Fine-grained Recognition Datasets for Biodiversity Analysis

Erik Rodner[1], Marcel Simon[1], Gunnar Brehm[3], Stephanie Pietsch[4]
J. Wolfgang Wägele[4], Joachim Denzler[1,2]

[1]Computer Vision Group, [2]Michael Stifel Center Jena, [3]Phyletisches Museum
Friedrich Schiller University Jena, Germany
[4]Zoological Research Museum Alexander Koenig, Bonn, Germany

Project website: http://www.inf-cv.uni-jena.de/fgvcbiodiv
Datasets available at: http://www.github.com/cvjena/{costarica,ecuador}-moths/

Abstract

In the following paper, we present and discuss challenging applications for fine-grained visual classification (FGVC): biodiversity and species analysis. We not only give details about two challenging new datasets suitable for computer vision research with up to 675 highly similar classes, but also present first results with localized features using convolutional neural networks (CNN). We conclude with a list of challenging new research directions in the area of visual classification for biodiversity research.

1. Introduction

Fine-grained visual recognition of birds and animals has come already a long way in the last years, starting from 10% recognition rate on the CUB200-2011 bird dataset in 2011 [10] to 85% recently achieved by [2]. Despite its obvious use as a benchmark for computer vision techniques, we argue that there is indeed a huge application potential for these approaches in the area of biodiversity research.

Currently, visual recognition techniques or even image analysis tools are rarely used by biologists, although an enormous amount of expert annotation is required to build large image datasets such as the ones of [3] and [5]. These datasets provide examples of highly diverse but poorly known tropical insect communities, which represent an important fraction of global biodiversity and which are functionally important in complex and endangered forest ecosystems. Furthermore, the datasets are important for understanding the changes of species composition in ecosystems caused by climate change and deforestation. Even when the majority of species are still unknown (as typical for tropical forests), visual discrimination allows inventorying for the goals of conservation biology. Therefore, there is a need for automated vision systems which are able to assist experts with discrimination and annotation as well as with systematic and quantitative analysis of species differences.

Interestingly, the expert-labeled datasets of [3] and [5] show that issues remain in fine-grained recognition which might have been underestimated by computer vision researchers; such as the lack of large-scale training data or detailed annotations as well as the need for approaches providing plausible models and visual features that can be interpreted by biologists and other experts. While we are briefly discussing several of these challenges at the end of the paper, we first introduce the datasets of [3] and [5], which we prepared for FGVC research, as well as results we were able to obtain with current techniques.

2. New FGVC biodiversity datasets

In the following, we present two datasets (Figure 2), which are ready to use for computer vision researchers. All images show moths and butterflies with artificially spread wings. While uncommon in natural photos, this is the way animals are prepared for scientific collections to expose the features of the hind wings, which normally are covered by the anterior wings in living specimens. In both datasets, species sorting was achieved by a combination of traditional sorting by specialists, according to external characters, and the use of so called DNA barcoding, *i.e.* the use of a standardized gene fragment of the mitochondrial gene which allows delineating species even in difficult, cryptic, and small taxa [8].

Ecuador moth dataset [3] The dataset of [3] includes only one single family of moths (Geometridae) quantitatively collected in montane tropical rainforests in southern Ecuador, the global diversity hotspot of this taxon. Our dataset covers 675 observed and genetically verified species in the area. It includes many closely related and look-alike species, most of them unknown to science, and is therefore particularly challenging. Since expert knowledge on these moths is very scarce, automated image analysis could

dataset	#classes	#images	#images for training	accuracy (global)	accuracy (pyramid)
Ecuador moth dataset [3]	675	2120	1445	55.7%	53.5%
Costa rica dataset [5][1]	331	2310	990	79.2%	81.3%

Table 1: Categorization results for the two biodiversity datasets (butterflies and moths) of [3] and [5].
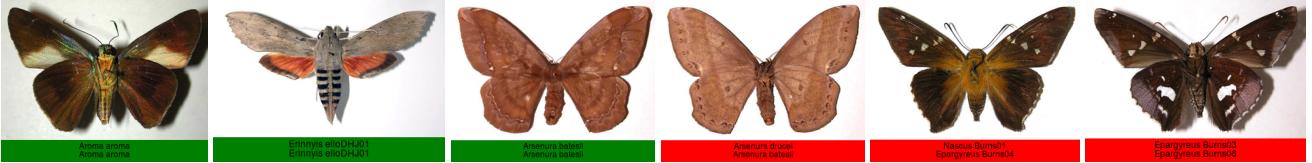
Figure 1: Example classification results for the Costa Rica dataset (input image, predicted label, ground-truth label). Images are directly obtained from [5] and have the following identifiers: 00-SRNP-1311-DHJ33001, 00-SRNP-1536-DHJ95316, 00-SRNP-4253-DHJ36384, 00-SRNP-4253-DHJ36385, 01-SRNP-16434-DHJ305668, 03-SRNP-20073-DHJ91439.

substantially contribute to species-sorting by untrained persons, or to monitoring schemes in endangered habitats. The images have been taken in a controlled environment with uniform background and canonical poses, which makes it easy to focus feature extraction on the important parts of the image. Since the dataset only includes a few images per species, we use male and female individuals within one category. We will release a challenging subset of the dataset to the public.

