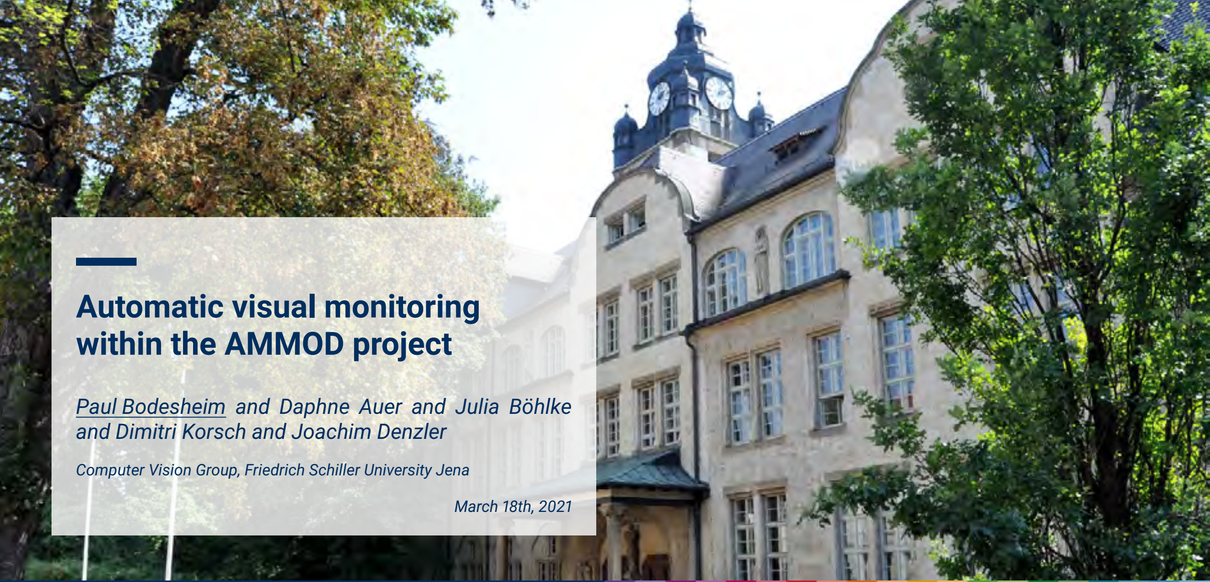


Automatic visual monitoring within the AMMOD project

Paul Bodesheim and Daphne Auer and Julia Böhlke and Dimitri Korsch and Joachim Denzler

Computer Vision Group, Friedrich Schiller University Jena

March 18th, 2021

Content

Introduction and motivation

The AMMOD project
- Project overview
- Moth scanner
- Wildlife camera traps

Summary

Content

Computer Vision Group @ Institute of Computer Science

https://www.inf-cv.uni-jena.de

- Head of the group: *Joachim Denzler*
- Four teams led by postdocs:

  Team: Computer vision and machine learning
  (Team leader: *Paul Bodesheim*)

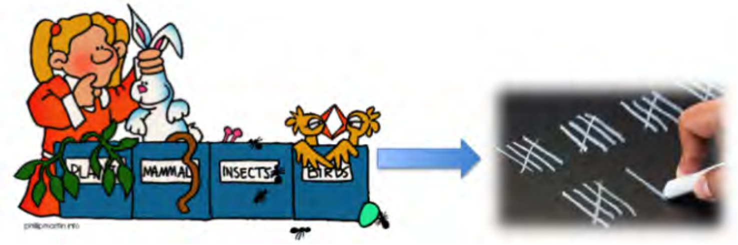
  Team: Event detection and causal reasoning
  (Team leader: *Maha Shadaydeh*)

  Team: Knowledge integration into machine learning
  (Team leader: *Björn Barz*)

  Team: Learning from 3D and unstructured data
  (Team leader: *Sven Sickert*)

