TenebrioVision: A Fully Annotated Dataset of Tenebrio Molitor Larvae Worms in a Controlled Environment for Accurate Small Object Detection and Segmentation

Angelos-Michael Papadopoulos[a], Paschalis Melissas[b], Anestis Kastellos[c], Panagiotis Katranitsiotis[d], Panagiotis Zaparas[e], Konstantinos Stavridis[f] and Petros Daras[g]

The Visual Computing Lab - Centre for Research and Technology Hellas/Information Technologies Institute, Thessaloniki, Greece

Keywords: TenebrioVision, Dataset, Benchmark, Instance Segmentation, Object Detection, Worms, Tenebrio Molitor, Edible Insects, Farming, Automation, Alternative Food Source.

Abstract: Tenebrio molitor worms have shown extreme nutritional benefits, as they contain useful natural compounds, making them worth as an alternative food source. It is beneficial for insect farms to have automated mechanisms that can detect these worms. Without an explicitly annotated dataset, the task of detecting tenebrio molitor worms remains challenging and underdeveloped. To address this issue, we introduce TenebrioVision, which is a fully annotated dataset, suitable for the detection and segmentation of tenebrio molitor larvae worms. The data acquisition is performed in a controlled environment. The dataset consists of 1,120 images, with a total of 53,600 worm instances. The 1,120 images are equally distributed on 14 distinct levels, each level containing a specific number of tenebrio monitor larvae worms. The dataset is validated in terms of mean average precision, memory allocation, and inference time, on several state-of-the-art baseline methods for both detection and segmentation purposes. The results unequivocally show that the detection and segmentation accuracy is high on both TenebrioVision and real farm images.

SCIENCE AND TECHNOLOGY PUBLICATIONS

1 INTRODUCTION

With the world's population expected to reach 9.7 billion by 2050 (Desa, 2019), the high demand for animal protein, without the detrimental environmental effects of animal husbandry, poses significant challenges for global food production. The available food stocks are limited and will eventually become insufficient to meet this demand. These core factors have initialized the development of the appropriate industrial production systems (Van Huis et al., 2013).

Insects have a high feed conversion efficiency, low greenhouse gas emissions, high-quality protein, and require overall fewer resources to produce than other animal proteins. Moreover, they can be produced on a larger scale than traditional livestock. For these reasons, several research studies (Ghaly et al., 2009; Oonincx and De Boer, 2012; Brandon et al., 2021)

Figure 1: TenebrioVision: Tenebrio molitor larvae worms are present in a dataset of 1,120 images, ranging from frames with 10 worms to frames with 100 worms.

[a] https://orcid.org/0009-0000-8117-312X
[b] https://orcid.org/0009-0000-2657-9557
[c] https://orcid.org/0009-0002-7812-1577
[d] https://orcid.org/0009-0004-6509-6338
[e] https://orcid.org/0009-0001-5407-7295
[f] https://orcid.org/0000-0002-3244-2511
[g] https://orcid.org/0000-0003-3814-6710

Papadopoulos, A., Melissas, P., Kastellos, A., Katranitsiotis, P., Zaparas, P., Stavridis, K. and Daras, P.
TenebrioVision: A Fully Annotated Dataset of Tenebrio Molitor Larvae Worms in a Controlled Environment for Accurate Small Object Detection and Segmentation.
DOI: 10.5220/0012295900003654
Paper published under CC license (CC BY-NC-ND 4.0)
In *Proceedings of the 13th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2024)*, pages 187-196
ISBN: 978-989-758-684-2; ISSN: 2184-4313
Proceedings Copyright © 2024 by SCITEPRESS – Science and Technology Publications, Lda.

have highlighted the benefits of using insects as an alternative source of animal protein for human consumption.

The yellow mealworm, also known as Tenebrio molitor, has demonstrated significant potential for human consumption. There are four development stages of this species: Eggs, larvae, pupae, and adults. Tenebrio molitor larvae is typically the preferred stage in many countries where insects are consumed (Stoops et al., 2016; Siemianowska et al., 2013). It is a valuable source of protein and minerals, easy to maintain, and can be harvested at an earlier stage of development. Several studies (Costa et al., 2020) and (Kröncke and Benning, 2022) have shown that Tenebrio molitor includes a nutritional composition similar to other conventional meat sources. Therefore, the European Commission implementing regulation 2023/58 (of the European Union, 2023) has officially authorized the placing on the market of frozen, paste, dried, and powdered larvae of Alphitobius diaperinus (minor mealworm) as a novel food.

Tenebrio molitor, in its larvae development stage, is therefore becoming increasingly popular for farming and raising edible insects, which highlights the need for standardized and cost-effective production techniques. There are already insect farms operating in numerous countries, but their production systems lack efficient and scalable automation processes (Grau et al., 2017), which can be supported by computer vision and machine learning technologies. Tenebrio molitor breeding procedures, such as feeding, wetting larvae, classifying larvae by size, harvesting chitinous moult, and finally harvesting larvae and separating them from impurities, should be automated and monitored in order to be profitable. The amount of manual labor required by current farming methods prevents them from being used for industrial-scale production.

