Creating datasets of moth morphology and behaviour from textual sources with large language models

Bartolome Ortiz-Viso[1,2,3][0000−0003−2181−0734], Jenna L. Lawson[2][0000−0001−5166−5510], and Tom August[2][0000−0003−1116−3385]

[1] Department of Computer Science and Artificial Intelligence, CITIC-UGR (Research Center for Information and Communication Technologies), University of Granada, Spain
[2] UK Centre for Ecology and Hydrology, United Kingdom
[3] Corresponding author: `bortiz@ugr.es`
`{barotr,tomaug,jenlaw}@ceh.ca.uk`

Abstract. The integration of language models into ecological workflows is opening new possibilities for automated species monitoring. Classification systems are especially relevant in this context, as the high volume of data generated by automated systems requires efficient tools to support expert curators. Multimodal approaches, which incorporate textual information alongside visual or acoustic data, have shown potential to improve classification performance and interpretability. However, for many insect taxa, structured and usable textual descriptions remain scarce or difficult to access. In this work, we present a tool for retrieving and merging textual information about moth species from official repositories and citable sources. The resulting descriptions can be used to enrich multimodal classification models across different taxonomic levels or to build structured databases for species comparison and discovery.

Keywords: Computational Entomology · Automated insect monitoring · Multimodal category discovery · Large Language Models · Biodiversity.

1 Introduction

In recent years, there has been significant growth in the development of automated insect monitoring systems [8]. These technologies are made possible by recent advances in image-based machine learning models and low-power electronics, which together enable the construction of specialised devices designed specifically for this task [11]. This trend reflects the growing need to understand the dynamics of insect populations and how they are being affected by multiple environmental factors, including climate change [26].

Camera-based automated systems for monitoring insects have proven to be highly effective in generating large volumes of insect monitoring data, which has in turn sparked a wide range of questions and applications in the field of computer science [9][12][20]. These range from how we collect and store ecological

data to how we design robust algorithms for species classification or detecting rare species within datasets.

The vast amount of data collected by such systems presents a considerable workload in terms of identifying the species in each image. This highlights the urgent need for machine learning tools that can alleviate the burden on taxonomists — experts who are already in short supply [13] — by identifying species where the model is able, and referring others that require expert attention.

In this context, classification systems play a crucial role in helping us to make sense of complex datasets. Among these, multimodal systems are particularly promising, as they enable the integration of multiple types of data, such as text and images, within a shared feature space. This not only facilitates the interpretation of machine decisions but also supports clustering and labelling processes, even when the exact taxonomic classification of a specimen is unknown.

Deploying multimodal classification systems in entomology requires the creation of high-quality multimodal datasets — a task that presents unique challenges. Despite the availability of trait and abundance data, textual descriptions of insects —covering features such as colour, shape, and size — remain scarce, fragmented, and complex to process automatically. This contrasts with domains such as ornithology, where rich textual data resources are already available [16]. In addition to the value of trait data for supporting multimodal classification models, trait data are also crucial for understanding the role of species in ecological systems. The notable lack of trait data for insects is highlighted [14][27].

In this work, we present our experience building a pipeline that leverages large language models (LLMs) to guide the extraction, standardisation, and formatting of species descriptions into usable forms for moth species. This contribution closes the gap between the potential of computational methods and the current limitations of available entomological data.

This paper is structured as follows: After the Introduction, Section 2 provides an overview of related work from a computational perspective, highlighting how textual information can support multimodal approaches in tasks such as classification and novel category discovery. In Section 3, we present our work on UK moths. Some experiments derived from the created dataset are given in Section 4, along with a brief discussion. Finally, we conclude the paper in Section 5 by reflecting on the outcomes of our work and outlining the next steps.

2   Related works

Image classification models continue to advance rapidly, with increasing capabilities being integrated into everyday applications. These models have recently been enhanced by the development of new multimodal architectures—a branch of machine learning that enables the joint representation of visual and textual data within a shared latent space. This shared space not only facilitates explainability but also supports cross-modal applications, such as image-to-text and text-to-image generation [7].