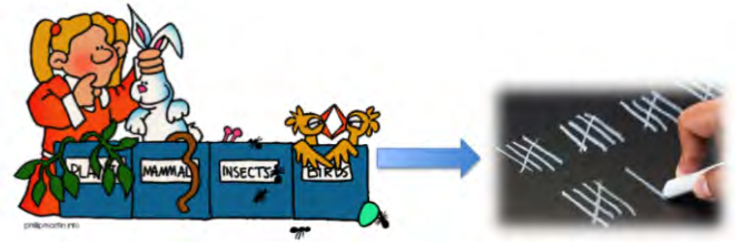
- Currently 16 PhD students allocated to these different teams

Goal: monitoring biodiversity



- Observing changes and trends (within and across species populations)
- This requires to maintain statistics (counting of individuals, species, ...)

Goal: monitoring biodiversity



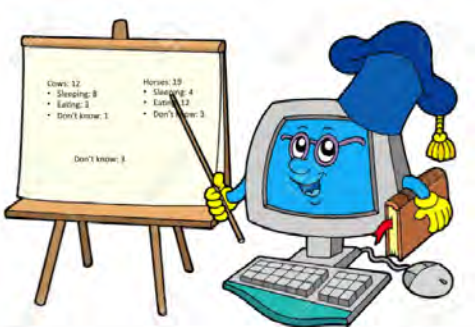
- Observing changes and trends (within and across species populations)
- This requires to maintain statistics (counting of individuals, species, ...)
- Counting requires recognition of individuals, species, ...
- Could be done by domain experts in the field, but time-consuming and limited to few locations

Visual monitoring with camera traps

- Solution: place many cameras / camera traps in the field
- Manual evaluation of image data is tedious / cumbersome / sometimes even impossible

Visual monitoring with camera traps

- Solution: place many cameras / camera traps in the field
- Manual evaluation of image data is tedious / cumbersome / sometimes even impossible



- Can the evaluation be done automatically?

Support by computer vision and machine learning

- Monitoring does not only mean to record and store data, but also to evaluate it
  $\Rightarrow$ Automatic analysis of sensor data

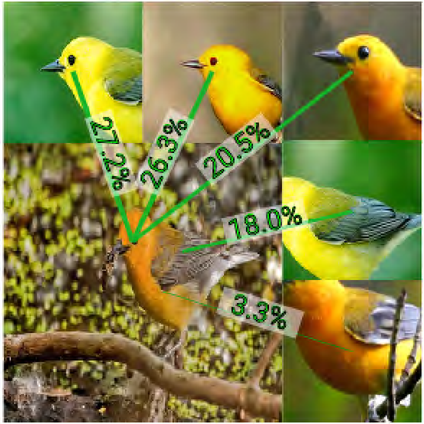
Support by computer vision and machine learning

- Monitoring does not only mean to record and store data, but also to evaluate it
  $\Rightarrow$ Automatic analysis of sensor data

- Millions of images (multiple images per camera station and day), GB of signals and recordings
- Presorting of errorneous or useless recordings
- Classification of expected species, events, etc.
- Activity detection (eating, sleeping, hunting, approaching, etc.)
- Finding the unexpected parts of the data (surprise sometimes drives research most)
- Open to changing setups, new species to be detected, etc.
- Possibility for humans to check, correct, and understand the automatic results
  $\Rightarrow$ Keep the human in the loop
- ...

Related work: fine-grained recognition

- Distinction of highly similar classes by small details
- Typical domain: bird species identification
- Common strategy: part-based approaches (allows for attribution of decisions)
- Part constellation models



Simon and Rodner: Neural Activation Constellations: Unsupervised Part Model Discovery with Convolutional Networks. ICCV 2015.

Simon et al.: Generalized orderless pooling performs implicit salient matching. ICCV 2017.

Korsch et al.: Classification-Specific Parts for Improving Fine-Grained Visual Categorization. GCPR 2019.

Simon et al.: The Whole Is More Than Its Parts? From Explicit to Implicit Pose Normalization. TPAMI 2020.

