

project-proposal-jsmets

September 4, 2019

1 Capstone proposal by Julien Smets

2 Machine learning for Lepus

2.1 1) The problem

By thinking about the capstone project, one emerged when discussing with prof. Yves Hausser from the nature management department at hepia (HES-SO Geneva) about a deep learning project of natural images. As follows, an introduction of this project is presented:

The Lepus software [1] was designed to help scientists analyze wildlife images acquired using photographic traps. At present, species recognition and individual identification are carried out manually, which is very time-consuming. Note that, to get the data, local people have to access to these cameras and are present on some images. Thus, humans are also a category in itself and must be classified correctly.

The objective of this project is to test Deep Learning technology to automate certain tasks such as

1. detecting the presence or not of an animal/human in the image
2. locating animals/humans using bounding boxes
3. identifying certain species given a fixed taxon level
4. ideally, identify each individual animal of a specific species with respect to physical characteristics (e.g. to help the Wild Life Conservation Society) e.g. [Computer Vision for Wildlife Conservation \(CVWC\)](#)

Project Proposal *Detection of the presence (#1): Classification of an animal, human and empty classes in the images (trinomial classification)*

Please let shortly motivate this choice. By solving this problematic, the saved time for nature management scientists could be very high (days of work) because only a small amount of images contains animals or humans (this will be more deeply detailed below). Moreover, the labelled data do not include the bounding boxes which excludes the second project (here I consider only supervised learning to ensure a validation metric) and the fourth one is limited to the significant inspection variance of the manual identification and the very low amount of data available. Although the final objective is to identify automatically animal species (third project), this is let to future works and this project aims solving the most time consuming part; the detection the presence or not of an animal/human in the image.

As follows, you can see the project cloud for more details:

- [1] <https://lepus.cloud>

This project aims analysing and developping an accurate method for identifying the presence of animal species or humans. The main question marks are: - How many data is available for each class (human/animal/empty) and in which proportion. - Which type of information is available, how can they be usefull for the objective and discuss potential issues.

This leads to a data exploration in order to answer these questions. Since this project has multiple problematics, several assumptions are needed.

In the end, the objective is to determine which information and method result to the best accuracy for the classification.

2.2 2) The data

2.2.1 (a) Clear overview of your data

Getting the data The data is shared by physical transfer in order to conserve privacy. Images are separated into several folders as explained above, their paths and other information are specified into some CSV files. All CSV files are grouped into a single file for simplicity.

Data information The labels and other images information are stored in a CSV file. Lets load this data.

```
[1]: import pandas as pd
      csv_file = '../data/Lepus/csv_files/DeepLearningExport.csv'
      data = pd.read_csv(csv_file)
      data.sample(frac=1).head()
```

```
[1]:      file_id      file_path session_dir      file_datetime \
7922      32361  2015/RW4/RW4_34/08250223.JPG      Dry 2015  2015-08-25 17:58:00
11525      3365   2015/M1/M1_34/03120025.JPG      M1 2015  2015-03-12 05:33:00
6443      30871  2015/RW3/RW3_33/08300137.JPG      Dry 2015  2015-08-30 02:40:00
5418      29845  2015/RW3/RW3_07/08220565.JPG      Dry 2015  2015-08-22 18:18:00
2950      27367  2015/RW1/RW1_12/09290448.JPG      Dry 2015  2015-09-29 09:00:00
```

```
      sun_angle file_period event_id prev_file_id session_id place_id \
7922      77.21         day      8587      32360.0          3      439
11525     -111.71        night       441       3364.0          5       22
6443     -154.09        night      8066          NaN          3      402
5418      82.27         day      7681      29844.0          3      377
2950     -54.95         day      6684      27366.0          3      311
```

```
      taxon_id taxon_tsn      taxon_name
7922         17      NaN  Phacochoerus africanus
11525         4      NaN   Sylvicapra grimmia
6443         4      NaN   Sylvicapra grimmia
5418         4      NaN   Sylvicapra grimmia
2950        39      NaN    Numida meleagris
```

where columns means: * **file_id** [unique integer]: the identifier (ID) of the file * **file_path** [string]: the path of the file image ({year}/{grid}/{camera}/{picture}) * **session_dir** [string]: the directory name of the file (called session) * **file_datetime** [timestamp]: the date time of the file * **sun_angle** [float]: angle of the sun with 0 meaning sunset * **file_period** [nominal/cyclic ordinal]: the period

