

Improving Bird Recognition using Pseudo-Labeled Recordings from the Target Location

Mario Lasseck

Museum für Naturkunde Berlin, Germany

Abstract

This paper presents a deep learning approach to identify bird species in soundscape recordings with Convolutional Neural Networks (CNNs). The proposed method employs an iterative process to create pseudo labels for a large number of unlabeled recordings from the target location and applies them during training to significantly improve model performance and address the domain shift between training and test data. The effectiveness of the approach is evaluated in the BirdCLEF 2024 competition hosted on Kaggle, where it achieves a macro-averaged area under the ROC curve (AUC) of 69 % on the official test set. This performance positions the method among the top two systems for identifying birds in wildlife monitoring recordings of the Western Ghats, a major biodiversity hotspot in India.

Keywords

Bird Species Recognition, Biodiversity Assessment, Soundscapes, BirdCLEF, Deep Learning, Domain Adaptation, Pseudo-Labeling, Semi-Supervised Learning, Kaggle Competition

1. Introduction

The BirdCLEF 2024 competition focuses on developing automated systems for detecting and classifying under-studied bird species in the Western Ghats. This mountain range, a global biodiversity hotspot in India, hosts a variety of endemic and endangered species, including many found nowhere else in the world. As the region faces drastic landscape and climatic changes, there's an urgent need for advanced conservation tools to assess and monitor its unique birdlife. The challenge aims to identify native species of the Western Ghats sky-islands, classify rare birds with limited training data and detect elusive nocturnal species. This year's edition introduces several challenges and unique aspects:

- Participants must address a significant domain shift between the training data, which consists of focal recordings from various locations, and the test data, which comprises soundscapes from the Western Ghats.
- The competition imposes a strict time limit for species identification in the test set, adding a practical constraint that mirrors real-world applications to assess and monitor biodiversity.
- To aid in bridging the domain gap, an additional unlabeled dataset from the target location is provided, allowing participants to explore un- and semi-supervised learning techniques.

CLEF 2024: *Conference and Labs of the Evaluation Forum, September 09–12, 2024, Grenoble, France*
EMAIL: Mario.Lasseck@mfn.berlin

© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

By improving the accuracy and efficiency of bird identification algorithms under these constraints, this initiative supports ongoing conservation efforts, such as those led by V. V. Robin's Lab at IISER Tirupati [1]. These innovations will empower researchers and practitioners to more effectively track avian population trends, evaluate threats and refine their conservation strategies in this ecologically crucial region.

Further details about the BirdCLEF 2024 competition are given in [2], [3] and [4]. The task is part of the LifeCLEF 2024 evaluation campaign [5,6] and the Conference and Labs of the Evaluation Forum [7,8].

2. Materials and Methods

The implementation of the machine learning based system for bird species recognition presented in this paper builds upon solutions for previous BirdCLEF competitions and similar tasks [9,10,11,12,13]. Further details on own past developments and implementation methods can be found for example in [14], [15], [16] and [17].

2.1. Datasets

The BirdCLEF 2024 training data consists of 24459 audio recordings provided by Xeno-canto [18], covering 182 different bird species. Unique to this year's task, an additional 8444 unlabeled recordings are provided from the same location as the test set soundscapes. Table 1 provides an overview of the individual datasets and their characteristics. All recordings are resampled to 32 kHz, converted to mono, and compressed to Ogg format.

Xeno-canto files are weakly labeled, meaning there is no precise information on the presence or absence of the labeled bird within the recording. However, there is a high probability of hearing the labeled bird at the beginning of each audio file, as recordists often trim their recordings accordingly before uploading them. To exploit this characteristic, only the first 5 seconds of recordings are used for training. For some recordings, one or more background species are also provided as secondary labels.

For cross-validation, the training dataset is split into 5 or 8 stratified randomized folds, ensuring that primary species are proportionally represented in each fold.

Table 1: Datasets overview and statistics

	Training set	Unlabeled set	Test set
Recording type	Focal	Soundscape	Soundscape
Source	Various locations (Xeno-canto)	Western Ghats	Western Ghats
# Recordings	24459	8444	1100
Min. duration per rec.	0.47s	20s	4m
Max. duration per rec.	1h 39m 24s	4m	4m
Acc. duration all rec.	11d 20h 50m 30s	23d 6h 19m 11s	3d 1h 20m
# Species / Classes	182	unknown	unknown
Min. # rec. per class	5	unknown	unknown
Max. # rec. per class	500	unknown	unknown

2.2. Feature Engineering

The public notebook [19] of Salman Ahmed [20] was used as a baseline for feature engineering and early model training, following discussions on the Kaggle forum [21] initiated by lihaoweicvch [22]. All models are trained on 5-second audio chunks represented as spectrograms. The raw 1D audio signal is converted to a 2D log Mel spectrogram image using the MelSpectrogram [23] and AmplitudeToDB [24] classes from the torchaudio.transforms library [25].

The baseline system uses:

- First 5 seconds of training files and no extra recordings or classes from other sources
- Model input: resized 3 channel Mel spectrogram images of size 256x256 pixel
- CNN backbone: eca_nfnet_10 [26] pretrained on ImageNet [27]
- Mel spectrogram parameters:
 - `n_fft = 2048`
 - `hop_length = 512`
 - `n_mels = 128`
 - `f_min = 20`
 - `f_max = 16000`
- Training parameters:
 - CosineAnnealingLR scheduler [28] with 5 warmup epochs [29]
 - Peak learning rate $1e-4$
 - 100 epochs with early stopping if AUC is not improving for 7 epochs
 - Batch size 64
 - Average of binary cross-entropy [30] and focal loss [31] as loss function
 - Generalized-Mean (GeM) pooling
- Augmentations:
 - HorizontalFlip [32]
 - CoarseDropout [33]
 - Mixup of Mel spectrogram images within training batches

This system achieves a maximum AUC of 66 % on the public test set. From this baseline, experiments were conducted with different CNN backbones, hyperparameter settings, augmentation methods and input image sizes. A major drawback of the initial model was its relative long submission time of over one hour. In addition to improving the score, one objective was to reduce inference time to fit more models in an ensemble without exceeding the 2-hour submission time limit. To address this, the CNN backbone was replaced with an EfficientNet B0 architecture (`tf_efficientnet_b0_ns` [34]) and the Mel spectrogram image was reduced to smaller dimensions. Results were initially unstable, with a public leaderboard score ranging from 62 % to 66 % AUC and were very sensitive to different combinations of Mel parameters and input image sizes. However, with further adjustments, it was possible to create single models with an inference time of around 12 minutes, still achieving a score of approximately 65 % AUC.

Main changes to the initial model included:

- CNN backbone: `tf_efficientnet_b0_ns`
- 5 dropout layers before the fully connected classification layer (inspired by models of BirdCLEF2021 2nd [35] and BirdCLEF2023 4th [36] place solutions)
- Higher learning rate ($1e-3$), less warmup epochs (3) and less training epochs (50)
- Different Mel parameters (`n_mels`, `hop_length`)
- Additional augmentation: local and global time and frequency stretching performed on Mel spectrogram images via resizing parts and/or the entire image
- Creating checkpoint soups instead of using early stopping

2.3. Training Methods

The training data is divided into 5 or 8 folds, stratified according to primary labels. Only the first 5 seconds of each audio file are used for training. The models are trained using Convolutional Neural Network (CNN) backbones, specifically `tf_efficientnet_b0_ns`, which are pretrained on ImageNet. The training process employs the AdamW [37] optimizer and a one-cycle CosineAnnealingLR scheduler with a peak learning rate of $1e-3$ and 3 warmup epochs. The average of binary cross-entropy and focal loss is used to optimize model performance.

For validation, the first 5 seconds of the files in the validation set are used to track learning progress through evaluation metrics Label Ranking Average Precision (LRAP) [38], cMAP [39], F1 [40] and AUC [41]. Background species are included with a target value of 1.0 and are treated equally to primary labeled species.

To enhance model stability and performance, "checkpoint soups" are used for single model inference. This follows the idea of model soups [42]. But here, weights from different checkpoints of the same model (typically from epochs 13-50) are averaged, provided there is an improvement in local cross-validation scores in at least one of the tracked metrics. This approach leads to more stable and occasionally better performance. For ensemble inference, predictions from several models are combined using simple mean averaging.

The above-described modifications to the baseline model allowed the creation of an ensemble of six models, achieving 70 % AUC. This ensemble was subsequently used to generate a first set of pseudo labels.

Performance Improvement with Pseudo Labels

Pseudo labels are created by applying the model ensemble on the unlabeled recordings from the test location. The predictions from all 5-second intervals of the 8444 unlabeled soundscapes form a large set of 401947 soft pseudo labels.

In the subsequent training stages, randomly selected audio segments from the pseudo-labeled recordings are mixed with the training samples at a probability of 25 to 45 percent. Before combining the audio signals, the amplitudes of both waveforms are multiplied by a random factor. The target vector of the training sample (with a value of 1.0 for primary and secondary species and 0 for others) is combined with the pseudo label vector (containing predicted probabilities) to form the new target vector by taking the maximum value of both.

Incorporating pseudo labels into training significantly improved scores for both single models and ensembles. The enhanced ensemble was then used to generate a new set of pseudo labels and this cycle was repeated multiple times to progressively improve model and ensemble performance. The iterative pseudo-labeling process is described in Figure 1. Its impact on public and private leaderboard scores is illustrated in Table 2 and visualized in Figure 2.

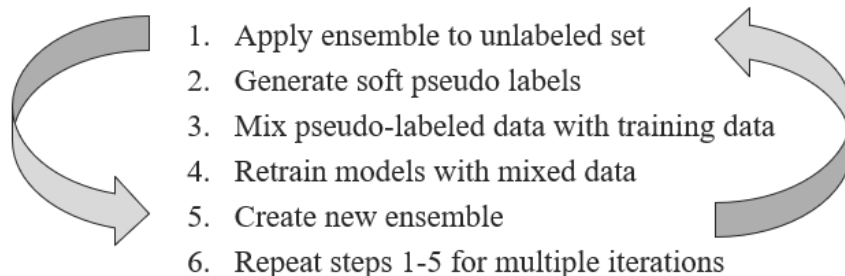
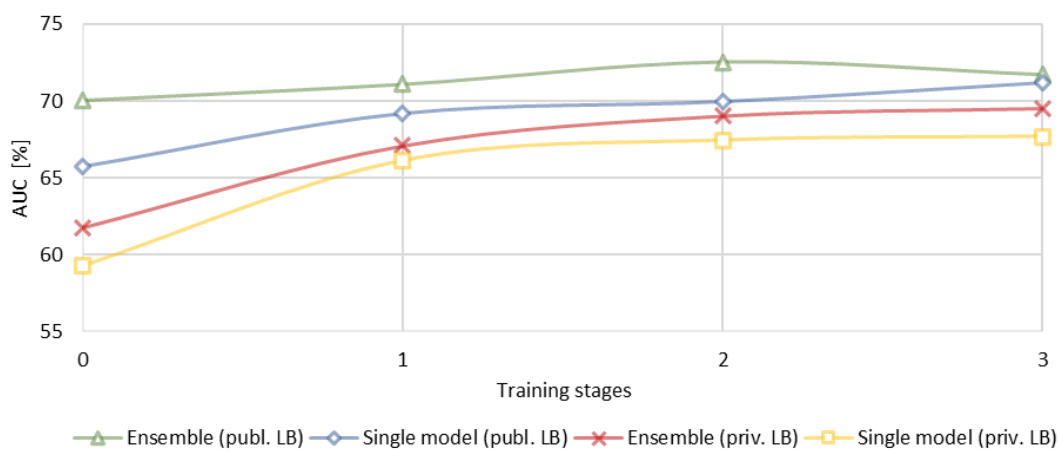


Figure 1: Iterative pseudo-labeling process to improve single model and ensemble performance

Table 2: Performance improvement using pseudo labels from different training stages

Stage	Pseudo labels	Single model (ID 4)		Ensemble	
		publ. priv. LB AUC [%]	publ. priv. LB AUC [%]		
0	-	65.735 59.270	70.065 61.738		
1	From stage 0 ensemble	69.165 66.119	71.090 67.084		
2	From stage 1 ensemble	69.936 67.445	72.528 69.035		
3	From stage 2 ens. (normalized)	71.154 67.683	71.716 69.527		

After the second iteration, pseudo label values became too large and required normalization by rescaling them back to the range [0,1] to allow stable model training. Unfortunately, the stage 3 ensemble was not selected for final ranking because public leaderboard score did not reveal the expected improvement.

**Figure 2:** Visualization of performance improvement using pseudo labels from different training stages

Post-Processing

Models are ensembled by simply taking the mean of predictions (probabilities from sigmoid outputs) of each individual model. As a final step, for each test file, predictions of a given time window are summed with those of the two neighboring windows using an aggregation factor of 0.5. This post-processing method was previously applied by Theo Viel and his team in the 3rd place solution [43] of the Cornell Birdcall Identification competition [44].

Inference Optimizations

To speed up inference, audio files from the test set are preprocessed in parallel using multithreading. Additionally, different versions of Mel spectrogram images are pre-calculated and reused for different models in the ensemble. By including models that work on smaller image sizes, ensembles of up to six models can run within the 2-hour limit to create predictions for all 1100 recordings in the test set.

Due to variations in the hardware provided by Kaggle for running inference notebooks, particularly in CPU types, the number of models that could be ensembled to identify all birds in the test set within the given time frame varied. To prevent submission errors, a timer is implemented in the notebook to ensure completion within the 2-hour limit. If the timer reaches approximately 118 minutes, inference is stopped and results are collected for all models and predicted file parts up to that point. Predictions from unfinished models or file parts are masked before averaging.

3. Results

The training and pseudo-labeling approach described in this paper secured 2nd place among a total of 974 participating teams. Final scores on the public and private leaderboards, as well as the ranking of the top 10 teams, are presented in Table 3. By combining several diverse models, a macro-averaged ROC-AUC of 69.035 % was achieved on the complete test set (see team 'adsr' in Table 3).

Table 3: Competition results of the top 10 teams (with solution of team 'adsr' described in this paper)

Rank	Team Name on Kaggle	AUC [%] (publ. LB)	AUC [%] (priv. LB)
1	Team Kefir	73.857	69.039
2	adsr	72.794	69.035
3	NVBird	74.212	68.997
4	Team Cerberus	74.691	68.777
5	coolz	74.396	68.717
6	penguin46	72.039	68.716
7	Team Unicorn	72.809	68.383
8	kapenon	69.660	67.928
9	Aphysict	71.453	67.891
10	Tamo	70.132	67.623

Parameters and performance of the six models from the 2nd place solution (2nd stage ensemble in Table 2) are detailed in Table 4. Model diversity in the ensemble is achieved by varying Mel parameters, data subsets, image sizes, the probability of adding pseudo labels and amplitude factors to adjust the volume ratio between training and pseudo-labeled data. The parameters ampExpMin and ampExpMax in Table 4 specify the range for the random amplitude factor applied to training and pseudo-label samples to adjust their volume in the mix:

$$\text{ampFactor} = 10^{**}(\text{random.uniform}(\text{ampExpMin}, \text{ampExpMax}))$$

Table 4: Single model parameters and performances of the 2nd place ensemble

Params. / Model ID	1	2	3	4	5	6
seed	42	42	42	42	70	42
n_folds	5	5	5	5	10	5
fold	4	1	4	4	0	4
dataset	bc24	bc24	bc24	bc24	bc24+	bc24
n_mels	128	128	128	64	64	64
hop_length	512	512	1024	1024	1024	1024
image_height	256	256	128	64	64	64
image_width	256	256	128	128	128	64
pseudoLabelChance [%]	35	40	45	30	30	25
ampExpMin	-0.5	-1.0	-0.5	-0.5	-0.5	-0.5
ampExpMax	0.1	0.2	0.1	0.1	0.1	0.1
Inference time	~ 50 min.	~ 50 min.	~ 17 min.	~ 12 min.	~ 12 min.	~ 11 min.
Public LB AUC [%]	73.270	71.975	71.104	69.936	69.124	69.309
Private LB AUC [%]	68.521	68.533	68.116	67.445	64.543	65.862

Model 5 in Table 4 is the only one utilizing external data. For this model, additional files for the 182 species in the competition were downloaded from Xeno-canto. The first 5 seconds of each file were added to the training set, with shorter files being padded with zeros to ensure a uniform length.

4. Discussion

As in previous editions of the BirdCLEF competition, the challenge was to use focal recordings from Xeno-canto to train a system capable of accurately identifying bird species in soundscapes. The inference time was again limited to 2 hours. However, compared to last year, over twice the amount of data had to be processed within that time (recordings with a total duration of 1 day, 9 hours and 20 minutes in 2023 vs. 3 days, 1 hour and 20 minutes in 2024). This placed even more constraints on the size and number of models that could be used to process all recordings in the test set. Other challenges included the extreme domain shift between training and test data, a significant class imbalance in the training samples (with some classes having only five example recordings per species) and the lack of diversity in the training material for many under-studied species in the target location.

Fortunately, a large set of unlabeled soundscapes from the same locations as the test data was provided this year. With this dataset, it was possible to create pseudo labels and find an effective method of incorporating them into training to significantly improve identification performance. The approach described in this paper, using pseudo-labeled data from soundscapes of the deployment location, combines several advantages:

1. **Noise augmentation:** By mixing training samples with samples from the target domain, the model learns how species sound within the environmental background noise of the test site habitat. This helps to address the domain shift between Xeno-canto recordings and test soundscapes.
2. **Training data extension:** The model receives more training samples representing the noise characteristics and species distribution of the deployment location.
3. **Knowledge distillation:** Since pseudo labels are derived from predictions of a stronger model (or ensemble of models in this case), its knowledge is transferred during training to the smaller model.

For pseudo-labeling, only ensembles that fit the time limit constraint were used for inference. Using larger ensembles or including models with stronger backbones (e.g. with a higher number of layers for feature extraction) would likely lead to better pseudo labels. It would be interesting to investigate in future experiments how much further scores can be improved if stronger pseudo labels are incorporated during training.

With only two of the best models from the 2nd place system (models 1 and 2 in Table 4), it is possible to achieve a private leaderboard score of 69.694 % AUC. The combination of these two models takes much less time for inference compared to using all six models. It surpasses the score of the entire ensemble and even the 1st place system of the competition (69.0391 % AUC). Another interesting finding is that, combined with pseudo-label training, the SED architecture with attention on frequency bands from last year [16] achieves the best single model score (69.701 % AUC on private leaderboard). This again proves that the feature engineering, network architecture, augmentation techniques and training methods of the BirdCLEF 2023 3rd place system [45] are quite robust and work well for the data and species sets of this year's task.

A customized version of the model to identify European bird species is available on GitHub [46]. It was successfully implemented in a number of tools and projects to assess and monitor avian biodiversity [47,48,49,50,51,52] and is also part of Naturblick [53], a smartphone application to discover and learn about nature in urban surroundings.

5. Acknowledgements

I would like to thank Stefan Kahl, Holger Klinck, Maggie, Sohier Dane, Tom Denton, Vijay Ramesh, Maximilian Eibl, Chiti Arvind, Harikrishnan C.P., Viral Joshi, V.V. Robin, Suyash Sawant, Alexis Joly, Henning Müller, Divya Mudappa, T.R. Shankar Raman, Meghana Srivathsa, Akshay V. Anand, Willem-Pier Vellinga and all involved institutions and individual contributors (Kaggle, Chemnitz University of Technology, Columbia University, Google Research, Indian Institute of Science Education and Research Tirupati, K. Lisa Yang Center for Conservation Bioacoustics, LifeCLEF, Nature Conservation Foundation, Parry Agro Industries Ltd., Project Dhvani, Tamil Nadu Forest Department, Tata Coffee Ltd., Tea Estates India Ltd., The Rufford Foundation, The University of Florida and Xeno-canto) for organizing this competition.

I also want to thank the Museum für Naturkunde and the team of the Animal Sound Archive Berlin [54] in particular Karl-Heinz Frommolt, Olaf Jahn and Benjamin Werner for supporting my work. The research was partly funded by the BMEL (Bundesministerium für Ernährung und Landwirtschaft) within the project “Machbarkeitsstudie - Integration (bio-)akustischer Methoden zur Quantifizierung biologischer Vielfalt in das Waldmonitoring” (FKZ: 2221NR050B).

6. References

- [1] <https://www.skyisland.in/>
- [2] Klinck H, Maggie, Dane S, Kahl S, Denton T, Ramesh V (2024) BirdCLEF 2024. Kaggle. <https://kaggle.com/competitions/birdclef-2024>
- [3] <https://www.imageclef.org/node/316>
- [4] Kahl S, Denton T, Klinck H, Ramesh V, Joshi V, Srivathsa M, Anand A, Arvind C, Harikrishnan CP, Sawant S, Robin VV, Glotin H, Goëau H, Vellinga WP, Planqué R, Joly A (2024) Overview of BirdCLEF 2024: Acoustic identification of under-studied bird species in the Western Ghats. In: Working Notes of CLEF 2024 - Conference and Labs of the Evaluation Forum
- [5] <https://www.imageclef.org/LifeCLEF2024>
- [6] Joly A, Picek L, Kahl S, Goëau H, Espitalier V, Botella C, Deneu B, Marcos D, Estopinan J, Leblanc C, Larcher T, Šulc M, Hružík M, Servajean M et al. (2024) Overview of lifeCLEF 2024: Challenges on Species Distribution Prediction and Identification. In: International Conference of the Cross-Language Evaluation Forum for European Languages, Springer, 2024
- [7] Faggioli G, Ferro N, Galuščáková P, García Seco de Herrera A (Ed.) (2024) Working Notes of CLEF 2024 - Conference and Labs of the Evaluation Forum
- [8] Goeuriot L, Mulhem P, Quénot G, Schwab D, Soulier L, Di Nunzio GM, Galuščáková P, García Seco de Herrera A, Faggioli G, Ferro N (Ed.) (2024) Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Fifteenth International Conference of the CLEF Association (CLEF 2024)
- [9] Sprengel E, Jaggi M, Kilcher Y, Hofmann T (2016) Audio based bird species identification using deep learning techniques. In: CEUR Workshop Proceedings.
- [10] Kahl S, Wilhelm-Stein T, Hussein H et al. (2017) Large-Scale Bird Sound Classification using Convolutional Neural Networks. In: CEUR Workshop Proceedings.
- [11] Grill T, Schlüter J (2017) Two Convolutional Neural Networks for Bird Detection in Audio Signals. In: 25th European Signal Processing Conference (EUSIPCO2017). Kos, Greece. <https://doi.org/10.23919/EUSIPCO.2017.8081512>
- [12] Sevilla A, Glotin H (2017) Audio bird classification with inception-v4 extended with time and time-frequency attention mechanisms. In: CEUR Workshop Proceedings.
- [13] Stowell D, Stylianou Y, Wood M, Pamuła H, Glotin H (2018) Automatic acoustic detection of birds through deep learning: the first Bird Audio Detection challenge. In: Methods in Ecology and Evolution
- [14] Lasseck M (2018) Audio-based Bird Species Identification with Deep Convolutional Neural Networks. In: CEUR Workshop Proceedings.
- [15] Lasseck M (2018) Acoustic Bird Detection with Deep Convolutional Neural Networks. In: Plumbley MD et al. (eds) Proceedings of the Detection and Classification of Acoustic Scenes and Events 2018 Workshop (DCASE2018), pp. 143-147, Tampere University of Technology.
- [16] Lasseck M (2019) Bird Species Identification in Soundscapes. In: CEUR Workshop Proceedings.

- [17] Lasseck M (2023) Bird Species Recognition using Convolutional Neural Networks with Attention on Frequency Bands. In: CEUR Workshop Proceedings.
- [18] <https://xeno-canto.org/>
- [19] <https://www.kaggle.com/code/salmanahmedtamu/training-0-65-0-66>
- [20] <https://www.kaggle.com/salmanahmedtamu>
- [21] <https://www.kaggle.com/competitions/birdclef-2024/discussion/497539>
- [22] <https://www.kaggle.com/lihaoweicvch>
- [23] <https://pytorch.org/audio/main/generated/torchaudio.transforms.MelSpectrogram.html>
- [24] <https://pytorch.org/audio/main/generated/torchaudio.transforms.AmplitudeToDB.html>
- [25] <https://pytorch.org/audio/main/transforms.html>
- [26] https://huggingface.co/timm/eca_nfnet_l0
- [27] Deng J et al. (2009) Imagenet: A largescale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition, 2009. pp. 248–255
- [28] https://pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.CosineAnnealingLR.html
- [29] <https://github.com/ildoonet/pytorch-gradual-warmup-lr>
- [30] <https://pytorch.org/docs/stable/generated/torch.nn.BCEWithLogitsLoss.html>
- [31] https://pytorch.org/vision/main/generated/torchvision.ops.sigmoid_focal_loss.html
- [32] https://albumentations.ai/docs/api_reference/augmentations/geometric/transforms/
- [33] https://albumentations.ai/docs/api_reference/augmentations/dropout/coarse_dropout/
- [34] <https://github.com/huggingface/pytorch-image-models/blob/main/timm/models/efficientnet.py>
- [35] <https://www.kaggle.com/competitions/birdclef-2021/discussion/243463>
- [36] <https://www.kaggle.com/competitions/birdclef-2023/discussion/412753>
- [37] <https://pytorch.org/docs/stable/generated/torch.optim.AdamW.html>
- [38] https://scikit-learn.org/stable/modules/generated/sklearn.metrics.label_ranking_average_precision_score.html
- [39] https://scikit-learn.org/stable/modules/generated/sklearn.metrics.average_precision_score.html
- [40] https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html
- [41] <https://www.kaggle.com/code/metric/birdclef-roc-auc>
- [42] Wortsman et. al (2022) Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time, arXiv:2203.05482
- [43] <https://www.kaggle.com/competitions/birdsong-recognition/discussion/183199>
- [44] <https://www.kaggle.com/competitions/birdsong-recognition>
- [45] <https://www.kaggle.com/competitions/birdclef-2023/discussion/414102>
- [46] <https://github.com/adsr71/BirdID-Europe254>
- [47] Stehle M, Lasseck M, Khorramshahi O, Sturm U (2020) Evaluation of acoustic pattern recognition of nightingale (*Luscinia megarhynchos*) recordings by citizens. In: Research Ideas and Outcomes 6: e50233. doi: 10.3897/rio.6.e50233
- [48] Wägele JW, Bodesheim P, Bourlat SJ, Denzler J et al. (2022) Towards a multisensor station for automated biodiversity monitoring. In: Basic and Applied Ecology (59), 105-138. doi: 10.1016/j.baae.2022.01.003
- [49] Wägele JW, Tschan GF et al. (2024) Weather stations for biodiversity: a comprehensive approach to an automated and modular monitoring system. Advanced Books, Pensoft, Sofia, 1-218. <https://doi.org/10.3897/ab.e119534>
- [50] <https://www.idmt.fraunhofer.de/en/institute/projects-products/projects/devise.html>
- [51] <https://www.museumfuernaturkunde.berlin/en/science/acoustic-forest-monitoring>
- [52] <https://www.thuenen.de/en/fachinstitute/waldoekosysteme/querschnittsgruppen/naturschutz/projekte/integration-bio-akustischer-methoden-fuer-die-quantifizierung-biologischer-vielfalt-in-das-waldmonitoring-akwamo-1-2>
- [53] <https://naturblick.museumfuernaturkunde.berlin/?lang=en>
- [54] <https://www.museumfuernaturkunde.berlin/en/science/animal-sound-archive>