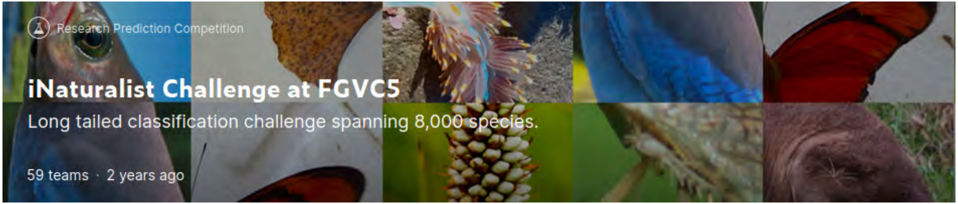
Related plant project (Ecotron)

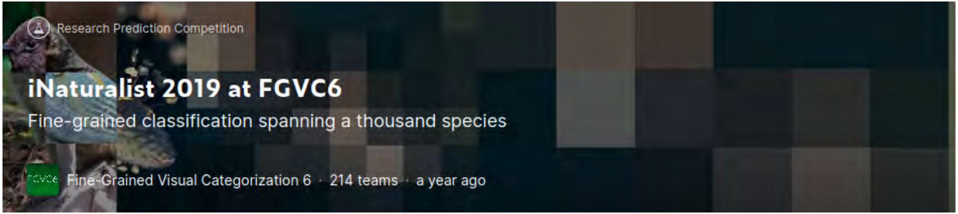
- 24 experimental chambers (controlled environment)
- 4 sensors for the measurement of soil temperature, moisture
- 4 tensiometers for measuring soil moisture tension in each of three different increments of soil depth
- Acrylic glass tubes inserted into the soil to monitor root development
- Cameras observe interactions between animals and plants (vegetation development and insect behaviour, *e.g.*, movement, habitat use, herbivory, predation and pollination)

- Investigate plant cover and phenology
- Joint work with Christine Römermann (Plant Biodiversity, FSU Jena)

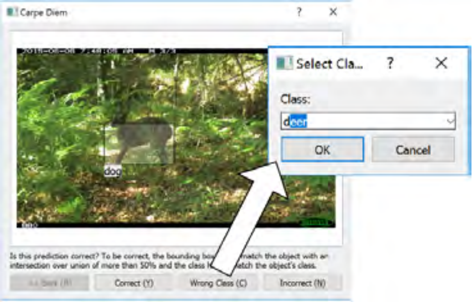
Körschens et al.: Towards Confirmable Automated Plant Cover Determination. CVPPP Workshop of ECCV 2020.

Annual challenges





Former monitoring task: herbivorous mammals from Portugal

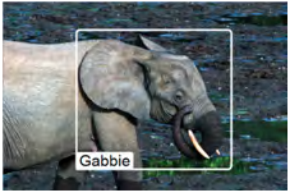


Joint work with Andrea Perino (iDiv)

Käding et al.: Large-scale Active Learning with Approximated Expected Model Output Changes. GCPR 2016.

Brust et al.: Active and Incremental Learning with Weak Supervision. KI 2020.

Further work on identifying individuals



- Chimpanzees:

  Freytag et al.: Chimpanzee Faces in the Wild: Log-Euclidean CNNs for Predicting Identities and Attributes of Primates. GCPR 2016.

  Käding et al.: Active Learning for Regression Tasks with Expected Model Output Changes. BMVC 2018.

- Gorillas:

  Brust et al.: Towards Automated Visual Monitoring of Individual Gorillas in the Wild. ICCV Workshop 2017.

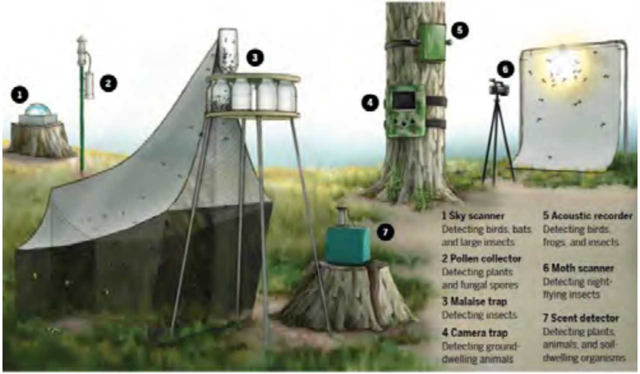
- Elephants:

  Körschens et al.: Towards Automatic Identification of Elephants in the Wild. AIWC Workshop 2018.