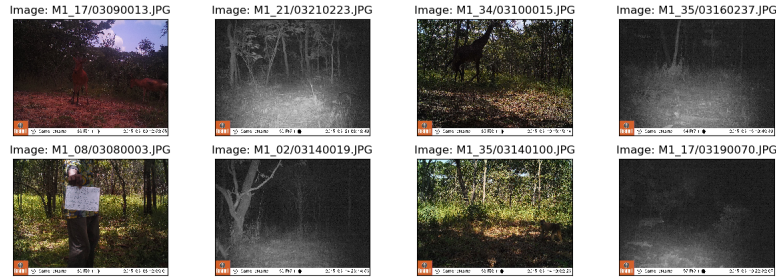
of the file (day/night/twilight) * **event_id** [integer]: ID of the independent capture evenment (ICE) * **pred_file_id** [float (integer+NaNs)]: ID of the previous file deduced from event with the same ID * **session_id** [integer]: ID of the session * **place_id** [integer]: ID of the camera's location * **taxon_id** [float (integer+NaNs)]: Id of the taxon (i.e. species or lower in animal ranking) * **taxon_tsn** [float (integer+NaNs)]: ID of the world official tsn (taxonomic serial number) * **taxon_name** [string]: common name of the taxon present in the image

```
[2]: print('The data has shape:', data.shape)
```

The data has shape: (32541, 13)

Images The data set is given as image files separated in several folders and subfolders. These folder are classified by year/grid/camera/picture where a grid correspond to a set of 36 cameras (6x6).

The total dataset represents a covered area of 10'000 km² with hundred of cameras placed in different nature spots in Tanzania. The entire dataset consist of **32k** color and grayscale images. All these cameras are placed in different nature spots in Tanzania. This data has been obtained since 2013 and until 2016 (more recent data remains unlabelled).



Examples of images from the M1 2015 dataset.



Examples of images from the M1 2015 dataset.

The given species can be very small depending on the animal size and its distance to the camera or very large taking a large part of the image. Animals can be **occluded** by background objects (trees) or even be **partially viewed** (especially for large animals such as giraffes or elephants). In rare cases, it is possible to have more than one individual or more than one species in some images.

2.2.2 (b) Plan to manage and process the data

Data information As a first step, the CSV files will be read into DataFrames and merged together (as shown above). Then this DataFrame will be cleaned and manipulated. After checking issues on formats and values of the data (nominal, ordinal and timestamp, ...), class labels will be grouped into human, animal and empty labels. This allows to compute the labels statistics as possible informative data (e.g day/night/twilight, solar angle w.r.t. human/animal/empty categories).

2.2.3 Images

The images have high and different resolution. Thus, a resizing is needed. Moreover a lower shape should be chosen to remove redundant information of pixels for classification models (especially convnets).

Data pre-processing Images also needs processing. First of all, image will be normalized: - resize each images by downsampling the large input image into an adapted smaller shape. This will be performed empirically to balance between complexity and information loss. - normalize pixel intensities. In this case, in addition to 0-1 scaling, an adaptative contrast adjustment (CLAHE from cv2) will be computed to adjust contrast and enhance locally the visibility of hidden (dark) species. Note that this also enhances the noise in low contrast areas.

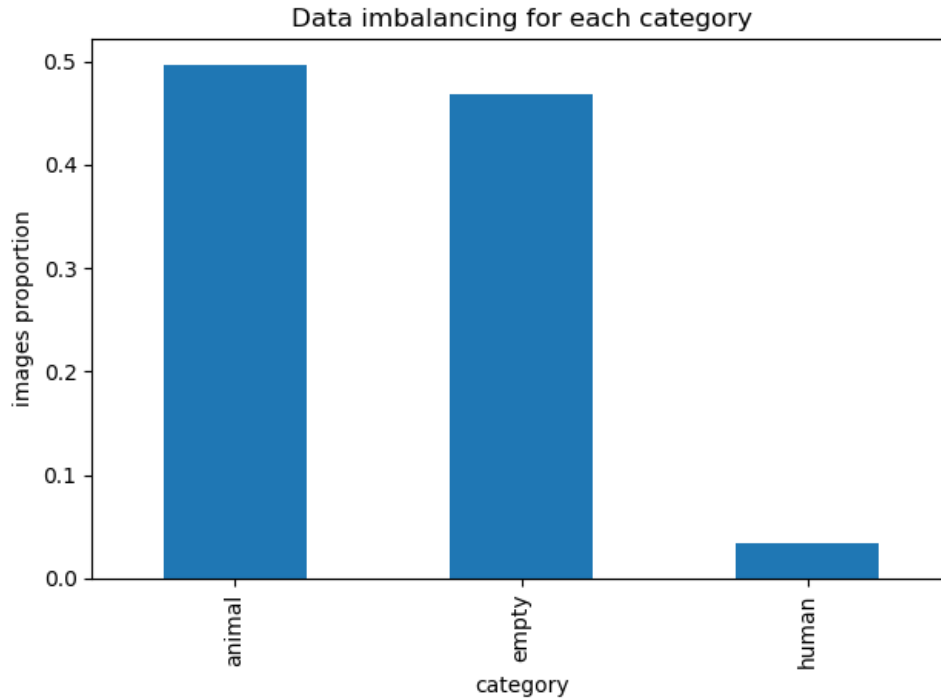
Data augmentation The animal often represent a sub part of the image, this can induce background learning instead of the animal. To reduce this by adding some variance to the data, some data modification is needed such as noise addition, horizontal flip, rotations or shear. Other potentially usefull methods but problematic will not be used because they could remove small species such as cut out or cropping.

