Transfer Learning From Youtube Soundtracks to Tag Arctic Ecoacoustic Recordings

Enis Berk Çoban, Dara Pir, Richard So, Michael I Mandel

ICASSP 2020

Problems

- Arctic-Boreal forests are warming at twice the global average

- Bird migrations and reproductive success are impacted

Problems

- Oil and Gas Extraction
- Frequency of vehicle usage is increasing

Solution

- Ecoacoustic monitoring
- Machine learning for data processing
- Transfer Learning

Data and Experiments

Outline

Data:

- 3 months of nature sounds from Alaska
- 8 categories for labeling

Experiments:

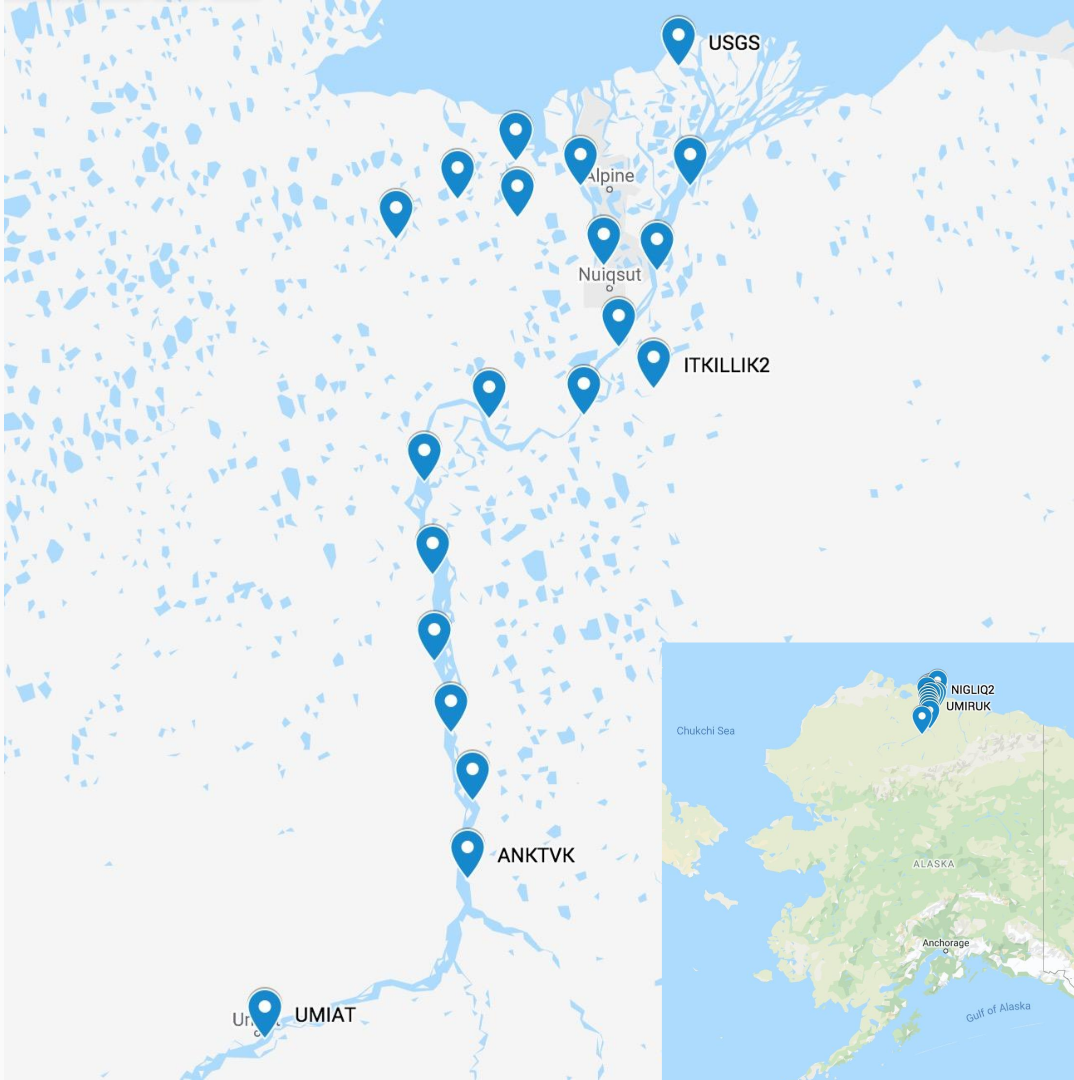
1. Out of box usage of an Audio classifier (0.7 AUC)
2. Training Classifiers with the Audio classifier results (0.77 AUC)
3. Training Classifiers with lower level embeddings (0.86 AUC)

Visualization:

- Songbird Predictions for 7 locations

Data

- Collected by Taylor Stinchcomb*
- The Colville River in Alaska
- 3 Months (June, July, and August of 2016)



Taylor R Stinchcomb, “Social-ecological soundscapes: examining aircraft-harvester-caribou conflict in arctic alaska,” M.S. thesis, University of Alaska Fairbanks, 2017.

