptability, rendering them a subject of significant interest within the field of biology. Insect recognition forms the fundamental basis for the study of insects and the management of pest populations. Nevertheless, existing research on insect recognition predominantly relies on the expertise of a limited number of taxonomic specialists to identify insects accurately based on morphological characteristics. Given the rapid progress in computer technology, there is a promising opportunity to employ computational methods for precise insect differentiation, thereby potentially supplanting the need for human

professionals in this domain. Arthropods play a pivotal ecological role and offer various advantages to humanity, encompassing their use in weed, harmful fungi, and bacteria control. Essential arthropod species such as bees, wasps, ants, butterflies, moths, flies, and beetles facilitate pollination, enabling the reproduction of numerous plant species. Pests, in turn, contribute vital nutrients to the soil and plants, which are subsequently transferred to humans and animals upon consuming these plants. Moreover, arthropods provide a multitude of human-produced goods. Bees produce honey and beeswax, caterpillars generate silk for cocoon protection, and spider webs are utilized in the production of fishing nets and surgical sutures. Additionally, many arthropods, including crabs, lobsters, shrimp, prawns, and crayfish, serve as food sources for human consumption. Consequently, the automated detection and classification of arthropods in their natural habitats hold significant importance for effective pest management, yielding profound economic implications. The scientific classification system employed by researchers follows a hierarchical structure, beginning with a broad category and progressing to increasingly specific categories. In the realm of biology, organisms are classified based on their kingdom, family, phylum, class, order, genus, and species. Notably, the phylum Arthropoda encompasses approximately 80% of all known animal species, making it the largest phylum within the animal kingdom [1].

Arthropods, being the only invertebrates capable of flight, represent the most abundant and diverse group of animals. They comprise over 1.3 million described species [1]. However, detecting and classifying arthropods at the order level presents challenges due to the varying sizes, shapes, and colors exhibited by objects within the same class. Object detection, a fundamental task in computer vision, involves localizing and classifying regions of interest (ROIs) by assigning them rectangular bounding boxes that indicate the confidence of their presence. Object detection serves as a cornerstone for applications such as object tracking, landmark detection, autonomous driving, and image segmentation. Over time, object detection has become a widely researched field, categorized broadly into two types: single-stage detectors and two-stage detectors.

Two-stage detectors, exemplified by the region-based convolutional neural network (R-CNN), employ an initial stage known as the region proposal network (RPN) to identify potential areas of interest and approximate bounding boxes within the image. The subsequent stage network then determines the class and refines the bounding box using the local features proposed by the RPN. In contrast, single-stage detectors like YOLO [2] perform object detection through fixed-grid regression. However, such detectors often struggle as a significant portion of grid cells and anchors tend to focus on the background rather than the actual objects of interest, thereby limiting the learning capabilities of the convolutional neural network (CNN).

To address the limitations of previous YOLO versions, YOLOX [3] introduces improvements such as the elimination of box anchors, resulting in enhanced inference speed and reduced computational cost. Furthermore, the YOLOX algorithm adopts a technique that separates the YOLO detection head into disassociated feature channels, enabling independent regression of box coordinates and object classification. This approach facilitates faster convergence rates and improved model accuracy.

The subsequent sections of the paper are organized as follows: the “[Related work](#)” section provides an overview of related work concerning the detection and classification of

insect species. The “[The dataset](#)” section introduces the Arthropod Taxonomy Orders Object Detection dataset. The “[Methods](#)” section presents the proposed YOLOX-based model as a methodology for object detection encompassing both localization and classification. Detailed experimental results and discussions are presented in the “[Results and discussion](#)” section. Finally, the “[Conclusions](#)” section concludes the paper by summarizing the study and highlighting potential directions for future research.

Related work

The field of object detection has seen significant advancements in various domains such as medical, industrial, and agricultural fields. Several models have been developed by computer architecture engineers, including YOLOv3 [4], Mask R-CNN [5, 6], YOLOv4 [7], YOLOv5 [8], and YOLOX [3]. In their study, Zhong, Gao, Lei, and Zhou [9] implemented an insect counting and recognition system using Raspberry PI as the platform. To capture real-time images of flying insects, a yellow sticky trap was installed, and a camera was utilized for data collection. The YOLO architecture was employed for object detection, while support vector machine (SVM) was used for classification purposes. The researchers evaluated the performance of their system by identifying six species of flying insects, namely chafer, mosquito, bee, fly, fruit fly, and moth. The obtained results exhibited an average counting accuracy of 92.50% and an average classification accuracy of 90.18% on the Raspberry PI platform, which is a promising outcome. In a separate study conducted by Cho et al. [10], an automatic identification system was developed for selected pest insects found in a greenhouse environment, specifically Aphids, Whiteflies, and Thrips. The researchers utilized a yellow sticky trap as a means to gather relevant data for analysis. Size and color components were employed as distinguishing features to classify the different insect classes. The experimental results demonstrated an average accuracy rate of 90.54% for Whitefly, 92.73% for Aphid, and 88.9% for Thrips, indicating the effectiveness of the proposed system in accurately identifying these pest insects in the greenhouse setting. In the study conducted by Kaya Y and Kayci L [11], an automated system for identifying butterfly species was presented. The system relied on the utilization of artificial neural networks (ANN). The researchers employed the grey level co-occurrence matrix (GLCM) technique to extract texture features using different angles and distances. The results of their experiment indicated a high accuracy rate of 92.85%. In their research paper, K. Li, J. Zhu, and N. Li [12] introduced a refined iteration of the YOLOv3 model to develop an automated system for insect detection and counting. They employed CSPDarknet-53 as the primary feature extraction network. Furthermore, to enhance the precision of network predictions, they employed the combined intersection ratio (CIoU) as the regression loss function. The enhanced YOLOv3 model demonstrated a notable accuracy rate of 90.62%, surpassing the original YOLOv3 model by a margin of 3%. Takimoto et al. [13] proposed a two-stage methodology to effectively detect and classify two specific species of flea beetles, namely *P. striolata* and *P. atra*, along with background objects in the field. Initially, they applied data augmentation techniques to expand the training dataset, encompassing rotational transformations, the addition of noise, cropping, flipping, scaling, and color transformations. Subsequently, they employed the YOLOv4 model as a single-stage approach for detection, yielding a precision score of 0.55. To further enhance model performance while maintaining efficiency, they integrated the YOLOv4 model as a region proposal network coupled