Additional information processing The additional data of the day/night/twilight is cyclic. Similarly to time data, it is possible to ensure locally constant difference by encoding them as day:1, twilight:0 and night:-1. This allows day and night being the highest difference possible. This idea will be analyzed in EDA over the entire dataset.

2.3 3) Exploratory data analysis (EDA)

2.3.1 (a) Preliminary EDA

Data imbalancing Depending on the type of environment and the acquisition sensitivity of the photographic traps, the distribution of the classes empty/human/animal can be significantly imbalanced (see figure as follows).

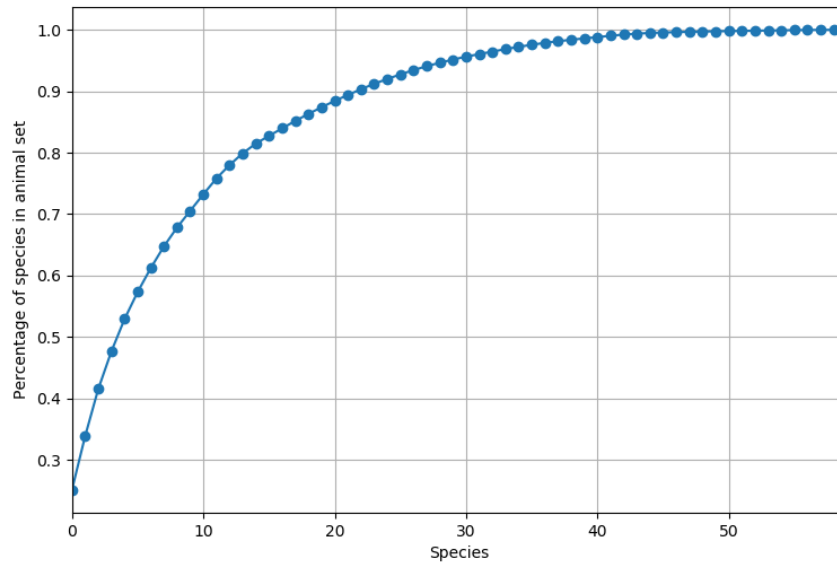


Example of data imbalancing of the full dataset.

The amount of species are in irregular proportion depending on the rarity of these species. Moreover, a majority of the given images doesn't contain any animals due to trees movement, dust tornado, butterflies, etc causing false positive captures. The proportion of empty pictures is ~40-100% depending on the device environment.

In the context of presence identification, the **data imbalancing** is an important issue. Indeed, if the data is used as is, the importance of very common label can be significantly overestimated. Indeed, the systematic identification of images into the label *empty* representing 47% of the data set results in an accuracy of 47%. The step of data balancing will be performed in order to induce no assumptions on the class distribution.

In addition to this class imbalancing, there is animal species imbalancing inside the animal class due to the rarity of certain species (cf. following figure). This imbalancing should also be considered.



Example of data imbalancing of the full dataset.

The first 5 species represent in quantity almost half of the species present in the data. Lets look at these species.



Five most common species of the full dataset.

These species will have the most impact on the animal classification and rare species will have minor significance.

