Overview of BirdCLEF 2020: Bird Sound Recognition in Complex Acoustic Environments

Stefan Kahl[1], Mary Clapp[2], W. Alexander Hopping[1], Hervé Goëau[3], Hervé Glotin[4], Robert Planqué[5], Willem-Pier Vellinga[5], and Alexis Joly[6]

[1] Center for Conservation Bioacoustics, Cornell Lab of Ornithology, Cornell University, NY, USA, {sk2487, wah63}@cornell.edu
[2] University of California, Davis, CA, USA, mkclapp@ucdavis.edu
[3] CIRAD, UMR AMAP, Montpellier, France, herve.goeau@cirad.fr
[4] Université de Toulon, Aix Marseille Univ, CNRS, LIS, DYNI team, Marseille, France, herve.glotin@univ-tln.fr
[5] Xeno-canto Foundation, The Netherlands, {wp,bob}@xeno-canto.org
[6] Inria/LIRMM ZENITH team, Montpellier, France, alexis.joly@inria.fr

Abstract. Passive acoustic monitoring is a cornerstone of the assessment of ecosystem health and the improvement of automated assessment systems has the potential to have a transformative impact on global biodiversity monitoring, at a scale and level of detail that is impossible with manual annotation or other more traditional methods. The BirdCLEF challenge—as part of the 2020 LifeCLEF Lab [12]—focuses on the development of reliable detection systems for avian vocalizations in continuous soundscape data. The goal of the task is to localize and identify all audible birds within the provided soundscape test set. This paper describes the methodology of the conducted evaluation as well as the synthesis of the main results and lessons learned.

Keywords: LifeCLEF, bird, song, call, species, retrieval, audio, collection, identification, fine-grained classification, evaluation, benchmark, bioacoustics, ecological monitoring

1 Introduction

Accurate knowledge of the identity, the geographic distribution and the evolution of bird species is essential for a sustainable development of humanity as well as for biodiversity conservation. Monitoring avian populations is one of the most important approaches to assess ecosystems in terms of conservation priority—especially in regions with high overall biological diversity which often face extinction. Acoustic monitoring using Autonomous Recording Units (ARU)

---

allows researchers to conduct point counts in almost any densely vegetated habitat over longer periods of time and has become a widely used sampling tool in ecological research and monitoring over the past decade [28].

While automated assessment of soundscapes in the hyper-diverse tropics presents significant challenges relative to less diverse systems, it also presents higher upside. The tropics harbor a hugely disproportionate percentage of Earth's biodiversity, more than three-quarters of all species and over 90% of the planet's terrestrial birds, and international biodiversity targets will be impossible to meet if these systems are not conserved [4]. Despite their critical importance, however, the tropics are widely neglected in biodiversity and ecosystem function literature [36], and conclusions from studies in temperate regions are often wrongly used as the basis for assumptions about tropical systems where they do not apply [6]. The avian species richness in regions like the Amazon is so high that conducting reliable avian surveys in the Amazon is too overwhelming for most field observers, particularly when combined with the logistical challenges and poor viewing conditions typical of dense tropical forests [25]. Biased research priorities and difficulties with field surveys in the tropics have led to overlooked biodiversity losses [31] and flawed baseline species occurrence data [5], which introduce significant problems for assessing biodiversity change. ARUs, a cost-effective way of collecting data in systems that are difficult to access and survey, are well suited to address many of these issues by generating detailed data that can be recorded simultaneously, stored permanently, reviewed by multiple observers, and that is far less impacted by open or closed habitats than point counts, which can lead to the relative overestimation of species richness and abundance at sites with more favorable viewing conditions [8].

The improvement of automated assessment systems has the potential to have a transformative impact on global biodiversity monitoring, at a scale and level of detail that is impossible with manual annotation or other more traditional methods. Building automated assessment programs that can handle the unique challenges of hyper-diverse tropical ecosystems must become a central priority for conservation organizations and research groups with an interest in protecting Earth's biodiversity.

2 BirdCLEF 2020 challenge description

The high amount of effort required to manually analyse recorded soundscapes means that fully analyzing large datasets is prohibitively time-intensive, neutralizing many of the advantages provided by continuous recording across multiple sites. The *LifeCLEF Bird Recognition Challenge* (BirdCLEF) focuses on the development of reliable detection systems for avian vocalizations in continuous soundscape data. Launched in 2014, it has become the largest bird sound recognition competition in terms of dataset size and species diversity with multiple tens of thousands of recordings covering up to 1,500 species [12,14].