with EfficientNet as a classifier. As a result of this hybrid approach, they achieved a significantly improved precision rate of 89%. The research conducted by [14] focused on the intricate task of insect identification and detection in outdoor images characterized by intricate backgrounds. In order to address the challenges posed by these backgrounds, the researchers employed a deep learning-based model to achieve multi-class object detection. Additionally, they introduced a novel approach that utilized a clustering algorithm for anchor box estimation, as opposed to relying on pre-defined anchor boxes. This approach resulted in improved precision and speed of the model. The effectiveness of the proposed method was successfully demonstrated through rigorous evaluation of a dataset consisting of insect images captured in natural environments. The authors of [15] conducted an extensive review that encompassed a wide array of techniques and the present state-of-the-art implementation of sensors employed for the purpose of automatic detection and monitoring of insect pests. Their scholarly publication placed particular emphasis on techniques that have proven to be effective in pest identification through the utilization of automatic traps, infrared sensors, audio sensors, and image-based classification. The review shed light on the diverse spectrum of available systems, showcased illustrative applications, and highlighted recent advancements such as machine learning and the Internet of Things, thereby providing a comprehensive overview of the subject matter. In the domain of security and surveillance, Rajagopal, B.G. [16] devised an Intelligent Surveillance system specifically designed for the purpose of vehicle detection and classification using real-time video recordings from road traffic. The primary objective of this system was to enhance vehicle safety and monitoring in challenging nighttime conditions and various weather scenarios such as rain, daytime, and nighttime. Additionally, the proposed system exhibits the capability to dynamically select the appropriate algorithm based on the prevailing weather conditions. The vehicle count and classification algorithm employed in this system incorporates image segmentation using a Laplacian of Gaussian edge detector (LoG), morphological filtering of edge map objects, and the categorization of vehicles into small, medium, and large sizes. A noteworthy advantage of this approach, in comparison to motion detection-based methods, is its applicability to both rapidly changing and static traffic scenarios. The proposed system achieved average classification and detection accuracies of 89.4% and 96.0% respectively, for rapidly changing traffic, while achieving accuracies of 83.8% and 82.1% respectively, for slow-moving traffic. In the realm of the manufacturing industry, the monitoring of industrial components holds immense significance. Sureshkumar, S., Mathan, G., RI, P. et al. [17] devised a computer vision-based system with the objective of detecting and classifying industrial components in an assembly line. The researchers conducted a thorough performance evaluation of three distinct object detection models, namely the faster R-CNN, single-shot detector (SSD), and YOLO. The experimental findings showcased the effectiveness of employing pre-processing techniques such as contrast enhancement, gamma correction, and canny-edge detection in augmenting the detection accuracy of the model. Leveraging the YOLOv4 model, the researchers achieved a commendable mean average precision (mAP) value of 0.95. Magnetic resonance imaging (MRI) has emerged as the preferred modality within the medical imaging domain for accurately assessing the severity of knee injuries. Nonetheless, the process of evaluating knee MRIs is time-consuming and susceptible to diagnostic errors, leading to an excessive number of unnecessary surgical interventions. In an endeavor to mitigate these challenges, Gupta, S., Pawar, P.M., and

Tamizharasan, P.S. [18] devised a deep learning-based framework for effectively classifying knee injuries into three distinct categories: meniscal tear, anterior cruciate ligament (ACL) tear, and abnormality. The researchers evaluated multiple deep learning models, including VGG19, VGG16, ResNet152V2, InceptionV3, and DenseNet201, and determined that the ResNet152V2-based model exhibited the highest accuracy rate of 78.33%. In a separate study [19], Sachar, S., and Kumar, A. developed a system grounded in transfer learning principles, with the aim of automating the classification of medical leaf images. The researchers conducted extensive training and evaluation procedures on the medicinal leaf dataset, which encompasses a comprehensive array of 30 distinct classes. To enhance the classification performance, the researchers proposed an ensemble learning approach that combines the predictive outputs of three component models, namely InceptionV3, MobileNetV2, and ResNet50. Employing threefold and fivefold cross-validation techniques, the Ensemble Deep Learning- Automatic Medicinal Leaf Identification (EDL-AMLI) classifier attained an exceptional accuracy of 99.66% on the test set, with an overall accuracy of 99.9%.

The dataset

This research paper utilizes the ArTaxOr dataset [20], which comprises arthropod images in JPEG format accompanied by object bounding boxes in JSON format. To prepare the dataset for analysis, the researchers employed Roboflow [21] to convert the annotations into the PASCAL Visual Object Classes (Pascal VOC) format and resize all images to a standardized resolution of 640×640 pixels. Each image contains between one and fifty objects, and the dataset is continuously updated with the addition of new orders on a regular basis. In the current version, the dataset covers seven orders, each containing a minimum of two thousand objects per order, as depicted in Fig. 1. Figure 2 further visualizes the class distribution and the corresponding number of images per class in the initial version of the ArTaxOr dataset, which encompasses a total of 15,374 images. Additionally, Fig. 3 provides insights into the size and aspect ratio distribution of the dataset images, with the purple box indicating the median width and height of an image (2048×1536 pixels).

To enhance the variability of the input data, the researchers propose the use of mosaic augmentation [8]. Figure 4 showcases samples of the mosaic data augmentation strategy, which involves combining multiple training images in specific ratios to enable the model to detect tiny objects effectively. The researchers applied mosaic augmentation to the dataset using Roboflow, resulting in a doubling of the number of dataset images. The augmented dataset, which incorporates mosaic augmentation techniques, consists of a total of 30,736 images. It was subsequently divided into a training set comprising 90% of the images and a validation set comprising the remaining 10%.

