Few-shot bioacoustic event detection: A new task at the DCASE 2021 challenge

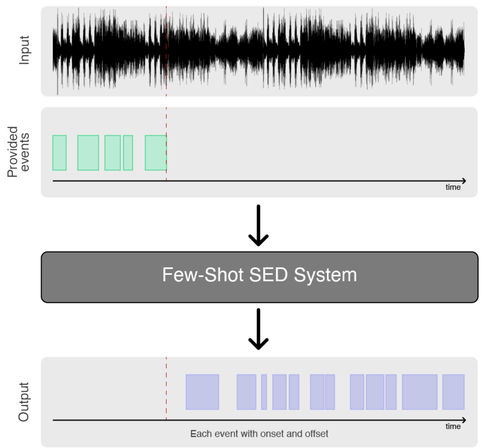
Veronica Morfi, Inês Nolasco, Vincent Lostanlen, Shubhr Singh,
Ariana Strandburg-Peshkin, Lisa Gill, Hanna Pamuła, David Benvent, Dan Stowell

Queen Mary University of London, CNRS, Laboratoire des sciences du numérique de Nantes (LS2N), University of Konstanz, Cornell University, Tilburg University, Max Planck Institute of Animal Behavior, BIOTOPIA Naturkundemuseum Bayern, AGH University of Science and Technology, Naturalis Biodiversity Centre

Bio
Task 5

An Important Problem

Few-shot bioacoustic event detection is a novel area of research that emerged from a need in monitoring biodiversity and animal behaviour: to annotate long recordings, that experts usually can only provide very few annotations for.



Baseline Methods

Template Matching:

- spectrogram cross-correlation method to find instances of a template in an image

Prototypical Network:

- N*-way-*k*-shot classification*, N is the number of classes and k the number of known samples per class
- class prototypes: Positive class: first 5 event annotations. Negative class: rest of the audio file.

Datasets

	Dataset	mic type	# audio files	total duration	# labels (excl. UNK)	# events (excl. UNK)
Development Set: Training	BV	fixed	5	10 hours	11	2,662
	HT	mobile	3	3 hours	3	435
	MT	mobile	2	70 mins	4	1,234
	JD	mobile	1	10 mins	1	355
Development Set: Validation	HV	mobile	2	2 hours	2	50
	PB	fixed	6	3 hours	2	260
Evaluation Set	ME	handheld	2	20 mins	2	70
	ML	various	17	20 mins	17	1,035
	DC	fixed	13	105 mins	3	967

Results

Rank	Team name	Evaluation set: F-score % (97.5% confidence interval)	Validation set: F-score %	DC F-score %	ME F-score %	ML F-score %
1	Zou_PKU [7]	**38.4** (36.2 - 40.6)	55.3	20.6	68.0	67.3
2	Tang_SHNU [6]	38.3 (36.1 - 40.5)	51.4	25.6	61.5	43.3
3	Anderson_TCD [1]	35.0 (33.1 - 37.0)	26.2	19.9	56.6	56.8
4	Baseline_TempMatch	34.8 (32.6 - 37.1)	2.0	**32.2**	47.1	29.5
5	Cheng_BIT [3]	23.8 (21.9 - 25.7)	46.3	10.6	53.5	**78.8**
6	Baseline_PROTO	20.1 (18.2 - 21.9)	41.5	8.5	**72.7**	55.7
7	Zhang_uestc [8]	16.8 (15.5 - 18.2)	54.4	8.1	45.1	29.9
8	Johannsmeier_OVGU [4]	15.2 (13.7 - 16.7)	**58.6**	6.5	64.3	35.8
9	Bielecki_SMSNG [2]	8.4 (7.1 - 9.7)	51.8	3.1	56.3	51.4

The best ranked system [7] applied a transductive inference method and adopted a mutual learning framework designed to make the feature extraction network more task dependent.

Second place [6] used data from Audioset to train a ResNet for feature extraction and adopted embedding propagation [5], for smoothing the decision boundaries as a way of increasing generalisation capabilities.

The third ranking system [1], followed the same approach as the prototypical network baseline, with the main differences being the use of data augmentation and reducing the size of the network.

Remarks

- All submitted systems adopted prototypical networks.
- Data augmentation was applied by the majority of the teams.
- All systems rely on some sort of post-processing mechanism designed to remove superfluous predictions.
- A popular choice was using Per-channel Energy Normalization (PCEN) as acoustic features.
- All systems are generally dataset sensitive.
- Drop in performance on the DC set of dawn chorus recordings.
- Better on ME and ML that include mainly mammal vocalisations.

References

[1] Mark Anderson and Naomi Harte. *Bioacoustic Event Detection with Prototypical Networks and Data Augmentation*. Tech. rep. DCASE2021 Challenge, June 2021.

[2] Radoslaw Bielecki. *FEW-SHOT BIOACOUSTIC EVENT DETECTION WITH PROTOTYPICAL NETWORKS , KNOWLEDGE DISTILLATION AND ATTENTION TRANSFER LOSS*. Tech. rep. DCASE2021 Challenge, June 2021.

[3] Hao Cheng, Chenguang Hu, and Miao Liu. *PROTOTYPICAL NETWORK FOR BIOACOUSTIC EVENT DETECTION VIA I-VECTORS*. Tech. rep. DCASE2021 Challenge, June 2021.

[4] Jens Johannsmeier and Sebastian Stober. *FEW-SHOT BIOACOUSTIC EVENT DETECTION VIA SEGMENTATION USING PROTOTYPICAL NETWORKS*. Tech. rep. DCASE2021 Challenge, June 2021.

[5] Pau Rodrıguez et al. “Embedding propagation: Smoother manifold for few-shot classification”. In: *European Conference on Computer Vision*. Springer. 2020, pp. 121–138.

[6] Tiantian Tang, Yunhao Liang, and Yanhua Long. *TWO IMPROVED ARCHITECTURES BASED ON PROTOTYPE NETWORK FOR FEW-SHOT BIOACOUSTIC EVENT DETECTION*. Tech. rep. DCASE2021 Challenge, June 2021.

[7] Dongchao Yang et al. *FEW-SHOT BIOACOUSTIC EVENT DETECTION: A GOOD TRANSDUCTIVE INFERENCE IS ALL YOU NEED*. Tech. rep. DCASE2021 Challenge, June 2021.

[8] Yue Zhang et al. *FEW-SHOT BIOACOUSTIC EVENT DETECTION USING PROTOTYPICAL NETWORK WITH BACKGROUND CLASSs*. Tech. rep. DCASE2021 Challenge, June 2021.

DCASE2021 CHALLENGE