Automated Visual Monitoring of Nocturnal Insects with Light-based Camera Traps

Dimitri Korsch [1] Paul Bodesheim [1] Gunnar Brehm [2] Joachim Denzler [1]

[1] Computer Vision Group, [2] Phyletic Museum
Friedrich Schiller University Jena, 07737 Jena, Germany

{dimitri.korsch, paul.bodesheim, gunnar.brehm, joachim.denzler}@uni-jena.de

Abstract

Automatic camera-assisted monitoring of insects for abundance estimations is crucial to understand and counteract ongoing insect decline. In this paper, we present two datasets of nocturnal insects, especially moths as a subset of Lepidoptera, photographed in Central Europe. One of the datasets, the EU-Moths dataset, was captured manually by citizen scientists and contains species annotations for 200 *different species and bounding box annotations for those. We used this dataset to develop and evaluate a two-stage pipeline for insect detection and moth species classification in previous work. We further introduce a prototype for an automated visual monitoring system. This prototype produced the second dataset consisting of more than* 27 000 *images captured on* 95 *nights. For evaluation and bootstrapping purposes, we annotated a subset of the images with bounding boxes enframing nocturnal insects. Finally, we present first detection and classification baselines for these datasets and encourage other scientists to use this publicly available data.*

1. Introduction

Climate change and the decline of species richness are severe challenges that influence the living conditions of humans around the world. Especially the dramatic loss of insects [6, 29] plays a crucial role in many ecological processes that affect agriculture and others. Hence, monitoring insect species populations becomes more important nowadays to better understand insect decline and long-term trends in species distributions. Furthermore, there are about one million named species on our planet [23], making manual counting of individuals unrealistic. Consequently, automated monitoring of insects is inevitably required to infer abundance estimations across larger regions. One possible way is to use camera traps to collect images of insects that computer vision algorithms can then process to recognize the depicted species automatically.

Figure 1. The prototype of the *moth scanner*: a white planar surface and an automated camera system. At night, UV light illuminates the surface to attract moths that land on the surface and the camera takes an image with flash every two minutes.

In this paper, we focus on nocturnal insects, mainly nocturnal moths (Lepidoptera). Even for this subset, there exist hundred thousands of different species worldwide and depending on the habitat, species lists can be narrowed down based on the study region. For example, image datasets containing hundreds of moth species from Ecuador and Costa Rica are publicly available and can directly be used for evaluating fine-grained recognition algorithms [18]. Here, we are interested in monitoring moth species in Central Europe. We present datasets of moth images we have collected so far and our analysis of algorithms for insect localization and species classification.

Our work is part of a larger project called AMMOD[1], which aims at developing self-sustaining multi-sensor stations for monitoring species diversity [31]. One component of these stations is a light-based camera trap for nocturnal insects, called the *moth scanner* [11,17]. It is a non-invasive monitoring system for automatically gathering images at nighttime. A UV-LED lamp illuminates a white planar surface to attract the insects that land on this surface. A high-resolution camera takes an image of the whole surface every two minutes. Our prototype is shown in Figure 1.

[1] AMMOD = **A**utomated **M**ultisensor Station for **M**onitoring **o**f Bio**d**iversity (https://ammod.de/)

With this setup, we can collect large-scale datasets of nocturnal insects over a long period that can then be used to develop and evaluate appropriate fine-grained species recognition algorithms. The moth scanner takes several hundred images during one night, and within five months, we collected more than 27 000 images with our prototype. In this paper, we refer to the resulting dataset as the *nocturnal insects dataset (NID)*, and more details are given in Section 3. Note that this dataset is supposed to be extended over time as our system will be in operation within the following years. We plan to maintain multiple sensor stations in parallel at different locations. Hence, it has the potential to become a valuable source for large-scale learning and continuous learning within a fine-grained domain.

Besides its impact on research in fine-grained recognition, our developments for automated visual monitoring of nocturnal insects are beneficial for ecologists. Until now, insect monitoring is mainly done by hand and supported by citizen scientists who manually take images of individual insects in their gardens. Previously, we published an image dataset of nocturnal moths captured manually by citizen scientists, called *EU-Moths* dataset at a local workshop [11]. This paper also includes a dataset description and our baseline results for insect localization and species classification. There are two reasons for this. First, we want to announce this dataset to a broader audience interested in fine-grained recognition because it can directly be used for algorithm development and evaluation. Second, we want to highlight the challenges for recognition algorithms that arise when processing automatically captured camera trap images compared to manually taken images with hand-held cameras.

In general, our paper aims to promote the application of moth species identification as a fine-grained visual recognition problem. We underpin this with existing datasets, results of baseline algorithms, and a light-based camera trap setup that will be used during the following years to automatically collect further large-scale image data. We believe that research on automated visual identification of hundreds to thousands of different nocturnal moth species can have a major impact on developing fine-grained recognition algorithms in general, and we, therefore, want to share our insights and datasets with the community.

2. Related work

Besides insect monitoring [1,8,11,17,24], there also exist automatic recognition systems for other animals. They are often used to re-identify individuals of a certain species, e.g., great apes [3, 5, 9, 20, 32], elephants [12, 13], or sharks [7], to name a few. Of course, the field of fine-grained recognition mainly benefited from bird species datasets like CUB [30] and NA-Birds [27].

For moth species identification, datasets exist with insects from Ecuador and Costa Rica containing images of 675 and 331 different species, respectively [18]. There is only a single individual in each image spanning the whole image area. In contrast to these datasets, we consider images of moth species from Central Europe recorded by light-based camera traps. First, the recorded insects are still alive and may take various positions, making it harder to analyze the image. Second, the captured image contains multiple insects, and individuals need to be localized before inferring the species. Furthermore, we extend our datasets over time due to continuous monitoring.

Figure 2. Example images from the *EU-Moths* dataset. The first two rows show two different moths species, whereas the third row shows images with more than one insect. These examples illustrate the versatility in the appearance of the moths in the dataset.

There are also similar camera trapping systems developed by other groups, e.g., as reported in [1]. So far, they only consider eight different moth species and do not provide ground truth bounding box annotations for the individual insects. In contrast, our EU-Moths dataset contains 200 different species, and we provide algorithms that can reliably recognize individuals from this larger set of species.

3. The datasets

In the following, we present two datasets that we are using within our project for moth detection and species recognition. They contain images of nocturnal insects, primarily moths, photographed on a white background. In contrast to other publicly available datasets constructed for species classification such as iNaturalist [28], standardized camera trap images lead to a more homogeneous setting: the insects are photographed on a uni-color planar surface.

European moths (EU-Moths) dataset[2]: This dataset consists of 200 species most common in Central Europe. Each of the species is represented by approximately 11 images. We consider a random but balanced split in eight training and three test images per species, resulting in roughly 1600 training and 600 test images in total. Furthermore, we manually annotated bounding boxes for each insect. For this dataset, citizen scientists photographed the insects man-

[2] https://inf-cv.uni-jena.de/eu_moths_dataset

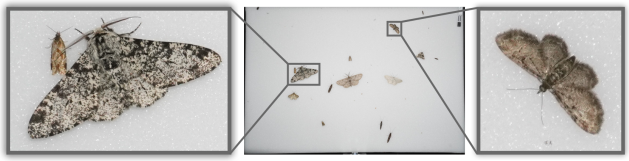

Figure 3. An example image of the *NID* dataset. Two of the bounding boxes and the corresponding image patches are shown. Note that even though these are relatively small parts of the original image, the visual species-related features in the patches have a high level of detail.

ually and mainly on a relatively homogeneous background. About $92\,\%$ of the images contain only a single individual like it is shown in the first two rows of Figure 2. The last row depicts images with more than one insect of interest.

Nocturnal insects dataset (NID)[3]: Our camera trap setup takes high-resolution images of the insects resting on an illuminated surface. We use a UV-LED lamp since this is the most attractive radiation for nocturnal insects than white light [2]. A 20-megapixel camera captured the images at an interval of two minutes (setup shown in Figure 1).

In five months (June - October 2021), the system captured images during 95 nights, and we removed empty images without any insects at the beginning and the end of every night. In total, we gathered more than 27 000 images.

We first selected images from ten nights equally distributed over the entire period and manually annotated bounding boxes around insects in 818 images to evaluate detection methods. As a result, we ended up with 9095 bounding box annotations. Figure 3 shows one of the images with two exemplary bounding boxes and the corresponding image patches. In our baseline experiments, we use the first five nights for training and parameter tuning and data from the last five nights for the evaluation.

4. Baseline methods

As presented in preliminary work [11], we deploy a two-stage pipeline for moth species detection and classification: (1) insect localization and (2) fine-grained moth species identification. This separation is vital since the later prototypes will operate autonomously in the field and transmit the gathered images to central storage. To reduce the amount of transmitted data, we will perform the detection directly at the moth scanner and transfer only the small image patches to the central storage.

4.1. Single-shot detector

We used a CNN-based state-of-the-art object detection model, namely the single-shot MultiBox detector (SSD) proposed by Liu *et al.* [15]. The authors utilize feature maps from multiple intermediate stages of the backbone CNN to predict location offsets and class confidences for a set of prior locations. For more details about the loss functions, we refer to the original paper of Liu *et al.* [15].

4.2. Fine-grained species classification

Neural networks, especially CNNs, yield state-of-the-art results in image classification tasks. As we presented in our previous work [10], one can utilize a linear classifier with a sparsity-inducing L1-regularization to identify the most informative feature subsets of a high-dimensional (e.g., 2048 in case of InceptionV3) feature vector. In combination with gradient maps [21], we use this subset of features to identify the regions of interest, the so-called saliency map, for an input image. Afterward, we estimate with k-means clustering the spatial extent of coherent regions based on the identified saliency map and place bounding boxes around each region. The image patches of these bounding boxes serve as an unsupervised part representation, i.e., each region corresponds to a single part. These detected parts are finally used as additional input for the CNN classifier. We refer to our previous work [10, 11] for more detail about the method and implementation details.

5. Baseline results

We performed detection and classification experiments to produce the first baseline results on the presented datasets. Species classification is only done on the EU-Moths dataset since the NID dataset has only bounding box annotations so far. For evaluating the species classifier, we utilized the ground-truth bounding box annotations and only used the cropped image patches as inputs.

We repeated each experiment ten times and provided in Tables 1 and 2 the mean and standard deviation of the evaluation metrics across different runs. We fine-tuned all models for 60 epochs and L2-regularization with a weight decay of 5×10^{-4}. For both models, we utilized standard image augmentation methods: random cropping, random horizontal and vertical flipping, and color jittering (contrast, brightness, and saturation).

The SSD model was trained with an AdamW [16] optimizer and a learning rate of 1×10^{-3} for all epochs. We used the VGG16 [22] backbone architecture pre-trained on the ImageNet [19] with an input size of 300px for the EU-Moths dataset and 512px for the NID dataset.

The classification model was trained with an RMSProp [26] optimizer with an initial learning rate of 1×10^{-4}, reduced by 0.1 after 20 and 40 epochs. Further, we utilized label smoothing [25] with a smoothing factor of 0.1. We used the InceptionV3 CNN architecture [25] with the default input size of 299px. Additionally, we used two different pre-training methods. Besides the typical ImageNet [19] pre-training, we used a pre-training

[3] https://inf-cv.uni-jena.de/nid_dataset

	MAP@0.75	MAP@0.50
EU-MOTHS	88.88 (±0.77)	99.01 (±0.09)
NID DATASET	26.19 (±5.64)	91.21 (±0.34)

Table 1. Detection results on the EU-Moths and NID datasets.

	IMAGENET PRE-TRAINING	INATURALIST PRE-TRAINING
NO PARTS	89.46 (±0.88)	90.54 (±1.10)
WITH PARTS	91.50 (±0.61)	93.13 (±0.76)

Table 2. Baseline classification results (accuracy in %) on cropped images of the EU-Moths dataset.

on the iNaturalist [28] dataset provided by Cui *et al.* [4]. Finally, we extract additional parts, as described in Sect. 4.2, and combine the predictions on these parts with the predictions on the entire image.

5.1. Insect detection

First, we report the detection performance on both datasets in Table 1 and we use mean average precision (mAP) as the evaluation metric. The precision is computed based on two different intersection over union (IoU) thresholds of the predicted and the ground-truth bounding boxes. The IoU-thresholds 0.5 and 0.75 (corresponding mAP denoted as *mAP@0.50* and *mAP@0.75*) are two typical choices used in the MC-COCO object detection benchmark [14].

Based on the results, the SSD method performs significantly better on the EU-Moths dataset. The baseline detector for the NID dataset achieves a stable mean-average precision of over 90 % for the *mAP@0.50* metric. Nevertheless, the detector performs much worse for the more precise metric, the *mAP@0.75*. A possible explanation for this may be many small insects (as seen in Figure 3), where minor discrepancies between the prediction and ground truth degrade the results. Even though we increased the input size for this dataset, the small sizes of some insects represent a challenge for the applied detection model.

5.2. Species classification

Table 2 shows the classification accuracies on the EU-Moth dataset for different setups. First, we can observe the effect of the pre-training on different datasets. Data used in the pre-training proposed by Cui *et al.* [4] is more related to the domain of insects, and we can see this benefit in our reported results.