Methods

The proposed methodology's flow diagram is illustrated in Fig. 5. It consists of three stages: dataset preprocessing, model training with the ArTaxOr dataset, and model evaluation with the test set. This paper proposes an exceeding YOLOX, one of the most advanced deep learning models for object detection. YOLOX is an anchor-free single-stage object detector that significantly improves training convergence time and model accuracy. YOLOX has eliminated the limitations of earlier YOLO versions through

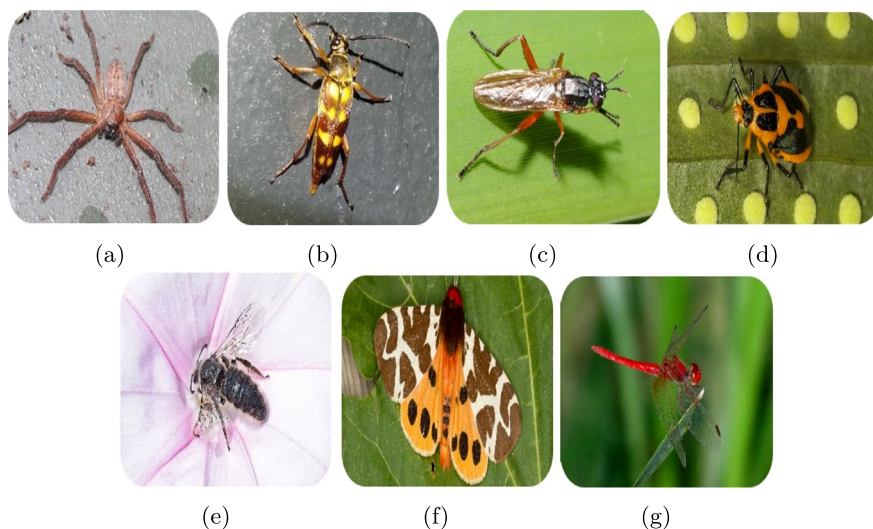


Fig. 1 Seven arthropod orders covered in the dataset. **a** Araneae (spiders), adults, juveniles. **b** Coleoptera (beetles), adults. **c** Diptera (true flies, including mosquitoes, midges, crane fly, etc.), adults. **d** Hemiptera (true bugs, including aphids, cicadas, planthoppers, shield bugs, etc.), adults and nymphs. **e** Hymenoptera (ants, bees, wasps), adults. **f** Lepidoptera (butterflies, moths), adults. **g** Odonata (dragonflies, damselflies), adults

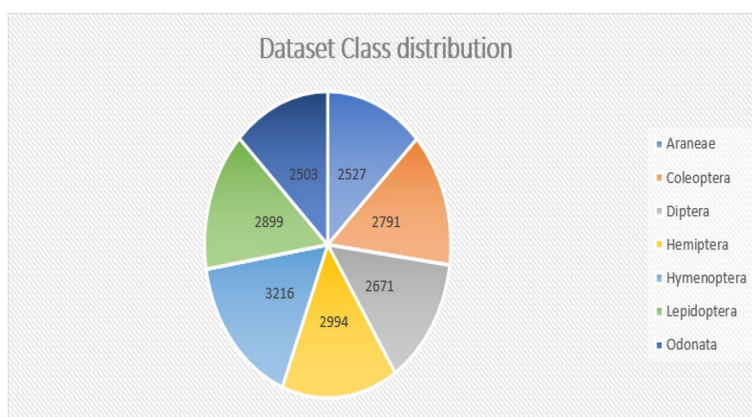


Fig. 2 Class distribution of the ArTaxOr dataset

dropping box anchors, which improve inference speed and computation cost. It also breaks down the YOLO detection head into separate feature channels for box coordinate regression and object classification, leading to faster convergence and higher accuracy, as shown in Fig. 6. The depicted figure serves as a visual representation of the contrasting attributes between the YOLOv3 head and the proposed decoupled head. Notably, for each level of FPN feature, the researchers initially employed a 1×1 convolutional layer to diminish the feature channel to 256. Subsequently, two parallel branches were introduced, with each branch comprising two 3×3 convolutional layers dedicated to the classification and regression tasks, respectively. Additionally, an IoU branch was incorporated within the regression branch. This IoU branch functions to capture Intersection over Union values, a metric central to evaluating the alignment between predicted and ground-truth bounding boxes. The key features of the YOLOX model are as follows:

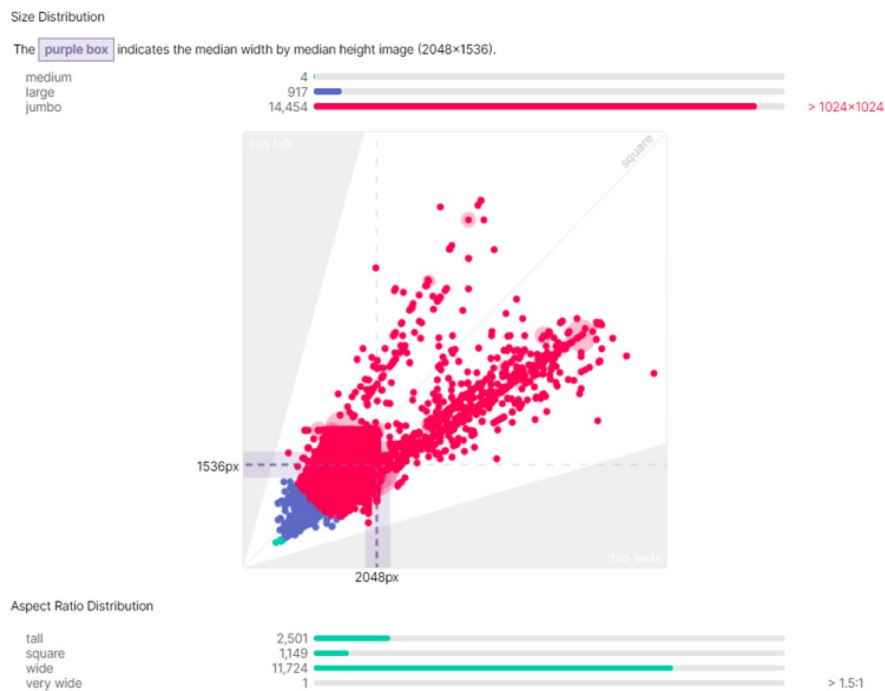


Fig. 3 Size and aspect ratio distribution of dataset images



Fig. 4 Samples of mosaic augmentation

1. Anchor-free design: YOLOX adopts a center-based approach, which eliminates the need for pre-defined boxes as object proposals. Instead, it directly localizes objects using centers or key points. This reduces the number of hyper-parameters and computational requirements associated with anchor-based detectors.
2. Decoupled head: YOLOX implements a decoupled head architecture for classification and regression tasks. This approach uses separate branches with convolutional