One of the most widely used models in this area is CLIP [18], which employs a contrastive learning framework. In this architecture, the model learns to align images and their corresponding textual descriptions by bringing similar pairs closer in the latent space and pushing dissimilar ones apart. This approach has led to significant improvements, particularly in the task of novel category discovery [23,25], where the goal is to identify previously unseen categories in an open-world setting. This discovery-type scenario has clear implications in ecology, although it is not widely known and therefore not integrated into the automatic insect monitoring studies.

A multimodal perspective offers several advantages: it enables more targeted and guided training, improves generalisation, and enhances the interpretability of resulting clusters. While CLIP includes a relatively small text transformer by default, longer and more informative descriptions can be utilised through extensions such as LongCLIP [29], which is better suited to handling detailed textual data. In addition, there are now CLIP models trained on natural data as BIO-CLIP [22] and InsectFoundation [24] that have shown the considerable potential these models have in the taxonomic / nature domain, creating a good starting point for fine-tuning tasks and other modalities alignment [28]. The trade-off, however, is that these models require a reliable source of textual descriptions corresponding to the images. While such descriptions often exist in the form of trait annotations and identification keys, they are not readily available in structured datasets suitable for machine learning. This is why studies like the present one contribute by proposing methods to generate and curate such datasets, serving as a foundation for the development of more accurate and practically useful multimodal resources.

3   Methodology

Our workflow is designed to extract, normalise, and synthesise textual descriptions of moth species from various online sources, supporting multimodal classification and structured trait extraction. The process involves several steps described below. A summary is presented in Figure 1. We started with a specific list of species based on a particular location (in this case, the United Kingdom) and, based on that, searched for databases of descriptions for UK and northern European species, as the overlap between those could offer a good complement. This is a key step that may vary between different countries and is also the main challenge to overcome, as we will discuss in Section 5.

3.1   Selection of Web Sources

We identified six web-based sources containing textual information about moth species/families/genus descriptions. These services include both official repositories and citable biodiversity platforms. The selected sources are:

- **Wikipedia** [4]: Wikipedia is a widely accessible and citable source that contains a substantial amount of biological information. It offers broad taxonomic coverage and is often the first point of reference for many users.
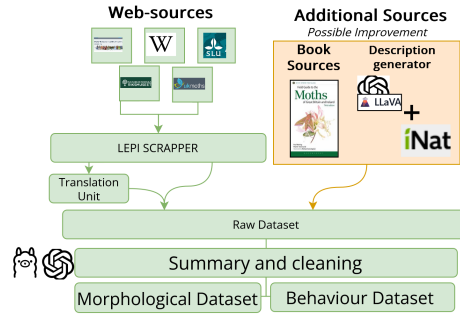
Fig. 1. Overview of the proposed workflow for textual data extraction, processing, and trait structuring.

However, its content is contributed by volunteers, and thus, the accuracy of descriptions can vary. While many pages do include cited references, there is a clear bias toward more popular or well-known species. Moreover, the textual information on species pages is typically unstructured, making it difficult to extract relevant data in a consistent manner. At higher taxonomic levels (e.g., genus or family), Wikipedia provides a high number of positives in terms of species coverage. However, the quality of this information tends to be low, as these pages often consist of long enumerations of species names without any accompanying morphological or behavioural descriptions.

– **Butterflies and Moths of North America**[2]: BAMONA is a curated online resource providing species information, distribution data, and user-submitted observations of Lepidoptera across North America. One of its strengths is the availability of concise species-level descriptions, often accompanied by high-quality images and confirmed locality records. Although its primary focus is North America, BAMONA can also be useful for particular species from Latin America, particularly those found in Central America. This is due to overlapping distributions and the presence of some records extending beyond the U.S. and Canada. Additionally, when species-level descriptions are not available, it is often possible to retrieve general information at the genus or family level. The data is curated by several volunteers mentioned in the acknowledgements [1], and some of the identification keys come from three specific manuals that need to be cited [6] [3] [17]

– **UKMoths**[10]: UKMoths is a reference website for moths in the United Kingdom. It provides species profiles that include a variety of information such as life cycle, host plants, and physical appearance. However, the content is unstructured and not all entries contain the same fields, making text processing essential for extracting relevant data.