In this paper, a publicly available dataset called "TenebrioVision" is introduced, which is a comprehensive, and fully annotated dataset for Tenebrio Molitor larvae insects in a controlled environment. TenebrioVision dataset consists of 1,120 images of 53,600 tenebrio molitor instances annotated for both object detection and instance segmentation purposes (figure 1). Every image is taken in a controlled setup by a UI camera (UI-3884LE, 2021). The resolution of each frame is 3088 x 2076 pixels. The dataset is distributed in 14 different levels, depending on the number of tenebrio molitor instances inside the crate, as shown in figure 4. It should be noticed that localizing insects in the visual scene at the level of the instance mask is a fundamental step that enables further quality analysis in the field. The significance of TenebrioVision dataset is then validated by running several state-of-the-art baseline methods for detection and instance segmentation. The intention for experimenting with several SoTA models is to compare both the detection/segmentation efficiency, but also their inference time on the TenebrioVision dastaset. A quick inference time can be a vital asset in the farm industries, thus a comprehensive comparison is necessary. Lastly, the high-value features learned from TenebrioVision are tested on farm images, taken from a variety of real farm industry environments, some of them containing a huge amount of tenebrio molitor insects, uncountable even by an insect expert's human eye.

By making this dataset available to the public, it is believed that the following aspects will be positively impacted:

- To the best of our knowledge, the TenebrioVision dataset is the first fully annotated dataset that includes the tenebrio molitor larvae insects for detection/segmentation tasks at a large scale.
- The TenebrioVision dataset can address challenges associated with automated insect breeding and production (Cadinu et al., 2020).
- It can also serve as a small object detection/segmentation benchmark for researchers and industry professionals for a variety of computer vision tasks.
- Rich features learned by state-of-the-art models from the augmented TenebrioVision dataset, proved to be substantial in real-case scenarios, surpassing expert human-eye capabilities.

2 RELATED WORK

2.1 Insect Related Datasets

Popular datasets like ImageNet (Deng et al., 2009) and COCO (Lin et al., 2014) that include different classes of animals have been developed over the years for a variety of computer vision tasks, such as image classification and object detection - segmentation, respectively. The interest in building specialized datasets that are only concerned with animal species has grown in recent years. Datasets like iNaturalist (Van Horn et al., 2018), which currently contains over 415,000 species of animals, Animal Kingdom dataset (Ng et al., 2022) containing 850 species across 6 major animal classes, and others (Beery et al., 2021; Gagne et al., 2021; Cao et al., 2019) have a major contribution to analyzing animal behavior. Some datasets are devoted exclusively to one or a few animal species

Table 1: Comparison of various domain-specific datasets.

Dataset	Worm Images	Tenebrio molitor larvae	Task		
		Images	Classification	Detection	Segmentation
iNaturalist(Van Horn et al., 2018)	436,200	1,620	X	-	-
IP102(Wu et al., 2020)	6,850	-	X	X	-
Larvae Dataset in roboflow(Probst, 2023)	179	179	-	X No. instances: 179	-
Mealworms Dataset in roboflow(egg detection mixed eggs, 2022)	518	518	X	X No. instances: 518	-
Multipurpose monitoring system(Majewski et al., 2022)	120	120	X	X	X No. instances: 1,026
TenebrioVision	1,120	1,120	-	X	X No. instances: 53,600

only (Wah et al., 2011; Fang et al., 2020; Labuguen et al., 2021; Nuthalapati and Tunga, 2021)

Regarding solely insects, there are datasets like (Van Horn et al., 2018; Hansen et al., 2020; Wu et al., 2020) that include large amounts of insect images. Yet, there have not been many attempts to study insect datasets resembling tenebrio molitor worms in the larvae stage, which is the nutrition-preferred one as the ingredient for the EU foods. Efforts have been made by (Hebert et al., 2021) to create synthetic images of worm posture, which would avoid the need for human-labeled annotation. Other researchers (Husson et al., 2018) and (Pereira et al., 2019) have developed tools for analyzing their pose and behavior. However, due to the nature and characteristics of real tenebrio-molitor worms, these approaches cannot be effective in real-world schemes.

Furthermore, regardless of whether Tenebrio molitor is present at its larvae stage or not, all of the existing datasets contain a small amount of the particular tenebrio molitor and are primarily used for animal classification tasks. The development of techniques for examining Tenebrio Molitor's general characteristics is severely hampered by this barrier. Thus, the need for tenebrio molitor images for both image detection and segmentation tasks emerges. There are two small, yet preferable datasets regarding tenebrio molitor worms in the larvae development stage (egg detection mixed eggs, 2022) and (Probst, 2023). The first one classifies the images of live or dead tenebrio molitor and also detects the worms. Both of them can be found in Roboflow (Dwyer,). However, they feature only a limited sample of tenebrio molitor images. These datasets are annotated solely for object detection and typically showcase just a single insect per frame, failing to represent authentic farming conditions. In a recent paper (Majewski et al., 2022), the authors made a multipurpose monitoring system for Tenebrio molitor breeding. Tenebrio molitor larvae, pupae, and beetles are all detected by the instance segmentation module (ISM), which also detects dead larvae and pests as anomalies. From the acquired image, the semantic segmentation module (SSM) extracted feed, chitin, and frass. Additionally, the Larvae Phenotyping Module (LPM) computes features for both the population as a whole and each individual larva (length, curvature, mass, segmentation, and classification). However, due to the difficult process of annotation and the need for a multipurpose monitoring system, they present only 120 total labeled images, including 1026 live tenebrio molitor larvae instances among others. These images do not all contain the tenebrio molitor insect in the larval stage. Also, these images are not publicly available.

A comparison of the aforementioned domain-specific datasets is presented in table 1. For each dataset, the total number of worm images, the total number of tenebrio molitor worm images and the purpose that this dataset accomplishes(classification, detection, segmentation) are provided. It is shown that even though there are large datasets that contain many worm images (iNaturalist (Van Horn et al., 2018), IP102 (Wu et al., 2020)) there is a limited number of tenebrio molitor images on the larvae development stage. On top of that, all these datasets either contain limited (Majewski et al., 2022) or no data at all for the segmentation task.