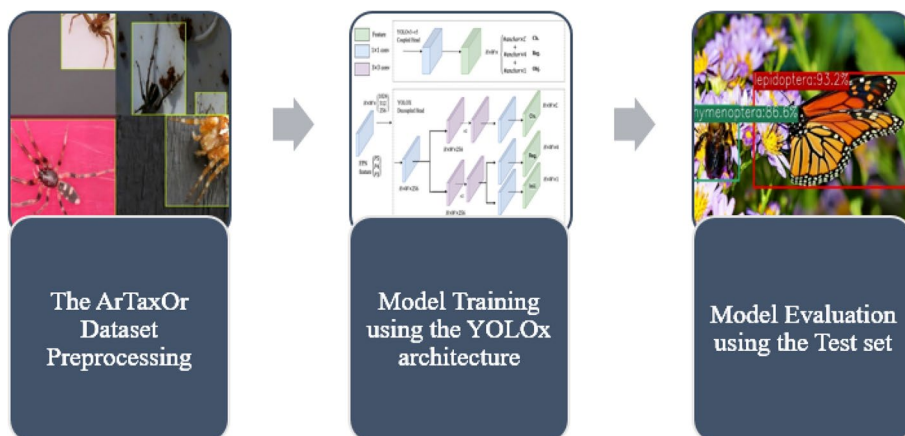


Fig. 5 Flow diagram of the proposed approach

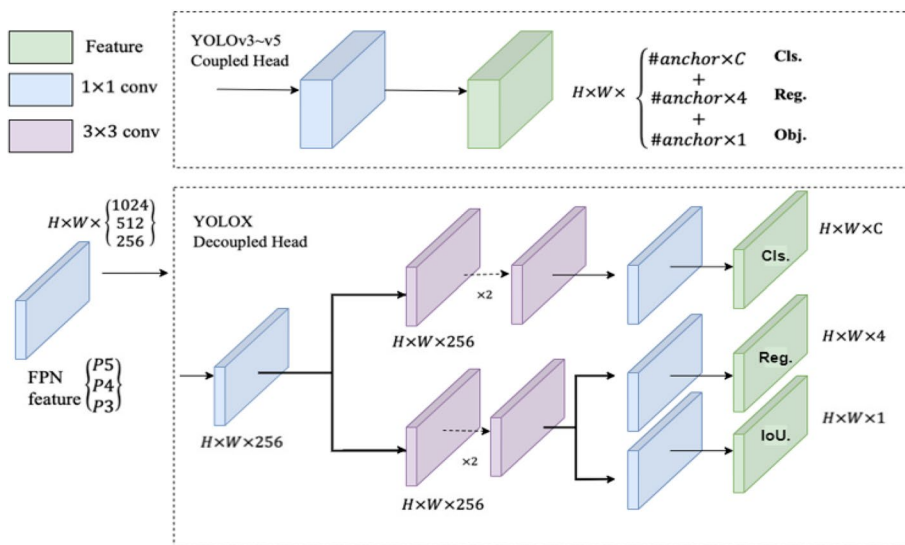


Fig. 6 Difference between YOLOv3 head and the proposed YOLOX decoupled head

layers to improve performance by addressing the misalignment of features between regression and classification as demonstrated in [22].

3. SimOTA label assignment strategy: YOLOX introduces a redesigned optimal transport assignment (OTA) strategy [23] called simOTA. It employs a Dynamic Top K strategy to estimate the number of positive anchors for each ground truth, reducing the number of iterations. This strategy improves the average precision (AP) without increasing training.
4. Advanced augmentations: YOLOX incorporates two advanced augmentation techniques, Mixup and Mosaic. Mixup augmentation involves the weighted addition of two images, while Mosaic augmentation combines four training images into one and crops them in a specific ratio. These augmentations enhance the network’s ability to detect smaller objects.

Overall, YOLOX's salient features include its anchor-free design, decoupled head architecture, simOTA label assignment strategy, and the use of advanced augmentations like Mixup and Mosaic. These design choices and techniques improve the performance and efficiency of object detection models. Due to memory limitations, the YOLOX model has been trained only for fifteen epochs with the mosaic-augmented dataset on the Kaggle NVIDIA TESLA P100 GPU. Training is based on the YOLOX repository by the Megvii Team [24]. The mosaic-augmented dataset includes 30736 images in total. The size of the input image is 640×640. Table 1 summarizes the training details of YOLOX.

Results and discussion

The trained model is applied to perform inference on the Arthropod Taxonomy Orders Object Detection Testset [25]. Figures 7, 8, 9, 10, 11, and 12 present a series of testing images (IMG01-IMG06) alongside the ground truth bounding boxes and class labels on the left-hand side and the predicted bounding boxes and class labels on the right-hand side. These visualizations demonstrate the robustness and substantial classification accuracy achieved by the proposed model. Notably, even though the second object in the IMG05 test image appears blurry, the model successfully detects both objects. Similarly, despite the similar texture and color properties of the two target objects in the IMG06 test image, the model effectively detects and classifies both, despite their close proximity.

Moving on to the "IMG07" test image (Fig. 13), which features five objects of varying sizes, colors, shapes, and classes against a flower-like background, our model successfully detects and assigns appropriate classes to four of them.

In Fig. 14, we encounter the only failure scenario where the suggested model fails to detect the target object. The target item possesses the same color and texture features as the background tree, making it challenging to distinguish. It is important to note that due to hardware limitations, the proposed model has only been trained for fifteen epochs. Further training with additional data and epochs would likely improve the detection performance in such challenging scenarios, leading to a higher classification rate.

To evaluate the quality of the object detection model, mean average precision (*mAP*) is employed. This metric measures the correspondence between the actual

Table 1 Values of YOLOX training parameters

Parameter	Value
num_classes	7
max_epoch	15
depth	0.33
width	0.5
batch size	16
data_num_workers	4
input_size	(640, 640)
scheduler	'yoloxwarmcos'
basic_lr_per_img	0.00015625
weight_decay	0.0005
momentum	0.9
min_lr_ratio	0.05

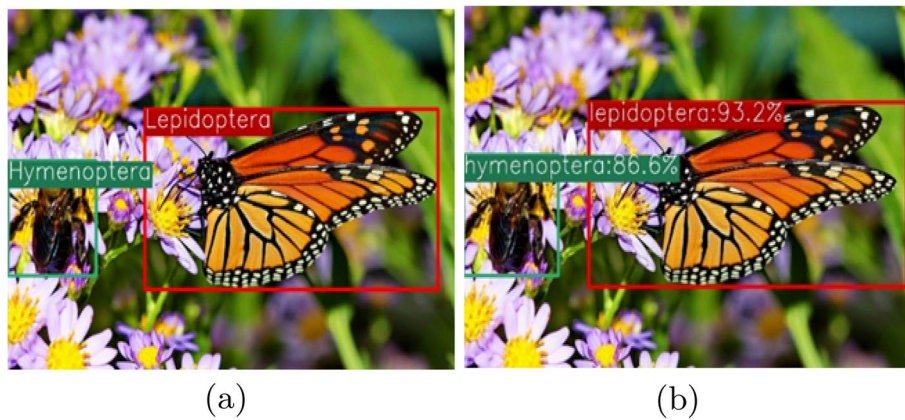


Fig. 7 The ground truth and the predicted bounding boxes and class labels for the IMG01 test image. **a** Ground truth. **b** Predicted