– **Animal Diversity**[15]: Animal Diversity offers species-level content with structured sections including descriptions and various biological data. Since much of the content is created by students, it is not always curated by domain experts. The platform is maintained by volunteers and is based in

the USA, which introduces both geographical and popularity-related biases. Nevertheless, for species with complete entries, the site offers a wealth of well-organised information.

– **Swedish Museum of Natural History (NRM)**[5]: The NRM provides a repository containing textual and specimen-level information for species housed in the museum's collections. For some species, general descriptive information is also available. We discovered this resource by following citation links from other biodiversity and abundance databases, where records of these specimens are maintained. There is an obvious geographic bias, but it has an insightful description of a wide selection of species.

– **Artfakta − SLU: Swedish Species Information Centre**[21]: Artfakta is another Swedish biodiversity repository frequently referenced in Scandinavian specimen abundance databases. The platform offers robust services and provides an API through which textual information, primarily written in Swedish, can be accessed. This facilitates automated extraction, and since experts curate the content, it is well-structured and organised. When descriptive text is available, it is typically easy to locate. Although the resource focuses on species native to Northern Europe, it has proven to be a valuable source of high-quality data.

3.2   LLM-Guided Summarization

In this section, we explain our pipeline step by step, highlighting the key aspects of each step. A diagram of this is shown in Figure 2 for a specific species.

1. **Grouping and Translation** The collected texts are grouped by species, genus, or family to facilitate subsequent processing. Descriptions written in languages other than English are translated using a machine translation system [4] that is based on the DeepTranslator Python library[5].

2. **Merging**
   The translated descriptions are merged and cleaned to remove formatted content (such as species lists, which are commonly found in Wikipedia) and special characters. From webpages that have a fixed schema, description fields are extracted. For those which do not have it, regular expressions are used to search for the most similar header to the description. This step ensures that most irrelevant information is discarded for subsequent summarisation, reducing the text size.

3. **Summarisation** We employ a large language model (LLM) with tailored prompts (see Appendix **??** for a full description of the prompts used) to generate three types of descriptions:

   – **Morphological description**: covering antennae, body structure, colouration, and other visual features.

---

[4] We avoided using LLMs for translation due to inconsistent performance on technical entomological vocabulary, though they remain a possible alternative.

[5] https://github.com/nidhaloff/deep-translator

- **Behavioural description**: including months of activity, larval behaviour, and host plant preferences.
- **Visual-only morphological description**: a refined version of the morphological description, restricted to strictly visual traits, intended for use in downstream multimodal models.

The model is guided by the specific keywords we encounter when analysing the descriptions (forewings, wingspan, antennae, abdomen, head), as well as some examples of the expected outcome (premade descriptions and the expected outcome).

4. **Dual textual datasets**

   The output of the LLM is assembled into a dataset containing cleaned and standardised descriptions per species, one for the morphological traits and the other for the behavioural characteristics (more broadly, including larval behaviour, eating patterns, and months). An additional human check is added to the loop to fix errors and remove short or useless descriptions that did not work correctly. The model is instructed to state when no info can be found, so short strings, such as '*Sorry, information missing*', are searched and removed. The result is a dataset that could be suitable for training multimodal classification systems or conducting ecological trait analysis via unsupervised trait extraction from the data.

5. **Structured Trait Extraction**

   In a final step, we optionally guide the LLM to transform the descriptive data into structured tables capturing individual traits (e.g., wing span, colouration patterns, activity periods). This structured dataset enables several downstream applications, such as:

   - Trait-based species comparison;
   - Integration with existing ecological trait databases;
   - Semi-automated validation of extracted descriptions through comparison with known trait values.