2.1 Goal and evaluation protocol

The goal of the evaluated task is to localize and identify all audible birds within the provided soundscape test set. Each soundscape is divided into segments of 5 seconds, and a list of species associated to probability scores has to be returned for each segment. The used evaluation metric is the classification mean Average Precision ($cmAP$), considering each class c of the ground truth as a query. This means that for each class c, all predictions with $ClassId = c$ are extracted from the run file and ranked by decreasing probability in order to compute the average precision for that class. The mean across all classes is computed as the main evaluation metric. More formally:

$$cmAP = \frac{\sum_{c=1}^{C} AveP(c)}{C}$$

where C is the number of classes (species) in the ground truth and $AveP(c)$ is the average precision for a given species c computed as:

$$AveP(c) = \frac{\sum_{k=1}^{n_c} P(k) \times rel(k)}{n_{rel}(c)}.$$

where k is the rank of an item in the list of the predicted segments containing c, n_c is the total number of predicted segments containing c, $P(k)$ is the precision at cut-off k in the list, $rel(k)$ is an indicator function equaling 1 if the segment at rank k is a relevant one (*i.e.* is labeled as containing c in the ground truth) and $n_{rel}(c)$ is the total number of relevant segments for class c.

2.2 Dataset

The 2020 BirdCLEF challenge featured the largest, fully-annotated collection of soundscape recordings from four different recording locations : in Peru, Germany and two in USA. With respect to real-world use cases, labels and metrics were chosen to reflect the vast diversity of bird vocalizations and high ambient noise levels in omnidirectional recordings.

Training data: Deploying a bird sound recognition system to a new recording and observation site requires classifiers that generalize well across different acoustic domains. Focal recordings of bird species from around the world form an excellent base to develop such a detection system. However, the lack of annotated soundscape data for a new deployment site poses a significant challenge. As in previous editions, training data was provided by the Xeno-canto community [1] and consisted of more than 70,000 recordings covering 960 species from three continents (South and North America and Europe). Participants were allowed to use this and other (meta) data to develop their systems. A representative validation dataset with two hours of soundscape data was also provided, but participants were not allowed to use this data for training—detection systems had to be trained on focal recordings only.

---

[1] https://www.xeno-canto.org

(a) PER recording habitat (b) HSN recording habitat

(c) SSW recording habitat (d) GER recording habitat

Fig. 1: Test data recording locations. ARU were used to collect audio data of targeted ecosystems at a large scale. Photos: W. A. Hopping, Mary Clapp, Brian Maley, Hendrik Reers.

Test data: In addition to the 2019 test data [14], soundscapes from three other recording sites were added in the 2020 edition of BirdCLEF. All audio data were collected with passive acoustic recorders from deployments in Germany (GER), Peru (PER), the High Sierra Nevada (HSN) of California-USA, and the Sapsucker Woods area (SSW) in New York-USA (locations are illustrated Fig. 1). In an attempt to lower the entry level of this challenge, the total amount of soundscape data was reduced to 153 recordings with a duration of ten minutes each (25.5 hours total). Expert ornithologists provided annotations for often extremely dense acoustic scenes with up to 8 species (1.3 on average) vocalizing at the same time (Fig. 3).

PER - Inkaterra Reserva Amazonica, Peru: This acoustic data was collected at the Inkaterra Reserva Amazonica (henceforth “ITRA”, 12°32’07.8”S, 69°02’58.2”W) between January 14 and February 2 2019, during the rainy season. ITRA is a 2km^2 lowland rainforest reserve on the banks of the Madre de Dios river, approximately 20km east of the frontier town of Puerto Maldonado.

(a) September 15, 2015 (b) September 11, 2018

Fig. 2: Satellite images (from Google Earth) showing active encroachment on Inkaterra Reserva Amazonica (green).

The reserve consists primarily of seasonally inundated Várzea forest. The western Amazon is the most biodiverse terrestrial system on the planet [11], and is part of the world's largest remaining intact tropical system [1]. The region's extraordinary biodiversity is threatened by accelerating rates of deforestation, degradation, and fragmentation, which are driven primarily by expanding road networks, mining, agriculture, and an increasing population [26]. ITRA is actively threatened by these processes, particularly encroachment from small-scale agriculture and selective logging, which has occurred illegally within its boundaries (Fig. 2). These threats are magnified by the site's proximity to Puerto Maldonado, the largest settlement in the region, which can be reached via an unmarked road that passes within 1km of the reserve.