Costa rica dataset [5] The dataset of [5], derived from long-term sampling and caterpillar rearing, includes a broad range of moth and butterfly taxa sampled in north western Costa Rica. Since we have a larger initial dataset, we reduced it to female individuals only and species with at least 2 images. The dataset is already publicly available and we plan to release converted meta data and links to ease its use for the computer vision community. Furthermore, a large part of it is already linked in the encyclopedia of life database [1], where additional meta information is likely to be published in future.

3. Global and pyramid-based CNN baseline

How well do current vision technologies perform on the datasets presented? Since most of the animals in the images of both datasets are already aligned, we computed global CNN features with AlexNet [6] (caffe reference network) using layer `pool5` and used a one-vs-all linear SVM for classification. For the Ecuador dataset [3], all except of one randomly selected image for each category has been used for training. Learning on the Costa Rica dataset was done with up to three training examples for each category.

Table 1 gives the accuracies for each of the datasets. At a first glance, although the number of classes is ex-

tremely high, we are able to achieve reasonable accuracies. The dataset is far more challenging than the Leeds butterfly dataset of [11] with 10 categories, where we are able to obtain an accuracy of 99.24% with the same techniques.

To focus on more subtle differences in just a few parts (different colors of parts of the wing for example), we calculate a spatial pyramid with two levels using CNN features. First, global features for the whole image are calculated. Then the image is divided into four equal-sized subregions and all features are concatenated. The spatial pyramid helps to improve the accuracy by 2.1% for the Costa Rica but not for the Ecuador dataset (Table 1). *Please note that both datasets contain a certain dataset bias, which is discussed in more detail on the project website (see header).*

4. Conclusions and upcoming challenges

As we have seen in the brief description of our first experiments, vision algorithms can already obtain a suitable accuracy for challenging species identification tasks. However, automated classification is not the only research direction in the area of computer-assisted biodiversity research and we list a few upcoming challenges:

1. Open-set recognition for counting known species and automatically detecting novel ones: biologists and citizen scientists need tools that allow them to detect animals that are likely going to belong to a new species. This would allow for a certain pre-filtering of animals prior to comprehensive DNA barcoding analysis. Furthermore, it could be also used to derive quantitative measures for biodiversity research [1].

2. Incorporating human-machine interaction not only for active classification [9] and learning [4]: There is a lot of expert knowledge already available which should be used to develop new models or actively guide the search for relevant features during learning.