   This dual use—as both a reference and a validation tool—adds robustness to the overall workflow and supports its integration into ecological monitoring pipelines.

3.3 Description Generation

In addition to textual sources, we incorporated a complementary approach based on visual information. Specifically, we used a pipeline that extracts and crops moth specimens from iNaturalist images. These cropped images were then processed using multimodal large language models (MLLMs), such as LLaVA and GPT-4-V, to generate descriptive captions—particularly focusing on visual traits such as coloration patterns, wing shapes, and distinctive morphological markers.

This technique is especially relevant for future applications where textual information is limited or absent. For instance, in hyper-diverse ecosystems or in museum datasets where many specimens lack detailed descriptions, this approach can provide a scalable method for generating representative visual descriptions.

Similarly, it holds promise for augmenting datasets used in training classification models, where consistent and diverse visual inputs are critical.

Moreover, this strategy opens the possibility of a feedback cycle. Once models are trained on descriptions derived from textual sources, they can be used to annotate or validate new images, reinforcing and expanding the dataset through iterative refinement and model adaptation.
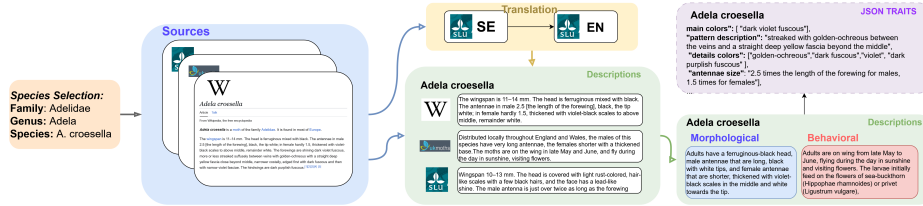

Fig. 2. Example of the pipeline for Adela croesella.

4   Results

In this section, we present the first results of our approach from the UK dataset. First, we will describe the species coverage that can be obtained from the dataset. This dataset is based on. It contains 2682 species from 213 genus, from 57 families of moths, with a variable length and distribution (see Table 1 and Figure 3 for additional details). This can be generalised to some extent to what could be found in the northern countries, in terms of similarity and species list (as some of the resources would be similar, so we can expect a similar amount of descriptions obtained for those). In the results, we present some insights into the species coverage and some results when applying textual models to the morphological descriptions in terms of similarity search. We offer insight on both aspects after the results are presented. Section 5 will outline the next steps based on the data presented here.

4.1   Species Coverage

To assess the quality and richness of the generated morphological and behavioral descriptions, we computed basic statistics across three taxonomic levels: species, family, and genus. These statistics include the average number of words, standard deviation, the number of descriptions with fewer than 20 words, and the corresponding percentage. Table 1 summarizes the results.

Overall, morphological descriptions tend to be longer than behavioural ones across all taxonomic levels. The species-level descriptions exhibit the highest average word count for morphology, while the genus-level behavioural descriptions show the highest proportion of short texts (less than 20 words).

Table 1. Descriptive statistics of morphological and behavioral descriptions across taxonomic levels.

Level	Type	Mean	Std.	Less than 20 words
Species	Morphological	57.56	38.47	19.35%
	Behavioral	42.36	18.55	9.26%
Family	Morphological	42.82	25.78	24.56%
	Behavioral	48.75	27.07	15.79%
Genus	Morphological	48.72	30.08	14.55%
	Behavioral	36.40	19.38	21.60%
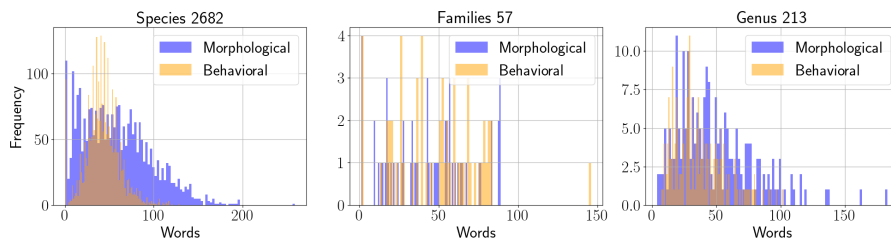
Fig. 3. Histogram of descriptions lengths for species, genus and families from the UK dataset.