The acoustic data from this site was collected as part of a study designed to assess spatio-temporal variation in avian species richness and vocal activity levels across intact, degraded, and edge forest, and between different days at the same point locations. ITRA is expected to experience permanent changes to its species composition and richness in coming years, as a result of the accelerating deforestation, degradation, and fragmentation in the region, as well as climate change [4]. Similar processes, including the extirpation of certain species, have been documented in forest fragments in other parts of the Amazon [33]. The impending changes to ITRA's soundscape mean that high-quality historic bioacoustic data is likely to be of particular value for future comparative studies, including those concerning primates, bats, katydids, and other non-avian taxa. Accordingly, collecting this data for archival reasons was an important incentive for the project.

Ten SWIFT recording units, provided by the Center for Conservation Bioacoustics at the Cornell Lab. of Ornithology[2], were placed at separate sites span-

---

[2] https://www.birds.cornell.edu/ccb/

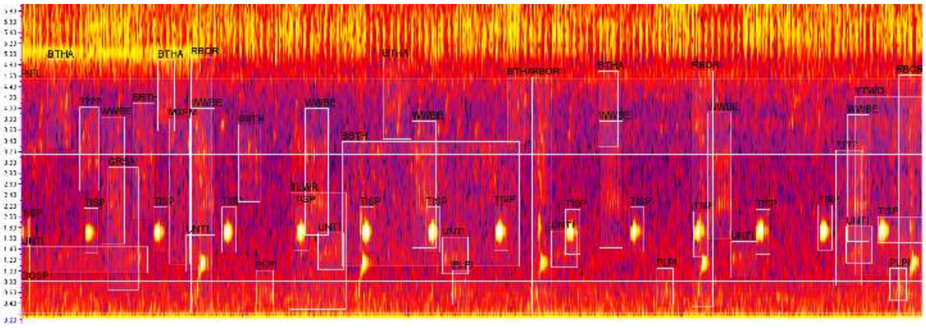

Fig. 3: Example of an annotated soundscape recording. South American soundscapes often have an extremely high call density. The 2020 BirdCLEF test data contains 48 fully annotated soundscapes recorded in Peru.

ning edge habitat, degraded forest, and intact forest within the reserve. These omnidirectional recorders were set to record continuously for the duration of their deployment, with a sampling rate of 48kHz, and a gain setting of 35dB. Sites were separated by approximately 350m to limit the chance of overlapping detection, and to ensure independence between sites [24]. Recorders were placed at a consistent height of approximately 1.5m above the ground. To minimize background noise, all sites used for data analysis were located at a minimum distance of 450m from the river.

A total of 21 dawn-hours, from 05:00-06:00 PET (10:00-11:00 UTC), representing 7 of the 10 sites on three randomly-selected dates, were manually annotated. Many neotropical bird species sing almost exclusively during the dawn hour [22], so this time window was selected to maximize the number of species present in the recordings. Roughly 20,000 individual bounding box annotations, representing 123 foreground species, were made for all audible bird vocalizations (Fig. 3). Annotations were particularly focused on vocalizations at frequencies below 5kHz, which were less likely to overlap with insect noise. Because these annotations included varying numbers of unidentified vocalizations, 9 of the 21 total dawn-hours, with the highest quality annotations and lowest share of unknown vocalizations, were used for BirdCLEF 2020. These 9 recordings featured 6,399 foreground annotations, 1.6% of which were unidentified, and 4.6% of which were either unidentified or grouped with other vocalizations of the same type, but were not connected to a known species or taxa.

HSN - Sierra Nevada/High Sierra, CA, USA: The recordings were made in Sequoia and Kings Canyon National Parks, two contiguous national parks on the southern end of the Sierra Nevada mountain range in California, USA. Ninety-six percent of the Parks' 865,964 acres is federally designated wilderness, which affords the highest level of protection to the land from human development. The Parks' boundary spans several biomes, from mid-elevation chaparral and conifer and Sequoia forest to alpine tundra and includes the highest peak in the lower 48 states, Mount Whitney/Tumanguya[3].

The focus of the acoustic study was the high-elevation region of the Parks; specifically, the headwater lake basins above 3km in elevation. These lake basins are comprised primarily of perennial graminoid vegetation (wet grasses and forbs), sparse evergreen woodland (foxtail and whitebark pine), riparian scrub (primarily willow/Salix sp.), and talus or boulderfield (Fig. 1). The avian community becomes less speciose above timberline due to decreased overall productivity compared to lowland and montane zones [29]. Alpine areas are highly seasonal, covered in deep snowpack for roughly half the year.