---

[1] http://eol.org
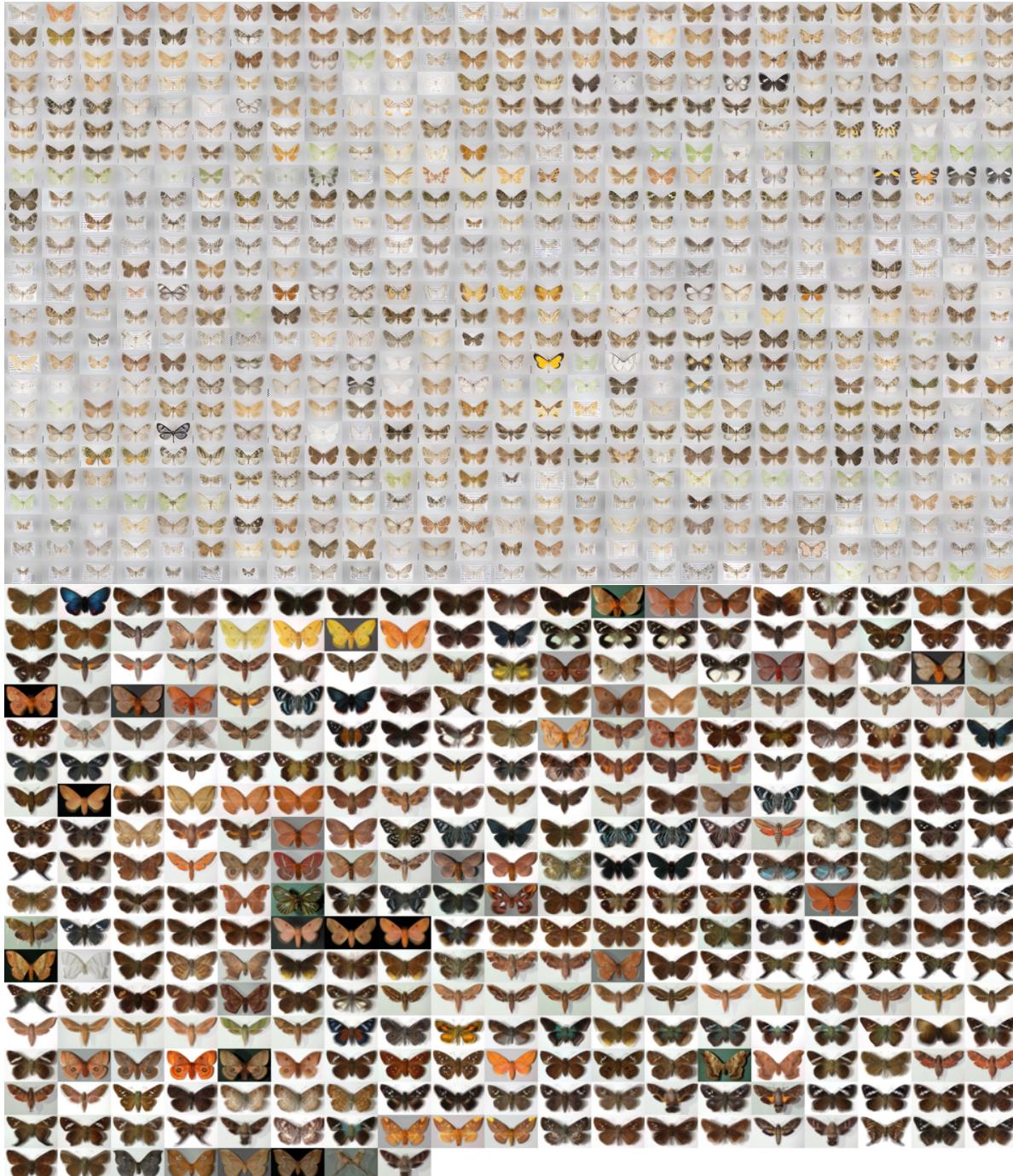
Figure 2: Average images of all categories of the Ecuador and the Costa Rica dataset.

3. Discovering interpretable features: automatically relating learned models to human-interpretable features would enable biologists to study especially hard to differentiate species in more detail.

4. Dealing with only a few training examples [7]: we need to build fine-grained recognition systems, which need to build fine-grained recognition systems, which

are especially able to deal with rare classes. This is important since currently available and important biodiversity datasets (see Section 2) are mostly comprised of classes with only up to 5 training examples.

5. Deriving compact textual and discriminative descriptions of the visual differences between the species.

Acknowledgements

References

[1] Paul Bodesheim, Alexander Freytag, Erik Rodner, and Joachim Denzler. Local novelty detection in multi-class recognition problems. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 813–820, 2015. 2

[2] Steve Branson, Grant Van Horn, Serge Belongie, and Pietro Perona. Improved bird species categorization using pose normalized deep convolutional nets. In *British Machine Vision Conference (BMVC)*, 2014. Preprint, http://arxiv.org/abs/1406.2952. 1

[3] Gunnar Brehm, Patrick Strutzenberger, and Konrad Fiedler. Phylogenetic diversity of geometrid moths decreases with elevation in the tropical andes. *Ecography*, 36(11):1247–1253, 2013. 1, 2

[4] Alexander Freytag, Erik Rodner, and Joachim Denzler. Selecting influential examples: Active learning with expected model output changes. In *European Conference on Computer Vision (ECCV)*, volume 8692, pages 562–577, 2014. 2

[5] D. H. Janzen and W. Hallwachs. Philosophy, navigation and use of a dynamic database (acg caterpillars srnp) for an inventory of the caterpillar fauna, and its food plants and parasitoids, of area de conservacion guanacaste (acg), northwestern costa rica, 2010. http://janzen.sas.upenn.edu. 1, 2

[6] Alex Krizhevsky, Ilya Sutskever, and Geoff Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1097–1105, 2012. 2

[7] Erik Rodner and Joachim Denzler. One-shot learning of object categories using dependent gaussian processes. In *Annual Symposium of the German Association for Pattern Recognition (DAGM)*, pages 232–241. Springer, 2010. 3

[8] David E Schindel and Scott E Miller. Dna barcoding a useful tool for taxonomists. *Nature*, 435(7038):17–17, 2005. 1

[9] C. Wah, S. Branson, P. Perona, and S. Belongie. Multiclass recognition and part localization with humans in the loop. In *International Conference on Computer Vision (ICCV)*, pages 2524–2531, 2011. 2

[10] 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. 1

[11] Josiah Wang, Katja Markert, and Mark Everingham. Learning models for object recognition from natural language descriptions. In *British Machine Vision Conference (BMVC)*, pages 2.1 – 2.11, 2009. 2