

ReadMe - V_rodents_cameratraps_image_classification_intensive_quadrats

Dataset responsible: Eeva Soininen
(eeva.soininen@uit.no)

01.02.2023

1 Protocol

Camera trapping of small mammals has been conducted using the COAT protocol 'protocol_camera_trapping_small_mammals_varanger'.

1.1 Spatial layout of sampling and changes in that.

Camera trapping of small rodents in meadow and heath habitats was started in summer 2020. All heath and meadow sites included in the currently active studydesign were equipped with a camera trap. That covers localities Komagdalen and Vestre Jakobselv.

More detailed information about which sites were included in the study design can be found in the auxiliary file 'V_rodents_cameratraps_image_classification_intensive_quadrats_aux.txt'

1.2 Timing of sampling

The camera traps in heath habitat are deployed the whole year. They are checked and images are downloaded once a year, usually in July. The camera traps in meadow habitat are deployed in July and collected in early September. This is done to avoid camera damage due to spring flood. Camera boxes are deployed permanently, and thus meadow cameras are always in the same place at each meadow site.

2 Description of the dataset

The dataset includes three different types of data files and all files are saved as ;-separated txt-files:

- One file per year and locality with data on presence or absence of small mammals and image quality (`_komagdalen_2021.txt`)
- One coordinate file with coordinates of all sites (`_coordinates.txt`)
- One auxiliary file with additional information about the sites (`_aux.txt`)

In addition to this readme-file, there is one file that describes the model used for automatic image classifications and one file per year that summarizes the quality check after automatic classification.

2.1 V_rodents_cameratraps_image_classification_intensive_quadrats.txt

These files contain presence (1) or absence (0) of the following small mammal species:

```
[1] "aves = Bird"
[2] "cricetidae = Vole (either Grey-sided vole or tundra vole, Red-backed vole very rare)"
[3] "lem_lem = Lemmus lemmus (Norwegian lemming)"
[4] "mus_erm = Mustela erminea (Stoat)"
[5] "mus_niv = Mustela nivalis (Least weasel)"
[6] "sor_sp = Sorex sp (Shrew)"
```

In addition, there is a category for empty images (empty = 1) and for images with bad quality. If it is not possible to recognize whether there is an animal on the image or not, the image is classified as bad quality (bad quality = 1). All other categories are set to NA if an image is classified as bad quality.

The images are classified automatically by a machine learning model (see 'small_mammal_classification_model_v2022_summary' for more information about the model).

After automatic classification, a subset of images was annotated manually for a quality check (see 'V_rodents_cameratraps_image_classification_intensive_quadrats_quality_check' for more information and the results of the quality check). The quality check showed that the model performance for the categories 'least_weasel' and 'stoat' and for classifications with low confidence is not satisfying. Therefore, all images classified as least weasel or stoat as well as all images classified with a confidence below 0.8 were annotated manually.

The manual annotations are included in the data files and can be used instead of the automatic classifications. We recommend to use the manual classifications for all images classified as least weasel or stoat and all images classified with a confidence below 0.8.

Example of the first rows of the classification file:

```
sn_region sn_locality      sn_section sc_type_of_sites_ecological
1  varanger  komagdalen komagdalen_midtre          heath_far
2  varanger  komagdalen komagdalen_midtre          heath_far
3  varanger  komagdalen komagdalen_midtre          heath_far
      sn_site      t_date      t_time      v_image_name      v_class_id
1 ko_km_hf_a 2020-09-03 17:55:24 ko_km_hf_a_2020-09-03_0001.JPG      aves
2 ko_km_hf_a 2020-09-03 17:55:24 ko_km_hf_a_2020-09-03_0001.JPG      bad_quality
3 ko_km_hf_a 2020-09-03 17:55:24 ko_km_hf_a_2020-09-03_0001.JPG      cricetidae
      v_presence_automatic v_presence_manual v_confidence_automatic
1              NA              NA              NA
2              1              NA              1
3              NA              NA              NA
      v_observer_manual v_type_manual_classification v_comment
1              <NA>              <NA>              <NA>
2              <NA>              <NA>              <NA>
3              <NA>              <NA>              <NA>
```

Description of the columns included in the classification file:

Column name	Description	Possible values
sn_region	Study region	varanger
sn_locality	Locality (within region)	komagdalen, vestre_jakobselv
sn_section	Section (within locality)	komagdalen_midtre, komagdalen_ovre, sandfjorddalen, bearalveaijohka, torvhaugdalen
sc_type_of_sites_ecological	Habitat type	heath_far, heath_near, meadow
sn_site	Unique Site ID	e.g. ko_km_hf_a, ko_km_hf_e, ko_km_hn_e, ko_km_m_d, ko_ko_hf_e
t_date	Date when the image was taken	YYYY-MM-DD
t_time	Time when the image was taken	HH:MM:SS
v_image_name	Image name	e.g. ko_km_hf_a_2020-09-03_0001.JPG, ko_ko_hf_a_2020-12-28_0002.JPG
v_class_id	Class ID (species/functional group or image quality)	aves, bad_quality, cricetidae, empty, lem_lem, mus_erm, mus_niv, sor_sp
v_presence_automatic	Presence of species/functional group and image quality (results from automatic classification of the images by a machine learning model)	NA, 1, 0
v_presence_manual	Presence of species/functional group and image quality (results from manual classification for a quality check and of images with unsatisfying results from automatic classification)	NA, 0, 1
v_confidence_automatic	Confidence of the machine learning model that the images belongs to the category	[numeric] values between 0 and 1
v_observer_manual	Initials of the person who manually classified the images	e.g. es (Eeva Soininen)
v_type_manual_classification	Reason for manual classification	komagdalen, vestre_jakobselv
v_comment	Comments	[character]

2.2 V_rodents_cameratraps_image_classification_intensive_quadrats_coordinates.txt

This file contains the coordinates of all camera sites included in the study design. Coordinates are given in decimal degrees and UTM 33 (WGS 84).

Example of the first rows of coordinate files:

```
sn_site e_dd n_dd e_utm33 n_utm33
16 ko_ko_hn_a 29.94898 70.32916 1056664 7871442
18 ko_ko_hn_c 29.97894 70.33058 1057721 7871871
20 ko_ko_hn_e 30.00741 70.32603 1058887 7871640
```

2.3 V_rodents_cameratraps_image_classification_intensive_quadrats_aux.txt

This file contains further information about the dataset such as old site names and the years when sites were first included in the study design and when sites were excluded from the study design.

Example of the first rows of auxiliary files:

```
sn_region sn_locality sn_section sn_site sn_site_old year_first
1 varanger komagdalen komagdalen_midtre ko_km_hf_a <NA> 2020
2 varanger komagdalen komagdalen_midtre ko_km_hf_b <NA> 2020
3 varanger komagdalen komagdalen_midtre ko_km_hf_c <NA> 2020
year_last v_comment
1 NA NA
2 NA NA
3 NA NA
```

* year_last is NA if the site is still included in the study design

3 Related datasets

3.1 V_rodents_cameratraps_image_metadata_intensive_quadrats

This dataset contains image metadata for each image such as trigger mode (motion sensor or timelapse) and temperature.

3.2 V_rodents_cameratraps_annual_metadata_intensive_quadrats

This datasets contains information of the annual visits of the camera sites when the cameras are checked and images are downloaded such as the serial number of the camera and whether the camera was functioning.

4 Data processing

1. Metadata is extracted and the images are renamed with a name consisting of the site-ID, the date when the image was taken and a number (e.g. ko_ga_sn_1_2020-01-01_0001.JPG)
2. The images are classified automatically using a machine learning model
3. A random subset of images is annotated manually for a quality check

4. Images belonging to categories with unsatisfying model performance are annotated manually
5. Information from metadata files, automatic and manual classifications are combined and formatted according to the requirements of the COAT data portal

The script ‘function_preprocessing_image_classifications_small_mamma_camara_traps.R’ (https://github.com/COATnor/data_preprocessing_scripts) can be used to preprocess the datasets for further analyses. The script selects one of the motion sensor images and either the automatic or the manual classification after the following criteria:

- If one of the images has a manual classification, this image will be kept
- If both images have a manual classification, the image that contains an animal will be kept, if both images contain an animal, the first image will be kept
- If both images have the same class, the image with the highest confidence will be kept
- If only one image contains an animal, the image with the animal will be kept
- If the images are classified as different animals, the image with the higher confidence will be kept

All data processing is performed in R.