Towards ML Methods for Biodiversity: A Novel Wild Bee Dataset and Evaluations of XAI Methods for ML-Assisted Rare Species Annotations

Teodor Chiaburu[1], Felix Bießmann[1, 2] and Frank Haußer[1]

[1)]Berlin University of Applied Sciences, [2)]Einstein Center Digital Future

Motivation

When delegated to assist entomologists in classifying insect species, AI systems need to deliver reliable explanations. However, **F**ine **G**rained **I**mage **A**nalysis (**FGIA**) is a challenging task, in particular combined with XAI.

Contribution

1) dataset of annotated images of wild bees (*body parts segments*)

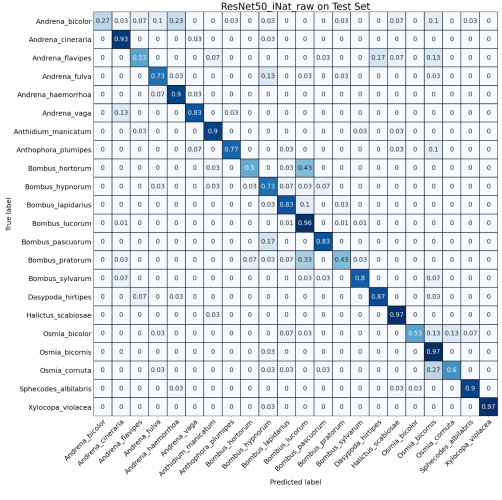
2) ResNet achieving **0.78 top-1 and 0.95 top-3 accuracy** in classifying 22 species

3) an investigation of XAI methods to support biologists in annotation tasks

New Annotated Dataset for FGIA & XAI

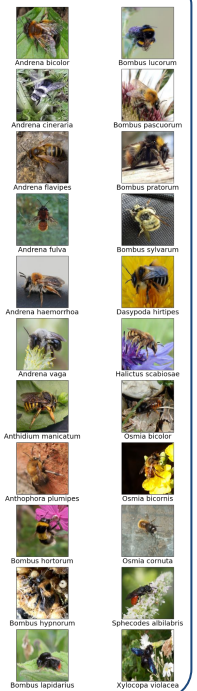
Raw vs. manually segmented (head, thorax, abdomen)

(a)

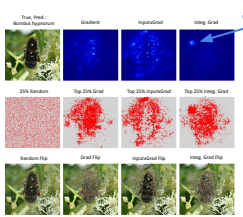


Confusion matrix for the ResNet50 [3] on the test set along with prototypical samples for every wild bee species.

(b)



Sample-wise XAI Evaluation: Pixel Flipping



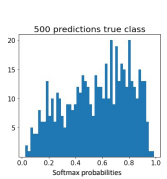
Pixels are masked iteratively according to relevance scores provided by saliency maps. Random flipping serves as baseline.

(c)

After every flipping step, a new inference is run with the model. It is expected that salient flipping decreases accuracy scores sooner and faster than random flipping.

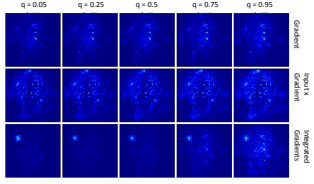
(d)

XAI Robustness: Monte Carlo Dropout



Monte-Carlo Dropout (MCD): activating the Dropout layer during inference produces a (true) class probability distribution instead of a point-estimation.

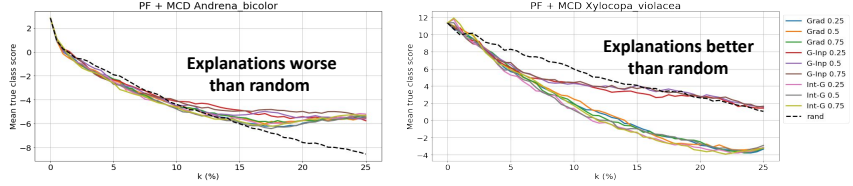
(e)



For every MCD prediction, new saliency maps are computed. Quantiles of the maps' distribution offer an idea of the uncertainty of explanations.

(f)

Class-wise XAI Evaluation: Pixel Flipping & MC-Dropout



Averaged Pixel Flipping on 100 MCD inferences for whole classes on the test set according to three quantile maps for three saliency methods each + random flipping [4].

(g)

Localisation Metrics for XAI

Metric	XAI Method		
	Gradient	Grad. x Input	Int. Grad.
Pointing Game	**0.9463**	0.9421	0.9298
Attribution Localisation	0.2984	0.2551	**0.3168**
Top-K Intersection	0.3320	0.2807	**0.3547**
Relevance Rank Accuracy	0.3047	0.2587	**0.3231**
AUC	0.7157	0.6671	**0.7297**

Evaluation of three XAI methods according to five localisation metrics [2]: **'coarse grained'** vs **'fine grained'**. The attribution maps were computed on the annotated test set for the ResNet50 w.r.t. the predicted label.

(h)

Summary & Outlook

- Evaluation of XAI methods on fine grained tasks is challenging
- Are poorly localized or unfaithful explanations also bad explanations?
- What are optimal XAI metrics given model, dataset and explanation?

References

[1] iNaturalist Database: https://www.inaturalist.org/

[2] Hedstrom et al., *Quantus: An explainable AI toolkit for responsible evaluation of neural network explanations*, 2022.

[3] Horn et al., *Benchmarking representation learning for natural world image collections.* In CVPR, 2021.

[4] Bykov et al., *How much can I trust you? – quantifying uncertainties in explaining neural networks*, 2020.

GitHub Repo