4.2   Textual similarity

In addition to descriptive statistics, we evaluated the semantic similarity among the morphological and behavioural descriptions using a pretrained language model. Specifically, we employed the *all-MiniLM-L6-v2*[6] model from the *sentence-transformers* library[19], which provides a lightweight yet practical transformer-based architecture for computing sentence embeddings.

Each description was embedded into a high-dimensional semantic space, and cosine similarity was used to measure pairwise similarities between descriptions within the species range. Results shown that morphological descriptions exhibit a similarity $mean = 0.6147$, $std = 0.1271$ while Behavior similarity $mean = 0.6258$ with $std = 0.125$. Some examples are shown in Figure 4, where a random species was selected and its similarity to all other species was computed. We then selected the top 3 most similar, the bottom 3 least similar, and 3 examples with mid-range similarity.

It is important to note that this analysis was conducted using a pretrained language model that was not specifically trained on insect-related descriptions. Therefore, this represents only a preliminary exploration of how general-purpose language models interpret and compare the semantic content of these descriptions.

---

[6] https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2

From the results, we observe that species with similar description structures tend to be marked as highly similar. For instance, the top three most similar descriptions often follow a standard descriptive scheme, highlighting comparable traits such as metallic colouration, head patterning, or wing texture. In contrast, the least similar descriptions are often brief and lack detail; they frequently refer to general information, such as size or resemblance to other species, without elaborating on specific morphological features. This hypothesis was tested by comparing morphological descriptions in short texts (less than 20 words), which exhibited a significantly lower mean similarity to other descriptions ($\mu = 0.5286$) compared to longer ones ($\mu = 0.6309$). A similar pattern was observed for behavioural descriptions, where short texts averaged 0.4923 in similarity versus 0.6340 for long ones. These differences were statistically significant, as shown by the results of independent t-tests: Morphological descriptions: $t = -20.19$, $p = 3.71 \times 10^{-65}$ and Behavioral descriptions: $t = -11.98$, $p = 1.34 \times 10^{-23}$.

These findings suggest that shorter descriptions tend to be less semantically aligned with the rest of the dataset. This may be due to reduced lexical richness, a lack of detailed traits, or template-like phrasing, which limits their usefulness in tasks that require semantic comparison.

Descriptions with mid-range similarity are typically more detailed and offer a more nuanced comparison. While not identical in content, they tend to share semantically rich elements, such as references to colours and textures commonly used in moth descriptions, which result in moderate similarity scores. These findings suggest that semantic similarity scores can reflect both structural and content-based resemblance, particularly when the descriptions are sufficiently informative and detailed. These models will enhance their discriminative power when trained alongside images in multimodal approaches, or we could further fine-tune them for a specific classification task, for example. This will improve the model's performance for those with complete descriptions, as well as for others that require further completion. At the same time, the question of a standardised or at least similar description pattern arises, and that's an interesting application of a structured trait/descriptive database, from which we can create those.