The original intent of the study was to monitor seasonal activity of birds and bats at lakes containing trout and lakes without trout. Recreational trout stocking in the historically fishless lakes of the Sierra Nevada began in the late 1800's, and though it was phased out in the National Parks in the 1970's and banned altogether in 1991, self-sustaining populations of trout persist in roughly 50% of the High Sierra's historically fishless watershed [16]. The disruptive effects of introduced trout within the aquatic system are well-studied: the diversity, distribution and abundance of many native aquatic fauna are vastly reduced in lakes containing trout [17,20]. However, the cascading impacts of trout on the adjacent terrestrial zone remain poorly understood, despite the importance of aquatic-terrestrial subsidies in the food web dynamics of many ecological systems [23,32].

Additional support for this project from the Parks was due to the fact that high-elevation areas are particularly vulnerable to climate change: range contractions for alpine specialists and drastic turnover to the composition of alpine communities are predicted due to changes in temperature and precipitation regimes linked to global climate change [18]. Most importantly, these changes are likely to happen heterogeneously across taxa as species distributions shift over both space and time to track their thermal niches, changing food supply, and/or habitat structures [34,2,30], making conservation decisions nuanced and difficult. High-quality data with large-scale replicability on the occurrence and distribution of alpine species is a pressing need for understanding extinction risk and conserving biodiversity as the climate changes. However, data collection in wilderness is limited by difficulty in accessibility: in the absence of roads, researchers must

---

[3] The United States government designated the land now known as Sequoia as a National Park in 1890, and what is now known as Kings Canyon was designated in 1940. Prior to their forced removal and relocation to reservations in surrounding areas, the land throughout the current Park boundary was inhabited and tended by the Me-wuk and Monache people, and the High Sierra in particular was traversed by the Me-wuk, Monache, Yokut, Western Mono Waksachi, and potentially other unrecognized tribes.

travel on foot for several kilometers simply to reach these remote alpine locations. Passive acoustic recorders are a promising tool to collect such data because they can be deployed at multiple locations, generate a permanent and unbiased record of vocal animal activity, and can be used into the future as both research needs and analytical technologies evolve [28].

Soundscapes were recorded for 24h continuously at 10 lakes (5 fishless, 5 fish-containing) throughout Sequoia and Kings Canyon National Parks during June-September 2015. Pilot data were collected at a subset of these lakes during the same months in 2014. SongMeter SM2+ units (Wildlife Acoustics, USA) powered by custom-made solar panels were used to obviate the need to swap batteries, due to the recording locations being extremely difficult to access. SongMeters recorded mono-channel 16-bits wav files continuously. This resulted in roughly 6TB of data, which were stored on external hard drives, servers at Cornell University, and backed up in Google Cloud Storage. To create an annotated subset of data used as evaluation data in the 2020 BirdCLEF challenge, 50 10-minute segments of audio between 9 and 12 July, 2015 from morning hours (05:10-09:10 PDT) from all 10 sites were selected at random. One expert annotated this subset: using RavenPro 1.5, a selection box was placed around each sound and annotated it to species. Every sound that could not be confidently assigned an identity was reviewed with 1-2 other experts in bird identification. When consensus on ID could not be reached, the sound was marked as “unknown.” To minimize observer bias, all identifying information about the location, date and time of the recordings was hidden. We used high-quality sound-cancelling headphones to minimize variation in the ambient environment that would interfere with hearing. We observed that it was difficult to reliably identify sounds with an absolute maximum amplitude less than 40dB; therefore, we only annotated sounds that exceeded this amplitude.

SSW - Sapsucker Woods, Ithaca, NY, USA: As part of the Sapsucker Woods Acoustic Monitoring Project (SWAMP), the Center for Conservation Bioacoustics at the Cornell Lab of Ornithology deployed 30 in-house developed acoustic recorders (called SWIFT) to the surrounding area. Each of the units records acoustic data at 48kHz sampling rate covering the frequency of all bird calls occurring in Sapsucker Woods. The ongoing study aims at investigating the vocal activity and diversity of local bird species as well as the impact of noise pollution (the local airport and highway 13 are in close proximity) on the behaviour of birds—especially focusing on changes in vocal output as anthropogenic sounds may alter the way that birds communicate. Over the past four years, more than 1 million hours of soundscape data have been recorded and stored. This amount of data is only fully accessible through automated analysis by using reliable detection systems that are robust against unforeseen environmental sounds and cope well with overlapping vocalizations. In 2018, expert birders annotated 20 complete days of audio data that were recorded between January and June of 2017 and provided almost 80,000 labels across randomly selected recordings (24 out of 30 recording sites per day with one hour per site). The 2019 edition of BirdCLEF used twelve of these days as test and three as validation data. This year, we limited the amount of test data to 48 10-minute recordings that also include previously unreleased audio from this deployment. This reduction became necessary to balance the test data and to reduce the bias towards one dataset. Six randomly selected recordings were provided as validation data to allow participants to fine-tune their systems. However, we chose to only release validation data for the SSW and PER dataset which forced participants to develop generic approaches that are location-independent.

