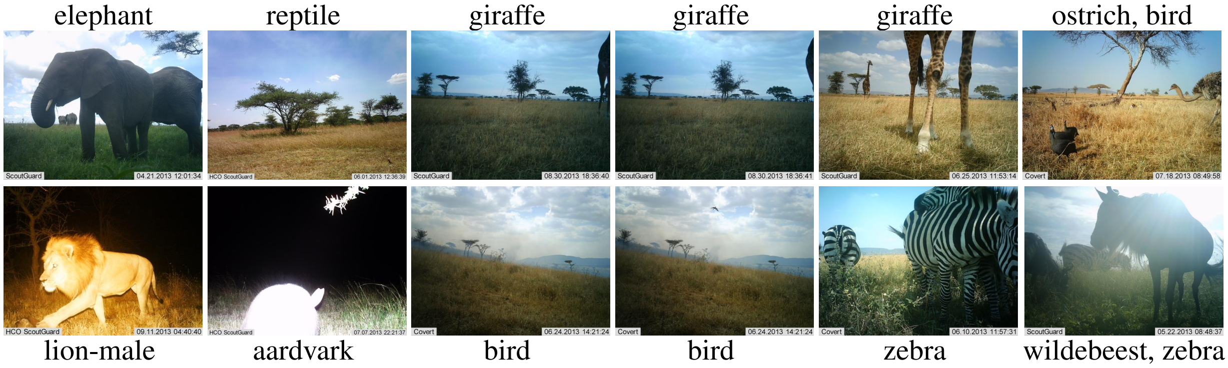
FAKULTA APLIKOVANÝCH VĚD
ZÁPADOČESKÉ UNIVERZITY
V PLZNI

STUDENTSKÁ VĚDECKÁ KONFERENCE 2021

# Hakuna Ma-Data – Animal Detection in the Wild

Lukáš Picek[1]




## 1 Introduction

Camera traps are essential for ecological research and conservation efforts; however, dealing with millions of images generated by camera traps every year is a tedious and time consuming task. For this reason, a system that can automatically analyze camera trap images while maintaining humans' accuracy would save years of manual work and unlocks new opportunities to study biodiversity and conservation. To address this issue, we propose a straightforward, lightweight, fast, and efficient method based on our winning submission to the *Hakuna Ma-data - Serengeti Wildlife Identification challenge* – Valan et al. (2020).

## 2 Methodology

The proposed system is based on 3 CNN architectures that were further fine-tuned from the publicly available ImageNet checkpoints. Namely: SE-ResNext-50, EfficientNet-B1 and EfficientNet-B3. All architectures share the Learning Parameters and used One Cycle Learning Policy. In contrast to standard training procedures, we decided to perform the training with just a random Horizontal Flip augmentation and trained the network just for one epoch. To address the multi-label problem we used a sigmoid function in the last Linear Layer of the CNN. Such activation function enables a "detection" of multiple animal categories in one image in a single feed-forward pass. Our most complex and best performing system was a weighted ensemble of 4 models (2*SE-ResNext-50, EfficientNet-B1, and EfficientNet-B3) with a total of 6 forward passes per test image. Two additional forward passes are generated from horizontally flipped image from both SE-ResNext-50 models. Prediction logits for each image were combined with geometrical mean. For sequences with two and more images, we obtained the final predictions by calculating the arithmetic mean of the predicted logits.

[1] PhD. student, University of West Bohemia, Faculty of Applied Sciences, Department of Cybernetics

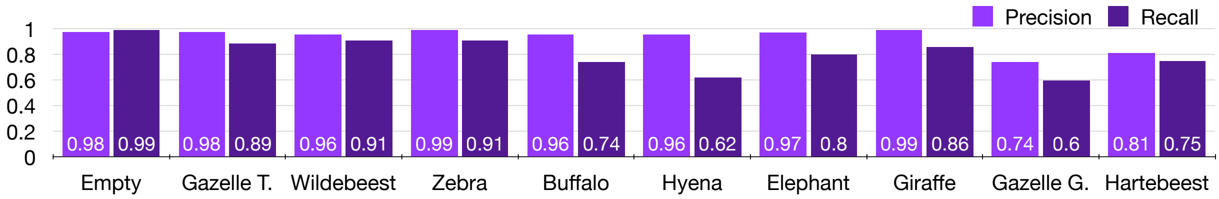



**Figure 1:** System performance on the 10 most common categories in the competition test set.

## 3 Results

The proposed system achieved the best performance in terms of Aggregated Logarithmic Loss in both the public and the private leader-board while outperforming other 811 participants. We further compared our system against previously published works on the earlier version of the dataset. Our system shows significant improvements (Table 1), reducing the previously reported error rates. Besides, our system performed better than human annotators, trailing only to the human-in-the-loop approach. At the same time, it achieved the state-of-the-art accuracy of 94.3% in recognizing 53 animal categories. Additionally, the proposed method based on one lightweight EfficientNet-B1 model achieved an accuracy comparable to human-level performance. Specifically, our model scored an accuracy of 96.5%,where humans achieved 96.6%. The source code and models' weights are freely available at *GitHub*.

| Study | Acc | Empty Acc | Epochs |
|---|---|---|---|
| Willi et al. (2019) | 93.4% | 96.0% | 70 |
| Norouzzadeh et al. (2018) | 93.8% | 96.8% | 50 |
| **(ours)** | **94.3%** | **97.0%** | **1** |
| Human – Swanson et al. (2016) | 96.6% | 96.6% | - |
| Human-Machine – Willi et al. (2019) | 99.8% | 99.8% | - |

**Table 1:** Comparison to other studies.

### Acknowledgement

LP was supported by the University of West Bohemia grant, project No. SGS-2019-027.

## References

Willi, M., Pitman, T., Cardoso, A. W., Locke, Ch., Swanson, A., Boyer, A., Veldthuis, M., Fortson, L. (2019), Identifying animal species in camera trap images using deep learning and citizen science, in *Methods in Ecology and Evolution*, pp. 80-91.

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M., Packer, C., Clune, J. (2018), Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning, in *Proceedings of the National Academy of Sciences*, vol. 115.

Swanson, A., Kosmala, M., Lintott, Ch., Packer, C., Clune, J. (2018), A generalized approach for producing, quantifying, and validating citizen science data from wildlife images, in *Conservation Biology*, vol. 30, no. 3, pp. 520–531.

Valan, M., Picek, L. (2020), Mastering Large Scale Multi-label Image Recognition with high efficiency over Camera trap images, in *arXiv preprint arXiv:2008.07828*.