5   Conclusions and future works

The pipeline presented in this work shows a full pipeline that is able to retrieve data from textual sources and summarize it in a way is useful for multimodal approaches as well as creation of new resources (structured and unstructured). One of the main aspect that is still an open issue is the validation of the final data. At any point in the pipeline, the descriptions stay perfectly traceable from the sources, so any check can be made in order to assess the summarisation performance. Additionally, the trait database can also serve as a double-check system, where previously known traits are cross-referenced with extracted ones to verify accuracy.
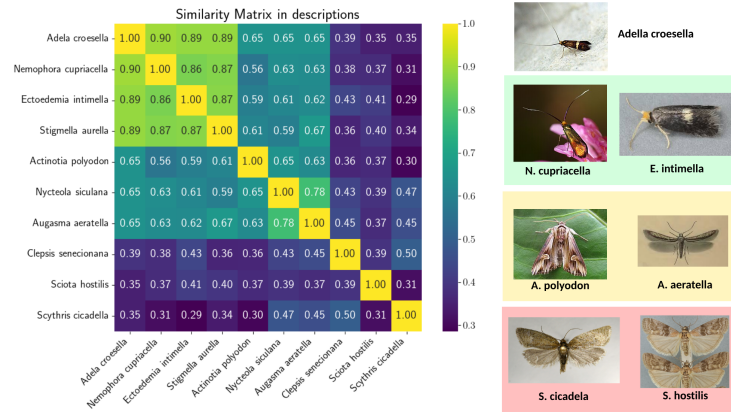
Fig. 4. Similarity matrix from Adela croesella selecting top 3, mid 3 and bottom 3 matching descriptions. Some photos of the species are highlighted in green (top 3), orange (mid-similarity), and red (low-similarity).

Another relevant issue we detected doing the similarity approach is the need ideally for a standardised collection of descriptions, as length and content may affect the similarities.

One of the most pressing issues is the lack of accessible digitalised records. Outside of Europe, structured textual descriptions of species are scarce, and many existing sources remain locked behind paywalls. Even within the UK, aside from some isolated trait datasets, there is no unified collection of morphological or behavioural textual descriptions. This gap has historically existed in part because textual descriptions were not considered machine-readable. However, this is changing: text-based digitalising is now feasible, and sufficiently curated training could even be used to generate coherent textual descriptions in reverse. This bidirectional potential underscores the importance of digitalising and semantically structuring taxonomic data.

We emphasise that biodiversity monitoring and taxonomy cannot be decoupled from domain expertise. While AI can support and accelerate specific tasks, it must not distract from the core problem: the underfunding and fragmentation of taxonomic knowledge. The models we proposed to build with this data should be designed to be open, shareable and always in close collaboration with entomologists and taxonomic experts.

Future work will involve expanding the current approach to incorporate more structured data and taxonomic nomenclature, particularly from underrepresented regions. We also plan to benchmark multimodal models against fine-tuned models with this data to assess their performance across ecological contexts (images vs text vs images + text). Ultimately, our goal is to develop adaptable, lightweight tools that serve entomologists and strengthen computational support for biodiversity conservation.

Acknowledgments. This publication is based upon work from COST action InsectAI CA22129, supported by COST (European Cooperation in Science and Technology)
    The authors would also like to acknowledge the support from:

– University of Granada Mobility plans programme.
– UK Centre for Ecology and Hydrology
– Amber Project (with its reference numbers)??

Disclosure of Interests. *The authors have no competing interests to declare that are relevant to the content of this article.*

Source code Source code, prompts and other supplementary material can be found at https://thebooort.github.io/ecodl2025-paper/