Given the aforementioned factors, it becomes evident that a direct and equitable comparison between these datasets and TenebrioVision, in terms of performance, is not feasible.

2.2 Vision Based Methods for Insect Analysis

Large strides have been made in the field of object detection in images, particularly with algorithms based on deep learning, which are typically divided into two categories: Two-stage detectors (methods based on the region proposal network), like Faster R-CNN (Girshick, 2015) and Mask R-CNN (He et al., 2017), and one-stage detectors like YOLO (Redmon et al., 2016), SSD (Liu et al., 2016).

Various models such as Mask R-CNN (He et al., 2017) and YOLO-V5 (Jocher, 2020) have been explored by researchers for the detection and segmentation of worms (Majewski et al., 2022). However, given the specific requirements of accuracy and rapid inference time in the context of insect farms, an extensive evaluation of five state-of-the-art baseline models is conducted in the TenebrioVision dataset. These models included Mask R-CNN (He et al., 2017), EfficientDet (Tan et al., 2020), YOLO-V7 (Wang et al., 2022), YOLO-V8 (Jocher et al., tics) and the current state-of-the-art YOLO-NAS (Aharon et al., 2021).

3 TenebrioVision DATASET

Detailed information about the TenebrioVision dataset, which comprises a large collection of high-quality images of Tenebrio molitor larvae worms in various poses and orientations are discussed in detail in this section.

3.1 Experimental Setup

The TenebrioVision dataset is collected using a custom-designed setup consisting of a crate in a controlled environment and a UI-3884LE-C-HQ camera (UI-3884LE, 2021) placed above it. The UI camera is chosen because it offers 3088 x 2076 pixel resolution at frame rates up to 58.0 fps even under low-light conditions and the focus can be conveniently adjusted. The tenebrio molitor worms are placed inside a crate with a spatial field of view of 20cm x 30cm, which is exactly a quarter of the crate's spatial field of view, as it is presented in figure 2. The quarter of the crate, as a spatial field of view, of the UI camera was chosen experimentally in order to capture the 3088x2076 pixel resolution. By taking into account the following equation (1) from (Fulton, 2015), it is concluded that 38 cm is the ideal height to set the camera from the crate:

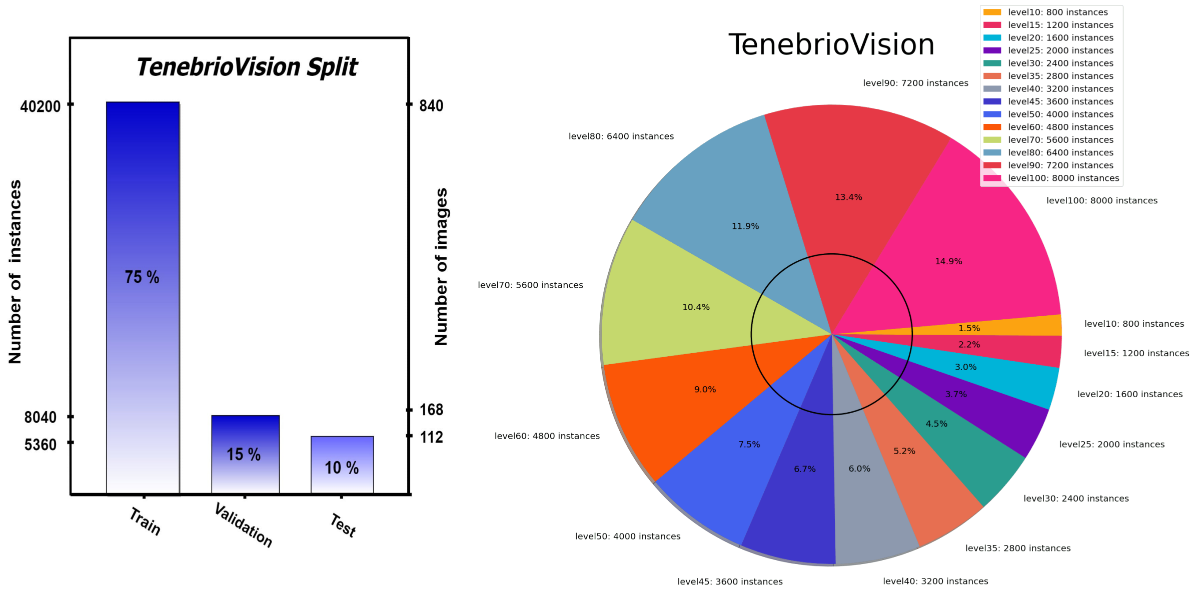
Figure 2: Experimental setup. The UI camera is placed at a 38 cm distance from the black crate, according to Equation 1. The spatial field of view is 20 x 30 cm. The camera is steadily placed on the desk during the whole experimental procedure.

$$D(cm) = \sqrt{\frac{Q \times F}{H}} \tag{1}$$

where D is the camera's distance from the crate in cm, Q is the Quarter crate's size: 20cm x 30cm, F is the Focal length of the camera: 9.6 mm and the H is the Object height on the sensor, which is actually the optical size of the camera sensor: 7.411mm x 4.982mm according to UI-3884LE-C-HQ specifications (UI-3884LE, 2021).

