Few-shot Long-Tailed Bird Audio Recognition

Advancing the science of bioacoustics and support ongoing research to protect endangered Hawaiian birds

Marcos V. Conde

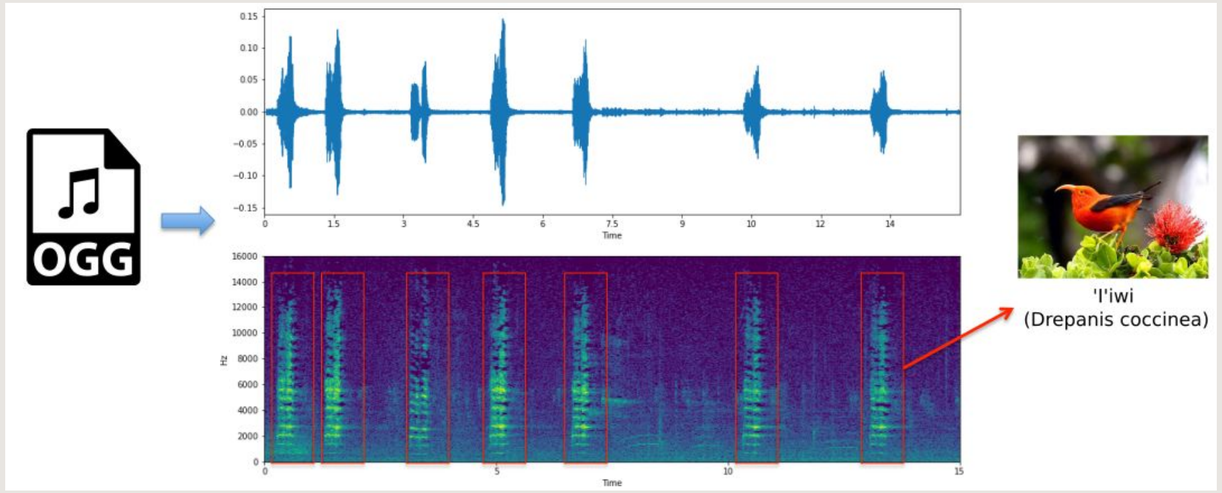
Introduction

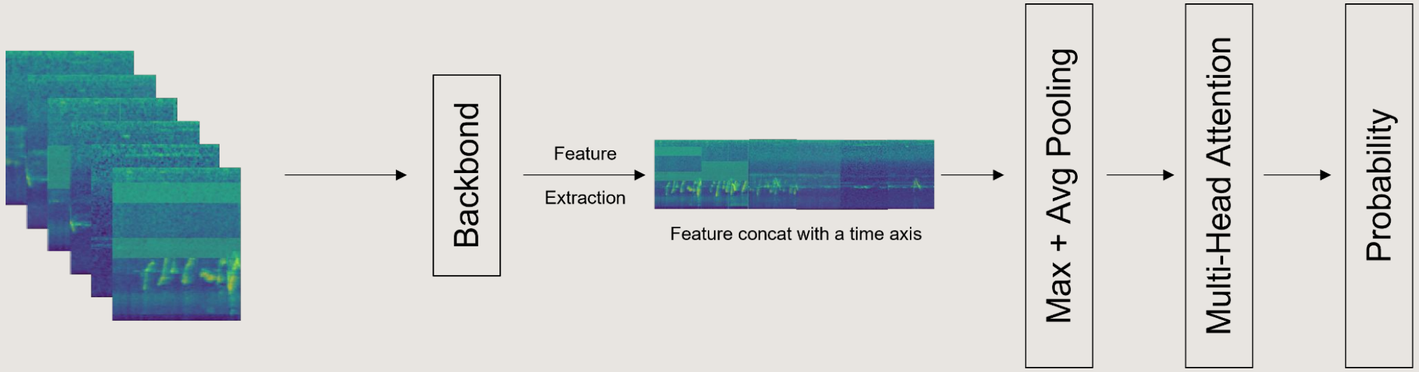
Use machine learning algorithms to detect and identify birds in soundscapes recorded in the wild.

Explore **new state-of-the-art** computer vision models (i.e. Transformers, NAFNets) and techniques applied to sound processing methods like Sound Event Detection and CNN-based Audio Recognition.

H2O.ai teams achieved top-10 in both competitions BirdCLEF 2021/22, over 800 teams. Our code, datasets and pre-processing are open-sourced.

General Approach

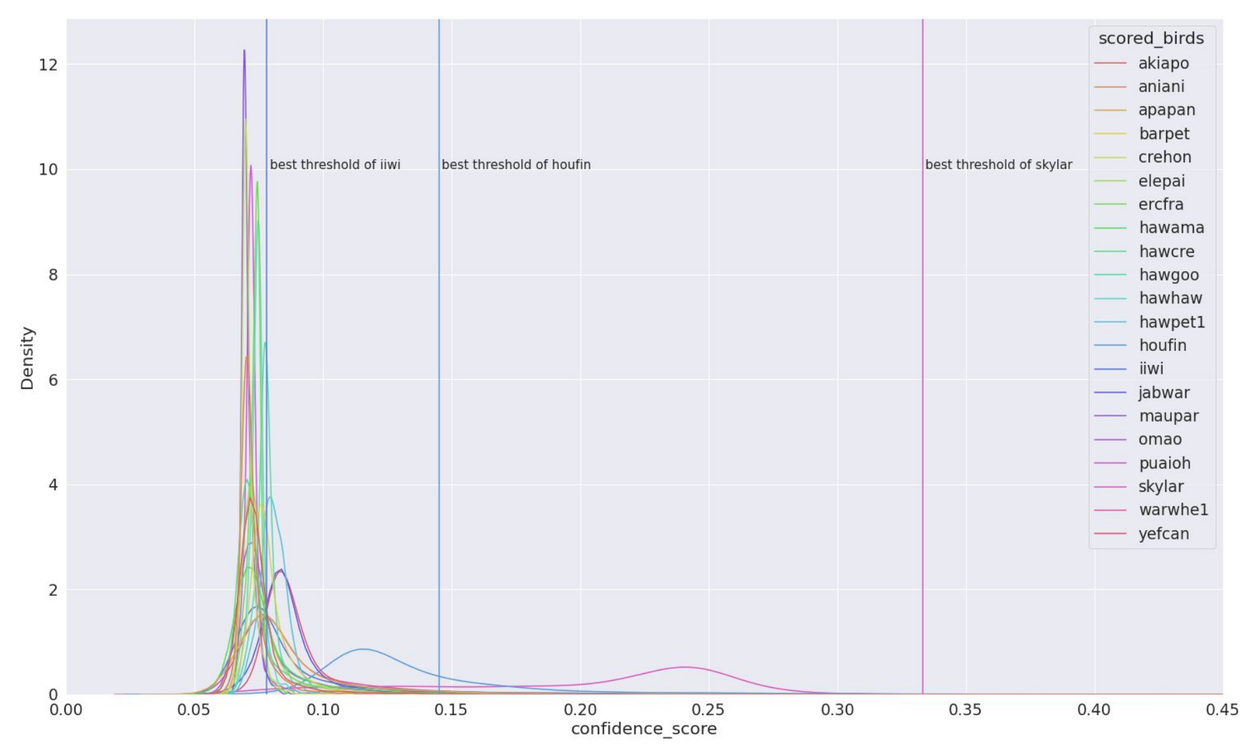


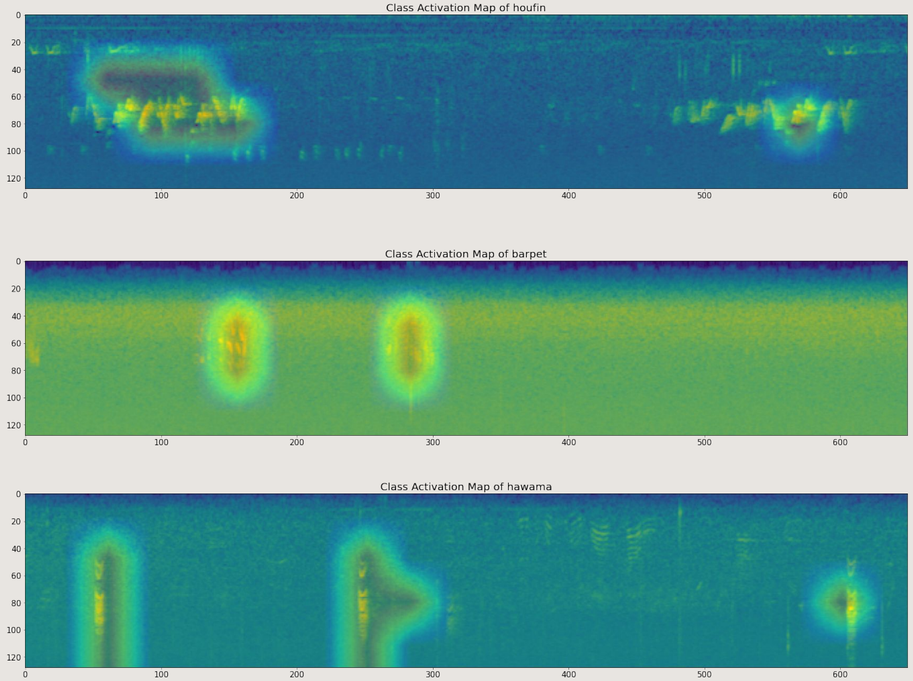


We present a sound event detection (**SED)** and classification pipeline for analyzing soundscape recordings **in the wild**.

Our models learn from few data and weak labels; they can accurately classify **fine-grained** bird vocalizations in 0.04s using a single mid-level GPU. Moreover, they show **robustness** against noisy sounds (e.g. rain, cars).

Our Approach



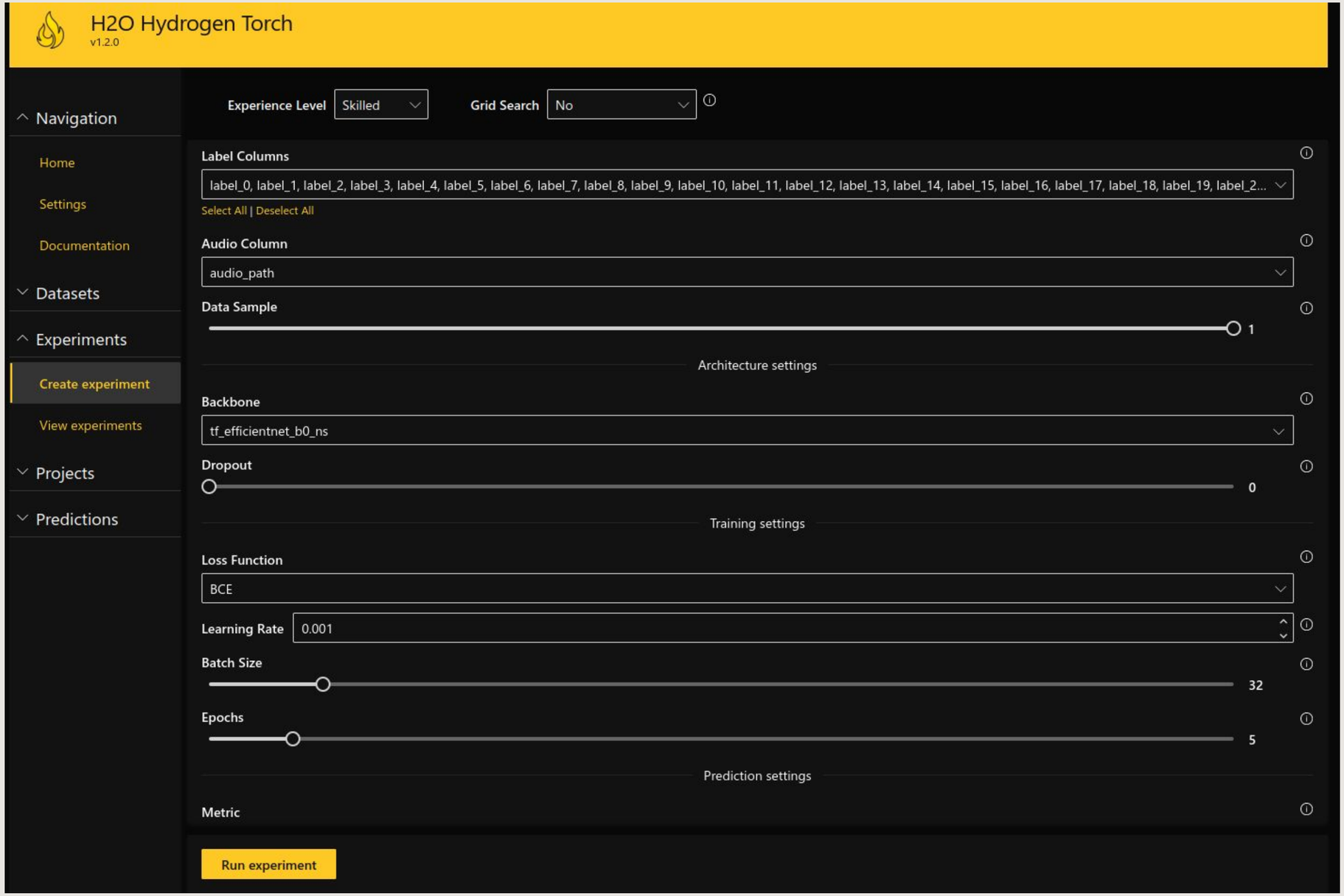


We use **adaptive class-wise thresholds** to obtain the maximum AUC-ROC score per bird class (21 birds of interest).

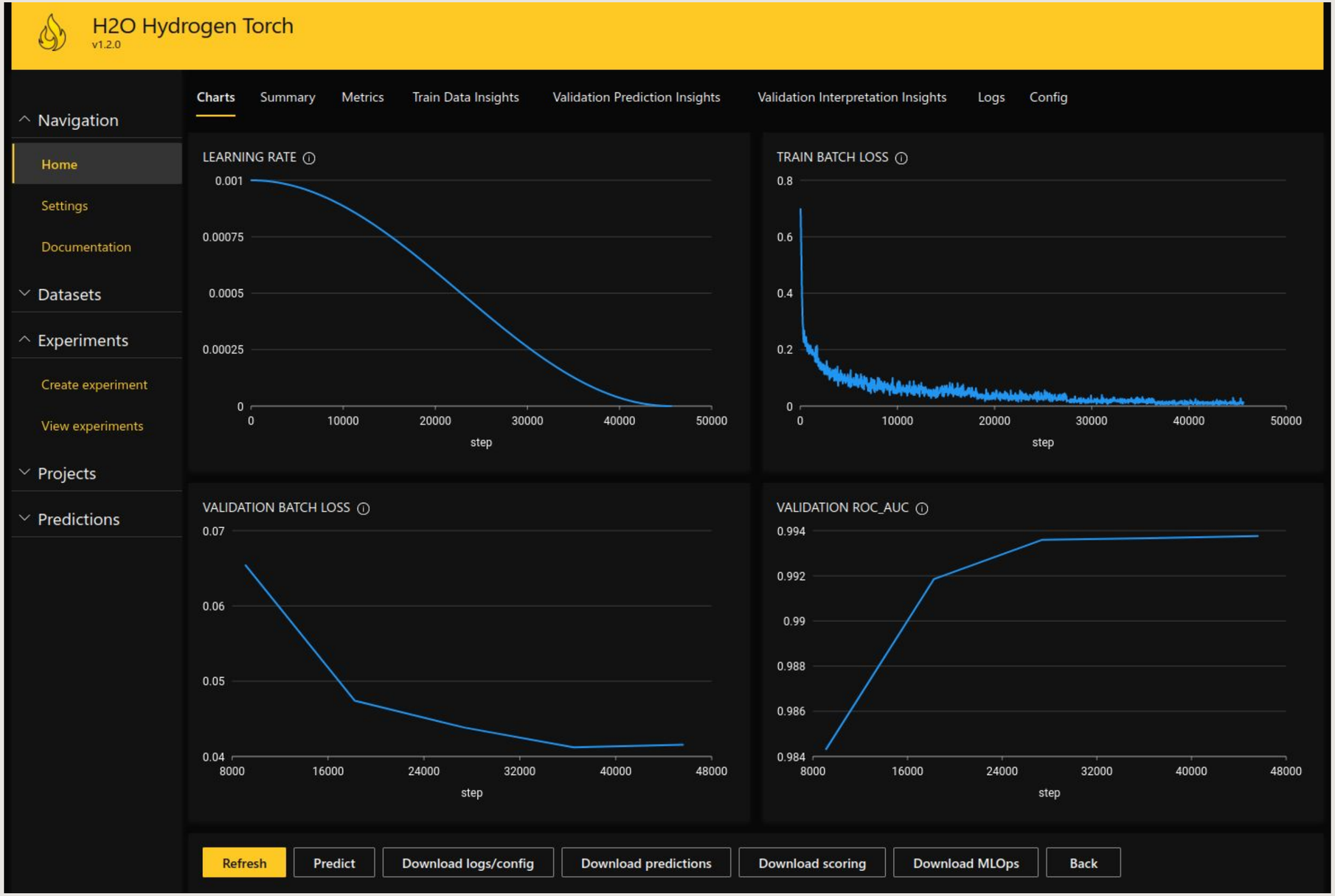
Our method learns from **weak labels and few data** and acoustically recognizes the bird species.

We also provide qualitative samples of our **Grad-CAM** maps that serve as an **interpretable** guide and soft segmentation/detection labels.

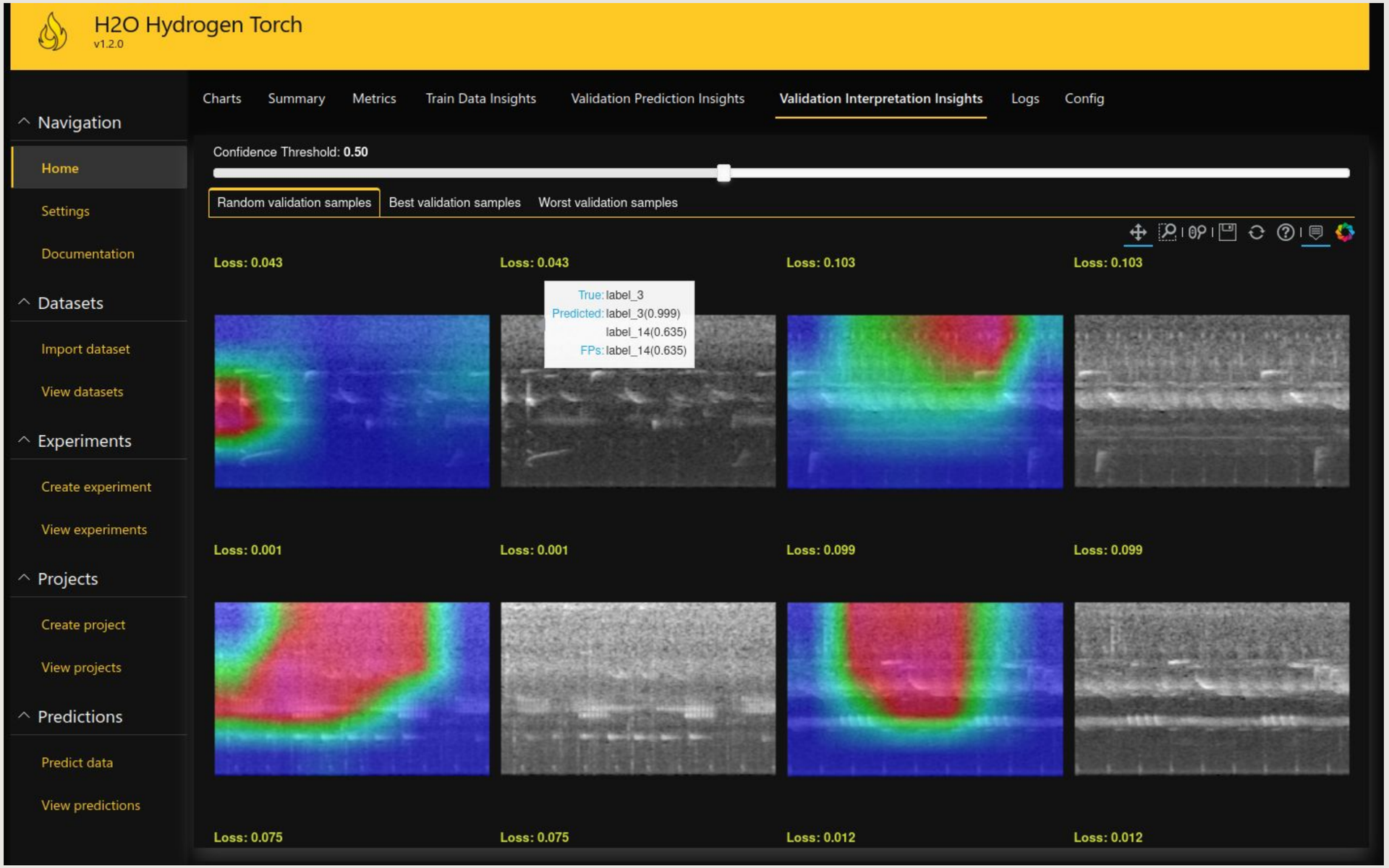
Our Framework: Hydrogen Torch



Our Framework: Hydrogen Torch



Our Framework: Hydrogen Torch



Our Reports

Few-shot Long-Tailed Bird Audio Recognition

Marcos V. Conde[1,2,†], Ui-Jin Choi[3,*,†]

[1]H2O.ai
[2]Computer Vision Lab, Institute of Computer Science, University of Würzburg, Germany
[3]MegaStudyEdu, South Korea

Backbone	CV	Public LB	Private LB
tf_efficientnet_b0_ns [13]	0.8745	0.7922	0.7240
tf_efficientnet_b0_ns [13]	0.8745	0.7817	0.7548
eca_nfnet_l0 [42]	0.8761	0.7510	0.7387
resnest50d [12]	0.8822	0.7550	0.7372
tf_efficientnet_b1_ns [13]	0.7843	0.7395	0.6946
tf_efficientnet_b2_ns [13]	**0.8848**	0.7277	0.7046
tf_efficientnet_b3_ns [13]	0.8561	0.7262	0.6640
tf_efficientnetv2_s_in21k [43]	0.8632	0.7620	0.7439
tf_efficientnetv2_b0 [43]	0.8762	0.7268	0.7070
Ours Ensemble	-	0.7971	**0.7733**
Ours Ensemble	-	**0.8359**	0.7630

Weakly-Supervised Classification and Detection of Bird Sounds in the Wild. A BirdCLEF 2021 Solution

Marcos V. Conde[1,4], Kumar Shubham[2,4], Prateek Agnihotri[3,4], Nitin D. Movva[4] and Szilard Bessenyei

Recognizing bird species in diverse soundscapes under weak supervision

Christof Henkel[1,4], Pascal Pfeiffer[2,4] and Philipp Singer[3,4]

H2O.ai

Thank you

Code

https://github.com/Choiuijin1125/bclef2022

Contact

Marcos V. Conde
Data Scientist
marcos.conde@h2o.ai