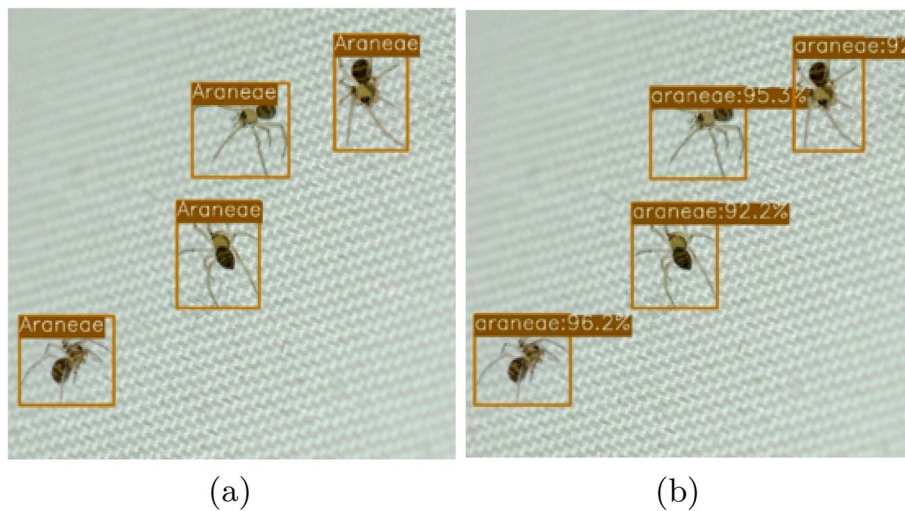


Fig. 8 The ground truth and the predicted bounding boxes and class labels for the IMG02 test image. **a** Ground truth. **b** Predicted

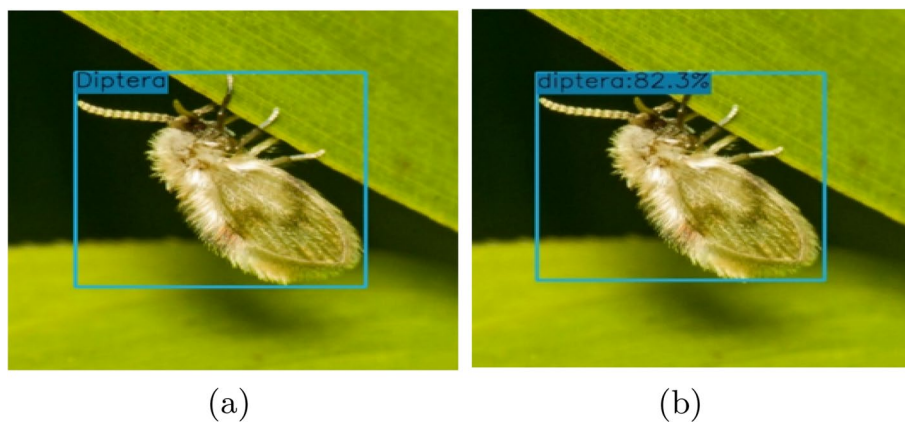


Fig. 9 The ground truth and the predicted bounding boxes and class labels for the IMG03 test image. **a** Ground truth. **b** Predicted

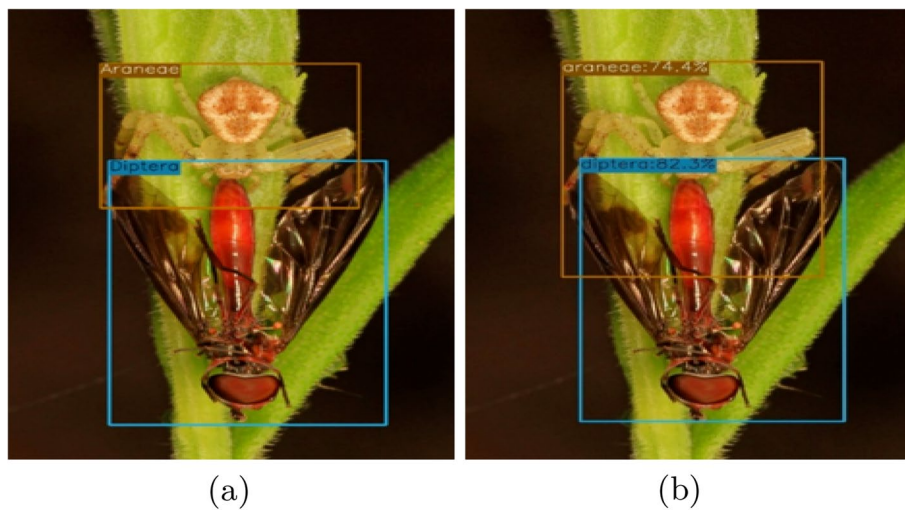


Fig. 10 The ground truth and the predicted bounding boxes and class labels for the IMG04 test image. **a** Ground truth. **b** Predicted

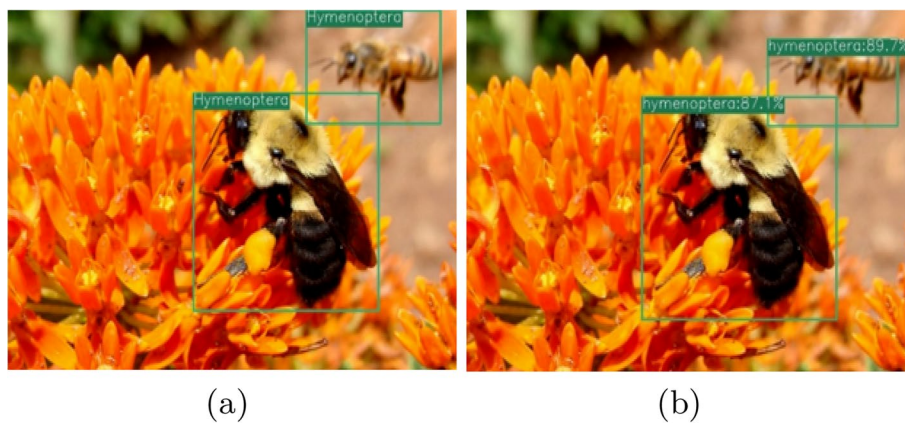


Fig. 11 The ground truth and the predicted bounding boxes and class labels for the IMG05 test image. **a** Ground truth. **b** Predicted