  Körschens and Denzler: ELPephants: A Fine-Grained Dataset for Elephant Re-Identification. ICCV Workshop 2019.

Content

One station, many sensors ("weather station for biodiversity")



Graphic by V.ALTOUNIAN/SCIENCE. From „Where have all the insects gone?" by Gretchen Vogel, SCIENCE, May 10, 2017 (doi:10.1126/science.aal1160).

- **Visual monitoring / camera traps**
- Smellscapes / scent detector
- Metabarcoding / Pollen collector and Malaise traps
- Acoustic monitoring / sound recordings

`https://ammod.de`

AMMOD modules

- Module 1: Management and coordination

- **Module 2: Automatized visual monitoring and image analyses**
- Module 3: Detection of smellscapes
- Module 4: Metabarcoding of environmental samples
- Module 5: Automated bioacoustic monitoring

- Module 6: The base station
- Module 7: Archiving, data management, and cross-platform analysis

Visual monitoring in AMMOD

1. Moth scanner (moth cam)
   - Light trap (illuminated white screen) to attract moths during night
   - High-resolution camera is taking images, e.g., one per minute

Visual monitoring in AMMOD

1. Moth scanner (moth cam)
    - Light trap (illuminated white screen) to attract moths during night
    - High-resolution camera is taking images, e.g., one per minute

2. Animal monitoring (site cam)
    - Typical camera trap with motion sensor
    - Stereo-camera setup allows for estimating depth

Visual monitoring in AMMOD

1. Moth scanner (moth cam)
   - Light trap (illuminated white screen) to attract moths during night
   - High-resolution camera is taking images, e.g., one per minute

2. Animal monitoring (site cam)
   - Typical camera trap with motion sensor
   - Stereo-camera setup allows for estimating depth

Hardware design and installation by our project partners!

Our tasks

- Developing algorithms for automatic analysis of recorded image data

Our tasks

- Developing algorithms for automatic analysis of recorded image data

- For the moth scanner:
  - Moth detection / localization by finding regions of interest (ROIs, bounding boxes)
  - Species classification for each ROI, probably part-based
  - Possibility to reject wrong ROIs (leafs, dirt)

Our tasks

- Developing algorithms for automatic analysis of recorded image data

- For the moth scanner:
  - Moth detection / localization by finding regions of interest (ROIs, bounding boxes)
  - Species classification for each ROI, probably part-based
  - Possibility to reject wrong ROIs (leafs, dirt)

- For images of the site cam:
  - Large-scale learning for recognizing different species
  - Applying lifelong learning methods to continuously improve recognition performance
  - Requires active learning with human in the loop and ...
  - ... novelty detection to handle previously unknown species

Content

Introduction and motivation

The AMMOD project

- Project overview
- Moth scanner
- Wildlife camera traps

Summary

Images from a prototypical camera setup

- Moth detection / localization and classification (100 species)

Images from a prototypical camera setup

- Moth detection / localization and classification (100 species)
- Visual results for a deep learning detector (SSD) that used an annotated dataset for training

Images from a prototypical camera setup

- Moth detection / localization and classification (100 species)
- Visual results for a deep learning detector (SSD) that used an annotated dataset for training
- Accuracy of deep learning classifier (InceptionV3): **93–94%**

Moth scanner and sample images

Joint work with Gunnar Brehm (FSU Jena and Phyletic Museum)

Moth scanner and sample images

Joint work with Gunnar Brehm (FSU Jena and Phyletic Museum)

- In general: limited moth images (with annotations) available per species
- But: classification systems need many sample images for learning to achieve satisfying accuracies during test

Moth scanner and sample images

Joint work with Gunnar Brehm (FSU Jena and Phyletic Museum)

- In general: limited moth images (with annotations) available per species
- But: classification systems need many sample images for learning to achieve satisfying accuracies during test

- Goal: exploit images from the web (image search engines and different websites: *lepiforum.de, inaturalist.org, gbif.org, boldsystems.org, ...*)
- Further question: usage of images from collections of museums?