3.2 Dataset Acquisition

The Tenebrio molitor larvae worms are placed inside the crate and allowed to move freely for a few seconds



(a) (b)

Figure 3: TenebrioVision dataset statistics. (a) Train, validation and test split of TenebrioVision dataset with respect to both the total number of instances and the total number of images. (b) This pie - chart portrays the distribution of TenebrioVision's levels among the entire dataset.

while being captured by the UI camera. More particularly, a 30-second video is recorded and processed to only obtain the first and last frames. This choice is made because worm position and orientation between the first and last images are likely to differ significantly ensuring that there are no identical images in the dataset, thus providing no bias. When the desired number of images has been obtained, the process is complete.

In order to capture the ground truth illumination inside the crate, which is crucial for accurately representing the colors of the worms, the specifications of the UI camera are modified. Specifically, the sensitivity is increased and the white balance settings are adjusted for the purpose of capturing the true colors of the larvae.

In order to guarantee that only live Tenebrio molitor larvae are included in the dataset throughout the experimental process, the worm specimens are consistently provided with adequate nutrition and temperature control. This procedure enables the collection of rich, representative data so that the deep learning approaches will learn features associated only with live tenebrio molitor worms.

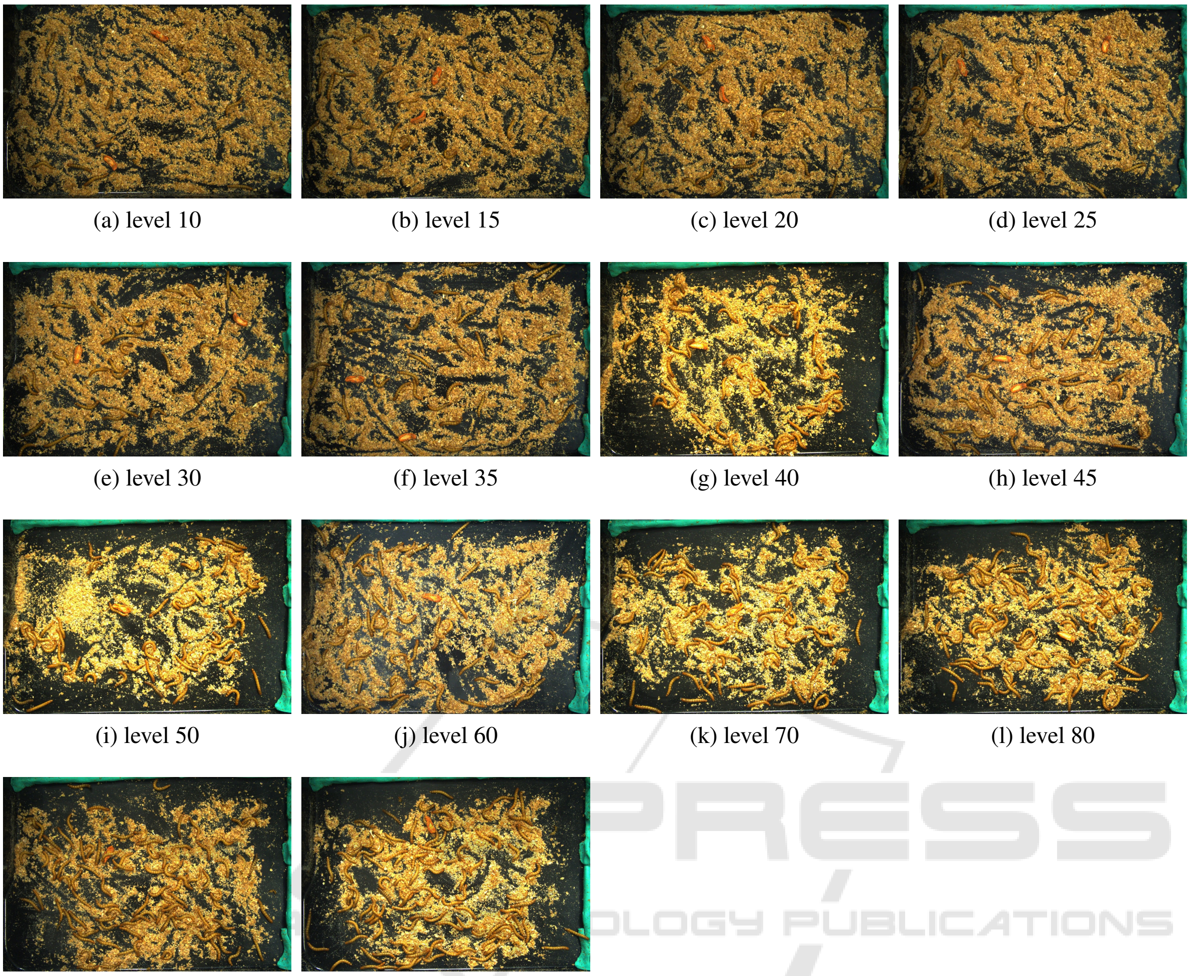
The captioning framework is initialized after the camera setup. In order to have a balanced distribution of the images across the dataset, 14 levels are formed, according to the number of worms inside each image. Each level contains 80 images. There are 14 levels: 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, and 100 as it is demonstrated in the figure 4. In essence, level 45 has 80 images, and each image includes 45 worm instances. Figure 3b depicts the proportion of each level to the total number of worm instances in the TenebrioVision dataset. In total, the TenebrioVision dataset includes 1,120 images, and each image has a resolution of 6 Megapixels.

3.3 Dataset Annotation

Annotation is performed using Datatorch (Nguyen, 2020) by drawing precise segmentation masks covering each tenebrio molitor worm in all frames. Tenebrio molitor larvae's high level of articulation and potential occlusion in crowded environments lead to an increased annotation effort. Without accounting for the overhead, annotating 1,120 images with segmentations masks and bounding boxes for the TenebrioVision dataset took more than 373 human hours of annotation time. The dataset is publicly available here: https://vcl.iti.gr/dataset/TenebrioVision/.