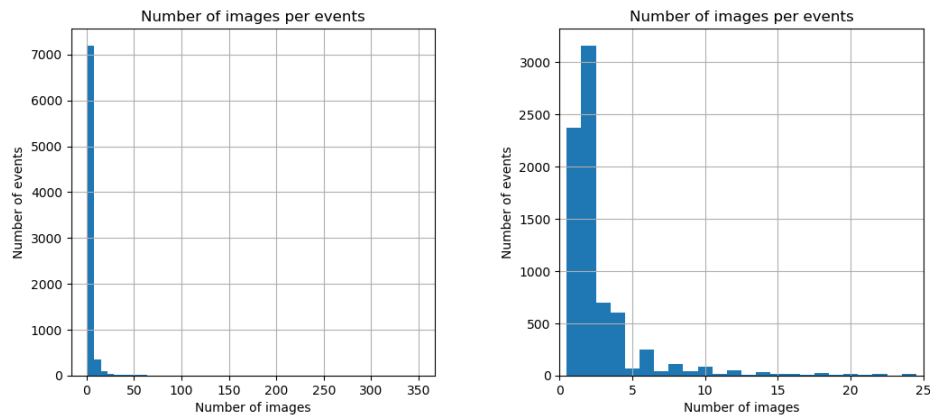
Temporal information (ICE) Some of the images are time correlated due to animals running and get captured several times, i.e. **small timelaps images**. These set of images are already grouped by the Lepus software. These grouped images are called **independent capture event (ICE)** and numbered from 3 up to ~300 if an animal stays in front of the camera a long part of day/night.

Since data is ordered by ICE (independent capture events), there is a temporal information that can be used. In a special case, it is possible to see animals and humans moving using the difference between successive images and can especially show their presence or not (cf. following figure). Each event identifier contains from one to hundreds of pictures. Based on the file data time, each file_id is linked to the previous file identifier. If the value is a NaN, the picture is the first of the event. To work with images differences, we need at least 2 images per independent event.



Example of timelaps images and their difference from the M1 2015 dataset.

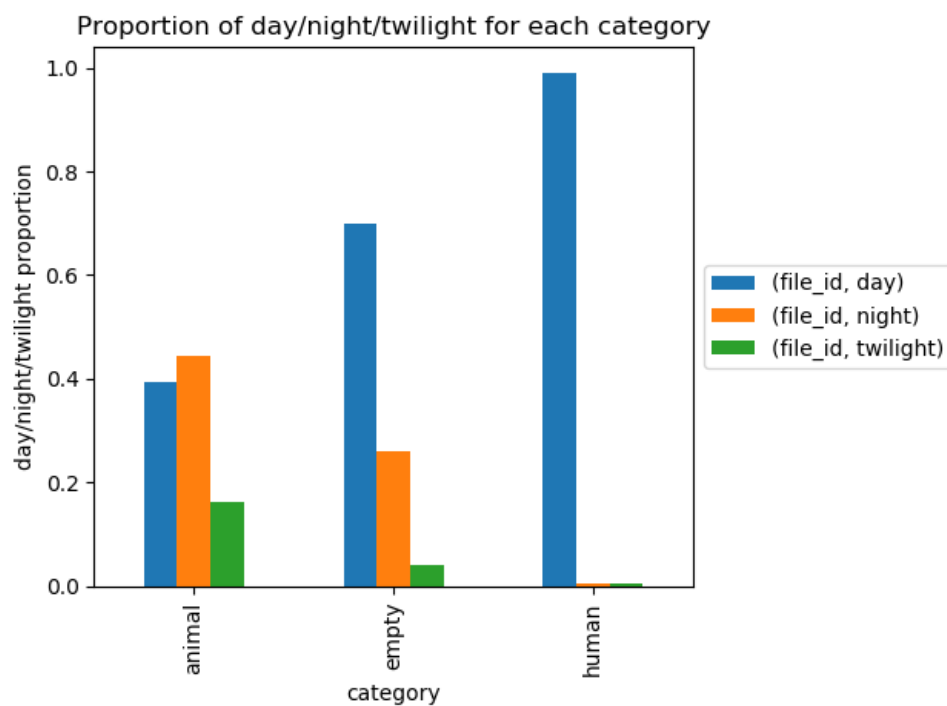
Another interesting this to observe is the amount of images per events (cf. figure below).



Histogram of amount of images per events of the full dataset.