References

1. Acknowledgments | Butterflies and Moths of North America, https://www.butterfliesandmoths.org/acknowledgments
2. Butterflies and Moths of North America | collecting and sharing data about Lepidoptera, https://www.butterfliesandmoths.org/
3. Western Butterflies (Peterson Field Guide) by Opler, P.a.; Wright, A.b., https://www.pemberleybooks.com/product/western-butterflies-peterson-field-guide/35828/
4. Wikipedia, the free encyclopedia, https://www.wikipedia.org/
5. Bert Gustafsson, T.M.: Lepidoptera - Svenska fjärilar - Naturhistoriska riksmuseet, http://www3.nrm.se/en/svenska_fjarilar/svenska_fjarilar.html
6. Charles V Covell, J.: A Field Guide to Moths of Eastern North America, https://www.nhbs.com/a-field-guide-to-moths-of-eastern-north-america-book
7. Guo, R., Wei, J., Sun, L., Yu, B., Chang, G., Liu, D., Zhang, S., Yao, Z., Xu, M., Bu, L.: A survey on advancements in image–text multimodal models: From general techniques to biomedical implementations. Computers in Biology and Medicine **178**, 108709 (Aug 2024). https://doi.org/10.1016/j.compbiomed.2024.108709, https://linkinghub.elsevier.com/retrieve/pii/S0010482524007947
8. Høye, T.T., Ärje, J., Bjerge, K., Hansen, O.L.P., Iosifidis, A., Leese, F., Mann, H.M.R., Meissner, K., Melvad, C., Raitoharju, J.: Deep learning and computer vision will transform entomology. Proceedings of the National Academy of Sciences **118**(2), e2002545117 (Jan 2021). https://doi.org/10.1073/pnas.2002545117, https://www.pnas.org/doi/full/10.1073/pnas.2002545117, publisher: Proceedings of the National Academy of Sciences
9. Jain, A., Cunha, F., Bunsen, M., Pasi, L., Viklund, A., Larrivée, M., Rolnick, D.: A machine learning pipeline for automated insect monitoring (Jun 2024). https://doi.org/10.48550/arXiv.2406.13031, http://arxiv.org/abs/2406.13031, arXiv:2406.13031 [cs]
10. Kimber, I.: UKMoths | Guide to the moths of Great Britain and Ireland, https://www.ukmoths.org.uk/
11. van Klink, R., August, T., Bas, Y., Bodesheim, P., Bonn, A., Fossøy, F., Høye, T.T., Jongejans, E., Menz, M.H.M., Miraldo, A., Roslin, T., Roy, H.E., Ruczyński, I., Schigel, D., Schäffler, L., Sheard, J.K., Svenningsen, C., Tschan,

G.F., Wäldchen, J., Zizka, V.M.A., Åström, J., Bowler, D.E.: Emerging technologies revolutionise insect ecology and monitoring. Trends in Ecology & Evolution **37**(10), 872–885 (Oct 2022). https://doi.org/10.1016/j.tree.2022.06.001, https://www.sciencedirect.com/science/article/pii/S0169534722001343

12. Korsch, D., Bodesheim, P., Brehm, G., Denzler, J.: Automated Visual Monitoring of Nocturnal Insects with Light-based Camera Traps (Jul 2023). https://doi.org/10.48550/arXiv.2307.15433, http://arxiv.org/abs/2307.15433, arXiv:2307.15433 [cs]

13. Löbl, I., Klausnitzer, B., Hartmann, M., Krell, F.T.: The Silent Extinction of Species and Taxonomists—An Appeal to Science Policymakers and Legislators. Diversity **15**(10), 1053 (Oct 2023). https://doi.org/10.3390/d15101053, https://www.mdpi.com/1424-2818/15/10/1053, number: 10 Publisher: Multidisciplinary Digital Publishing Institute

14. Meier, R., Hartop, E., Pylatiuk, C., Srivathsan, A.: Towards holistic insect monitoring: species discovery, description, identification and traits for all insects. Philosophical Transactions of the Royal Society B: Biological Sciences **379**(1904), 20230120 (May 2024). https://doi.org/10.1098/rstb.2023.0120, https://royalsocietypublishing.org/doi/10.1098/rstb.2023.0120, publisher: Royal Society

15. Myers, P., Espinosa, R., Parr, C.S., Jones, T., Hammond, G.S., Dewey, T.A.: The animal diversity web (online) (2025), https://animaldiversity.org, [Accessed 2025-06-06]

16. of Ornithology, C.L.: Search, All About Birds, Cornell Lab of Ornithology, https://www.allaboutbirds.org/guide/search

17. Powell, J.A., Opler, P.A.: Moths of Western North America. Univ of California Press (May 2009), google-Books-ID: USoWEQAAQBAJ

18. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning Transferable Visual Models From Natural Language Supervision (Feb 2021), https://arxiv.org/abs/2103.00020v1

19. Reimers, N., Gurevych, I.: Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks (Aug 2019). https://doi.org/10.48550/arXiv.1908.10084, http://arxiv.org/abs/1908.10084, arXiv:1908.10084 [cs]