Moth scanner and sample images

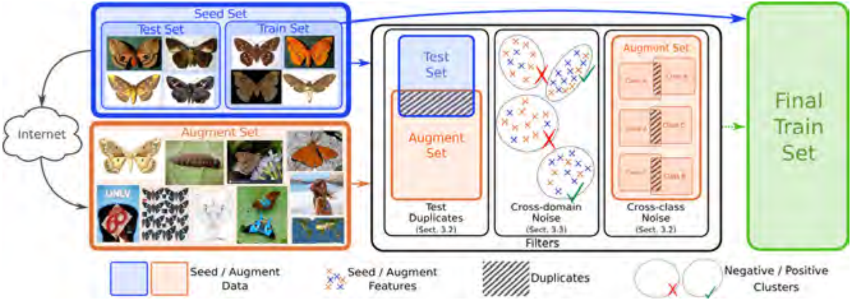
Joint work with Gunnar Brehm (FSU Jena and Phyletic Museum)

- In general: limited moth images (with annotations) available per species
- But: classification systems need many sample images for learning to achieve satisfying accuracies during test

- Goal: exploit images from the web (image search engines and different websites: *lepiforum.de, inaturalist.org, gbif.org, boldsystems.org, ...*)
- Further question: usage of images from collections of museums?

- Incorporate citizen scientists to collect additional training data!

Augmentation of the training dataset with Internet images



- Additional training images using Internet image search and species names as keywords (weak annotations)
- Filtering cross-class noise and cross-domain noise, improving classification performance to: **95–96%**

Böhlke et al.: Lightweight Filtering of Noisy Web Data: Augmenting Fine-grained Datasets with Selected Internet Images. VISAPP 2021.

Content

Camera trap images from Bavarian Forest National Park


Red deer


Red fox


Red squirrel


Wild boar


Two wild boars


European badger

- Daytime vs. nighttime
- Small vs. large animals
- Occlusion and truncation
- Scene / background clutter

Camera trap images from Bavarian Forest National Park

Red deer

Red fox

Red squirrel

Wild boar

Two wild boars

European badger

- Daytime vs. nighttime
- Small vs. large animals
- Occlusion and truncation
- Scene / background clutter

- Filter empty images
- Binary task (empty or not)

Camera trap images from Bavarian Forest National Park

Red deer

Red fox

Red squirrel

Wild boar

Two wild boars

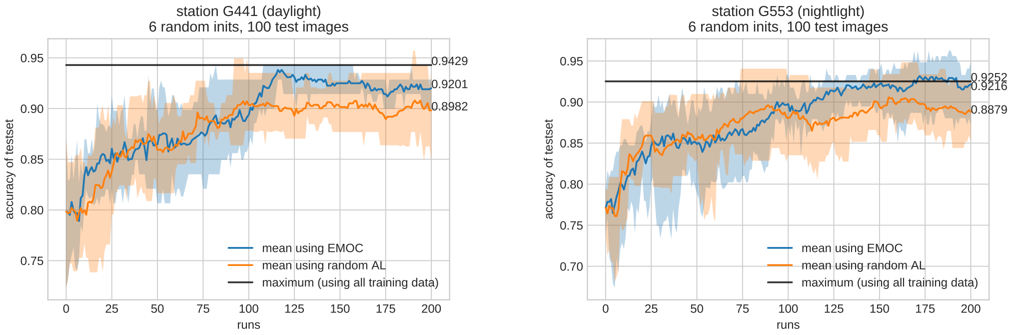
European badger

- Daytime vs. nighttime
- Small vs. large animals
- Occlusion and truncation
- Scene / background clutter