Data Labeling

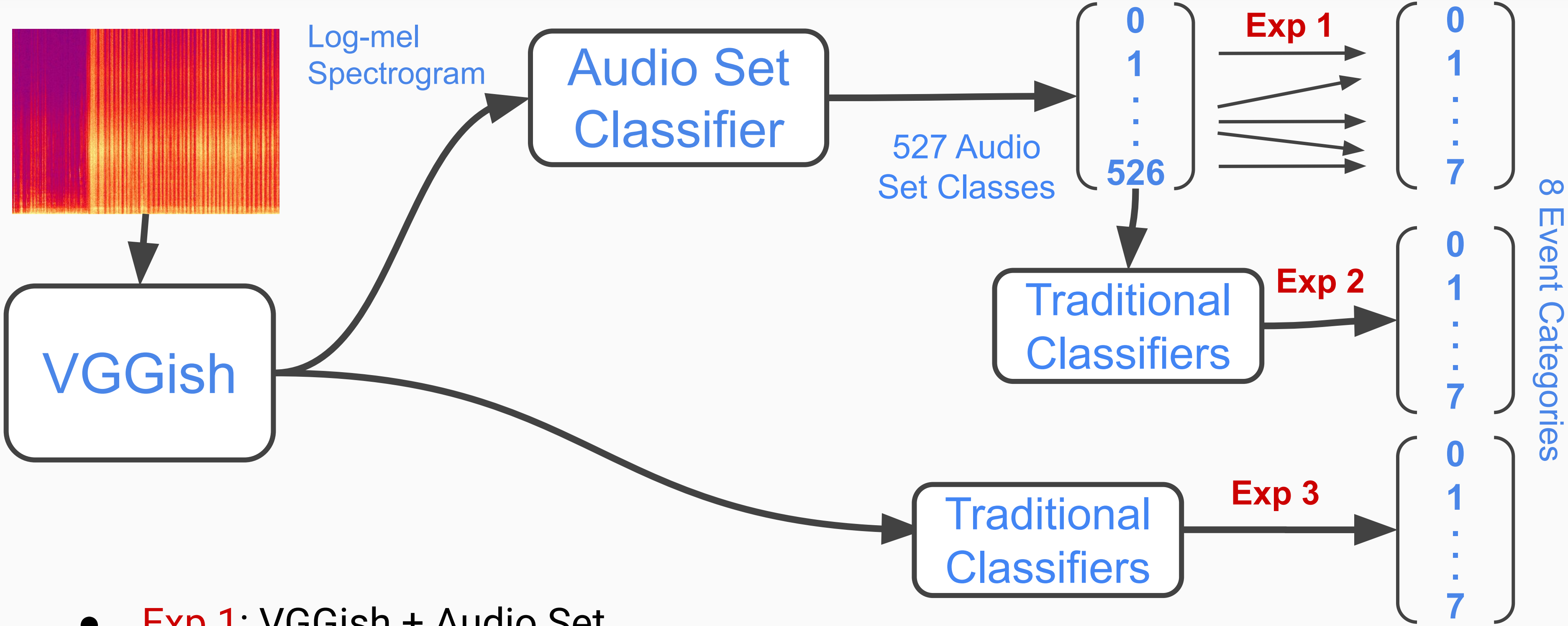
Tag	Sample Count
Wind	641
Cable Noise	456
Songbird	409
Running Water	210
Water Bird	196
Insect	190
Rain	102
Aircraft	28

Songbird

Aircraft

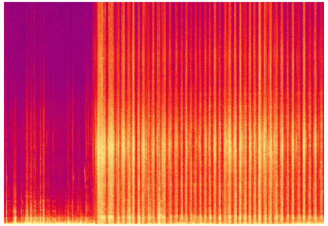
Insect

System Diagram



- Exp 1: VGGish + Audio Set
- Exp 2: VGGish + Audio Set + Traditional Classifiers
- Exp 3: VGGish + Traditional Classifiers