3.4 Dataset Split

TenebrioVision. To enable the reproduction of the experimental results, a 75:15:10 split is followed, for the training, validation, and test set, respectively, as shown in figure 3a.



(a) level 10 (b) level 15 (c) level 20 (d) level 25

(e) level 30 (f) level 35 (g) level 40 (h) level 45

(i) level 50 (j) level 60 (k) level 70 (l) level 80

(m) level 90 (n) level 100

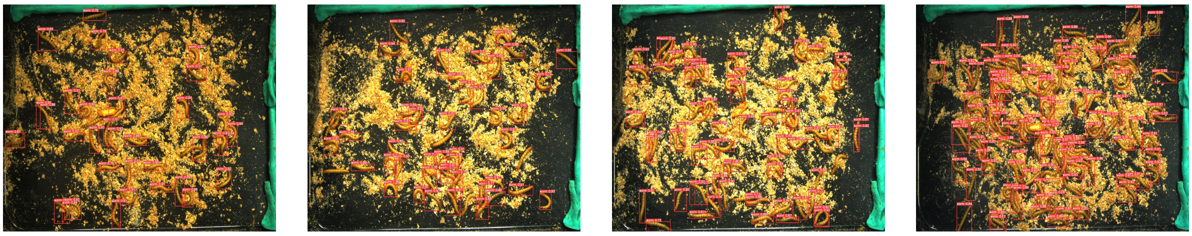
Figure 4: Levels of the TenebrioVision dataset. There are 14 levels. The numerical value assigned to each level indicates the cumulative count of tenebrio molitor worm instances captured in the corresponding images at that level. Each level has 80 images, totaling 1,120 images. For example, we have 80 images for level 90, and every image at level 90 contains only 90 tenebrio monitor worms.

4 EXPERIMENT AND EVALUATION

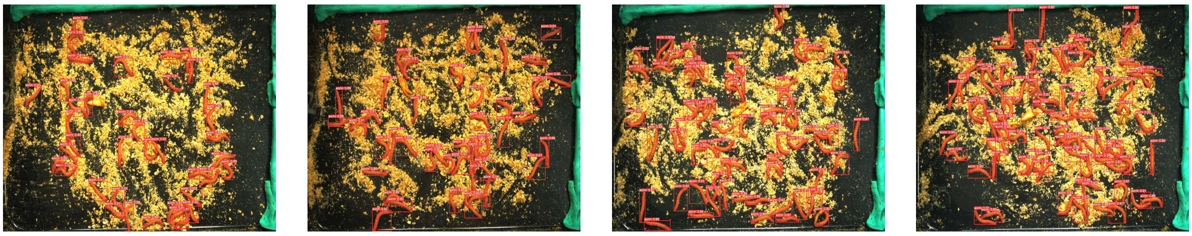
In this section, the performance of TenebrioVision dataset is evaluated on various state-of-the-art object detection and instance segmentation models. The images in the training set are resized to each baseline setup, always trying, if possible, to maintain the aspect ratio of 6 Megapixels. The performance of each model is evaluated based on mean average precision (mAP), inference time, and memory allocation. These evaluation metrics are chosen because they are essential indicators of the final product's quality in automated farming. Finally, a qualitative evaluation of images provided by real insect farms is presented. All experiments are conducted on the NVIDIA GeForce RTX 3090 GPU with 24GB memory.

4.1 Quantitative Evaluation

The five baselines baseline models used are Mask R-CNN (He et al., 2017), EfficientDet (Tan et al., 2020), YOLO-V7 (Wang et al., 2022) and YOLO-V8 (Jocher et al., tics) for both object detection (OD) and instance segmentation (IS) tasks and the YOLO-NAS (Aharon et al., 2021). Mask R-CNN is conducted using the Detectron2 (Wu et al., 2019) framework. Detectron2 is preferred over the original implementation (Matterport Mask R-CNN (Abdulla, 2017)) due to its improved performance and flexibility. The implementa-



(a) object detection for a level 35 image (b) object detection for a level 50 image (c) object detection for a level 70 image (d) object detection for a level 100 image



(e) instance segmentation for a level 40 image (f) instance segmentation for a level 45 image (g) instance segmentation for a level 70 image (h) instance segmentation for a level 80 image

Figure 5: Experiment results on TenebrioVision's dataset test set. On top, there are predictions regarding object detection on different levels of the dataset, and below there are predictions regarding instance segmentation on different levels of the dataset.

Table 2: Detection and segmentation results on the test set of TenebrioVision using different state-of-the-art models. We evaluate the mAP@0.5:0.05:0.95 and mAP@0.75 for the bounding box and the mask. *Empty objects on the table indicates that the model performs only object detection.*

Methods	mAP_bbox	mAP@75_bbox	mAP_mask	mAP@75_mask
Mask RCNN	0.781	0.954	0.654	0.776
EFFICIENTDET	0.787	0.926	-	-
YOLO_V7_OD	0.842	0.93	-	-
YOLO_V7_IS	0.848	0.93	0.632	0.802
YOLO_V8_OD	0.874	0.96	-	-
YOLO_V8_IS	0.88	0.965	**0.729**	0.83
YOLO_NAS	**0.892**	0.972	-	-

tion of the other baselines is taken from their official repositories.

