



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

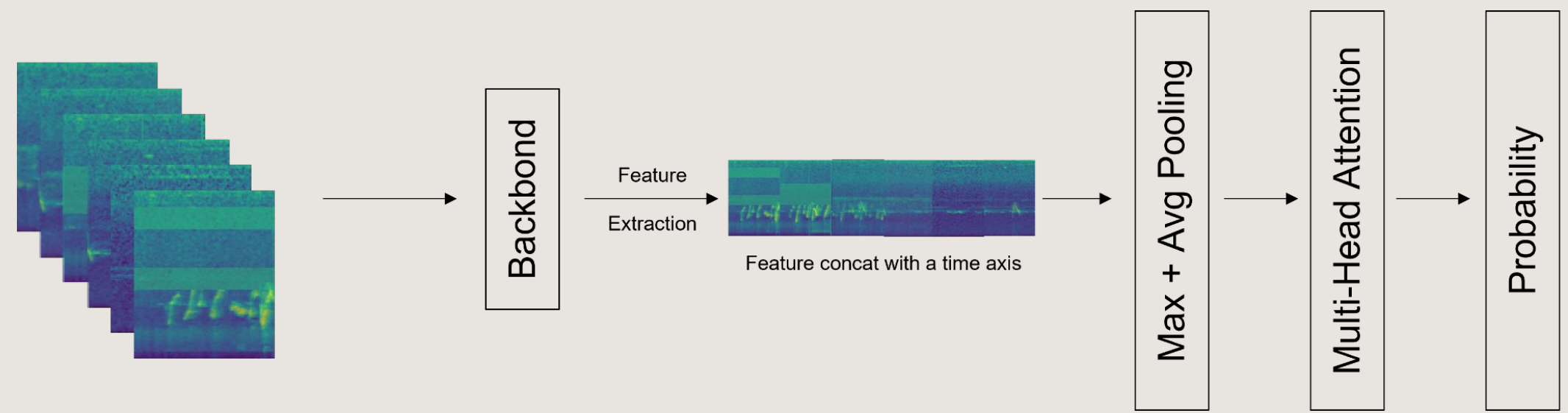
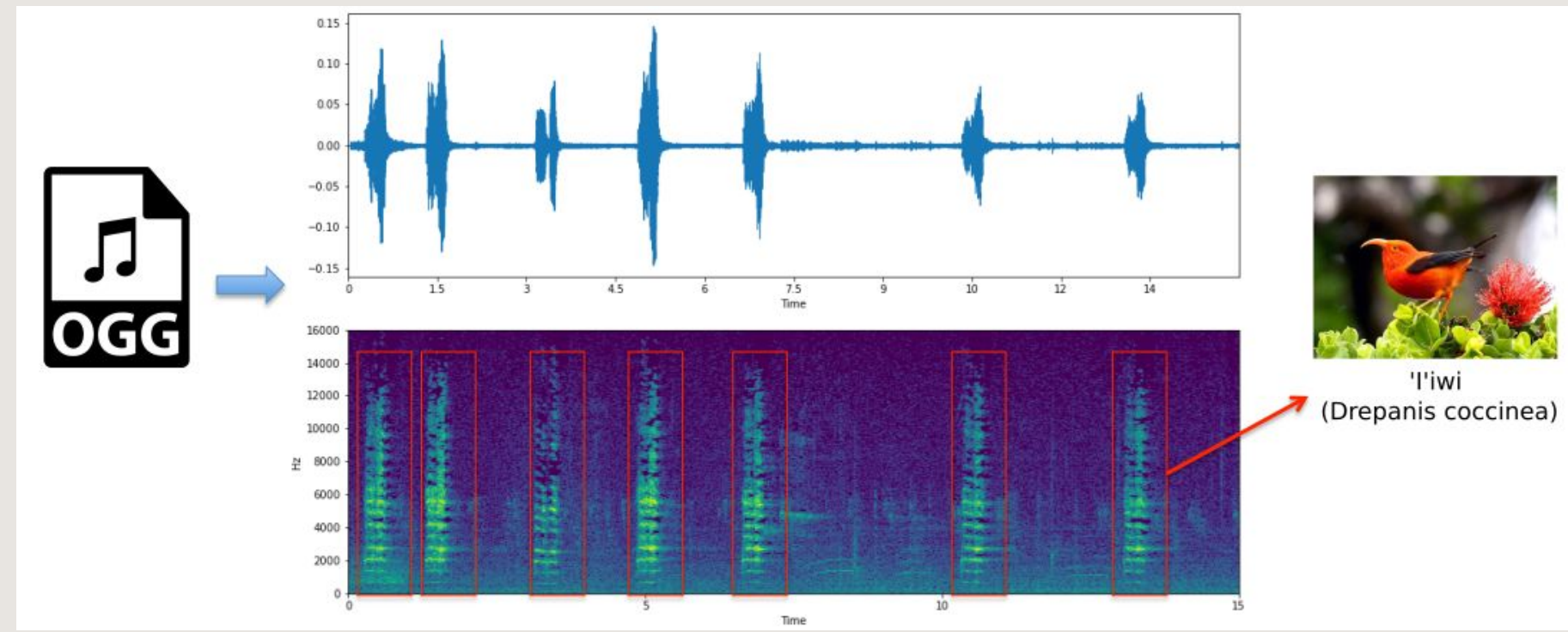


Use machine learning algorithms to detect and identify birds in soundscapes recorded in the wild.

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.

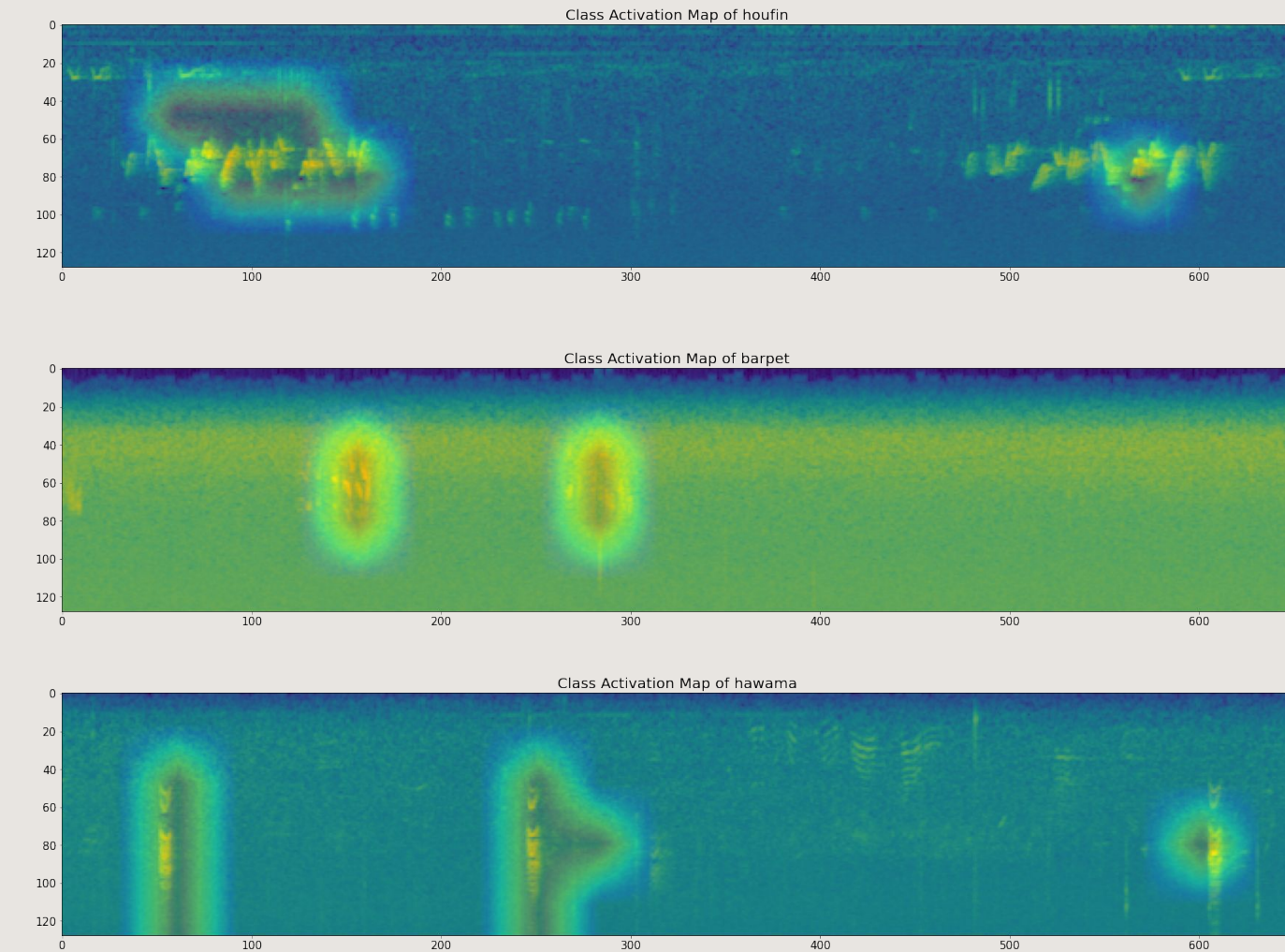
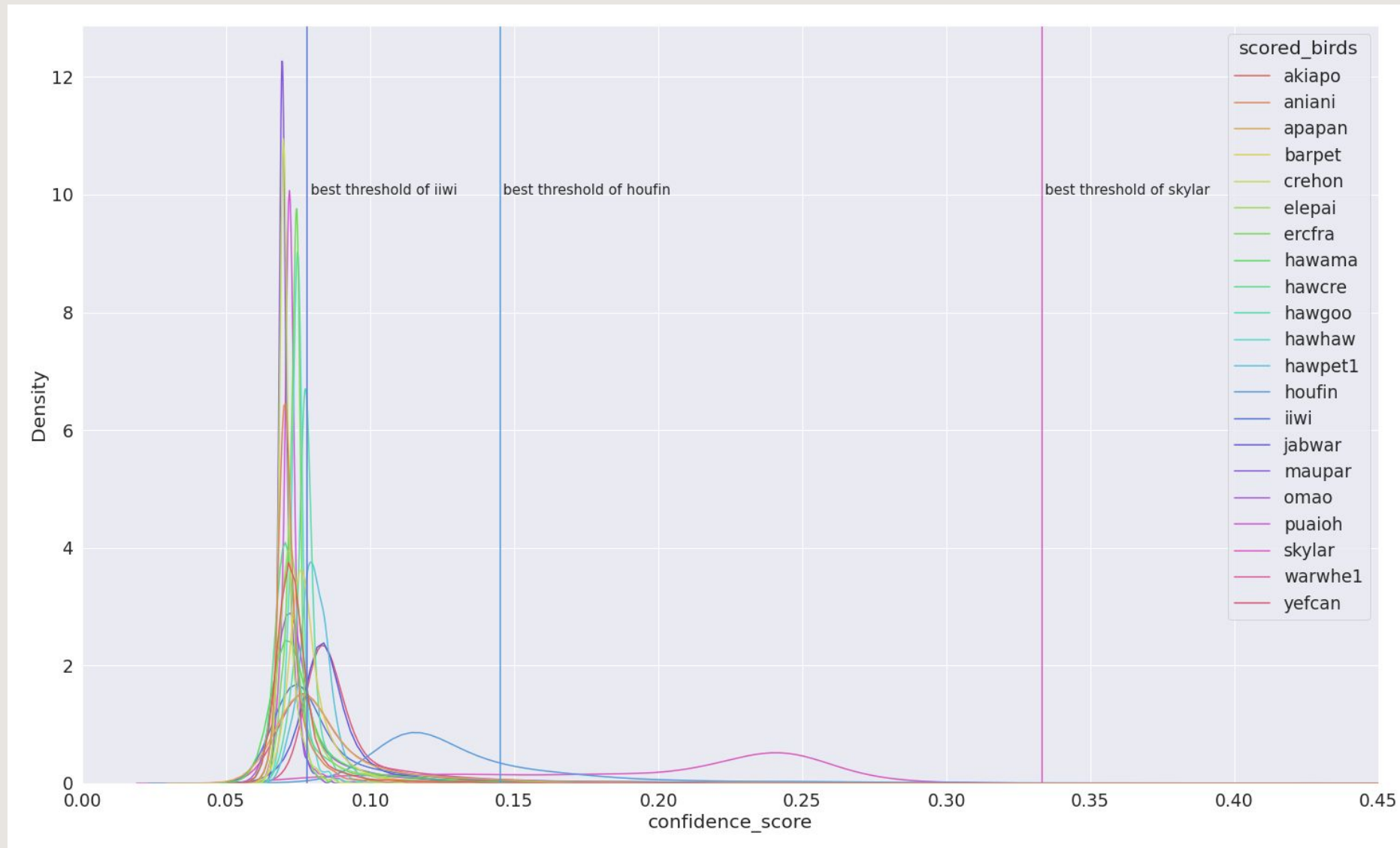
General Approach



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.

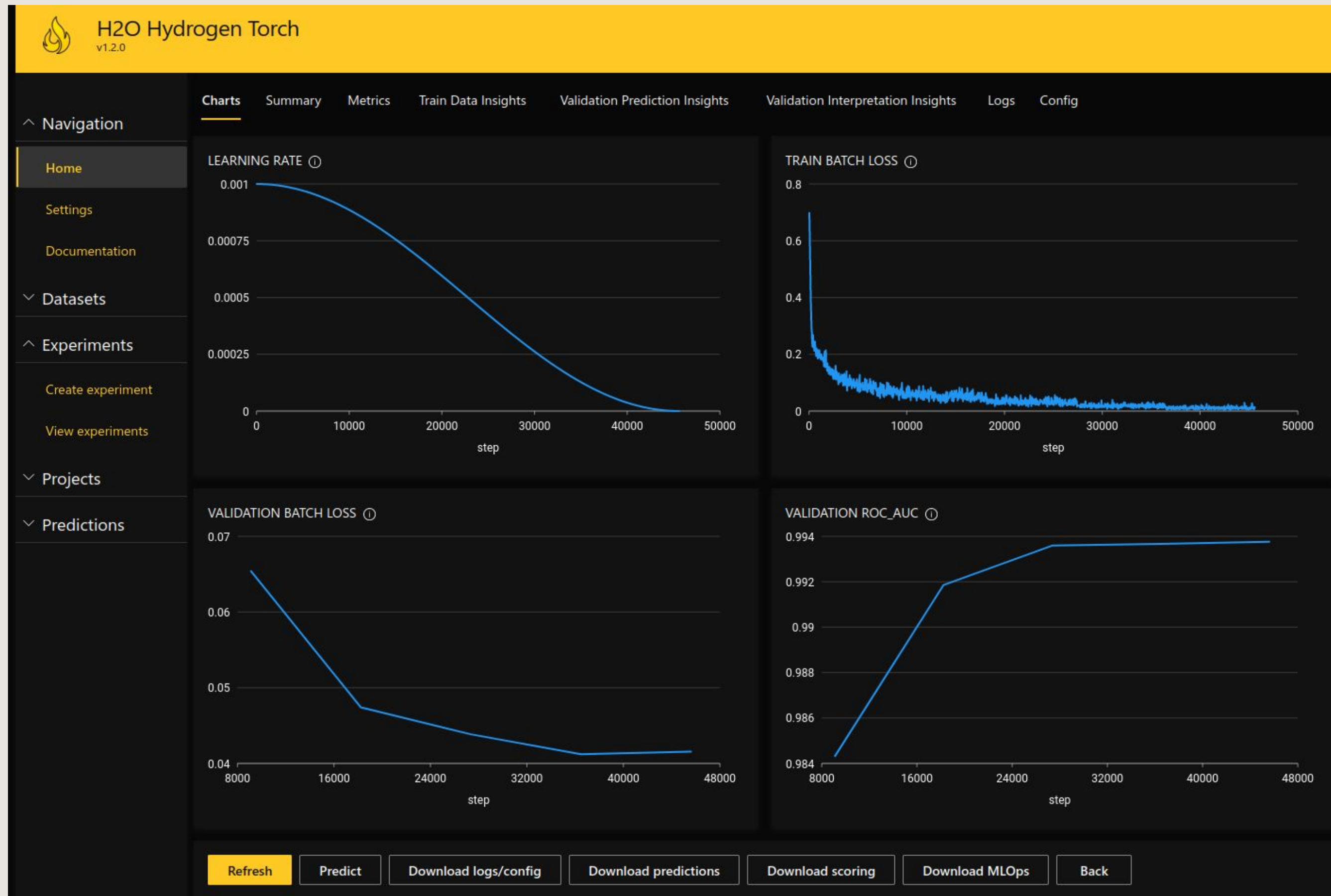
Our Framework: Hydrogen Torch

The screenshot displays the H2O Hydrogen Torch v1.2.0 interface. On the left is a navigation sidebar with categories: Navigation (Home, Settings, Documentation), Datasets, Experiments (Create experiment, View experiments), Projects, and Predictions. The main area is titled 'H2O Hydrogen Torch v1.2.0' and contains several configuration sections:

- Experience Level:** Skilled (dropdown)
- Grid Search:** No (dropdown)
- Label Columns:** A list of labels from label_0 to label_19, with a 'Select All | Deselect All' option.
- Audio Column:** audio_path (dropdown)
- Data Sample:** A slider set to 1.
- Architecture settings:**
 - Backbone:** tf_efficientnet_b0_ns (dropdown)
 - Dropout:** A slider set to 0.
- Training settings:**
 - Loss Function:** BCE (dropdown)
 - Learning Rate:** 0.001 (input field)
 - Batch Size:** A slider set to 32.
 - Epochs:** A slider set to 5.
- Prediction settings:**
 - Metric:** (input field)

A yellow 'Run experiment' button is located at the bottom of the configuration area.

Our Framework: Hydrogen Torch



Our Framework: Hydrogen Torch

The screenshot displays the H2O Hydrogen Torch v1.2.0 interface. The top navigation bar includes tabs for Charts, Summary, Metrics, Train Data Insights, Validation Prediction Insights, Validation Interpretation Insights (selected), Logs, and Config. A left sidebar contains navigation options: Home, Settings, Documentation, Datasets, Experiments, Projects, and Predictions. The main content area features a 'Confidence Threshold: 0.50' slider and three tabs: 'Random validation samples' (selected), 'Best validation samples', and 'Worst validation samples'. A grid of 12 pairs of images is shown, each pair consisting of a heatmap and a corresponding grayscale image. The loss values for each pair are: Row 1: 0.043, 0.043, 0.103, 0.103; Row 2: 0.001, 0.001, 0.099, 0.099; Row 3: 0.075, 0.075, 0.012, 0.012. A tooltip for the first pair in the first row shows: True: label_3, Predicted: label_3(0.999), label_14(0.635), FPs: label_14(0.635). The interface also includes a toolbar with icons for zoom, pan, and other visualization controls.

Our Reports

Few-shot Long-Tailed Bird Audio Recognition

Marcos V. Conde^{1,2,†}, Ui-Jin Choi^{3,*,†}

¹H2O.ai

²Computer Vision Lab, Institute of Computer Science, University of Würzburg, Germany

³MegaStudyEdu, 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. Movva⁴ 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}



Thank you

Code



<https://github.com/Choiuijin1125/bclef2022>

Contact

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