GER - Laubach, Hesse, Germany: Forests, the habitat of woodpeckers, owls and other bird species, are increasingly being used as locations for wind farms in the wake of the transformation of energy production. Noise, movement or scenery effects could have a disturbing impact on these forest bird species and lead to the avoidance of forest areas near wind farms. Using automated acoustic recorders, this study aimed to investigate whether wind farms lead to a change in the use of forest habitats by selected bird species. In order to produce reliable and objectively recorded data in sufficient quantities to be able to investigate relatively small effects, passive acoustic monitoring is extremely valuable and above all cost-effective, which is what makes this kind of study possible in the first place. On the other hand, acoustic monitoring is also suitable for the detection and monitoring of rare, endangered or particularly sensitive animals, which could not be monitored by human observers. Yet, only automatic processing of the recordings allows us to handle such data quantities.

Over the course of two seasons (March to June 2019 and 2020), 100 solar-powered passive recorders (based on a Raspberry Pi 3 A+ with USB soundcard and mono microphone) were deployed across 11 wind parks with close-proximity forest (Fig. 1). Each device started to record for two hours around sunrise (one before, one after) and one hour after sunset. The entire data collection comprised 25,000 hours of audio data. We randomly selected 9 soundscapes with a duration of 10 minutes each and annotated all audible bird vocalizations using Audacity. The selection included dawn chorus recordings as well as dusk and night-time soundscapes with low or no bird activity adding to the overall diversity of this year's test data.

3 Results

A total of 69 participants registered for the BirdCLEF 2020 challenge and downloaded the dataset. Four teams succeeded in submitting runs, Fig. 4 shows individual scores achieved by each team sorted by rank in 2020. Details of the methods and systems used in the runs are synthesized in this overview and further developed in the individual working notes of the participants. Most submitted runs scored best for the High Sierra data which has the lowest call density of all subsets (0.48 species per 5-second interval, best cmAP=0.33). The GER

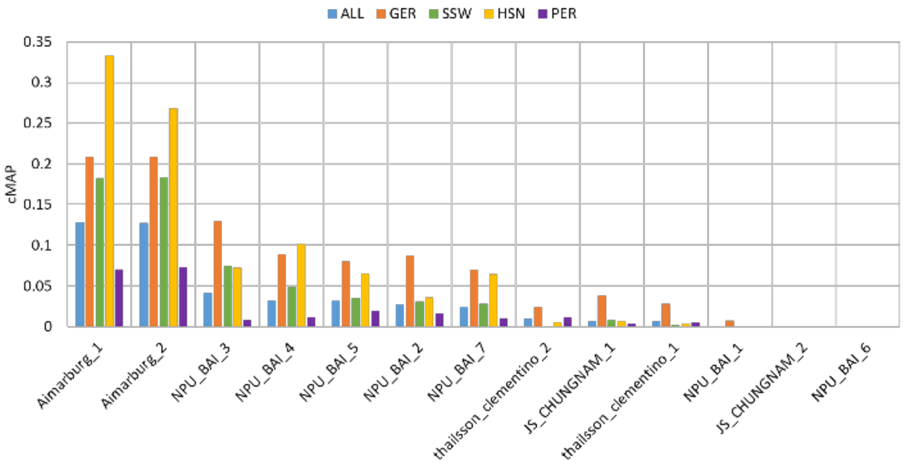

Fig. 4: Scores achieved by all systems evaluated within the bird identification task of LifeCLEF 2020.

subset has the second highest vocal density and medium score (1.77 species per 5-second interval, best cmAP=0.21), scores for SSW are significantly lower in most runs (0.73 species per 5-second interval, best cmAP=0.18). All teams struggled with the Peruvian soundscape dataset that has a significantly higher species diversity which appears to pose a significant challenge (2.05 species per 5-second interval, best cmAP=0.07). Interestingly, the two first competitions on another Peruvian high-density soundscape dataset demonstrated the same trends (best cmAP=0.08 at BirdCLEF2017 [10], best cmAP=0.12 at BirdCLEF2018 [9]). However, it appears that a high vocal activity alone does not suffice to predict how well a recognition system will perform. In contrast to the PER data, GER audio recordings contain species that vocalize frequently over longer periods of time which helps classifiers to identify them eventually. A high species diversity and only occasionally vocalizing species seem to prevent recognition systems from achieving higher cmAP scores.