For each method, the smallest backbone version of the model is used. For Mask R-CNN the backbone is Resnet-50-FPN, for EfficientDet the EfficientDet-D0, for Yolov7-(OD) the yolov7-tiny, for Yolov7-(IS) yolov7-seg, for yolov8-(OD) the yolov8n, for Yolov8-(IS) yolov8n-seg and finally for Yolo-NAS the YOLO-NAS-s. The reason that the largest backbones of those baselines are not employed is twofold: firstly, there is no discernible difference in the output mAP, and secondly, inference time decreases. A basic augmentation scheme is followed for these baselines. The basic augmentations are Horizontal and Vertical flips. For Mask R-CNN and Yolo-V7 the training im-

Table 3: Comparison of memory allocation in Gigabytes(GBs) and inference time (FPS) during the testing phase on several state-of-the-art models.

Methods	Memory(GBs)	FPS
Faster R-CNN	2.27	20
Mask R-CNN	5.06	9
EfficientDet	1.75	25
YOLO_V7_OD	2.37	220
YOLO_V7_IS	3.44	40
YOLO_V8_OD	1.59	150
YOLO_V8_IS	**2.12**	**80**
YOLO_NAS	**1.48**	**416**

age has a 1024 x 688 resolution with respect to the camera's aspect ratio. For EfficientDet the training image size is 512 x 512 in order to get the smallest EfficientDet version (EfficientDet-D0). Finally, the training image size for Yolo-V8 and Yolo-NAS is 1280 x 1280.

The results are presented in table 2, where the mAP scores for all methods tested are presented. In table 3 the memory allocation and the inference time of each method are reported. The memory allocation refers to the Gigabytes (GBs) required to infer an image from the TenebrioVision dataset and the inference time to the frames per second (FPS). The optimal per-

SCIENCE AND TECHNOLOGY PUBLICATIONS

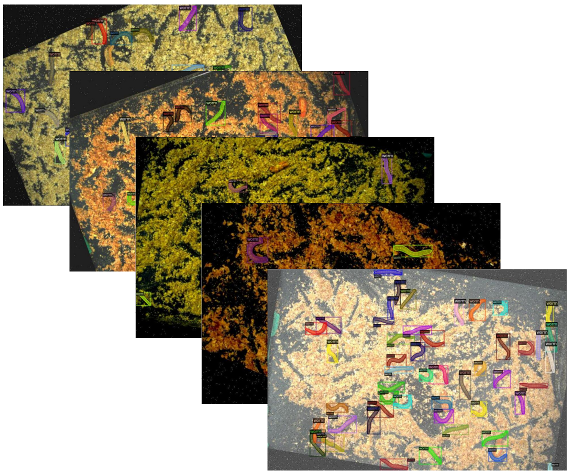


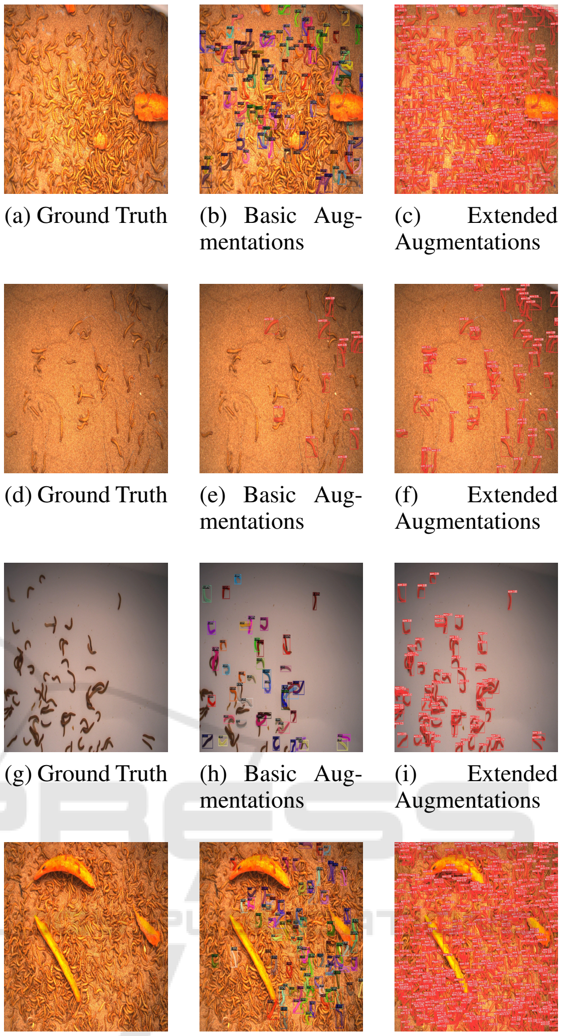
Figure 6: Extended augmentation scheme for TenebrioVision dataset. The applied augmentations are horizontal-vertical flips, random rotation, adding noise, zooming-in, brightness, and exposure. The annotated segmentation masks are also presented.

formance is achieved using the current SoTA Yolo-V8(IS) (Jocher et al., tics) with a mAP of 0.729 for the segmentation mask and the Yolo-NAS (Aharon et al., 2021) with a mAP of 0.892 for the bounding box, both evaluated at IoU=0.5:.05:0.95. Considering the inference time, the YOLO-NAS model exhibits the lowest memory allocation of 1.48 GB for object detection, while YOLO-V8 (IS) requires 2.12 GB for the instance segmentation task. Additionally, the highest FPS values of 416 and 80 for detection and segmentation tasks are attained by Yolo-NAS and Yolo-V8(IS) respectively, on the TenebrioVision dataset.

4.2 Qualitative Evaluation

Some experimental results regarding object detection and instance segmentation can be seen in figure 5. This figure clearly shows that small and highly articulated objects like those of tenebrio molitor can be accurately detected. A precise segmentation mask can provide a wealth of information about these worms. Information like color, length, width and size of the tenebrio molitor larvae can help experts gain knowledge about the health and life cycle of each worm. This analysis is really vital for farm workers as it saves them a lot of time when trying to figure out if a crate filled with worms is healthy or unhealthy in a production pipeline scheme. These data will aid the competent scientific community and hasten the breeding of insects for insect farming.

Further evaluation of the models is conducted on actual farm images, which contain only tenebrio molitor worms at the larvae stage. It is found that even



(a) Ground Truth (b) Basic Augmentations (c) Extended Augmentations

(d) Ground Truth (e) Basic Augmentations (f) Extended Augmentations

(g) Ground Truth (h) Basic Augmentations (i) Extended Augmentations

(j) Ground Truth (k) Basic Augmentations (l) Extended Augmentations

Figure 7: Expreriments conducted on real farm images. The left column depicts the Ground Truth farm images, the center column the predictions of the SoTA models, trained on TenebrioVision with Basic Augmentations, and the right column the prediction of the SoTA models, trained on TenebrioVision with an extended augmentation scheme.

though the current SoTA baseline methods achieved a satisfying mAP at the test set, their performance on real farm images is poor and inadequate as it is demonstrated in figure 7 (center column). Genuine farm images, in contrast to TenebrioVision, show a significant amount of Tenebrio molitor larvae inside the inspection crates, frequently beyond the limit of human counting capacity. Real farm images typically have a range of backgrounds, colors, camera angles, and many other aspects. To tackle this problem the best models are trained with an extended augmentation scheme. More specifically, these augmentations

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

are horizontal-vertical flips, random rotation, adding noise, zooming-in and out, brightness, and exposure. The noise augmentation technique applies a Gaussian distribution to 8% of the pixels in each image. Samples from this extended augmentation scheme are demonstrated in figure 6. The results from this extended augmentation scheme are fascinating, as seen in figure 7 (right column), since YOLO-V8(IS) now detects and captures a very large amount of tenebrio molitor larvae worms, that are even uncountable by an expert's human eye.

It is believed that the process of inference on real images and adding them back to the TenebrioVision will further improve the detection/segmentation tasks, and enhance the automation process of the farms.

5 CONCLUSIONS

In this paper, TenebrioVision was introduced, a dataset that contains tenebrio molitor worms in the larvae development stage. The TenebrioVision dataset contains 1,120 fully annotated images for detection and segmentation tasks, at a resolution of 3088x2076 pixels. The total number of worm instances is 53,600. The performance of several state-of-the-art object detection and instance segmentation models was evaluated on TenebrioVision. The experiments' findings demonstrate that, despite the SoTA algorithms' robustness, they perform poorly in real-world situations, necessitating augmentations. By extending TenebrioVision with augmentations, astonishing results were achieved, surpassing the expert human-eye detection accuracy. By making TenebrioVision publicly available our aim is to assist and enhance the automation process on real farms, thus helping the scientific community produce valuable data and knowledge on this high-nutrition worm. This dataset can also be utilized, by the computer vision community, as a benchmark for small object detection and segmentation tasks.

ACKNOWLEDGMENTS

The work presented in this paper was supported by the European Commission under contract H2020 - 101016953 CoRoSect.

REFERENCES

Abdulla, W. (2017). Mask r-cnn for object detection and instance segmentation on keras and tensorflow. https://github.com/matterport/Mask_RCNN.

Aharon, S., Louis-Dupont, Ofri Masad, Yurkova, K., Lotem Fridman, Lkdci, Khvedchenya, E., Rubin, R., Bagrov, N., Tymchenko, B., Keren, T., Zhilko, A., and Eran-Deci (2021). Super-gradients.

Beery, S., Agarwal, A., Cole, E., and Birodkar, V. (2021). The iwildcam 2021 competition dataset. *arXiv preprint arXiv:2105.03494*.

Brandon, A. M., Garcia, A. M., Khlystov, N. A., Wu, W.-M., and Criddle, C. S. (2021). Enhanced bioavailability and microbial biodegradation of polystyrene in an enrichment derived from the gut microbiome of tenebrio molitor (mealworm larvae). *Environmental science & technology*, 55(3):2027–2036.

Cadinu, L. A., Barra, P., Torre, F., Delogu, F., and Madau, F. A. (2020). Insect rearing: Potential, challenges, and circularity. *Sustainability*, 12(11):4567.

Cao, J., Tang, H., Fang, H.-S., Shen, X., Lu, C., and Tai, Y.-W. (2019). Cross-domain adaptation for animal pose estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9498–9507.

Costa, S., Pedro, S., Lourenço, H., Batista, I., Teixeira, B., Bandarra, N. M., Murta, D., Nunes, R., and Pires, C. (2020). Evaluation of tenebrio molitor larvae as an alternative food source. *NFS journal*, 21:57–64.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255.

Desa, U. (2019). World population prospects 2019: Highlights. *New York (US): United Nations Department for Economic and Social Affairs*, 11(1):125.

Dwyer, B., N. J. . S. J. e. a. R. V. . S. A. f. u.

egg detection mixed eggs (2022). Mealworms-detection dataset. https://universe.roboflow.com/egg-detection-mixed-eggs/mealworms-detection. visited on 2023-02-23.

Fang, C., Zhang, T., Zheng, H., Huang, J., and Cuan, K. (2020). Pose estimation and behavior classification of broiler chickens based on deep neural networks. *Comput. Electron. Agric.*, 180:105863.