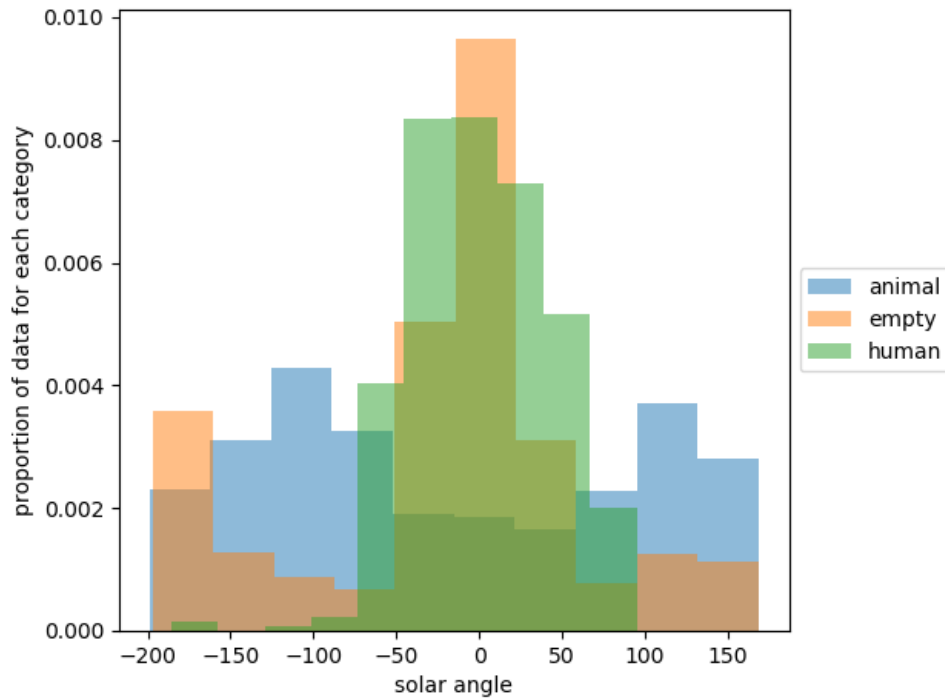
Zooming on the histogram, the majority of the events contain 2 images. 2400 of them contain 1 image and are not usable for image differences.

Additionnal information The information of animals lifestyle is a usefull assumption used by scientists to identify species in practice. Since the camera traps are well located and include **timesteps, the day, night and twilight** can be known (see below for further details). This can be used by adding this knowledge to infer the classification e.g. humans are (almost) always present during the day (see the following figure). To this end, it is possible to experiment and determine if this information can be interesting and how it could be in the context of human/animal/empty classification.



Day/night/twilight information of the full dataset.

Note that the day/night/twilight are deduced from the solar angle which is more precise. This more precise information can be used also.



Solar angle information of the full dataset.

2.3.2 (b) How does the EDA inform your project plan?

The EDA shows that, in addition to color and grayscale images, their temporal difference and additional information can help the prediction task. To this aim, questions arise.

Is there a difference in the final result if RGB or grayscale images are used? Even more, what is the gain or the loss obtained after using the difference between 2 images following in a sequence? Grayscale images differences or RGB images differences? These are the main project ideas and goals. Indeed, if 2 images follow one after the other with a relatively short amount of time separating them, the background would subtract and the result image would focus on the differences between them. For example, the movement of an animal.

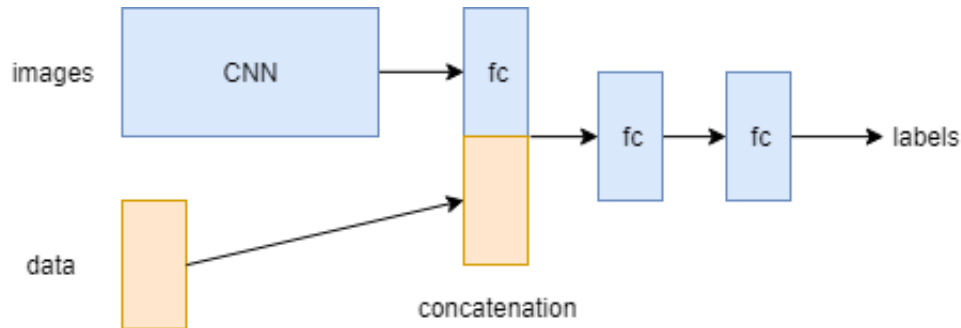
To perform these analyses, our machine learning models will be trained on :

- Grayscale images
- Colour images
- Grayscale images differences
- Colour images differences
- Day/night/twilight as additional information
- Solar angle as additional information

Additional Information as Improvement for Machine Learning Since the photographic traps are precisely located with timestamp, there is the information of solar elevation angle, thus the time of the day (day, night, twilight), which can be a powerful information because certain species have specific lifestyle.

The solar elevation angle should give more information since it is a (cyclic) continuous variable, in contrary to day,night,twilight which is a (cyclic) nominal feature.

The following example shows the structure of the data concatenation (e.g. here with fc layers).



Example of the additional information utilisation as improvement on a grid (M1 2015) with 518 species or humans.

2.3.3 (c) What further EDA do you plan for project?

Further analysis of methods could be used to improve the data.

2.4 4) Machine learning

2.4.1 (a) Phrase your project goal as a clear machine learning question

The proposed project focuses on the classification of animal, human and empty in the pictures. To perform this and by looking at the data, one question arises.

- How can we combine machine learning models to obtain an accurate prediction one ?

Main plan As a final challenge, can we combine some of the best trained models together and some additional information (i.e. to further improve prediction results. In this last part, several methods will be experimented.

- combine
 - best image model
 - image difference model
 - additional information model

2.4.2 (b) What models are you planning to use and why?

A second question would be which model to use ? Is there a model which is more efficient with images differences than raw images ? What about grayscale and RGB ? For this purpose several models will be used.

- Most frequent (as baseline)
- K-nearest neighbours
- Decision trees and random forest
- Support vector machines
- Dense (fully connected) neural networks
- Convolutional neural network

- small not pretrained model*
- bigger pretrained model*

*for time optimization

ConvNet training strategy Instead of training on a randomly initialized model, it is common to perform transfer learning on a large challenging task pre-trained model (e.g. 1000-class [ImageNet](#)). In addition of reducing the time of training, it also helps generalization (cf. [this paper](#)).

For transfer learning, one very promising model has recently been published, on June 10th 2019, and shared, on [tensorflow](#), by Google Research called [EfficientNet-B0](#). This model can be scaled depending on the computer learning capability and has already been trained on 1000-class [ImageNet](#). This can be an ideal starting point for ConvNets.

As we saw during the fourth project, training over well defined features increase the overall accuracy of the machine learning models. For this purpose and to keep the scope of the project on the main goals, the pixels values will be used as input for each model.

2.4.3 (c) Please tell us your detailed machine learning strategy

1. Data Loading, Cleaning and Manipulation:

- Load CSV file and get images pathnames in several folder and subfolders using recursive search.
- Clean data by removing unlabelled data and normalize labels and values.
- Group the data into each class labels: human, animal and empty.
- Manage the remaining data for easy access and processing (e.g. using flow from directory).

2. Images Preprocessing:

- Select state of the art preprocessing methods adapted to our problematic for images (e.g.* histogram equalisation, denoising, resizing, data augmentation such as noise, horizontal flip, rotations or shear) and for the additional information (e.g. day/night/twilight encoding).
- Implement them and make them easy to use such as using flow from directory.

*based on iWildCam 2019 challenge [Top7 report](#).

3. Splitting the data:

- Separate the data into a stratified train/validation/test sets in order to conserve the labels proportions, thus by keeping its global imbalance.

4. Training models:

- Train the selected models on the dataset.
- Validate the trained models.

5. *Analyse trained models results:*

- Analyse trained models results and discuss about incentives
- If needed, adapt the structure of the models by tuning parameters or adjusting layers.
- Identify most accurate trained models to combine them together.

6. *Obtain a final model:*

- Use the knowledge learned by all the experiments to define a final model combining best of the tested models.
- Compute its accuracy and discuss about the results.

2.5 5) Additional information

2.5.1 Information

The code sample will be implemented in several python scripts (.py) for convenience and simplicity (e.g. separation of pre-processing, training and testing). An additional notebook (.ipynb) will be included with analysis details and visualization figures as a small report (similarly to this document).

Contribution of the proposed project The proposed project of classifying animal/human/empty classes in image. This work will differ on several points w.r.t. the state of the art [1,2].

In [1], the authors focused on image processing methods for classifying and locating animal/human presence and absence in images where the proposed project ignores this and assumes that machine learning can do even better.

References:

- [1] Yousif, Hayder, et al. "Animal Scanner: Software for classifying humans, animals, and empty frames in camera trap images." Ecology and Evolution (2019).

Others Note that the data can need to be confidential with a DNA (standard Non-Disclosure Agreement) due to some very rare species which are often hunted for money. If this is the case, it will be demanded to the EPFL soon. But anyways the precise location of the picture will not be transmitted.