Overview of BirdCLEF 2020

Bird Sound Recognition in Complex Acoustic Environments

Stefan Kahl | Center for Conservation Bioacoustics | Cornell Lab of Ornithology | Cornell University

Introduction

Passive acoustic monitoring is a cornerstone of the assessment of ecosystem health.

The improvement of automated assessment systems has the potential to have a transformative impact on global biodiversity monitoring.

→ *Involve the scientific community*

Introduction

The BirdCLEF challenge evaluates the state-of-the-art of audio-based bird identification systems at a very large scale.

The goal is to design, train and apply an automated detection system that can reliably recognize bird sounds in diverse soundscape recordings.

Large scale?

- Multiple hundreds of species covering three continents
- Tens of thousands of audio recordings
- Hundreds of hours of annotated test data

BirdCLEF is the largest evaluation campaign focusing on avian biodiversity monitoring.

Soundscapes?

Focal recordings

- High signal-to-noise-ratio
- Often non-overlapping vocalizations of a
- Single bird species

Photo: Jay McGowan

Soundscapes?

Soundscape recordings

- Low signal-to-noise-ratio
- Often overlapping vocalizations of
- Multiple bird species
- High levels of anthropogenic sounds

Photo: Center for Conservation Bioacoustics

Focal vs. soundscape

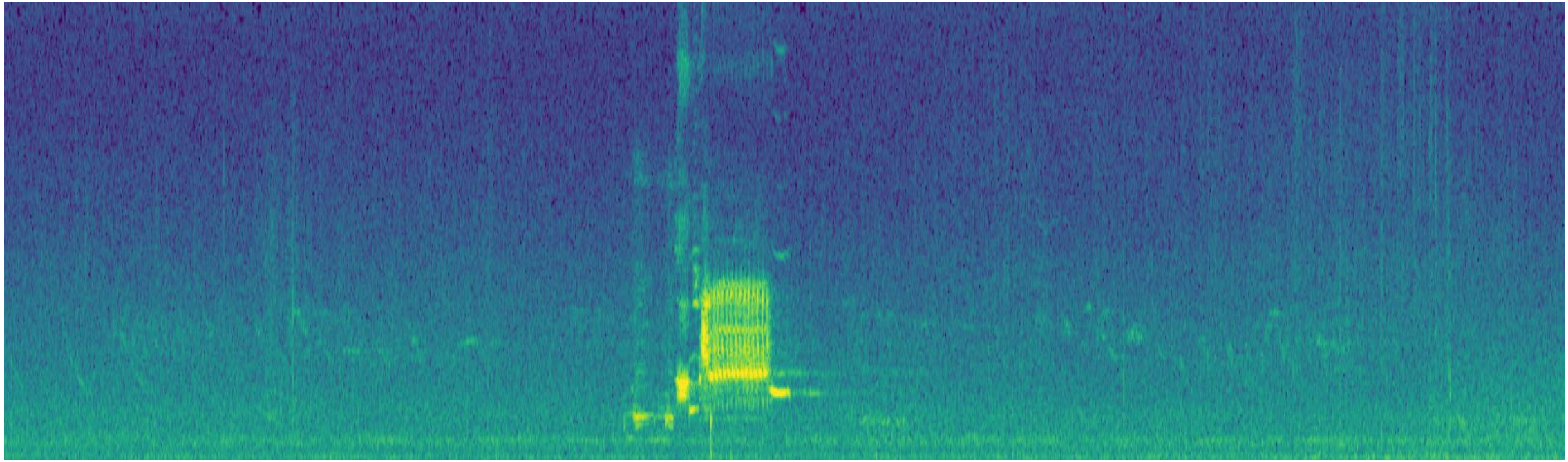


Photo: Phil Kahler | Audio: Russ Wigh (XC135705)

Focal vs. soundscape

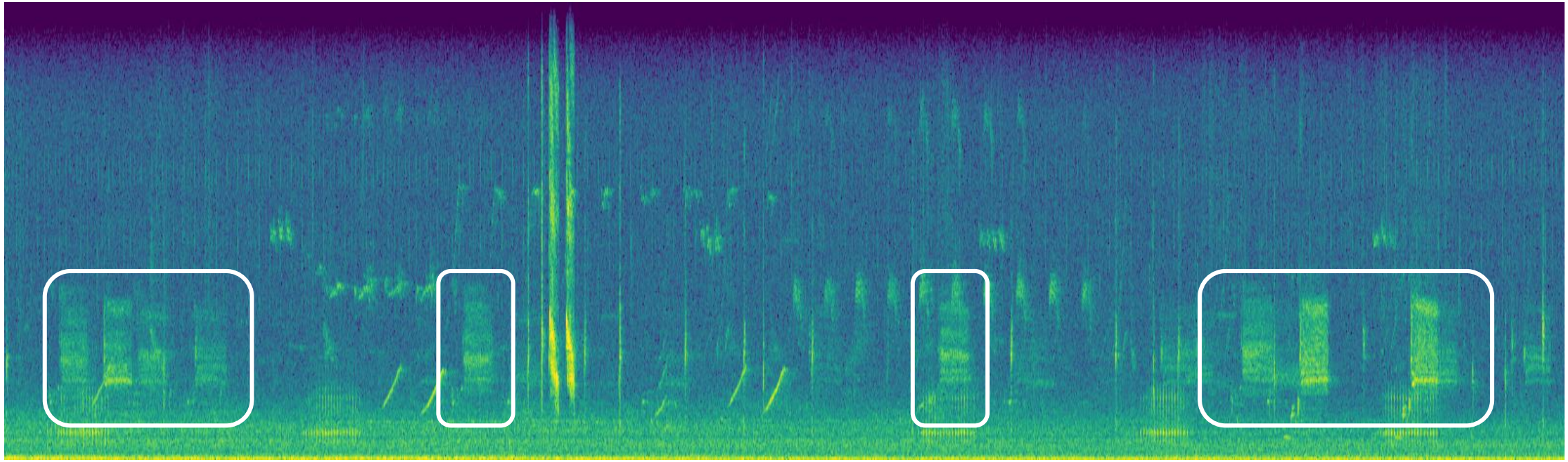


Photo: Phil Kahler | Audio: Center for Conservation Bioacoustics

Task

- Train a detection system on focal recordings only
- Apply the detection system to soundscape data
- Return a ranked list of detected species for each 5-second interval

Task

- Example results:

```
49845;00:00:00-00:00:05;compau;0.85555845
49845;00:00:10-00:00:15;compau;0.87857926
49845;00:00:10-00:00:15;lobher;0.8531503
49845;00:00:15-00:00:20;lobher;0.8800533
49845;00:00:15-00:00:20;grepot1;0.8581267
49845;00:00:25-00:00:30;lobher;0.8580585
```

- Evaluation uses ranking metrics (rmAP and cmAP)

Training data

xeno-canto

eBird

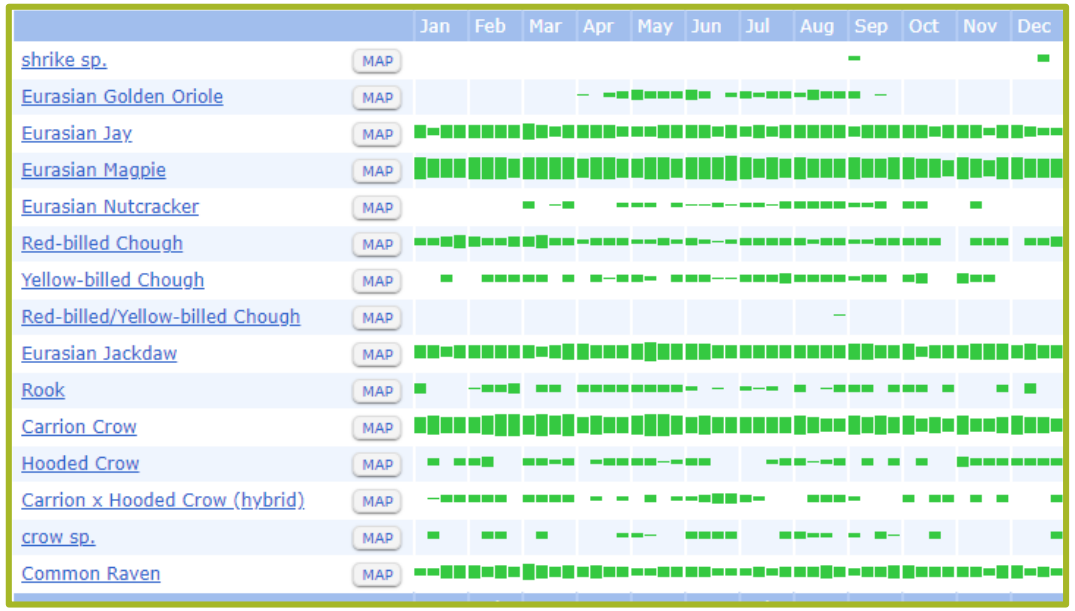
Xeno-canto:

- 960 species from South and North America and Europe
- ~950 hours across ~70,000 focal recordings

eBird:

- Year-round frequency data
- Based on checklists for each recording location

Participants were <u>**not**</u> allowed to use other than the provided audio and metadata.

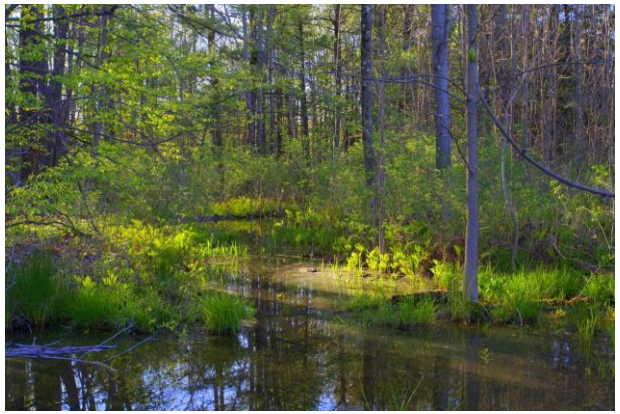


Test data

- 153 soundscapes of 10 minute duration recorded at 4 sites (25.5 hours total)
- Up to 8 species vocalizing at the same time (1.3 on average)

Peru

USA

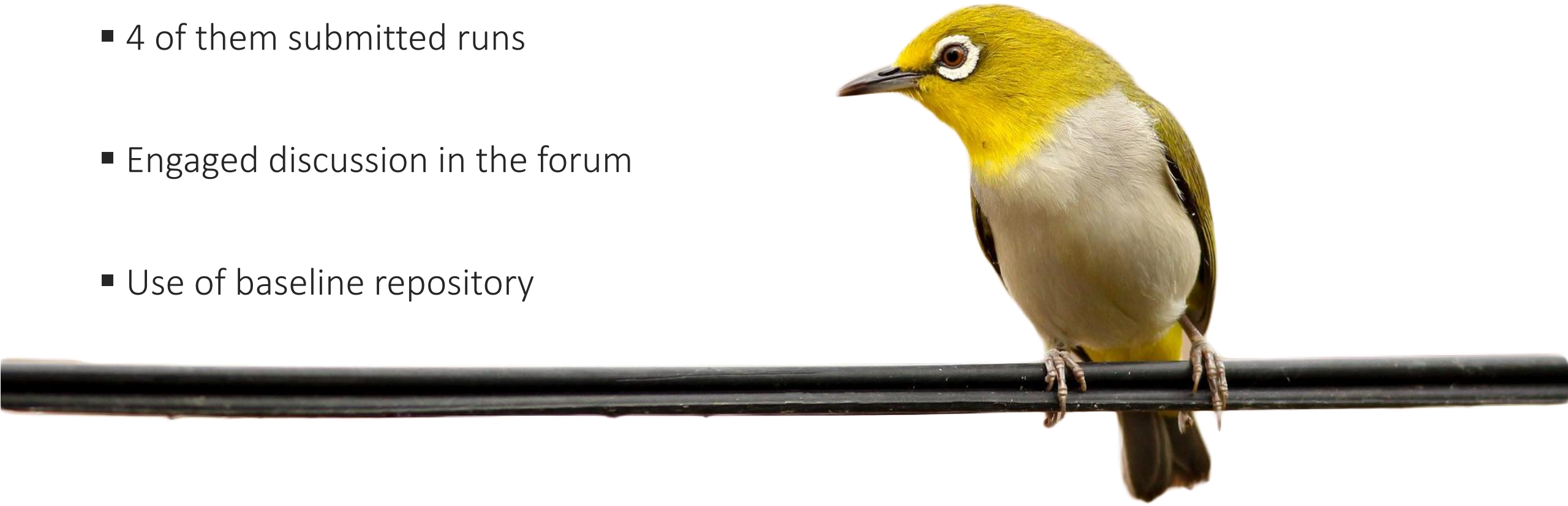


USA

Germany

Participants

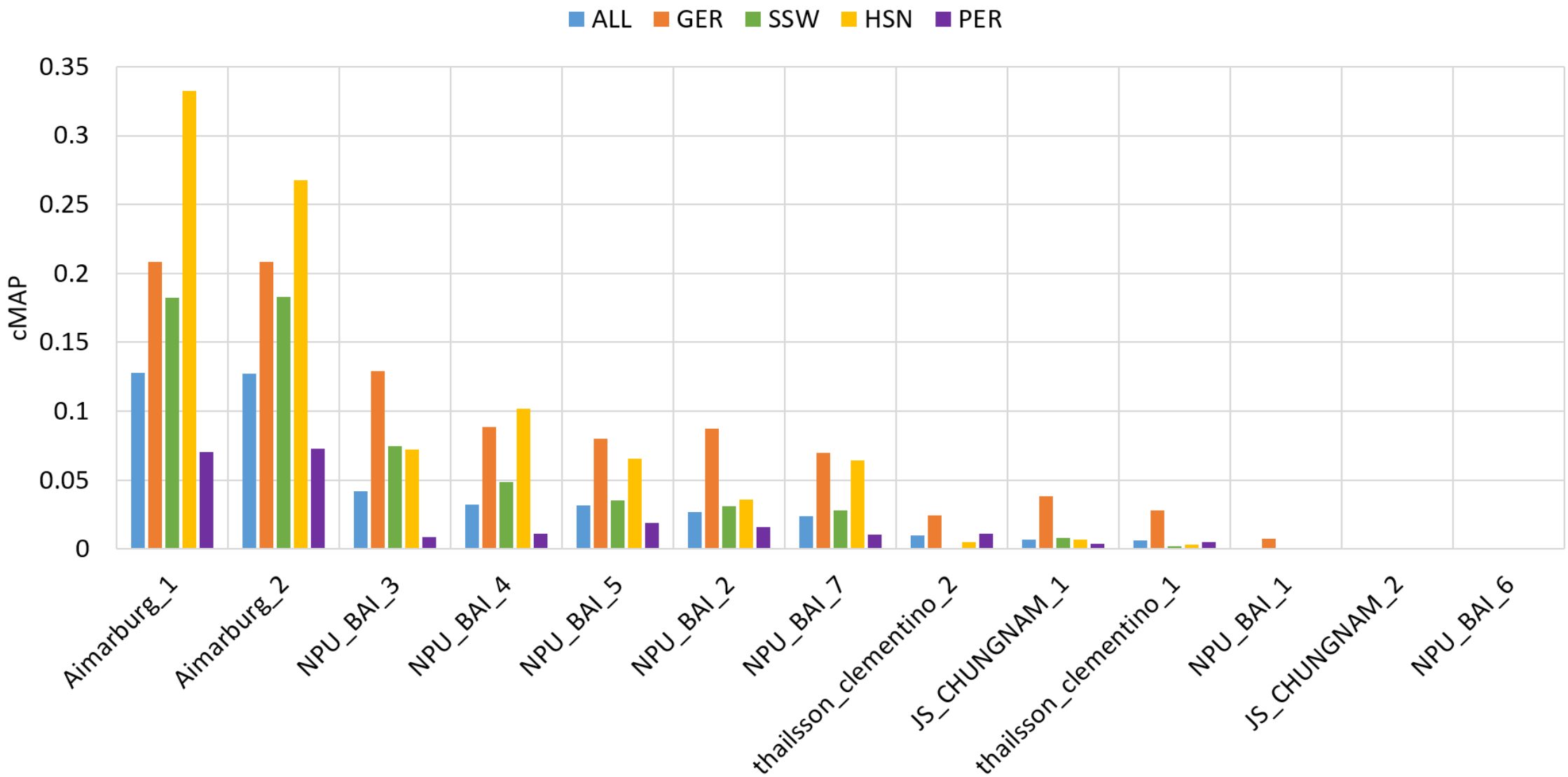
- 69 teams registered on aicrowd.com
- 4 of them submitted runs
- Engaged discussion in the forum
- Use of baseline repository



Results

- Deep artificial neural networks only
- Spectrograms as input
- Established architectures and custom designs (neural architecture search)
- Augmentation and training follow best practices

Results


ALL
GER
SSW
HSN
PER
cMAP
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Aimarburg_1
Aimarburg_2
NPU_BAI_3
NPU_BAI_4
NPU_BAI_5
NPU_BAI_2
NPU_BAI_7
thailsson_clementino_2
JS_CHUNGNAM_1
thailsson_clementino_1
NPU_BAI_1
JS_CHUNGNAM_2
NPU_BAI_6

Results

Site	Best score (cmAP)	Species per 5-second interval
High-Sierra Nevada, USA	0.33	0.48
Hesse, Germany	0.21	1.77
Ithaca, USA	0.18	0.73
Inkaterra Reserva, Peru	0.07	2.05

Detection performance vs. soundscape complexity

Lessons learned

- DNNs are the go-to tool for bird sound recognition
- Participants struggled with the amount of data
- Publicly available code repositories are worth the effort
- Results leave considerable room for improvements

Outlook

Habitat loss and the destruction of critical environmental niches pose a serious threat to many species.

Biodiversity assessments may only be possible for archived records of long destroyed areas.

Thank you!

Life Clef

xeno-canto

The Cornell Lab of Ornithology

eBird

UC DAVIS
UNIVERSITY OF CALIFORNIA

TECHNISCHE UNIVERSITÄT CHEMNITZ

OekoFor
Ökologische Datenerfassung und Forschung GbR

AIcrowd