Finally, utilizing methods for additional information extraction in the form of parts also improves the classification performance by approximately 2 %, as Table 2 shows. We achieved the best results using the part-based approach: 91.50 % and 93.13 % with ImagenNet and iNaturalist pre-training, respectively.

6. Challenges of automated insect monitoring

As we worked with the images captured by our *moth scanner*, we faced some challenges that may interest others. First, we obtained many images in a short period, and each image contains many insects of different sizes. Annotating this huge amount of gathered data is very time-consuming. Especially the manual species identification that requires expert knowledge is still ongoing, even though we use an annotation tool that supports with automatically inferred suggestions. We further encountered typical challenges in real-world datasets like a long-tailed species distribution.

In our experiments, we also observed that current state-of-the-art detection models have problems with detecting tiny objects, as mentioned in Sect. 5.1. Although the uniform background might suggest that insect detection is an easy task in this scenario, fallen leaves and dirt on the surface and the large variability of insect sizes pose further challenges. In contrast to the images recorded by citizen scientists with hand-held cameras that can focus on individual resting insects, automatically captured images of the light-based camera trap also contain motion blur and insects flying around that also (partially) occlude others.

Finally, the most challenging task is to perform the detection directly at the *moth scanner* (edge computing). As mentioned previously in Sect. 4, we plan to operate the camera traps autonomously in the field. To reduce the amount of transmitted data, it is desirable to transfer only the small image patches instead of the entire image. Unfortunately, this implies that current CNN-based state-of-the-art detection methods cannot be deployed on the hardware of the *moth scanner* with limited computational power and restricted energy budgets. Hence, we will need to consider detection methods based on basic computer vision algorithms, like the blob detector presented by Bjerge *et al.* [1].

7. Conclusions

In this paper, we presented a prototype of an automatic light-based camera trap for monitoring nocturnal insects. The so-called *moth scanner* allows for capturing large-scale image datasets that can be used for moth localization and fine-grained species recognition. Hence, this application domain can become an important area for research on fine-grained recognition with a large impact on ecology. Besides the presented datasets, we also provided baseline results of a two-stage pipeline for detecting and classifying insects in images.

References

[1] Kim Bjerge, Jakob Bonde Nielsen, Martin Videbaek Sepstrup, Flemming Helsing-Nielsen, and Toke Thomas Høye. An automated light trap to monitor moths (lepidoptera) using computer vision-based tracking and deep learning. *Sensors*, 21(2):343, 2021. 2, 4

[2] Gunnar Brehm. A new led lamp for the collection of nocturnal lepidoptera and a spectral comparison of light-trapping lamps. *Nota lepidopterologica*, 40:87, 2017. 3

[3] Clemens-Alexander Brust, Tilo Burghardt, Milou Groenenberg, Christoph Käding, Hjalmar Kühl, Marie Manguette, and Joachim Denzler. Towards automated visual monitoring of individual gorillas in the wild. In *ICCV Workshop on Visual Wildlife Monitoring (ICCV-WS)*, pages 2820–2830, 2017. 2

[4] Yin Cui, Yang Song, Chen Sun, Andrew Howard, and Serge Belongie. Large scale fine-grained categorization and domain-specific transfer learning. In *Proceedings of CVPR*, 6 2018. 4

[5] Alexander Freytag, Erik Rodner, Marcel Simon, Alexander Loos, Hjalmar Kühl, and Joachim Denzler. Chimpanzee faces in the wild: Log-euclidean cnns for predicting identities and attributes of primates. In *German Conference on Pattern Recognition (GCPR)*, pages 51–63, 2016. 2

[6] Caspar A Hallmann, Martin Sorg, Eelke Jongejans, Henk Siepel, Nick Hofland, Heinz Schwan, Werner Stenmans, Andreas Müller, Hubert Sumser, Thomas Hörren, et al. More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PloS one*, 12(10):e0185809, 2017. 1

[7] Benjamin Hughes and Tilo Burghardt. Automated visual fin identification of individual great white sharks. *International Journal of Computer Vision*, 122(3):542–557, 2017. 2

[8] Dennis Jonason, Markus Franzn, and Thomas Ranius. Surveying moths using light traps: Effects of weather and time of year. *PLOS ONE*, 9(3):1–7, 03 2014. 2

[9] Christoph Käding, Erik Rodner, Alexander Freytag, Oliver Mothes, Björn Barz, and Joachim Denzler. Active learning for regression tasks with expected model output changes. In *British Machine Vision Conference (BMVC)*, 2018. 2