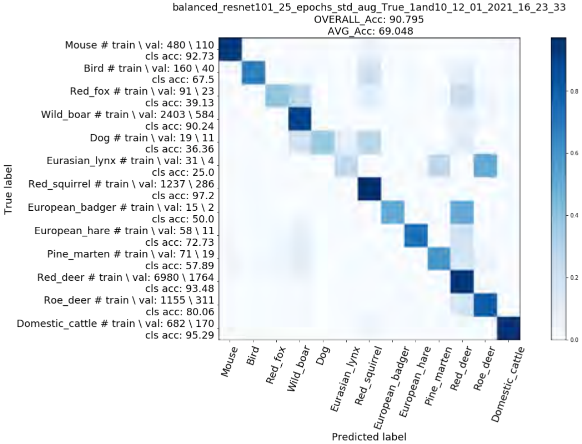
- Filter empty images
- Binary task (empty or not)

- Species classification in a lifelong learning scenario
- Including novelty detection and active learning with human-in-the-loop

Filtering empty images from camera traps



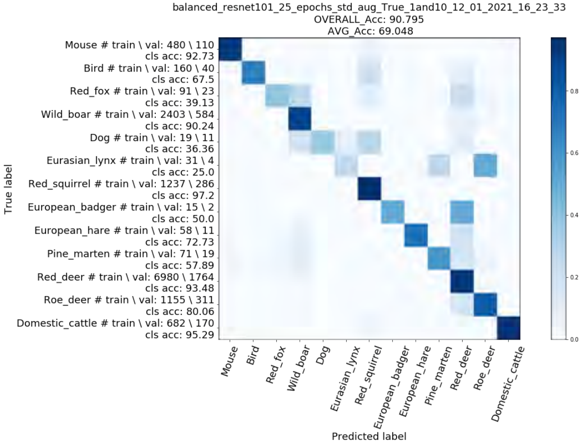
Species classification in camera trap images



- Trained a deep learning classifier (ResNet-101) using only non-empty images of all stations
- 13 classes (some species are merged):

Mouse	*Eurasian lynx*	*Red deer*
Bird	*Red squirrel*	*Roe deer*
Red fox	*European badger*	*Domestic cattle*
Wild boar	*European hare*	
Dog	*Pine marten*	

Species classification in camera trap images



- Trained a deep learning classifier (ResNet-101) using only non-empty images of all stations
- 13 classes (some species are merged):

Mouse	*Eurasian lynx*	*Red deer*
Bird	*Red squirrel*	*Roe deer*
Red fox	*European badger*	*Domestic cattle*
Wild boar	*European hare*	
Dog	*Pine marten*	

- Average accuracy per image: $\approx$**91%**
- Average accuracy per class: $\approx$**69%**
- Huge class imbalance (strongly varying number of sample images per species in the training set)

Content

Recap

Computer Vision Group

AMMOD project overview

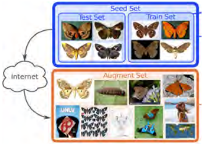
Recap

Computer Vision Group

Moth scanner

AMMOD project overview

Localize + classify / exploit web images

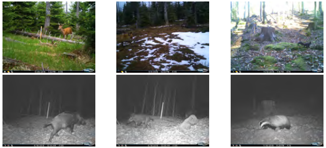


Recap

Computer Vision Group

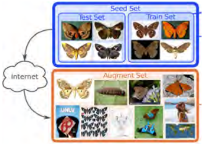
Moth scanner

Wildlife camera traps

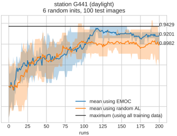


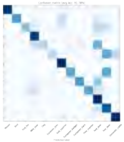
AMMOD project overview

Localize + classify / exploit web images



Filter empty images / classify species





Thank you for your attention!

Contact:

Paul Bodesheim (`paul.bodesheim@uni-jena.de`)

Computer Vision Group: `https://www.inf-cv.uni-jena.de`
The AMMOD project: `https://ammod.de`

Acknowledgments for project funding:

GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

Thank you for your attention!

Contact:

Paul Bodesheim (`paul.bodesheim@uni-jena.de`)

Computer Vision Group: `https://www.inf-cv.uni-jena.de`
The AMMOD project: `https://ammod.de`

Acknowledgments for project funding:

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung