Two-phase training mitigates class imbalance for camera trap image classification with CNNs

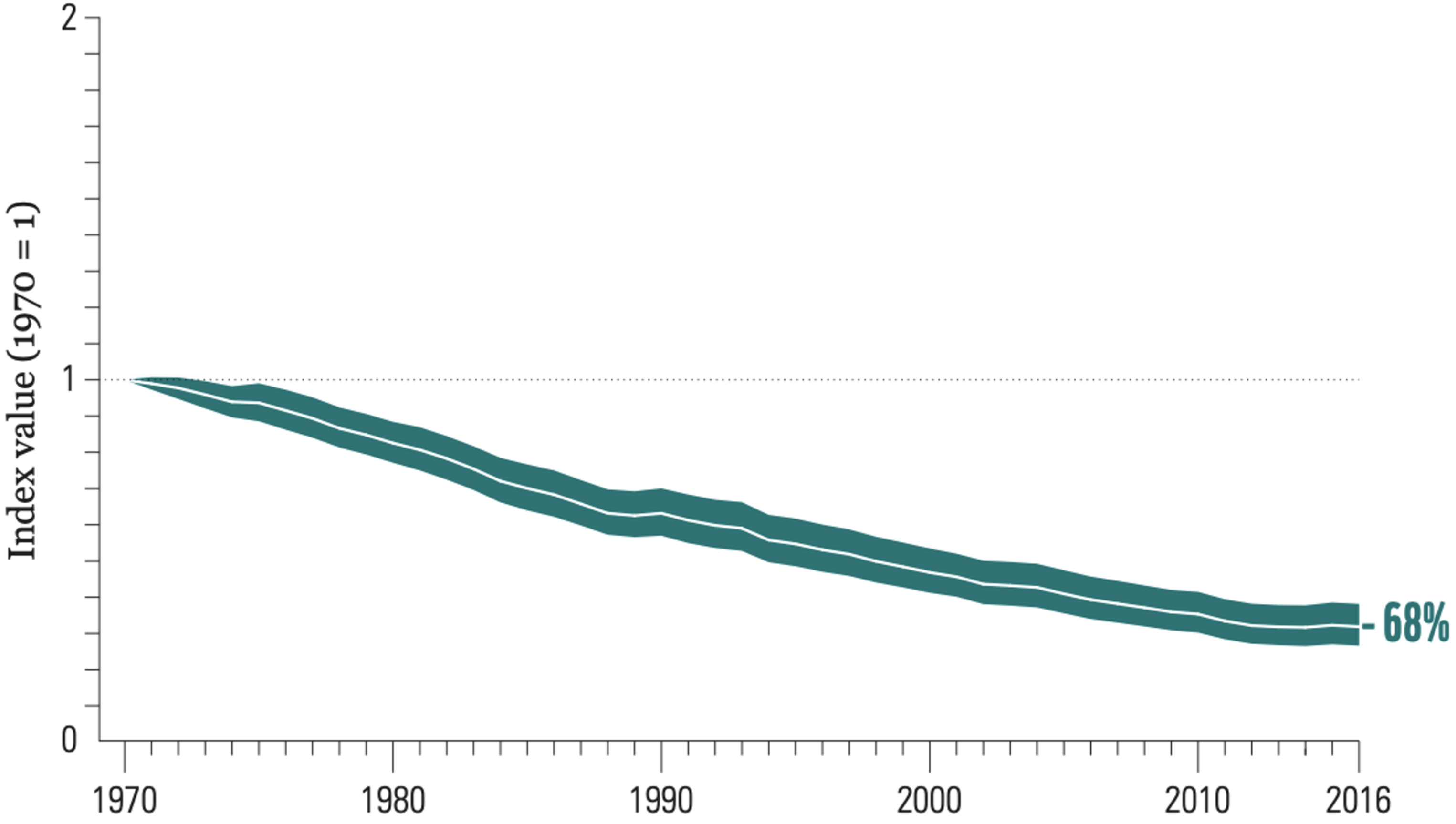
Farjad Malik, Simon Wouters,
Ruben Cartuyvels, Erfan Ghadery,
Marie-Francine Moens

Department of Computer Science,
KU Leuven

Background: biodiversity decline

- Causes
  - Habitat loss & degradation
  - Species overexploitation
  - Invasive species & diseases
  - Climate change

- Importance
  - Water quality
  - Air quality
  - Climate
  - Food production
  - Spread of infectious diseases



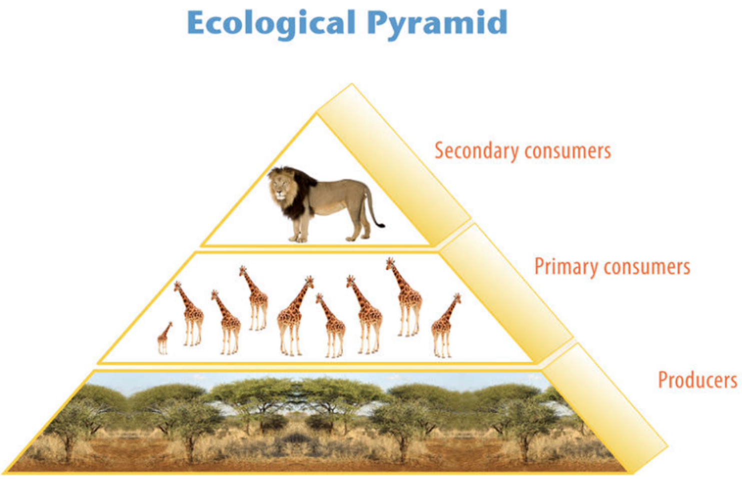
Evolution of the Living Planet Index since 1970

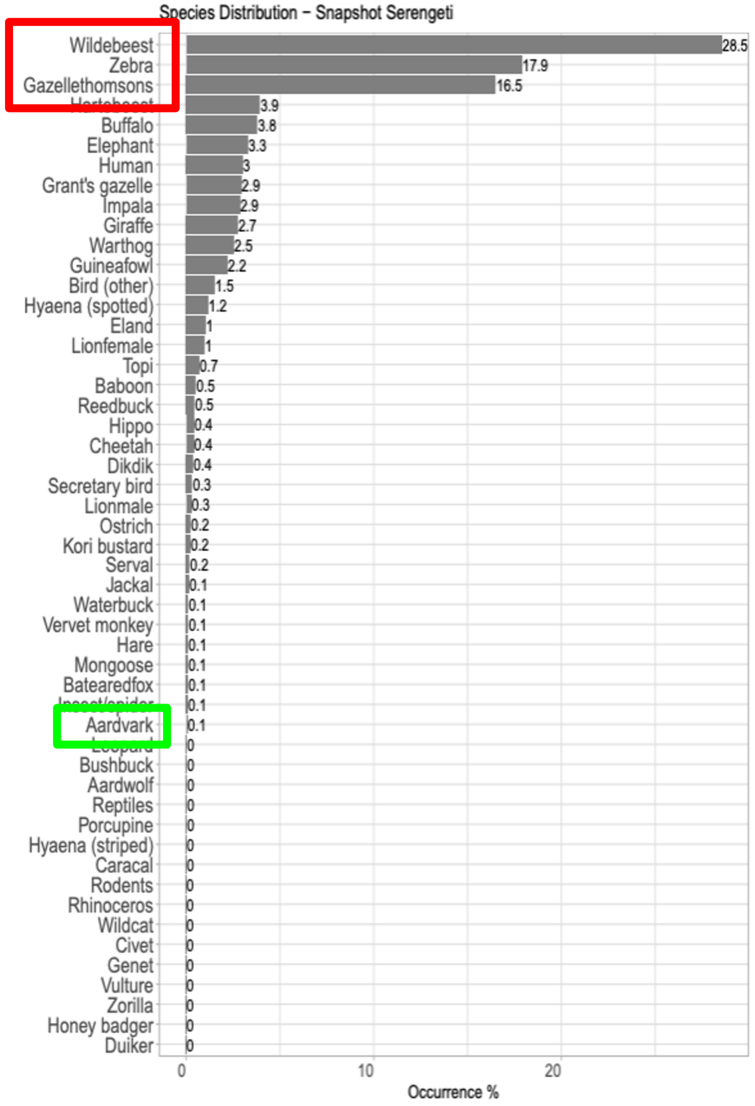
Background: ML for biodiversity monitoring

- **Camera trap images**
  - Automatic species classification
  - Increase duration & scope of studies

Literature: main challenges

1. Insufficient / bad training data
2. Generalisation (to new locations)
3. Class imbalance
   - Ecological pyramid
   - Size/activity differences
   - Ecosystem deterioration





KU LEUVEN

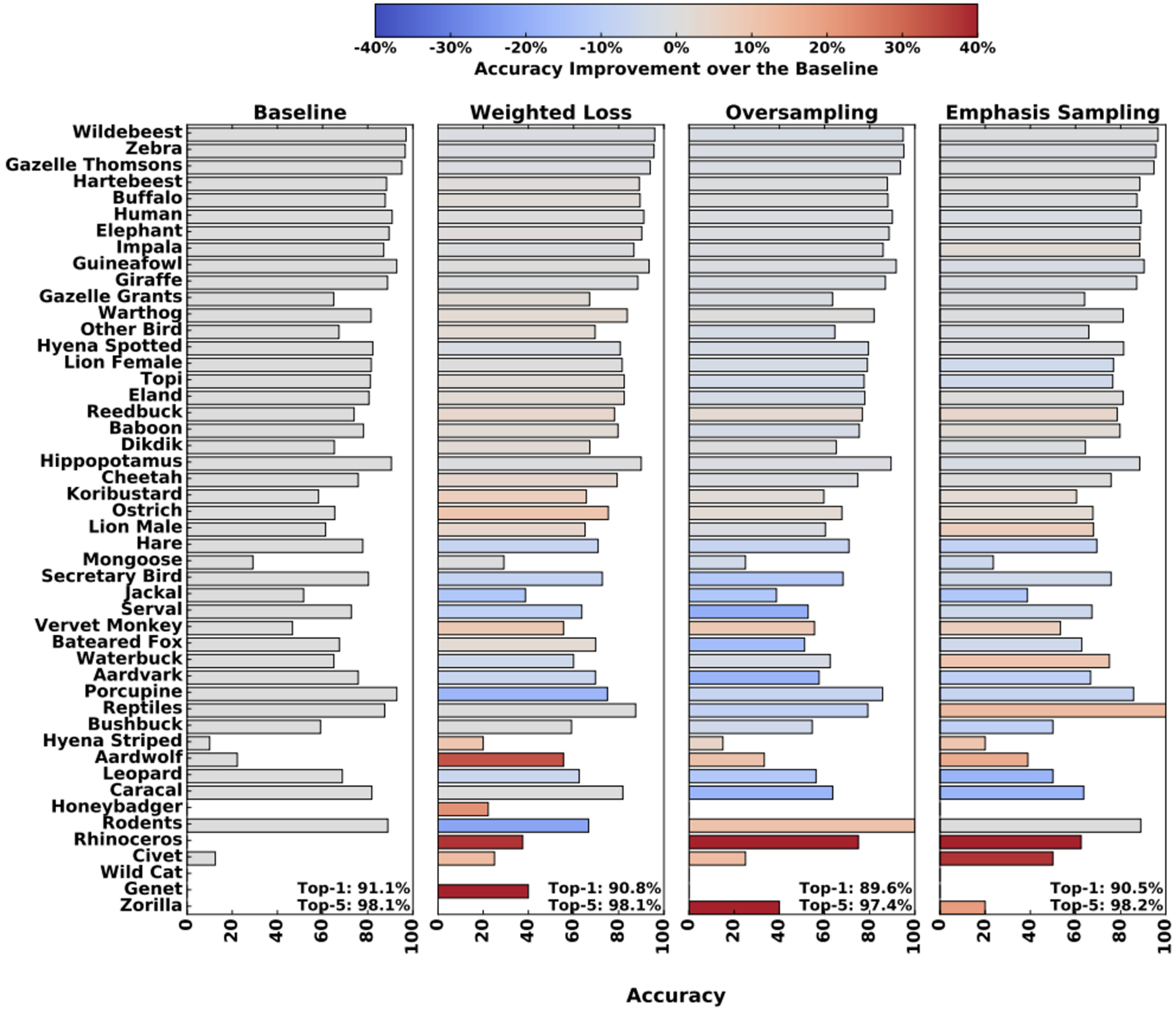
Literature: mitigating class imbalance

Observations:

- High overall accuracy
- Poor performance for minority classes

Efforts:

- Removing the rare classes
- Review uncertain classifications
- Cost-sensitive learning
- Oversampling
- Novel sampling methods



Literature: mitigating class im

Data-level techniques

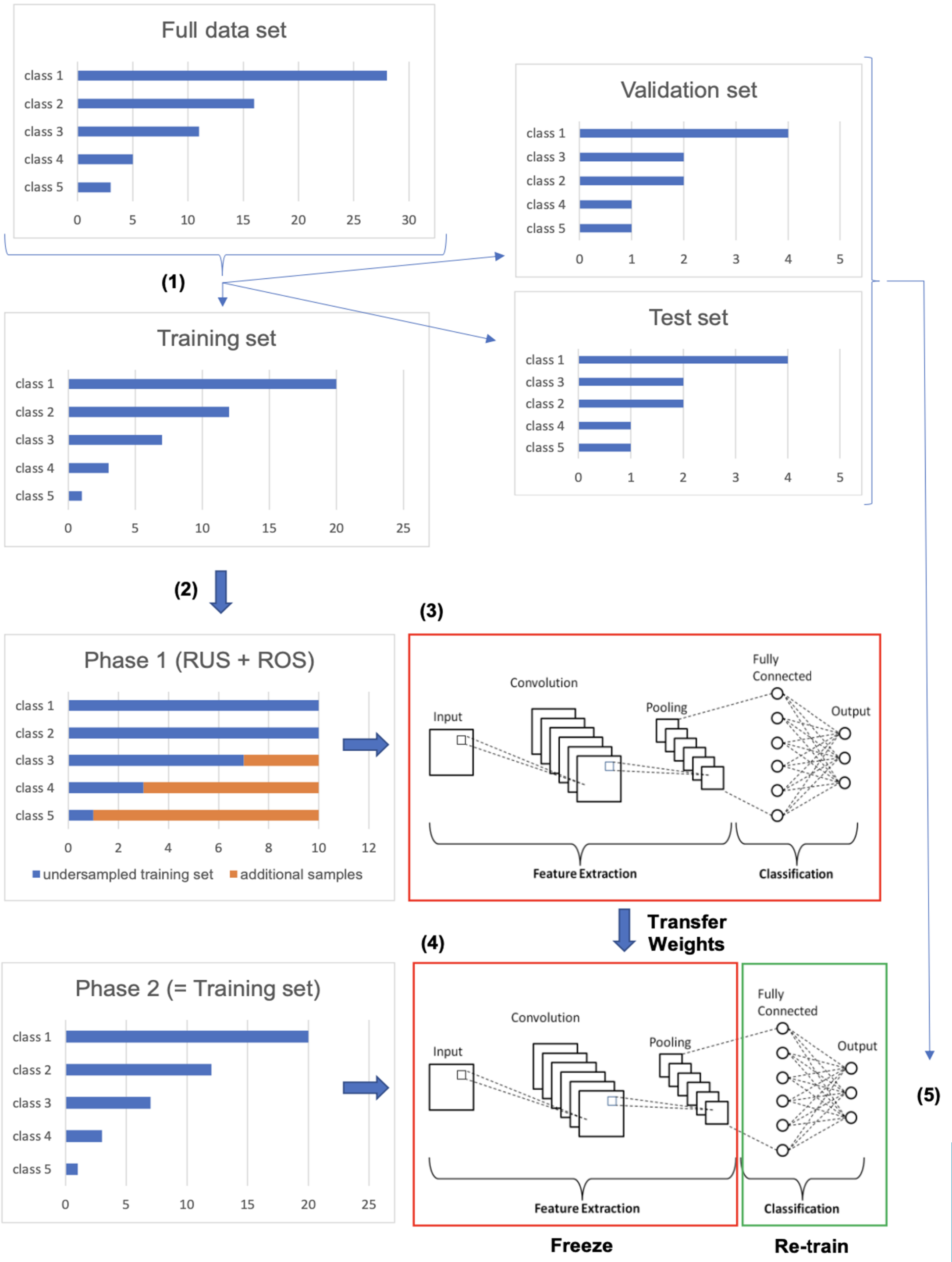
- Random minority oversampling (ROS)
- Random majority undersampling (RUS)

Algorithm-level techniques

- Loss-function, cost-sensitive learning

Hybrid techniques

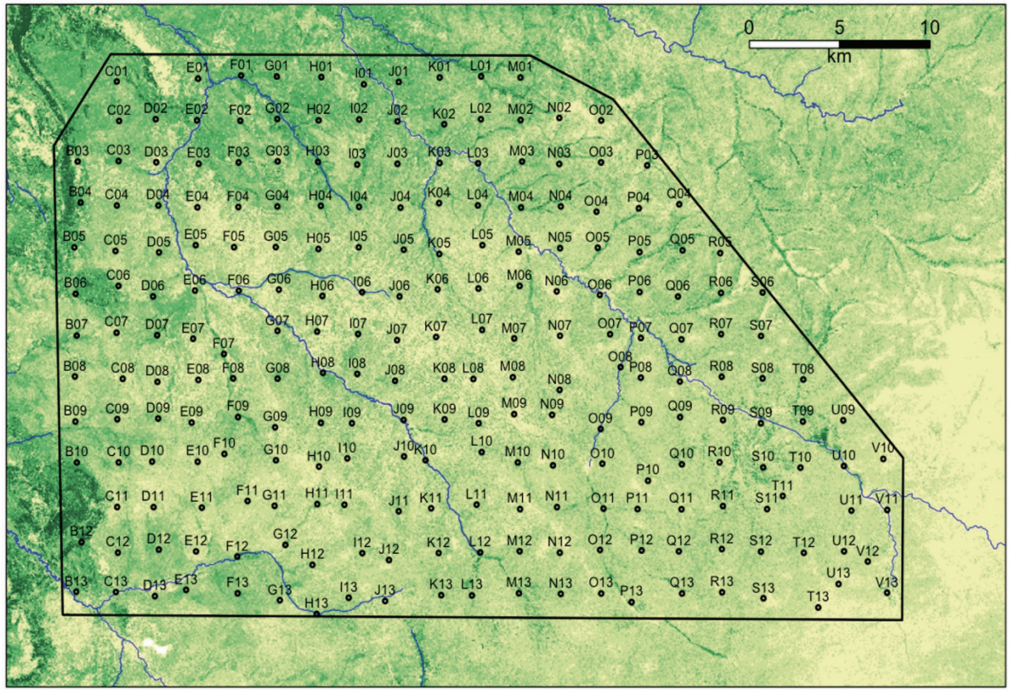
- **Two-phase training**

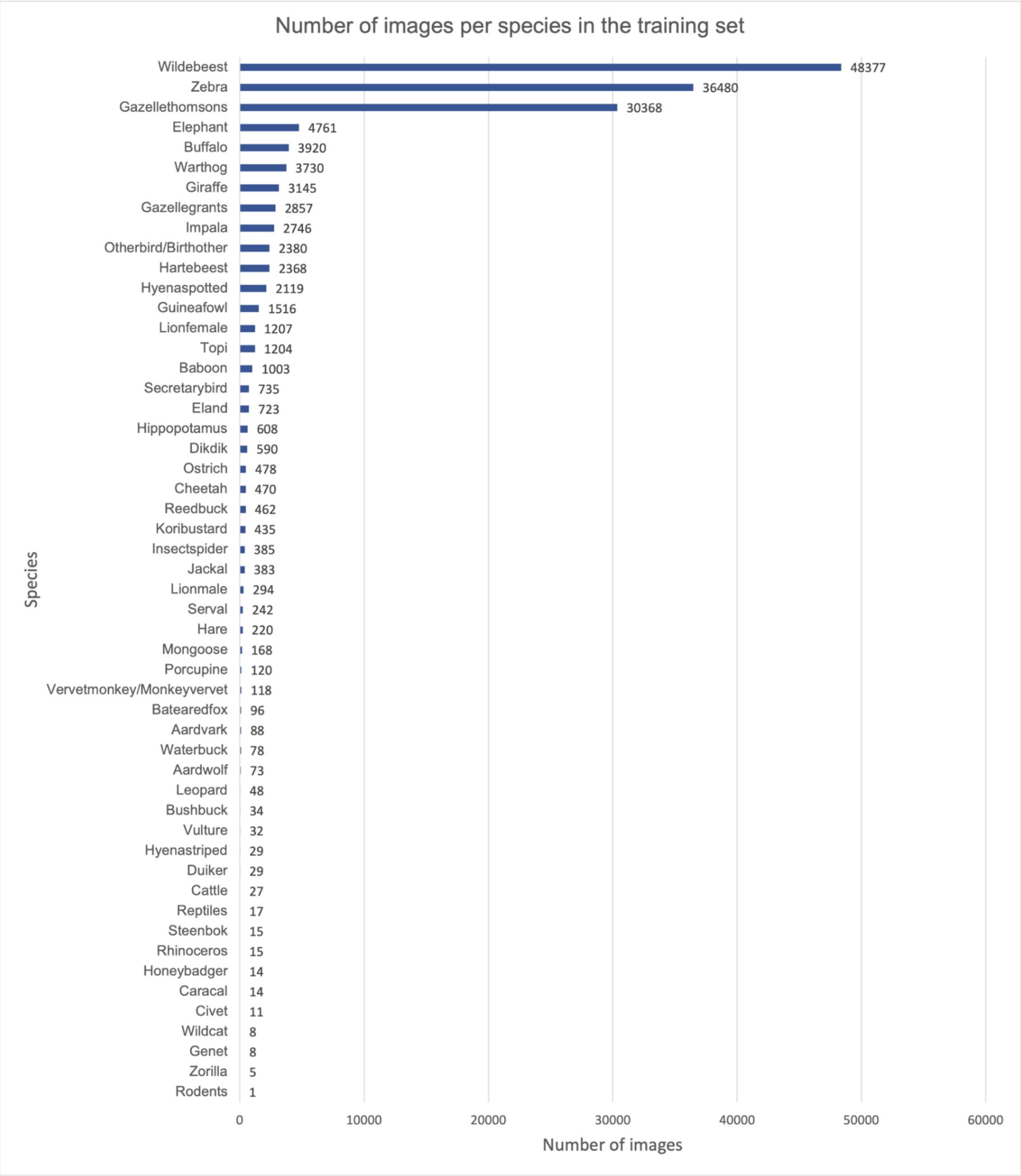


Methodology: data set

9th season of Snapshot Serengeti data set

- 80%-10%-10% train, validation, test split





Methodology: experiments

Baselines:

- ResNet-18
- ROS, RUS, ROS&RUS trained without 2nd phase

Two-phase training models:

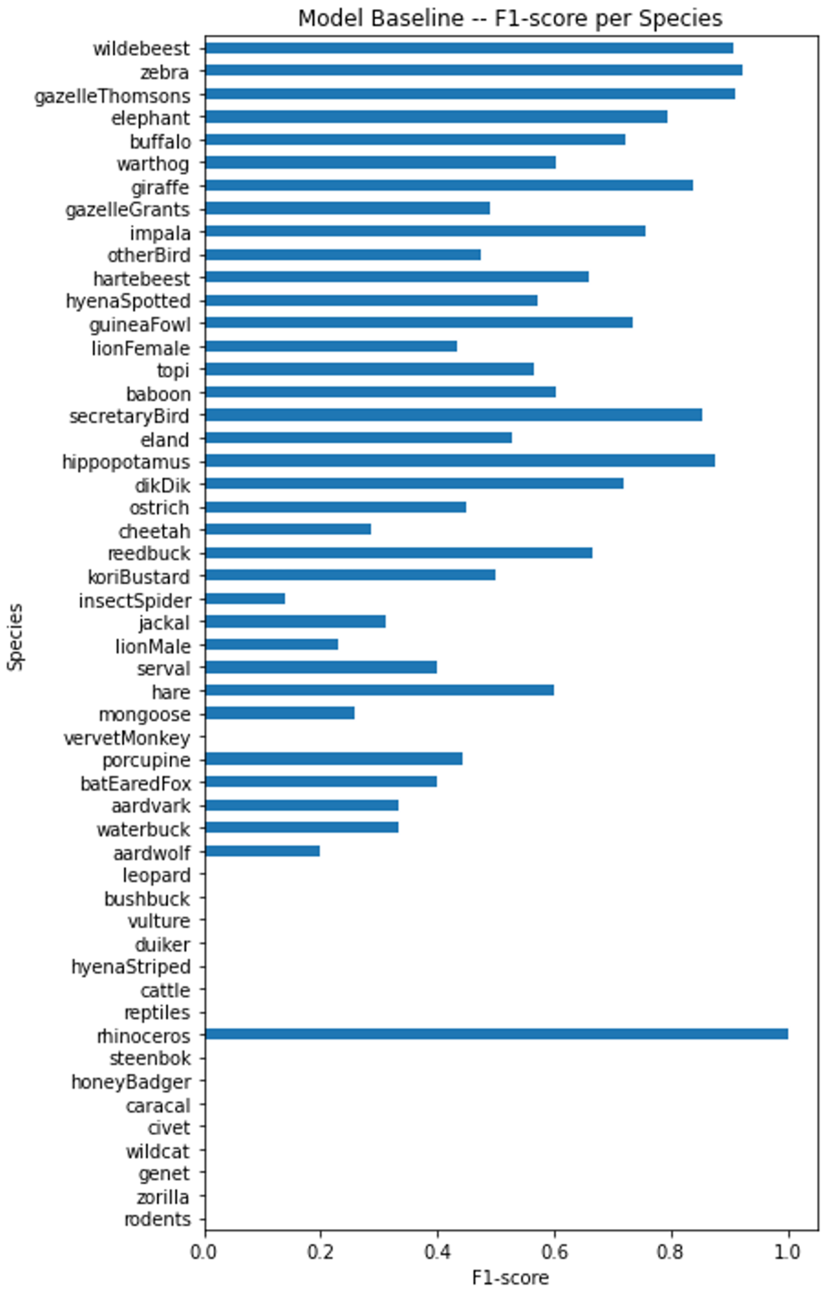
- ROS
- RUS
- ROS&RUS (15K)
- ROS&RUS (5K)

Models	**Oversampling**	**Undersampling**
Baseline	No	No
ROS	Yes, up to 5K	No
RUS	No	Yes, until 15K
ROS&RUS (15K)	Yes, up to 5K	Yes, until 15K
ROS&RUS (5K)	Yes, up to 5K	Yes, until 5K

Results: baseline model

Baseline Model
- Top-1 Accuracy = 85.27%
- Macro F1-score = 39.44%

Class specific performance:
- Better for majority classes
- Majority classes: recall > precision



KU LEUVEN

Results: models comparison

- Accuracy vs. baseline
  - Drops in phase 1 because of balanced data sets
  - Increases again to same value in phase 2
- Macro F1 vs. baseline
  - Drops in phase 1
  - Increases to higher value in phase 2

Model	Phase 1: Acc.	Phase 2: Acc.
Baseline	**0.8527**	/
ROS	0.8326	**0.8528**
RUS	0.8012	0.8491
ROS&RUS(15K)	0.8346	0.8454
ROS&RUS(5K)	0.7335	0.8066

Model Comparison - Top-1 Accuracy

Model	Phase 1: F1	Phase 2: F1
Baseline	0.3944	/
ROS	0.3843	0.4012
RUS	0.3681	**0.4147**
ROS&RUS(15K)	**0.4179**	0.4094
ROS&RUS(5K)	0.3620	0.4001

Model Comparison - F1 score

Discussion: limitations

- Overall accuracy lower than most relevant literature due to
  - Smaller number of data samples
  - Larger number of classes
  - Multiple images per capture event
- Results for smallest minority classes are less robust and need to be interpreted with care
- More robust results could be obtained by averaging over several runs

General conclusions

- ML can help to promote biodiversity conservation
- State-of-the-art camera trap image classifiers suffer from a majority class bias
- Two-phase training can be used to (partly) mitigate this bias
- Two-phase training leads to a better performance than only applying sampling techniques

Thank you!