UNIVERSITY OF COPENHAGEN
DEPARTMENT OF COMPUTER SCIENCE

14 Apr 2023

Dear Colleague,

Title of the Doctoral Thesis: AI-Based Species Identification in the Wild
Name of the Candidate: Lukáš Picek
Name of the Reviewer: Serge Belongie

Significance of the Doctoral Thesis
The doctoral thesis addresses the development of novel computer vision methods and datasets for the automatic identification and localization of biological specimens in their natural environment. The significance of this work for the given field is evident in its potential impact on biodiversity monitoring, conservation, and snakebite mortality prevention. By tackling unique challenges related to visual similarity, varying observation conditions, unbalanced species distributions, and decision-making consequences, this thesis contributes valuable insights and techniques to the field of AI-based species identification.

Approach to Problem Solving, Methods Used, and Fulfillment of the Objective
Lukáš Picek employed a range of machine learning and computer vision methods tailored to different use cases, such as coral reef annotation, Varroa destructor infestation rate monitoring, and automatic snake, plant, and fungi species identification. The candidate has also created novel datasets to support the development and evaluation of the proposed algorithms. The thesis demonstrates a thorough understanding of the challenges involved in "in the wild" species identification and effectively addresses these issues through the proposed techniques.

Results and Original Contribution
The results of the thesis show that the newly proposed datasets and state-of-the-art deep neural network architectures provide sufficient robustness and recognition accuracy for "in the wild" species identification. Additionally, the candidate's original contributions include the development of advanced optimization strategies, novel techniques utilizing metadata, and innovative loss functions. These contributions have led to the algorithms presented in this work ranking among the top places in several international competitions focused on the automatic identification and localization of biological species.

Systematic Approach, Clarity, Appropriateness of Form and Language
The thesis is well-organized and systematically presented. The candidate has provided a clear structure with a comprehensive abstract, methodology, results, and discussion sections, which enables the reader to follow the arguments and the progression of the research. The language used in the thesis is appropriate, and the overall form is coherent and professional.

Publications
Lukáš Picek's extensive list of publications is a testament to the high quality and impact of his research work. The range of publication venues is impressive, encompassing respected machine learning conferences, such as IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), as well as domain-specific workshops and internationally recognized journals like Sensors and PLOS Neglected Tropical Diseases. Picek's work has consistently been well-received by the research community, with many of his publications amassing a significant number of citations. This not only demonstrates the strong scientific contributions of the thesis but also underlines the relevance of Picek's research to both machine learning and the specific application domain of species identification in the wild.

Recommendation
Considering the significance of the doctoral thesis, the candidate's approach to problem-solving, the results and original contributions, and the clarity and appropriateness of form and language, I enthusiastically recommend the doctoral thesis for defence.

Sincerely,

Serge Belongie
Professor and Director
Pioneer Centre for Artificial Intelligence
(+45) 93 58 87 86
s.belongie@di.ku.dk

Jan Čech
Center for Machine Perception
Department of Cybernetics, Faculty of Electrical Engineering
Czech Technical University in Prague

Doctoral Thesis Review

Prague, April 6, 2023

Title: AI-Based Species Identification in the Wild
Author: Ing. Lukáš Picek
Supervisor: prof. Ing. Luděk Müller, Ph.D.
Date received: 10/03/2023

The thesis addresses the problem of natural species identification. The methods are demonstrated on recognizing species of Fungi, Snakes, Plants, Corals, and on monitoring Varroa disease. The problem is challenging in several respects: Fine-grained categories counting thousands of species typically having high intra-class and low inter-class variance. In-the-wild setup, where images are not acquired in the controlled laboratory condition but may be shot by a cell phone in various quality, scale, orientation, and on a diverse background. The multidisciplinary nature, where the problem extends technical disciplines (computer vision/machine learning) and involves a certain level of understanding of a language and methodology of another (biology) community, is always a challenge.

The manuscript is generally well written and organized. The thesis consists of eight chapters. The introduction that presents the challenges and formulates the problems is given in Chapter 1. Chapters 2 and 3 present the Funghi recognition methods. Chapter 4 deals with the recognition of snake species. Chapter 5 provides methods on flora recognition and a comparison between recognition and retrieval based methods. Chapter 6 describes a study on monitoring the rate of infestation of Varroa mites. Chapter 7 proposes semantic segmentation to automatically annotate coral reefs. Finally, Chapter 8 concludes the thesis.

The thesis presents several novel contributions. A significant contribution is that the author proved that it is possible to identify species automatically from in-the-wild images with high accuracy. The methods developed by the thesis author won several international competitions, which confirms the state-of-the-art level of accuracy. This finding extends the impact of the thesis beyond computer vision to biology, ecology, citizen science, etc. A systematic comparison of various deep neural network architectures including classical CNNs and recent ViT and SWIN transformers, together with several recipes for training, is certainly valuable for the community. Several advanced techniques have been developed or adapted to the problem, e.g., using meta-data (GPS, substrate, etc.) with the images for recognition, estimating, and adjusting priors on the test set, introducing specialized loss functions to decrease edible/poisonous fungi confusion. Comparing k-NN semantic retrieval and classical cross-entropy recognition (for flora identification) provides excellent insight together with a nice side-effect of interpretability of the system.

The author clearly demonstrated a competence in the field. Excellent results and a quantity of methods and problems addressed are impressive. A nice bonus are practically

working systems – a mobile app for fungi recognition, or an end-to-end system for Varroa mite infestation rate monitoring.

To name a few weaknesses of the thesis:

1. The thesis certainly provides many great engineering results. On the down side, it is somewhat unclear how the findings and recipes leading to the state-of-the-art accuracy generalize to different problems. It would be great if the author writes all the universal recipes up in "the lessons learned" chapter.

2. More elaborate ablation studies are sometimes missing on various factors. For instance, the focal loss and a standard cross-entropy loss, in Sec. 4.3.1., should be compared. It is unclear how important are pre-training of the deep net models (Sect. 2.4.1, 2.4.3), or using super-resolution in Varroa monitoring system (Sec. 6.2.5).

3. The statistical significance of the proposed improvements is never tested. All results show absolute error scores without estimates of error variance over, e.g., training/test splits, random initialization of the model, etc. For some results, the improvement is only by units/fraction of percent points, and the significance remains unclear.

4. Certain details are missing in the description of the experiments. For example: Human in the loop in Chapter 3 should be detailed, or weak (noisy) labels in Sec. 4.3.2 should be presented (size of the dataset, protocol, cleanliness estimation, etc.). The details on noisy labels in Sec. 5.4.4 are missing, as well as pseudo labels in Sec. 7.2 are reported only they were "added sensitively".

5. Related research questions could have been further investigated. For instance, in case of Poison Loss defined Eq. (3.5), it is unclear why using it simultaneously with the standard CE-loss improves Top1 accuracy, as seen in Tab. 3.12. Could edible/poisonous recognition be implemented as a Neyman-Pearson problem? Is there a way to provide some confidence score that could be used for a reject option? It might be interesting to investigate a sequential analysis where a user is asked to provide other photos, e.g. capturing a specific part of the specimen to resolve ambiguity. I would have appreciated an attempt to visualize the distinguishing features that the deep net classifier learned. See the "Questions for the defense" section of the review.

6. The text is sometimes repeated within the thesis. For example, Eq. (1.2) and (3.3) are identical, and Eq. (3.1) and (5.1) are very similar. Discussion on challenging high intra/low inter class variability occurs multiple times. There are missing images in Fig. B.2, B.3, B.4 in Appendix B.

Nevertheless, the weaknesses listed above are not so important and do not compromise overall high-quality of the thesis. The author published his work regularly (3 impacted journals, 3 conferences, and several workshop papers). The publications are cited 251 times (based on Google Scholar, in April 2023), which is exceptional. The author reports that he got first place in 9 and second place in 1 international competitions/challenges since 2018, which is utterly outstanding. All these achievements confirm the impact of the research on the community.

In conclusion, I do recommend the thesis for presentation with the aim of receiving the Ph.D. degree.

Ing. Jan Čech, Ph.D.

Questions for the defense

1. Concerning the poison loss in Eq. (3.5), it is unclear why the convex combination with a standard CE loss iproves the Top1 accuracy over the CE loss only. The table 3.1.2 should go ad extremum, by increasing w_{nll} until Top1 accuracy drops.

2. Can you formulate the edible/poisonous species recognition as an instance of the Neyman-Pearson problem? Did you consider an alternative definition, where the toxicity is not a binary property, but there exist different levels of toxicity, e.g., based on LD50 units?

3. How would you obtain a confidence score for the class decision that would be used for a reject option? The system should warn the user that classification is not reliable, due to a potential confusion between classes or a low quality of input data. In such a case, the system might demand a user to provide additional images. This could be implemented as a sequential analysis, where if the decision based on an input is not conclusive, the system would demand another image, at best of a specific part of the specimen. How would you implement this functionality?

4. Combining meta-data (location, substrate, etc.) with the images is implemented by a late fusion, assuming statistical independence. How exactly were the softmax outputs calibrated for this problem? Is there a way to implement the combination as an early-level fusion or mid-level fusion instead?

5. Chapter 5 on Flora Recognition compares a retrieval-based semantic embedding system with the softmax CE-loss trained recognition network. Could you comment on the prospects of a combination of these two approaches?

6. For the trustworthiness of the system, the interpretability of the results is very important. The only steps in this direction are made by proposing the retrieval system, where a user gets semantically similar images with a ground-truth label for a querry image. This is surely helpful, but is there a way to visualize what a trained deep net actually learned to distinguish certain species? Mycologists know the typical morphological differences between confusing classes. Is it true that the same classes that are easily confused for humans are confusing for the model too? Do you have any observation on that?

7. Recent approach to recognition of various classes is a zero shot prediction based on CLIP (OpenAI's model that connects text and images). CLIP was trained on a giant dataset and seems to understand broad semantic text and image concepts. Could it be an alternative for fine-grained species recognition, or only for the most common species, or there is not enough semantic granularity in the model?