[10] Dimitri Korsch, Paul Bodesheim, and Joachim Denzler. Classification-specific parts for improving fine-grained visual categorization. In *Proceedings of the German Conference on Pattern Recognition*, pages 62–75, 2019. 3

[11] Dimitri Korsch, Paul Bodesheim, and Joachim Denzler. Deep learning pipeline for automated visual moth monitoring: Insect localization and species classification. In *INFORMATIK 2021, Computer Science for Biodiversity Workshop (CS4Biodiversity)*, pages 443–460, 2021. 1, 2, 3

[12] Matthias Körschens, Björn Barz, and Joachim Denzler. Towards automatic identification of elephants in the wild. In *AI for Wildlife Conservation Workshop (AIWC)*, 2018. 2

[13] Matthias Körschens and Joachim Denzler. Elpephants: A fine-grained dataset for elephant re-identification. In *ICCV Workshop on Computer Vision for Wildlife Conservation (ICCV-WS)*, 2019. 2

[14] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *ECCV*, pages 740–755. Springer, 2014. 4

[15] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *ECCV*, pages 21–37. Springer, 2016. 3

[16] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 3

[17] Bernd Radig, Paul Bodesheim, Dimitri Korsch, Joachim Denzler, Timm Haucke, Morris Klasen, and Volker Steinhage. Automated visual large scale monitoring of faunal biodiversity. *Pattern Recognition and Image Analysis*, pages 477–488, 2021. 1, 2

[18] Erik Rodner, Marcel Simon, Gunnar Brehm, Stephanie Pietsch, J. Wolfgang Wägele, and Joachim Denzler. Fine-grained recognition datasets for biodiversity analysis. In *CVPR Workshop on Fine-grained Visual Classification (CVPR-WS)*, 2015. 1, 2

[19] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015. 3

[20] Faizaan Sakib and Tilo Burghardt. Visual recognition of great ape behaviours in the wild. *arXiv preprint arXiv:2011.10759*, 2020. 2

[21] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*, 2013. 3

[22] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 3

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

[24] Cecilie S Svenningsen, Diana E Bowler, Susanne Hecker, Jesper Bladt, Volker Grescho, Nicole M van Dam, Jens Dauber, David Eichenberg, Rasmus Ejrnæs, Camilla Fløjgaard, et al. Contrasting impacts of urban and farmland cover on flying insect biomass. *bioRxiv*, 2020. 2

[25] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of CVPR*, June 2016. 3

[26] Tijmen Tieleman and Geoffrey Hinton. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, 4(2):26–31, 2012. 3

[27] G. Van Horn, S. Branson, R. Farrell, S. Haber, J. Barry, P. Ipeirotis, P. Perona, and S. Belongie. Building a bird recognition app and large scale dataset with citizen scientists: The fine print in fine-grained dataset collection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 595–604, June 2015. 2

[28] Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona, and Serge Belongie. The inaturalist species classification and detection dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8769–8778, 2018. 2, 4

[29] David L Wagner, Eliza M Grames, Matthew L Forister, May R Berenbaum, and David Stopak. Insect decline in the anthropocene: Death by a thousand cuts. *Proceedings of the National Academy of Sciences*, 118(2), 2021. 1

[30] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The Caltech-UCSD Birds-200-2011 Dataset. Technical Report CNS-TR-2011-001, California Institute of Technology, 2011. 2

[31] J. Wolfgang Wgele, Paul Bodesheim, Sarah J. Bourlat, Joachim Denzler, Michael Diepenbroek, Vera Fonseca, Karl-Heinz Frommolt, Matthias F. Geiger, Birgit Gemeinholzer, Frank Oliver Glckner, Timm Haucke, Ameli Kirse, Alexander Klpin, Ivaylo Kostadinov, Hjalmar S. Khl, Frank Kurth, Mario Lasseck, Sascha Liedke, Florian Losch, Sandra Mller, Natalia Petrovskaya, Krzysztof Piotrowski, Bernd Radig, Christoph Scherber, Lukas Schoppmann, Jan Schulz, Volker Steinhage, Georg F. Tschan, Wolfgang Vautz, Domenico Velotto, Maximilian Weigend, and Stefan Wildermann. Towards a multisensor station for automated biodiversity monitoring. *Basic and Applied Ecology*, 2022. 1

[32] Xinyu Yang, Majid Mirmehdi, and Tilo Burghardt. Great ape detection in challenging jungle camera trap footage via attention-based spatial and temporal feature blending. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019. 2