3.1 Aimarburg [21], Best run overall

Due to the restrictions of the training data, this team decided to implement a network architecture search (NAS) to cope with the diversity of the classification task. The authors argue that an optimal network architecture is vital when the use of pre-trained networks is prohibited. The authors use pre-selected 5-second audio snippets from the training data as input of the neural network. Each segment was analyzed for the presence of bird sounds using the heuristic developed for the 2018 baseline system [15]. A number of augmentations were applied to the data before passing it to the first layer of the network—a 1-D convolution acting as Gabor wavelet transformation. The remaining layers of the network were established by performing the NAS with a restricted search space (i.e., specific layer types and operations) and an evolutionary algorithm. A number of output heads based on recurrent layers conclude the network topology. Species lists for each recording location of the test data were used to filter the detections. The best scoring run submitted by this team achieved a cmAP of 0.128 and a rmAP of 0.193 thus being the best overall result.

3.2 NPU-BAI [3]

This team implemented a more traditional approach building on the results of previous editions of BirdCLEF. The participants decided to base their classifier architecture on the Xception neural network—an advanced version of the Inception-v3 model which performed well in the past years [27,19]. Again, preprocessed spectrograms were used as input for the network and a number of augmentations were applied to the input samples. Most notably, mixup training is used to simulate multiple birds vocalizing by overlaying samples of focal recordings. This method led to a significant improvement compared to other trials and can be considered the most important addition to the training regime. This observation is backed by other attempts in the same domain [13]. This team managed to achieve a cmAP of 0.042 and rmAP of 0.067 with their best submission.

3.3 Thailsson Clementino [7]

Triplet loss has been shown to perform well in different classification scenarios [35] and focusing on similarities instead of categories can help to cope with limited amounts of training data (which is the case for many South American bird species). This team decided to implement a Siamese network with triplet loss to generate unique features for each bird vocalization from input mel spectrograms. A kNN classifier with Euclidean distance and a multilayer perceptron acted as classification instances based on extracted feature embeddings. An AlexNet-like architecture with 5 convolutional and 3 dense layers was used to train the feature extractor that provided the best features and thus the best scoring submission. The participants achieved a cmAP of 0.063 and a rmAP of 0.108 with their (unofficial) post-deadline submission. In their working note, the authors argue that data augmentation and other variations of the triplet loss might help to improve the performance of this attempt.

4 Conclusion

Passive acoustic monitoring is an important sampling tool for habitat assessments—especially for highly endangered environments with often extraordinarily high biodiversity. Despite the fact that automated tools for analyzing soundscape recordings are far from perfect, the manual examination and annotation of field recordings is extremely labor-intensive and often negates the benefits of acoustic monitoring compared to human point counts. Habitat loss and the destruction of critical environmental niches pose a serious threat to many species, and biodiversity assessments may only be possible for archived records of long destroyed areas. The annual BirdCLEF sound recognition challenge is the largest evaluation campaign that specifically aims at developing state-of-the-art classifiers to help researchers to cope with conservation challenges of our time. Deep neural networks provide good overall baselines in many domains and adapting architectures and training regimes to suit the domain of acoustic event recognition will be a major focal point of future editions. Bird sounds are an extremely diverse class of acoustic events and entering this domain has become increasingly challenging for new participants of BirdCLEF. We will strive to further lower this barrier to allow more teams to develop and test their ideas so that they can contribute to this high-impact field of research.

Acknowledgements: The organization of the BirdCLEF task is supported by the Xeno-canto Foundation, the European Union and the European Social Fund (ESF) for Germany, Jake Holshuh (Cornell class of '69), the Arthur Vining Davis Foundations, as well as by the French CNRS projects SABIOD.ORG, SEAMED, EADM GDR MADICS, BIOSA STIC-AmSud, ANR-18-CE40-0014 SMILES and ANR-20-CHIA-0014 ADSIL. We want to thank all expert birders who helped to annotate SSW soundscapes with incredible effort: Cullen Hanks, Jay McGowan, Matt Young, Randy Little, and Sarah Dzielski. We would also like to thank OekoFor for providing soundscapes and annotations for the GER dataset, which was funded by the German Federal Agency for Nature Conservation.