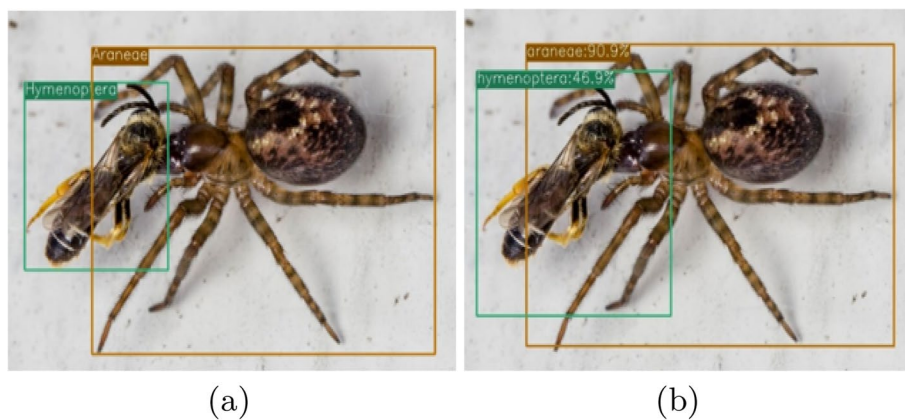


Fig. 12 The ground truth and the predicted bounding boxes and class labels for the IMG06 test image. **a** Ground truth. **b** Predicted

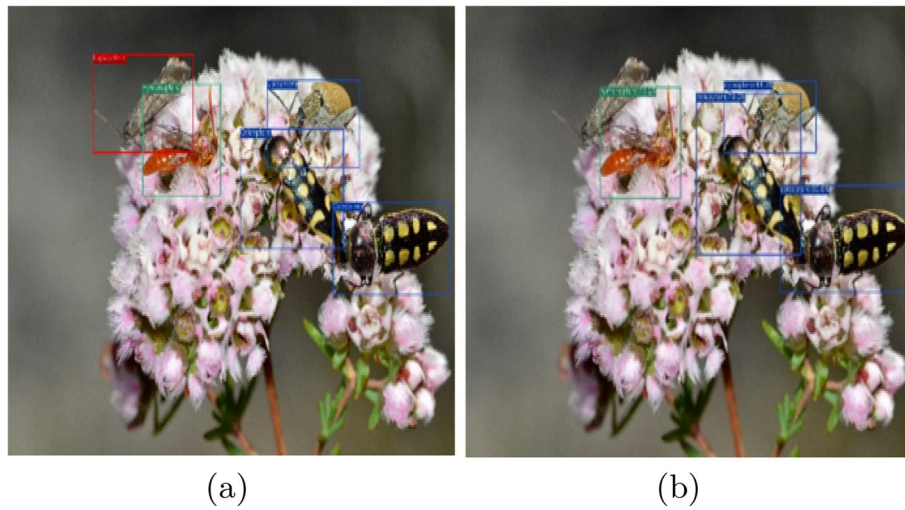


Fig. 13 The ground truth and the predicted bounding boxes and class labels for the IMG07 test image. **a** Ground truth. **b** Predicted

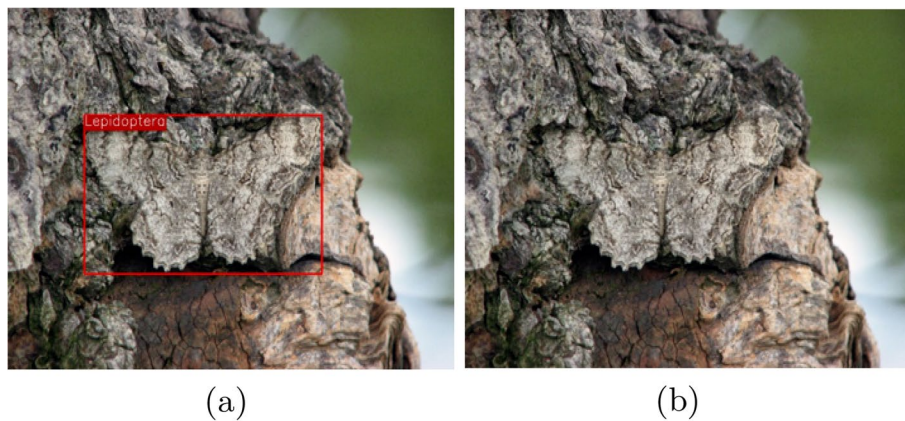


Fig. 14 The ground truth and the predicted bounding boxes and class labels for the IMG08 test image. **a** Ground truth. **b** Predicted

bounding boxes and the predicted bounding boxes, yielding a score that indicates the model’s accuracy in detecting objects. Intersection over Union (*IoU*) is a quantitative measure utilized to determine if a region contains an object or not. *IoU* is computed based on the formulation specified in Eq. 1.

$$IoU = \frac{IntersectionArea}{UnionArea} \tag{1}$$

[[The *IoU* value spans from zero, indicating no overlap between the actual bounding box and the predicted bounding box, to one, indicating that the actual bounding box and the predicted bounding box precisely coincide in terms of their coordinates.

To compute the *mAP*, the average precision (*AP*) is initially calculated for each individual class. Subsequently, the *mAP* is obtained by taking the mean of the *AP* values

across all seven classes. The mathematical expression for mAP in the context of “ n ” classes is defined by Eq. 2.

$$mAP = \frac{1}{n} \sum_{k=1}^n AP_k \quad (2)$$

where AP_k is the AP of class “ k ” and “ n ” is the number of classes. To evaluate our object detector, the AP is computed for each of the seven classes and then averaged across all classes. This provides a comprehensive assessment of the detector’s performance.

In the Pascal VOC challenge, the AP is calculated at a single IoU threshold of 0.5, resulting in the mean average precision at 0.5 IoU ($mAP@50$).

