# BirdVox-full-night:

# A dataset and benchmark for avian flight call detection

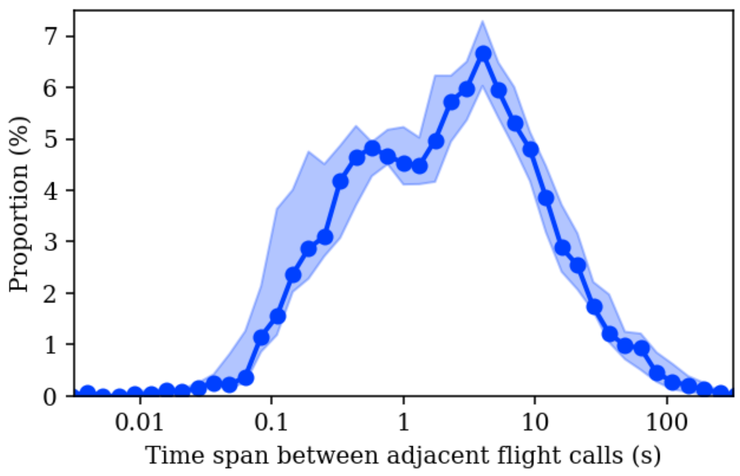
Vincent **Lostanlen**[1,2,3], Justin Salamon[2,3], Andrew Farnsworth[1], Steve Kelling[1], Juan Pablo Bello[2,3]

[1] Cornell Lab of Ornithology
[2] Music and Audio Research Laboratory, New York University
[3] Center for Urban Science and Progress, New York University

**wp.nyu.edu/birdvox/birdvox-full-night**
**vincent.lostanlen@nyu.edu**

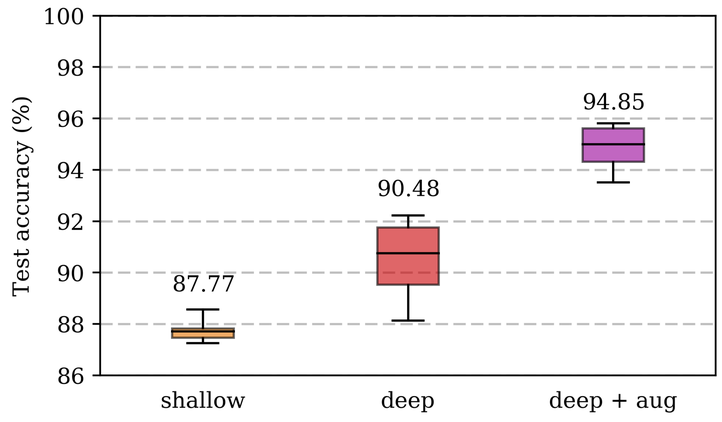
## The first dataset of bird calls in continuous audio

- Context: **bioacoustic monitoring** of bird migration.
- Problem: **lack of data** for evaluating automatic event detection.
- Existing datasets were either extracted with:
  1. an off-the-shelf detector of **unknown recall** (e.g. CLO-SWTH)
     - impossible to know how many events were missed.
  2. crowdsourced annotations (e.g. BirdCLEF)
     - annotating flight calls requires **expert knowledge**.
  3. time scales ranging from 1 to 10 seconds (e.g. PolandNFC)
     - 80% of adjacent calls are **100 ms to 10 seconds apart**.




- Our solution is the BirdVox-full-night dataset:
  - **6 sensors** in various urbanized areas near Ithaca, NY, USA.
  - 1 full night in the fall migration: Sep. 23rd, 2015.
  - **35k flight calls** from 25 species.
  - **62 hours** of continuous, single-channel audio (6x10 hours).
  - An **expert** (A. Farnsworth) spent 102 hours annotating these calls.
  - Annotation **in time and frequency**, but not per species.

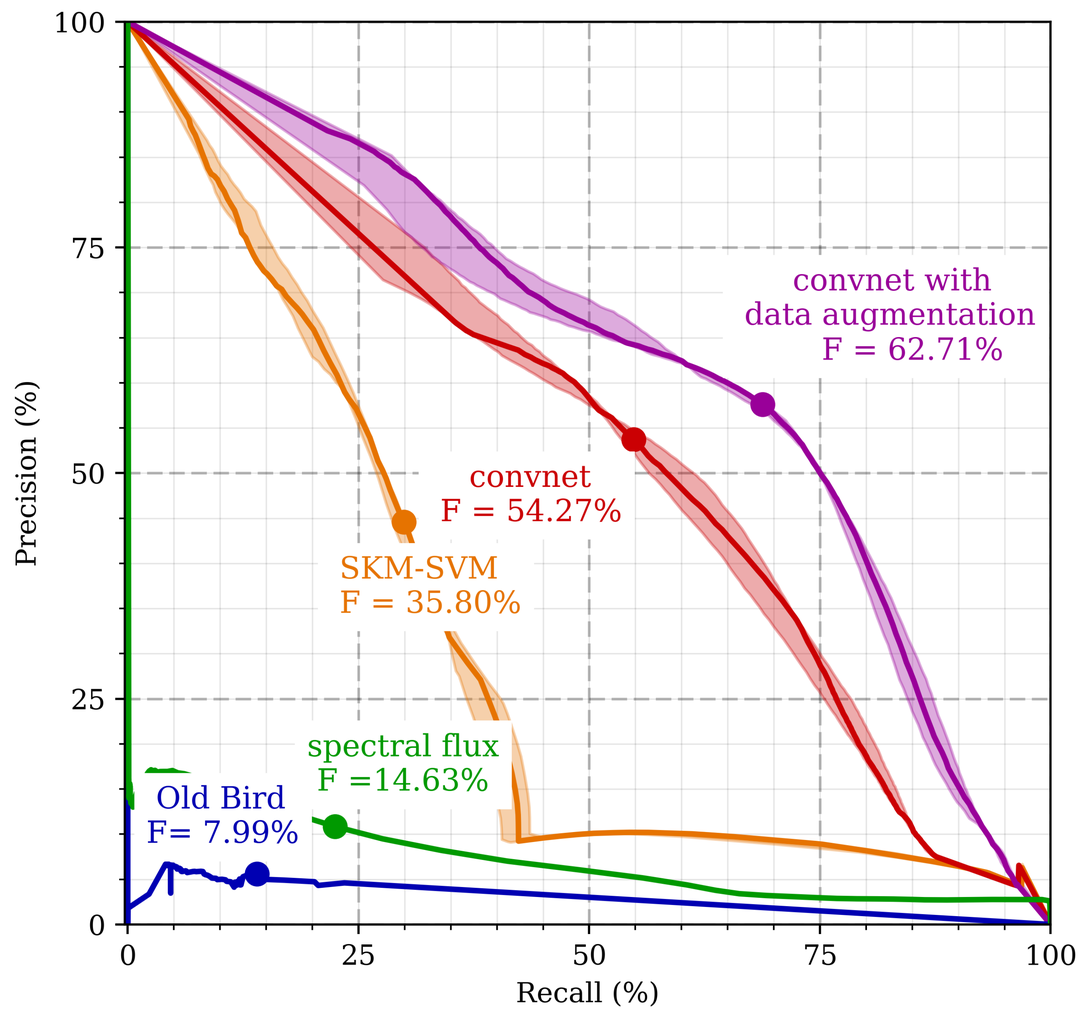
## From 15% $F_1$-score to 63% with deep learning




- **Data augmentation**:
  - pitch shifting
  - time stretching
  - background noise

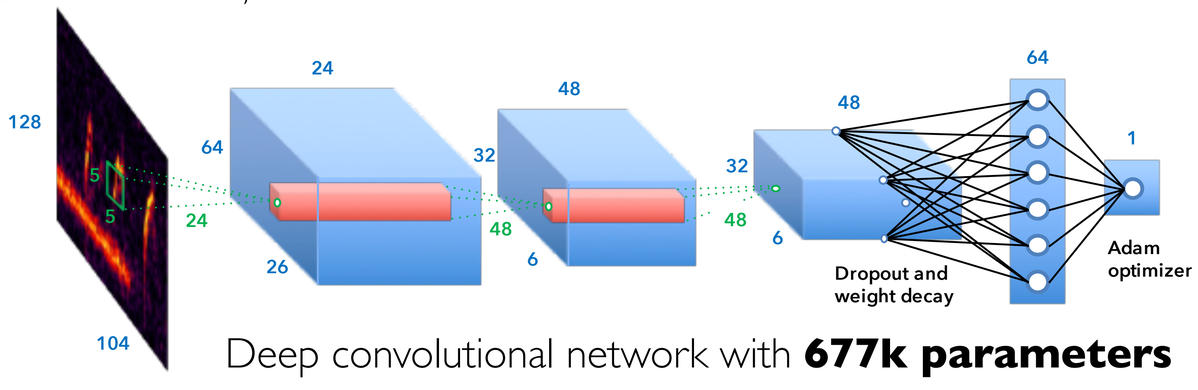
improves ConvNet accuracy on BirdVox-70k.

- We predict on BirdVox-full-night after training on BirdVox-70k:




## Binary classification on a balanced subset

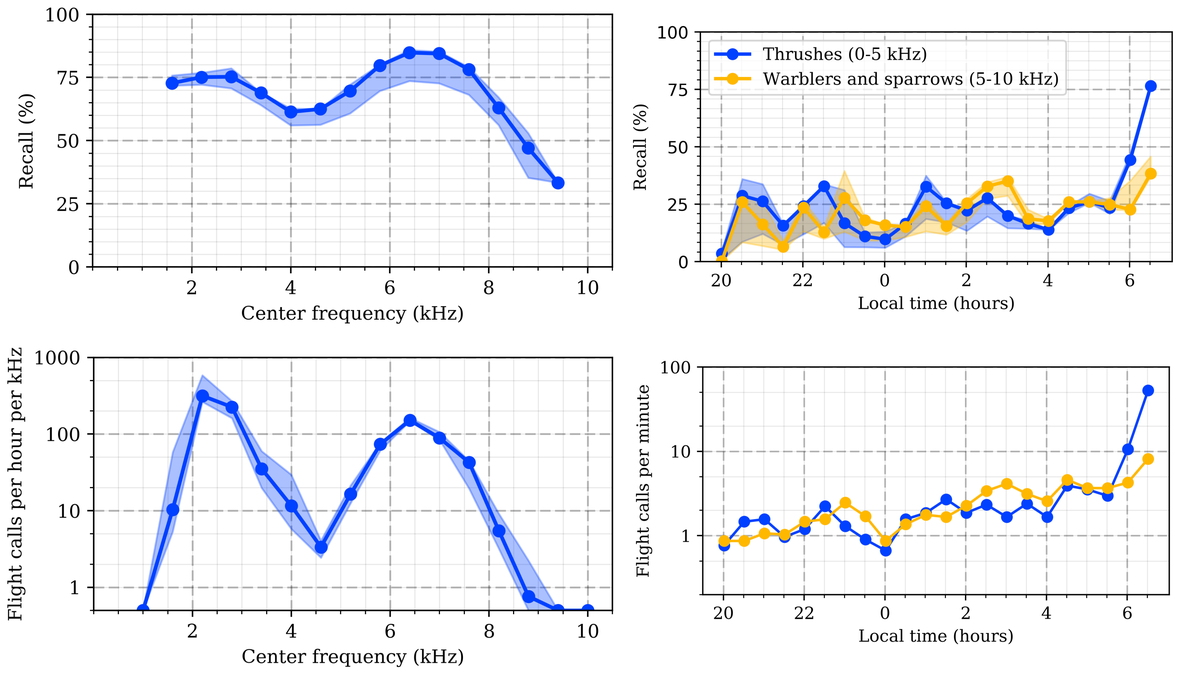
- Formulation: presence vs. absence on **clips (duration 150 ms)**.
- Problem: BirdVox-full-night is **unbalanced** (4.5M clips, 0.7% positives).
- Solution: **BirdVox-70k**, a subset of 70k clips (50% positives).
- To extract challenging **confounding factors**, we run an off-the-shelf shallow classifier (**PCA-SKM-SVM** on log-mel-spectrogram input) which was trained on external data (different years and locations), and derive negative clips in BirdVox-70k as false alarms of the detector.
- We train the shallow classifier and a deep classifier (**ConvNet** with **log-mel-spectrogram** input) on BirdVox-70k: 3 sensors for training, 2 for validation, 1 for test.




Deep convolutional network with **677k parameters**

## Recall correlates with data availability

- Rare flight calls (e.g. at dusk) are less likely to be retrieved.
- Correlations **in time and frequency** (R=0.89 for both).




- Future work will address **background noise adaptation** and deriving insights from detection on the full season (**6600 hours**).

NEW YORK UNIVERSITY

The Cornell Lab of Ornithology