System Diagram



- Log mel spectrograms of 960 ms sound excerpts
- Size 96 × 64

VGGish

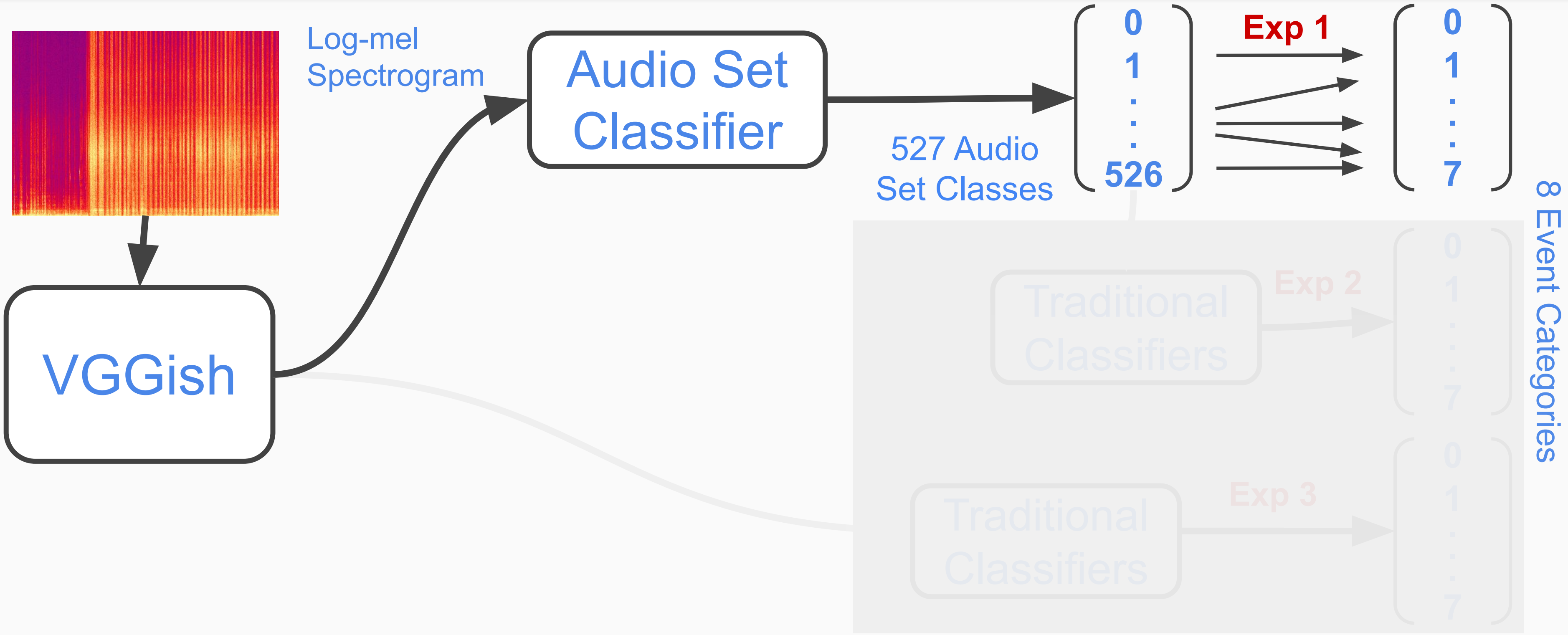
- Deep convolutional network adapted to audio from VGG object recognizer
- Trained on the Youtube-100M dataset
- 128-dimensional embedding vectors

Audio Set Classifier

- Trained on hierarchically organized sound events from Youtube-100M*
- Attention-based

* Qiuqiang Kong, Changsong Yu, Yong Xu, Turab Iqbal, Wenwu Wang, and Mark D. Plumbley, "Weakly labelled audioset tagging with attention neural networks," IEEE Tr. Aud., Spch., & Lang. Proc., vol. 27, no. 11, pp. 1791–1802, Nov. 2019.

Exp 1 - Manual Audio Set Mapping



- Combining multiple Audio Set labels into each event category

Exp 1 - Manual Audio Set Mapping

Songbird: Bird; Owl; Bird vocalization, bird call, ...

WaterBird: Duck; Goose; Quack; Frog; ...

Insect: Fly, housefly; Insect; Bee, wasp, etc.; ...

Aircraft: Engine; Fixed-wing aircraft, airplane; ...

Running Water: Waterfall; Waves, surf

Cable: Bang; Slap, smack; Whack, thwack; ...