In contrast, the Common Object Context (COCO) challenge considers a range of IoU threshold values. The AP is computed for each IoU threshold within the range of 0.5 to 0.95, with a step size of 0.05. The individual AP values are then averaged to obtain the final mean Average Precision ($mAP@50 : 95$). This approach provides a more comprehensive evaluation by considering varying levels of overlap between the predicted and ground truth bounding boxes. The results of the $mAP@50$ and $mAP@50 : 95$ metrics across epochs are depicted in Fig. 15. The $mAP@50$ metric exhibited an initial value of 61.1% in the first epoch and steadily increased to reach a superior performance of 90% in the final epoch. This achievement of 90% $mAP@50$ is particularly noteworthy considering the challenging nature of the task. On the other hand, the $mAP@50 : 95$ metric commenced with a value of 44.99% in the first epoch and concluded at 75.41%.

The total number of epochs conducted was 15, with each epoch consisting of 3150 steps. Figure 16 illustrates the loss curves in relation to the number of steps. The total loss encompasses the summation of iou_loss , $l1_loss$, $conf_loss$, and cls_loss . For instance, after ten steps, the total_loss amounted to eleven and ultimately converged to a value of 2.1 at the final step.

For improved visualization, Fig. 17 presents the loss curves across epochs. The total_loss decreased to 5.5 after the initial epoch, followed by numerous fluctuations over time. The final value of the total_loss across the last fifteen epochs was 2.1.

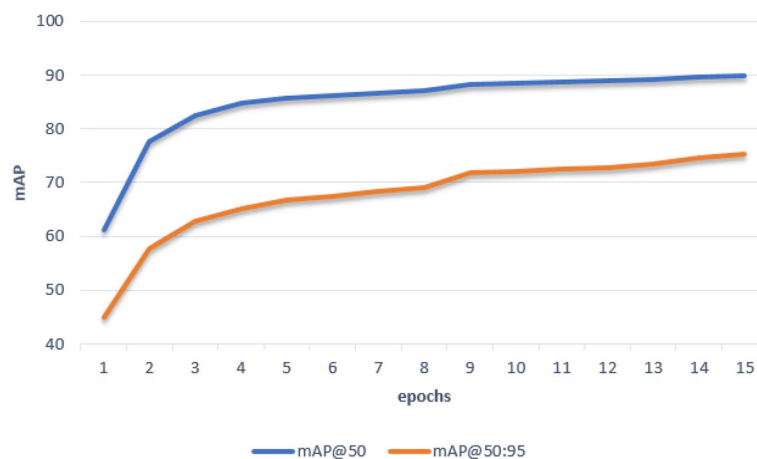


Fig. 15 mAP@50 and mAP@50:95 vs epochs

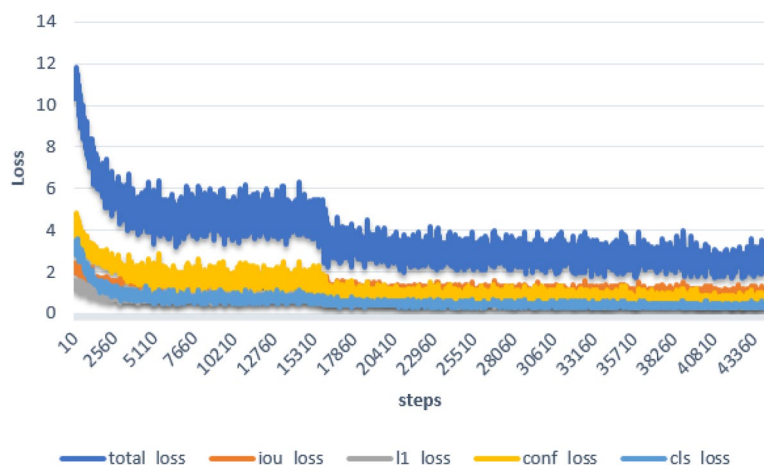


Fig. 16 Loss curves vs. steps

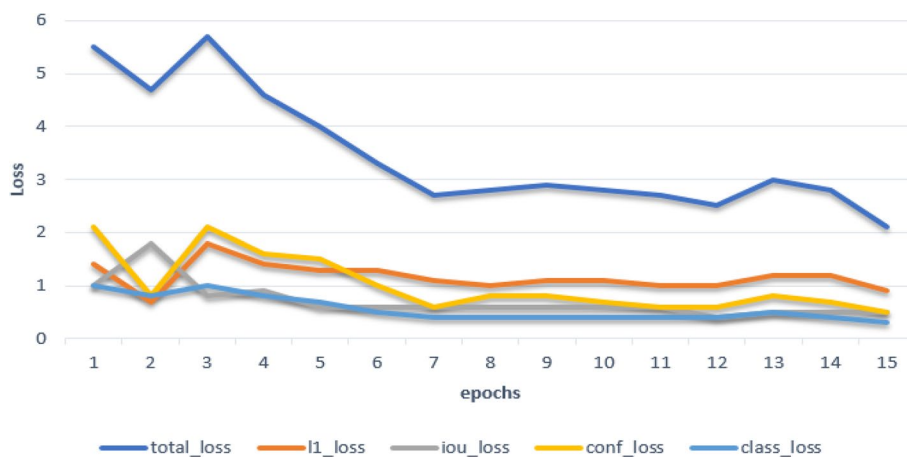


Fig. 17 Loss curves vs. epochs

Conclusions

This research paper introduces an automated system designed to detect and classify Arthropods against complex backgrounds. A modified version of the ArTaxOr dataset, referred to as “Pascal VOC” ArTaxOr, was created using Roboflow to serve as the input dataset for training the YOLOX model. The model was trained on a Kaggle NVIDIA TESLA P100 GPU for fifteen epochs, enabling it to effectively detect Arthropods and classify them into seven distinct classes: Araneae, Coleoptera, Diptera, Hemiptera, Hymenoptera, Lepidoptera, and Odonata.

Experimental results demonstrate that the model achieves a high level of accuracy in recognizing Arthropods within complex environments. The implementation of mosaic data augmentation significantly enhances the model’s recognition performance. It is capable of accurately identifying Arthropods in images captured under diverse and intricate environmental conditions, successfully classifying multiple insect species in a single instance. The performance evaluation of the model is based on the *mAP*, which is calculated as the average precision across all seven classes. The developed model achieves an

outstanding *mAP* of 90% at an *IoU* threshold of 0.5 and an *mAP* of 75% when considering *IoU* values ranging from 0.5 to 0.95. In the future, the proposed model holds the potential for deployment as a real-time mobile application for Arthropod identification and categorization. Its simplicity and accurate recognition capabilities make it a valuable asset for the development of a productive and commercially viable mobile application. This study serves as a significant contribution, showcasing the potential of automated Arthropod detection and classification systems to enhance and streamline the taxonomy process. Additionally, the developed model has the potential to be utilized for effective training on datasets containing harmful insects in insect monitoring devices, thereby mitigating the reliance on pesticides and other potentially hazardous methods of insect control. By leveraging the model's capabilities, alternative and more environmentally friendly approaches can be explored to address the challenges associated with insect management. The findings of this study serve as a benchmark against which future research in this domain can be compared and evaluated.

Abbreviations

YOLOX	Exceeding You Only Look Once
ArTaxOr	Arthropod Taxonomy Orders Object Detection
ROIs	Regions-of-interests
R-CNN	Region-based convolutional neural network
RPN	Region proposal network
SVM	Support vector machine
ANN	Artificial neural networks
GLCM	Grey level co-occurrence matrix
CIOU	Combined intersection ratio
LoG	Laplacian of Gaussian
SSD	Single-shot detector
MRI	Magnetic resonance imaging
mAP	Mean average precision
ACL	Anterior cruciate ligament
EDL-AMLI	Ensemble Deep Learning- Automatic Medicinal Leaf Identification
Pascal VOC	PASCAL Visual Object Classes
IoU	Intersection over Union
AP	Average precision
simOTA	Simplified optimal transport assignment

Acknowledgements

Not applicable.

Authors' contributions

F.M.A confirms that she was responsible for the conception and design of the study, literature review, preprocessing and analysis of data, performing all experiments, interpretation of results, drafting of the manuscript, and critical revision of the manuscript. She also reviewed and approved the final version to be published.

Funding

Not applicable.

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

The author agrees.

Consent for publication

The author agrees.

Competing interests

The author has not received any research grants from any company. The author has not received speaker honorarium from any company. The author does not have any stock for any company. The authors are not members of any committee. The author declares that she has no conflict of interest. The author has no competing interests to declare relevant to the content of this article.

Received: 29 May 2023 Accepted: 10 September 2023

Published online: 30 September 2023

References

1. Encyclopedia britannica (2023). <https://www.britannica.com/animal/arthropod>. Accessed 29 Sept 2023.
2. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. CVPR, pp 779–788
3. Ge Z, Liu S, Wang F, Li Z, Sun J (2021) YOLOX: exceeding YOLO series in 2021. arXiv preprint [arXiv:2107.08430](https://arxiv.org/abs/2107.08430)
4. Adarsh P, Rathi P, Kumar M (2020) YOLO v3-Tiny: object detection and recognition using one stage improved model. In: 2020 6th international conference on advanced computing and communication systems (ICACCS). IEEE, pp 687–694
5. He K, Gkioxari G, Dollár P, Girshick R (2017) Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. IEEE, pp 2961–2969
6. Yu C, Wang J, Peng C, Gao C, Yu G, Sang N (2018) Bisenet: bilateral segmentation network for real-time semantic segmentation. In: Proceedings of the European conference on computer vision (ECCV), arXiv. pp 325–341
7. Wu D, Lv S, Jiang M, Song H (2020) Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. *Comput Electron Agric* 178:105742
8. Ultralytics yolov5 (2022). <https://github.com/ultralytics/yolov5>. Accessed 29 Sept 2023.
9. Zhong Y, Gao J, Lei Q, Zhou Y (2018) A vision-based counting and recognition system for flying insects in intelligent agriculture. *Sensors* 18(5):1489
10. Cho J, Choi J, Qiao M, Ji C, Kim H, Uhm K, Chon T (2007) Automatic identification of whiteflies, aphids and thrips in greenhouse based on image analysis. *Red* 346(246):244
11. Kaya Y, Kayci L (2014) Application of artificial neural network for automatic detection of butterfly species using color and texture features. *Vis Comput* 30:71–79
12. Li K, Zhu J, Li N (2021) Insect detection and counting based on YOLOv3 model. In: 2021 IEEE 4th International Conference on Electronics Technology (ICET). IEEE, pp 1229–1233
13. Takimoto H, Sato Y, Nagano AJ, Shimizu KK, Kanagawa A (2021) Using a two-stage convolutional neural network to rapidly identify tiny herbivorous beetles in the field. *Ecol Inform* 66:101466
14. Pang HW, Yang P, Chen X, Wang Y, Liu CL (2019) Insect recognition under natural scenes using R-FCN with anchor boxes estimation. In: Image and Graphics: 10th International Conference, ICIIG 2019, Beijing, China, August 23–25, 2019, Proceedings, Part I 10. Springer, pp 689–701
15. Cardim Ferreira Lima M, de Almeida Damascena, Leandro ME, Valero C, Pereira Coronel LC, Gonçalves Bazzo CO (2020) Automatic detection and monitoring of insect pests—a review. *Agriculture* 10(5):161
16. Rajagopal BG (2020) Intelligent traffic analysis system for Indian road conditions. *Int J Inf Technol*. 14. <https://doi.org/10.1007/s41870-020-00447-3>.
17. Sureshkumar S, Mathan G, RI P, Govindarajan M (2022) Deep learning framework for component identification. *Int J Inf Technol*. 14. <https://doi.org/10.1007/s41870-022-00895-z>.
18. Gupta S, Pawar PM, Tamizharasan P (2022) Intelligent detection of knee injury in MRI exam. *Int J Inf Technol* 14(4):1815–1821
19. Sachar S, Kumar A (2022) Deep ensemble learning for automatic medicinal leaf identification. *Int J Inf Technol* 14(6):3089–3097
20. Arthropod taxonomy orders object detection dataset. <https://doi.org/10.34740/KAGGLE/DSV/1240192>
21. Dwyer B, Nelson J, Solawetz J, et al (2022) Roboflow (Version 1.0) [Software]. Available from <https://roboflow.com>.
22. Wu Y, Chen Y, Yuan L, Liu Z, Wang L, Li H, Fu Y (2020) Rethinking classification and localization for object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. arXiv, pp 10186–10195
23. Ge Z, Liu S, Li Z, Yoshie O, Sun J (2021) Ota: Optimal transport assignment for object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, pp 303–312
24. Yolox (2021). <https://github.com/Megvii-BaseDetection/YOLOX>. Accessed 29 Sept 2023.
25. Arthropod taxonomy orders object detection testset (2020). <https://www.kaggle.com/datasets/mistag/arthropod-taxonomy-orders-object-detection-testset>. Accessed 29 Sept 2023.

Publisher's Note

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