Note: The challenge is open for post-deadline submissions at aicrowd.com and the data will be available for further download and use. Please do not hesitate to contact the organizers if you have any questions or would like to use the data to evaluate your system.

https://www.aicrowd.com/challenges/lifeclef-2020-bird-monophone

References

1. Allan, J.R., Venter, O., Watson, J.E.: Temporally inter-comparable maps of terrestrial wilderness and the last of the wild. Scientific data 4, 170187 (2017)
2. Auer, S.K., King, D.I.: Ecological and life-history traits explain recent boundary shifts in elevation and latitude of western n orth a merican songbirds. Global Ecology and Biogeography 23(8), 867–875 (2014)
3. Bai, J., Chen, C., Chen, J.: Xception based system for bird sound detection. In: CLEF working notes 2020, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2020, Thessaloniki, Greece. (2020)
4. Barlow, J., França, F., Gardner, T.A., Hicks, C.C., Lennox, G.D., Berenguer, E., Castello, L., Economo, E.P., Ferreira, J., Guénard, B., et al.: The future of hyperdiverse tropical ecosystems. Nature 559(7715), 517–526 (2018)

5. Boakes, E.H., McGowan, P.J., Fuller, R.A., Chang-qing, D., Clark, N.E., O'Connor, K., Mace, G.M.: Distorted views of biodiversity: spatial and temporal bias in species occurrence data. PLoS Biol 8(6), e1000385 (2010)
6. Clarke, D.A., York, P.H., Rasheed, M.A., Northfield, T.D.: Does biodiversity–ecosystem function literature neglect tropical ecosystems? Trends in ecology & evolution 32(5), 320–323 (2017)
7. Clementino, T., Colonna, J.G.: Using triplet loss to bird species recognition on birdclef 2020. In: CLEF working notes 2020, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2020, Thessaloniki, Greece. (2020)
8. Darras, K., Batáry, P., Furnas, B., Celis-Murillo, A., Van Wilgenburg, S.L., Mulyani, Y.A., Tscharntke, T.: Comparing the sampling performance of sound recorders versus point counts in bird surveys: A meta-analysis. Journal of applied ecology 55(6), 2575–2586 (2018)
9. Goëau, H., Glotin, H., Vellinga, W.P., Planqué, R., Joly, A.: LifeCLEF Bird Identification Task 2017. In: CLEF: Conference and Labs of the Evaluation Forum. vol. CEUR Workshop Proceedings. Dublin, Ireland (Sep 2017), `https://hal.archives-ouvertes.fr/hal-01629175`
10. Goëau, H., Kahl, S., Glotin, H., Planqué, R., Vellinga, W.P., Joly, A.: Overview of BirdCLEF 2018: monospecies vs. soundscape bird identification. In: CLEF: Conference and Labs of the Evaluation Forum. vol. CEUR Workshops Proceedings. Avignon, France (Sep 2018), `https://hal.archives-ouvertes.fr/hal-02189229`
11. Jenkins, C.N., Pimm, S.L., Joppa, L.N.: Global patterns of terrestrial vertebrate diversity and conservation. Proceedings of the National Academy of Sciences 110(28), E2602–E2610 (2013)
12. Joly, A., Goëau, H., Kahl, S., Deneu, B., Servajean, M., Cole, E., Picek, L., Ruiz De Castañeda, R., é, Lorieul, T., Botella, C., Glotin, H., Champ, J., Vellinga, W.P., Stöter, F.R., Dorso, A., Bonnet, P., Eggel, I., Müller, H.: Overview of lifeclef 2020: a system-oriented evaluation of automated species identification and species distribution prediction. In: Proceedings of CLEF 2020, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2020, Thessaloniki, Greece. (2020)
13. Kahl, S.: Identifying Birds by Sound: Large-scale Acoustic Event Recognition for Avian Activity Monitoring. Ph.D. thesis, Chemnitz University of Technology (2019)
14. Kahl, S., Stöter, F.R., Glotin, H., Planqué, R., Vellinga, W.P., Joly, A.: Overview of birdclef 2019: Large-scale bird recognition in soundscapes. In: CLEF working notes 2019, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2019, Lugano, Switzerland. (2019)
15. Kahl, S., Wilhelm-Stein, T., Klinck, H., Kowerko, D., Eibl, M.: Recognizing birds from sound - the 2018 birdclef baseline system. arXiv preprint arXiv:1804.07177 (2018)
16. Knapp, R.A.: Non-native trout in natural lakes of the sierra nevada: an analysis of their distribution and impacts on native aquatic biota. In: Sierra Nevada ecosystem project: final report to Congress. vol. 3, pp. 363–407. Centers for Water and Wildland Resources (1996)
17. Knapp, R.A., Matthews, K.R., Sarnelle, O.: Resistance and resilience of alpine lake fauna to fish introductions. Ecological monographs 71(3), 401–421 (2001)
18. La Sorte, F.A., Jetz, W.: Projected range contractions of montane biodiversity under global warming. Proceedings of the Royal Society B: Biological Sciences 277(1699), 3401–3410 (2010)
19. Lasseck, M.: Bird species identification in soundscapes. In: CLEF working notes 2019 (2019)