20. Roy, D.B., Alison, J., August, T.A., Bélisle, M., Bjerge, K., Bowden, J.J., Bunsen, M.J., Cunha, F., Geissmann, Q., Goldmann, K., Gomez-Segura, A., Jain, A., Huijbers, C., Larrivée, M., Lawson, J.L., Mann, H.M., Mazerolle, M.J., McFarland, K.P., Pasi, L., Peters, S., Pinoy, N., Rolnick, D., Skinner, G.L., Strickson, O.T., Svenning, A., Teagle, S., Høye, T.T.: Towards a standardized framework for AI-assisted, image-based monitoring of nocturnal insects. Philosophical Transactions of the Royal Society B: Biological Sciences **379**(1904), 20230108 (May 2024). https://doi.org/10.1098/rstb.2023.0108, https://royalsocietypublishing.org/doi/10.1098/rstb.2023.0108, publisher: Royal Society

21. SLU Artdatabanken: Artfakta (2024), https://www.artfakta.se, [Accessed 2025-06-06]

22. Stevens, S., Wu, J., Thompson, M.J., Campolongo, E.G., Song, C.H., Carlyn, D.E., Dong, L., Dahdul, W.M., Stewart, C., Berger-Wolf, T., Chao, W.L., Su, Y.: BioCLIP: A Vision Foundation Model for the Tree of Life (May 2024). https://doi.org/10.48550/arXiv.2311.18803, http://arxiv.org/abs/2311.18803, arXiv:2311.18803 [cs]

23. Troisemaine, C., Lemaire, V., Gosselin, S., Reiffers-Masson, A., Flocon-Cholet, J., Vaton, S.: Novel Class Discovery: an Introduction and Key Concepts

(Feb 2023). https://doi.org/10.48550/arXiv.2302.12028, http://arxiv.org/abs/2302.12028, arXiv:2302.12028 [cs]

24. Truong, T.D., Nguyen, H.Q., Nguyen, X.B., Dowling, A., Li, X., Luu, K.: Insect-Foundation: A Foundation Model and Large Multimodal Dataset for Vision-Language Insect Understanding (Feb 2025). https://doi.org/10.48550/arXiv.2502.09906, http://arxiv.org/abs/2502.09906, arXiv:2502.09906 [cs]

25. Vaze, S., Han, K., Vedaldi, A., Zisserman, A.: Generalized Category Discovery (Jun 2022). https://doi.org/10.48550/arXiv.2201.02609, http://arxiv.org/abs/2201.02609, arXiv:2201.02609 [cs]

26. Wagner, D.L., Grames, E.M., Forister, M.L., Berenbaum, M.R., Stopak, D.: Insect decline in the Anthropocene: Death by a thousand cuts. Proceedings of the National Academy of Sciences **118**(2), e2023989118 (Jan 2021). https://doi.org/10.1073/pnas.2023989118, https://www.pnas.org/doi/full/10.1073/pnas.2023989118, publisher: Proceedings of the National Academy of Sciences

27. Wong, M.K.L., Guénard, B., Lewis, O.T.: Trait-based ecology of terrestrial arthropods. Biological Reviews **94**(3), 999–1022 (2019). https://doi.org/10.1111/brv.12488, https://onlinelibrary.wiley.com/doi/abs/10.1111/brv.12488, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/brv.12488

28. Zhai, X., Wang, X., Mustafa, B., Steiner, A., Keysers, D., Kolesnikov, A., Beyer, L.: LiT: Zero-Shot Transfer with Locked-image text Tuning (Nov 2021), https://arxiv.org/abs/2111.07991v3

29. Zhang, B., Zhang, P., Dong, X., Zang, Y., Wang, J.: Long-CLIP: Unlocking the Long-Text Capability of CLIP (Jul 2024). https://doi.org/10.48550/arXiv.2403.15378, http://arxiv.org/abs/2403.15378, arXiv:2403.15378 [cs]