Wind: Wind; Howl

Rain: Rain; Raindrop; Rainonsurface

Exp 1 - Manual Audio Set Mapping

Songbird: Bird; Owl; Bird vocalization, bird call, bird song; Pigeon, dove; Coo; Chirp, tweet; Squawk; Bird flight, flapping wings; Gull, seagull; Chirp tone; Hoot

WaterBird: Duck;Goose;Quack;Frog;Croak;Caw

Insect: Fly, housefly; Insect; Bee, wasp, etc.; Buzz; Mosquito; Cricket; Rustle

Aircraft: Engine; Fixed-wing aircraft, airplane; Aircraft engine, Propeller, airscrew; Aircraft; Helicopter

Running Water: Waterfall;Waves,surf

Cable:Bang; Slap,smack; Whack,thwack; Smash,crash; Breaking; Knock; Tap; Thump, thud; Whip; Flap; Clip-clop

Wind: Wind;Howl

Rain: Rain;Raindrop;Rainonsurface

Exp 1 - Manual Audio Set Mapping

Tag	NPos	Bulbul	Manual	Audio Set	VGGish10	VGGish1
Wind	641	0.70	0.66	0.85 (gp)	0.90 (gp)	**0.91** (nn)
Cable noise	456	0.70	0.65	0.80 (rbf)	**0.87** (gp)	0.86 (gp)
Songbird	409	**0.86**	0.70	0.77 (gp)	0.83 (nn)	**0.86** (nn)
Running water	210	0.70	0.57	0.85 (gp)	**0.92** (nn)	0.89 (nn)
Water bird	196	0.65	0.59	0.74 (gp)	0.76 (nn)	**0.77** (rbf)
Insect	190	0.58	0.66	0.79 (nn)	**0.87** (lsvm)	0.82 (lsvm)
Rain	102	0.56	0.44	0.81 (rbf)	**0.85** (gp)	0.82 (gp)
Aircraft	28	0.66	0.52	0.78 (nn)	**0.86** (ab)	0.52 (gp)

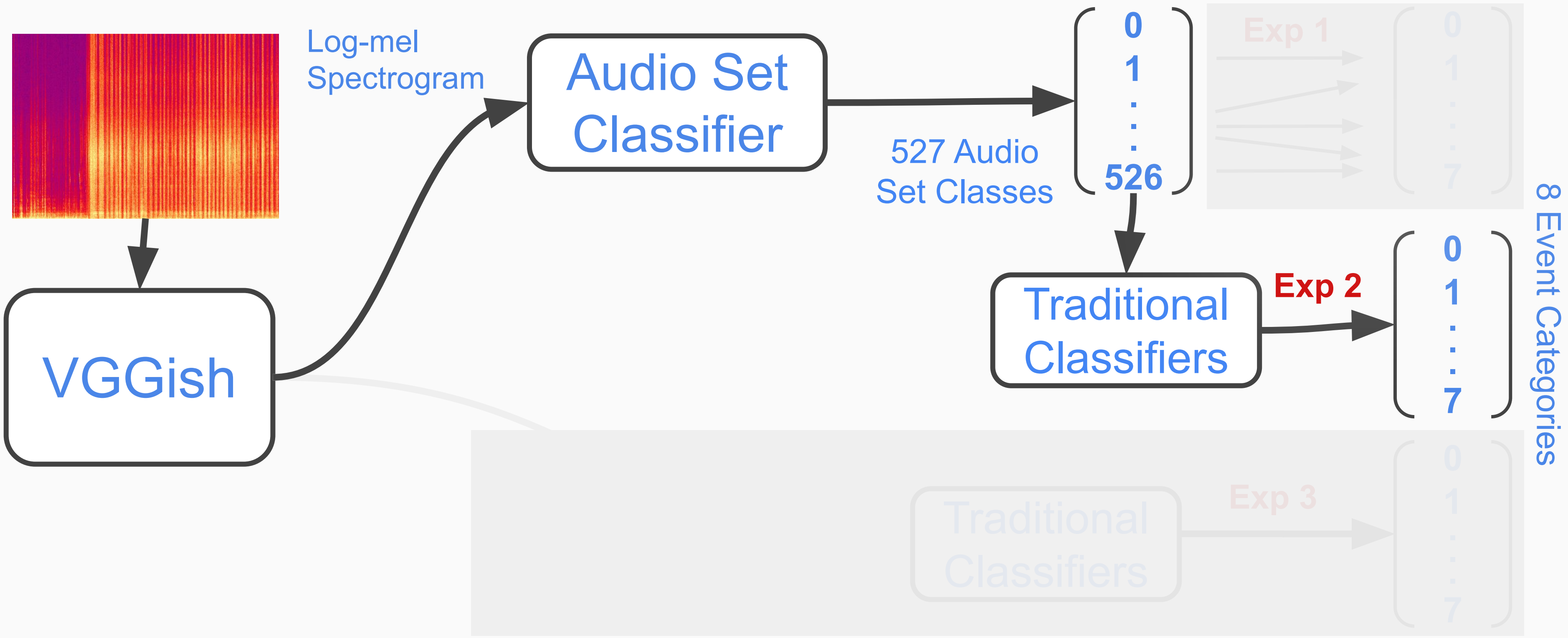
- VGGish + Audio Set
- **Bulbul***: attention-based state-of-the-art dedicated bird detector

* Thomas Grill and Jan Schlüter, “Two convolutional neural networks for bird detection in audio signals,” in Proc. EUSIPCO, Aug. 2017, pp. 1764–1768.

Exp 2 - Heatmap

- The AUC score for each category given predictions of a single Audio Set label
- Unrelated labels predict certain categories successfully

Exp 2 - Traditional Classifiers on Top of Audio Set Labels

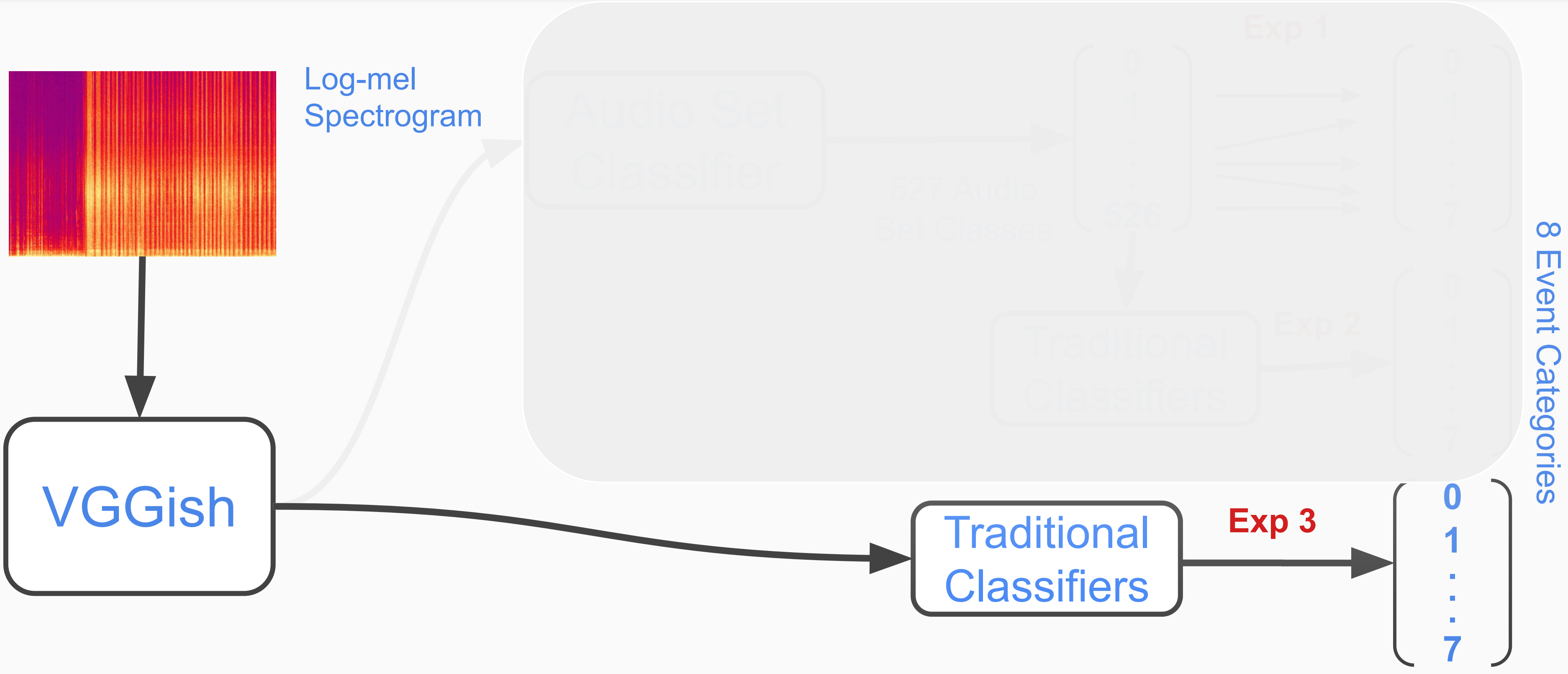


Exp 2 - Traditional Classifiers on Top of Audio Set Labels

Tag	NPos	Bulbul	Manual	Audio Set	VGGish10	VGGish1
Wind	641	0.70	0.66	0.85 (gp)	0.90 (gp)	**0.91** (nn)
Cable noise	456	0.70	0.65	0.80 (rbf)	**0.87** (gp)	0.86 (gp)
Songbird	409	**0.86**	0.70	0.77 (gp)	0.83 (nn)	**0.86** (nn)
Running water	210	0.70	0.57	0.85 (gp)	**0.92** (nn)	0.89 (nn)
Water bird	196	0.65	0.59	0.74 (gp)	0.76 (nn)	**0.77** (rbf)
Insect	190	0.58	0.66	0.79 (nn)	**0.87** (lsvm)	0.82 (lsvm)
Rain	102	0.56	0.44	0.81 (rbf)	**0.85** (gp)	0.82 (gp)
Aircraft	28	0.66	0.52	0.78 (nn)	**0.86** (ab)	0.52 (gp)

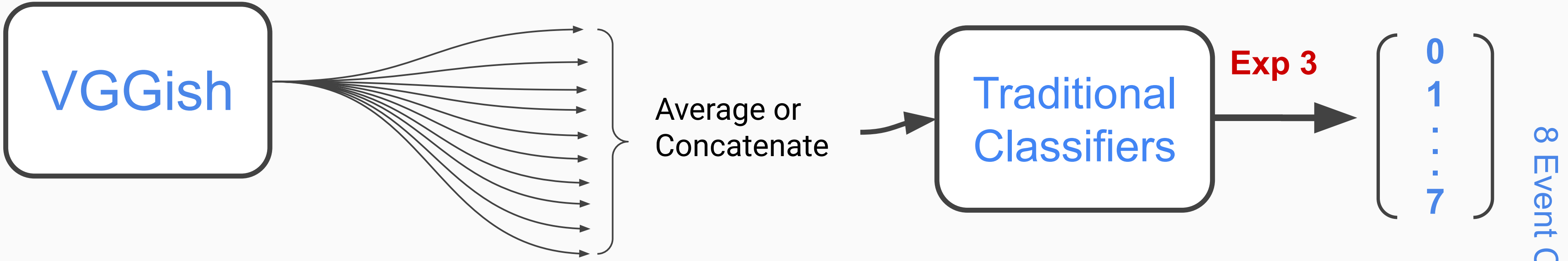
- Test set results using classifiers with best validation set performance

Exp 3 - Traditional Classifiers on VGGish Embeddings



Exp 3 - Traditional Classifiers on VGGish Embeddings

VGGish-1: Combining embeddings



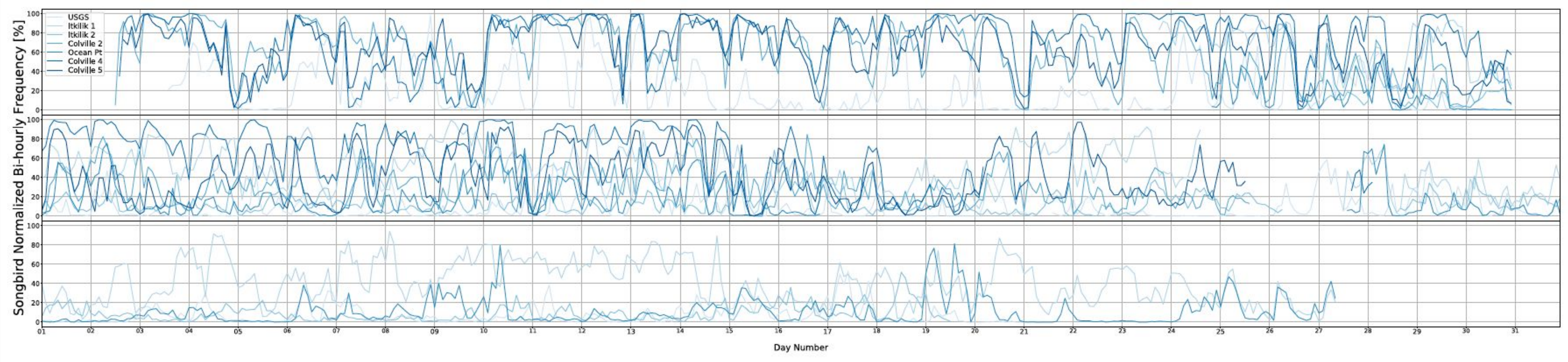
VGGish-10: Weakly supervised

Exp 3 - Traditional Classifiers on VGGish Embeddings

Tag	NPos	Bulbul	Manual	Audio Set	VGGish10	VGGish1
Wind	641	0.70	0.66	0.85 (gp)	0.90 (gp)	**0.91** (nn)
Cable noise	456	0.70	0.65	0.80 (rbf)	**0.87** (gp)	0.86 (gp)
Songbird	409	**0.86**	0.70	0.77 (gp)	0.83 (nn)	**0.86** (nn)
Running water	210	0.70	0.57	0.85 (gp)	**0.92** (nn)	0.89 (nn)
Water bird	196	0.65	0.59	0.74 (gp)	0.76 (nn)	**0.77** (rbf)
Insect	190	0.58	0.66	0.79 (nn)	**0.87** (lsvm)	0.82 (lsvm)
Rain	102	0.56	0.44	0.81 (rbf)	**0.85** (gp)	0.82 (gp)
Aircraft	28	0.66	0.52	0.78 (nn)	**0.86** (ab)	0.52 (gp)

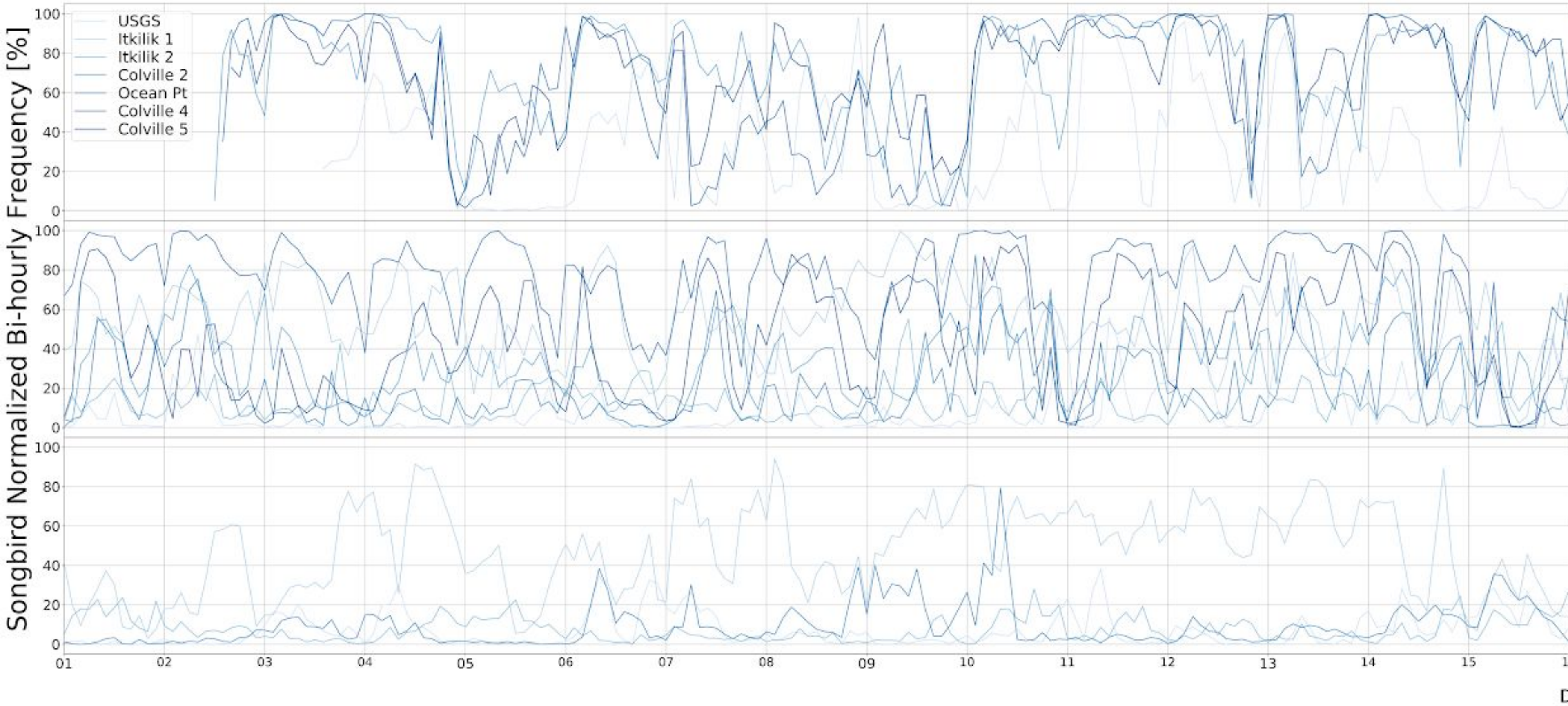
- These models perform well enough that we can use them with a certain confidence
- VGGish1 averaging version is reported

Songbird Predictions Over 7 Sites

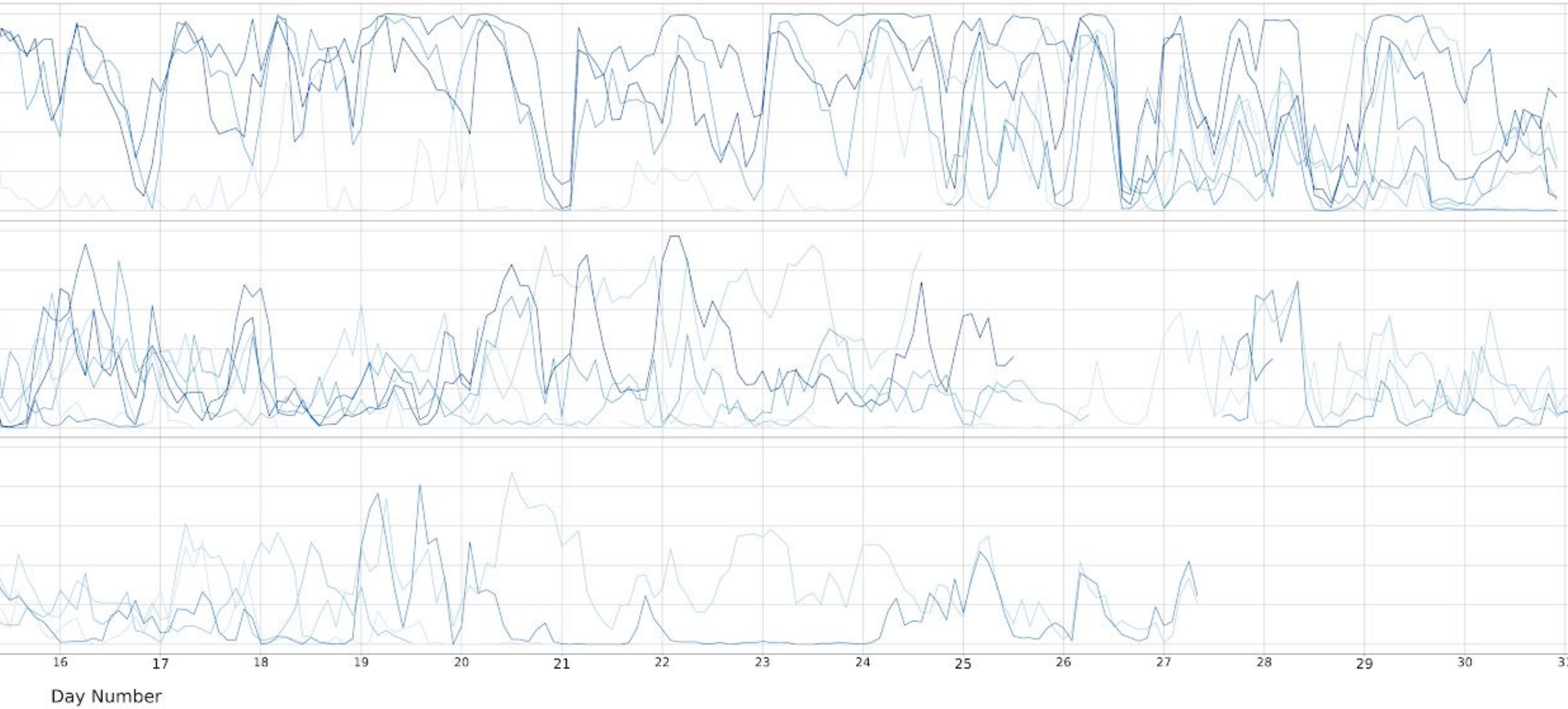


- “Songbird” Neural Network model trained on VGGish raw embeddings
- Top: *June*, Middle: *July*, Bottom: *August*.

Songbird Predictions Over 7 Sites



Songbird Predictions Over 7 Sites



Conclusions

- **Best technique -** Classical ML models with VGGish embeddings as input
- **Results** - AUC above 80% for all categories except one
- **Exception** - Water birds (we grouped waterfowl together with shorebirds)
- **General -** This general model performs on par with Bulbul, which is specialized for songbird, but much better on the other tags

Future Work

- Break categories down into a finer granularity, species level
- Identify important events in phenology of bird communities
- Measure human-generated noise affecting caribou herds

Acknowledgements

This material is based upon work supported by the National Science Foundation (NSF) grant OPP-1839185. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

Thank you