20. Matthews, K.R., Knapp, R.A., Pope, K.L.: Garter snake distributions in high-elevation aquatic ecosystems: is there a link with declining amphibian populations and nonnative trout introductions? Journal of herpetology pp. 16–22 (2002)
21. Mühling, M., Franz, J., Korfhage, N., Freisleben, B.: Bird species recognition via neural architecture search. In: CLEF Working Notes 2020, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2020, Thessaloniki, Greece. (2020)
22. Parker, Theodore A., I.: On the Use of Tape Recorders in Avifaunal Surveys. The Auk 108(2), 443–444 (04 1991), `https://doi.org/10.1093/auk/108.2.443`
23. Polis, G.A., Anderson, W.B., Holt, R.D.: Toward an integration of landscape and food web ecology: the dynamics of spatially subsidized food webs. Annual review of ecology and systematics 28(1), 289–316 (1997)
24. Ralph, C.J., Droege, S., Sauer, J.R.: Monitoring bird populations by point counts, vol. 149, chap. Managing and monitoring birds using point counts: Standards and applications, pp. 161–168. US Department of Agriculture, Forest Service, Pacific Southwest Research Station (1995)
25. Robinson, W.D., Lees, A.C., Blake, J.G.: Surveying tropical birds is much harder than you think: a primer of best practices. Biotropica 50(6), 846–849 (2018)
26. Sánchez-Cuervo, A.M., de Lima, L.S., Dallmeier, F., Garate, P., Bravo, A., Vanthomme, H.: Twenty years of land cover change in the southeastern peruvian amazon: implications for biodiversity conservation. Regional Environmental Change 20(1), 8 (2020)
27. Sevilla, A., Glotin, H.: Audio bird classification with inception v4 joint to an attention mechanism. In: Working Notes of CLEF 2017 (Cross Language Evaluation Forum) (2017)
28. Shonfield, J., Bayne, E.: Autonomous recording units in avian ecological research: Current use and future applications. Avian Conservation and Ecology 12(1) (2017)
29. Siegel, R., Wilkerson, R., Saracco, J., Steel, Z.: Elevation ranges of birds on the sierra nevada's west slope. Western Birds 42(1), 2–26 (2011)
30. Socolar, J.B., Epanchin, P.N., Beissinger, S.R., Tingley, M.W.: Phenological shifts conserve thermal niches in north american birds and reshape expectations for climate-driven range shifts. Proceedings of the National Academy of Sciences 114(49), 12976–12981 (2017)
31. Socolar, J.B., Valderrama Sandoval, E.H., Wilcove, D.S.: Overlooked biodiversity loss in tropical smallholder agriculture. Conservation Biology 33(6), 1338–1349 (2019)
32. Soininen, J., Bartels, P., Heino, J., Luoto, M., Hillebrand, H.: Toward more integrated ecosystem research in aquatic and terrestrial environments. BioScience 65(2), 174–182 (2015)
33. Stouffer, P.C.: Birds in fragmented amazonian rainforest: Lessons from 40 years at the biological dynamics of forest fragments project. The Condor (2020)
34. Tingley, M.W., Beissinger, S.R.: Cryptic loss of montane avian richness and high community turnover over 100 years. Ecology 94(3), 598–609 (2013)
35. Weinberger, K.Q., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. Journal of Machine Learning Research 10(2) (2009)
36. Wilson, K.A., Auerbach, N.A., Sam, K., Magini, A.G., Moss, A.S.L., Langhans, S.D., Budiharta, S., Terzano, D., Meijaard, E.: Conservation research is not happening where it is most needed. PLoS Biology 14(3), e1002413 (2016)