Fulton, W. (2015). Calculate distance or size of an object in a Photo Image. https://www.scantips.com/lights/subjectdistance.html.

Gagne, C., Kini, J., Smith, D., and Shah, M. (2021). Florida wildlife camera trap dataset. *arXiv preprint arXiv:2106.12628*.

Ghaly, A. E., Alkoaik, F., et al. (2009). The yellow mealworm as a novel source of protein. *American Journal of Agricultural and Biological Sciences*, 4(4):319–331.

Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448.

Grau, T., Vilcinskas, A., and Joop, G. (2017). Sustainable farming of the mealworm tenebrio molitor for the production of food and feed. *Zeitschrift für Naturforschung C*, 72(9-10):337–349.

Hansen, O. L., Svenning, J.-C., Olsen, K., Dupont, S., Garner, B. H., Iosifidis, A., Price, B. W., and Høye,

T. T. (2020). Species-level image classification with convolutional neural network enables insect identification from habitus images. *Ecology and Evolution*, 10(2):737–747.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969.

Hebert, L., Ahamed, T., Costa, A. C., O'Shaughnessy, L., and Stephens, G. J. (2021). Wormpose: Image synthesis and convolutional networks for pose estimation in c. elegans. *PLoS computational biology*, 17(4):e1008914.

Husson, S. J., Costa, W. S., Schmitt, C., and Gottschalk, A. (2018). Keeping track of worm trackers. *WormBook: The Online Review of C. elegans Biology [Internet]*.

Jocher, G. (2020). YOLOv5 by Ultralytics.

Jocher, G., Chaurasia, A., and Qiu, J. (2023., {https://github.com/ultralytics/ultralytics}). Yolo by ultralytics.

Kröncke, N. and Benning, R. (2022). Self-selection of feeding substrates by tenebrio molitor larvae of different ages to determine optimal macronutrient intake and the influence on larval growth and protein content. *Insects*, 13(7):657.

Labuguen, R., Matsumoto, J., Negrete, S. B., Nishimaru, H., Nishijo, H., Takada, M., Go, Y., Inoue, K.-i., and Shibata, T. (2021). Macaquepose: a novel "in the wild" macaque monkey pose dataset for markerless motion capture. *Frontiers in behavioral neuroscience*, 14:581154.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C. (2016). Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pages 21–37. Springer.

Majewski, P., Zapotoczny, P., Lampa, P., Burduk, R., and Reiner, J. (2022). Multipurpose monitoring system for edible insect breeding based on machine learning. *Scientific Reports*, 12(1):1–15.

Ng, X. L., Ong, K. E., Zheng, Q., Ni, Y., Yeo, S. Y., and Liu, J. (2022). Animal kingdom: A large and diverse dataset for animal behavior understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19023–19034.

Nguyen, M. (2020). Datatorch annotation tool. https://datatorch.io/.

Nuthalapati, S. V. and Tunga, A. (2021). Multi-domain few-shot learning and dataset for agricultural applications. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1399–1408.

of the European Union, O. J. (2023). Commission implementing regulation (eu) 2023/58. https://eur-lex.europa.eu/eli/reg_impl/2023/58/oj. *Official Journal of the European Union*, 66.

Oonincx, D. G. and De Boer, I. J. (2012). Environmental impact of the production of mealworms as a protein source for humans–a life cycle assessment. *PloS one*, 7(12):e51145.

Pereira, T. D., Aldarondo, D. E., Willmore, L., Kislin, M., Wang, S. S.-H., Murthy, M., and Shaevitz, J. W. (2019). Fast animal pose estimation using deep neural networks. *Nature methods*, 16(1):117–125.

Probst, A. (2023). larvae dataset. https://universe.roboflow.com/andre-probst-egt9g/larvae-bbsvy. visited on 2023-02-23.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788.

Siemianowska, E., Kosewska, A., Aljewicz, M., Skibniewska, K. A., Polak-Juszczak, L., Jarocki, A., and Jedras, M. (2013). Larvae of mealworm tenebrio molitor l. as european novel food.

Stoops, J., Crauwels, S., Waud, M., Claes, J., Lievens, B., and Van Campenhout, L. (2016). Microbial community assessment of mealworm larvae (tenebrio molitor) and grasshoppers (locusta migratoria migratorioides) sold for human consumption. *Food Microbiology*, 53:122–127.

Tan, M., Pang, R., and Le, Q. V. (2020). Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10781–10790.

UI-3884LE (2021). Ids. ui-3884le-c-hq camera (ab00979). https://en.ids-imaging.com/IDS/datasheet_pdf.php?sku=AB00979.

Van Horn, G., Mac Aodha, O., Song, Y., Cui, Y., Sun, C., Shepard, A., Adam, H., Perona, P., and Belongie, S. (2018). The inaturalist species classification and detection dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8769–8778.

Van Huis, A., Van Itterbeeck, J., Klunder, H., Mertens, E., Halloran, A., Muir, G., and Vantomme, P. (2013). *Edible insects: future prospects for food and feed security*. Number 171. Food and agriculture organization of the United Nations.

Wah, C., Branson, S., Welinder, P., Perona, P., and Belongie, S. (2011). The caltech-ucsd birds-200-2011 dataset.

Wang, C.-Y., Bochkovskiy, A., and Liao, H.-Y. M. (2022). Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*.

Wu, X., Zhan, C., Lai, Y.-K., Cheng, M.-M., and Yang, J. (2020). Ip102: A large-scale benchmark dataset for insect pest recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8787–8796.

Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., and Girshick, R. (2019). Detectron2. https://github.com/facebookresearch/detectron2.