Segment-level Metric Learning for Few-shot Bioacoustic Detection

Detection and Classification of Acoustic Scenes and Events (DCASE) Workshop 2022

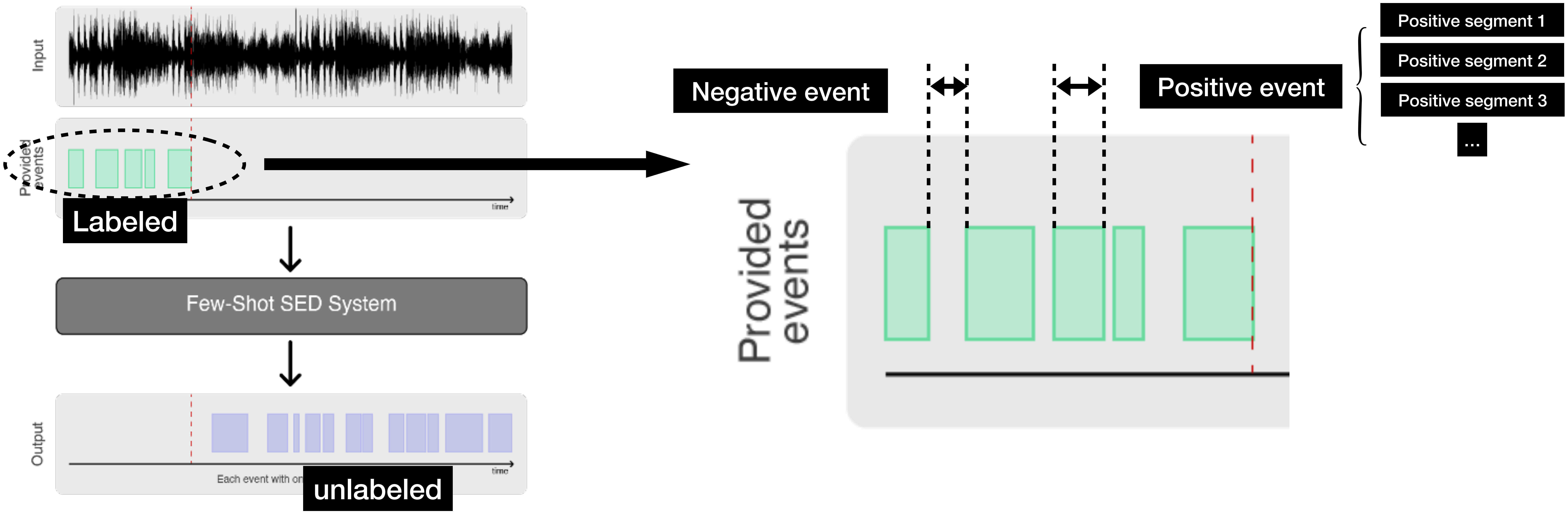
Haohe Liu[1] , Xubo Liu[1] , Xinhao Mei[1] , Qiuqiang Kong[2] , Wenwu Wang[1] , Mark D. Plumbley[1]

[1]Centre for Vision, Speech and Signal Processing (CVSSP), University of Surrey, UK
[2]Speech, Audio, and Music Intelligence (SAMI) Group, ByteDance, China

Introduction

Few-shot bioacoustic detection

- **Detecting** the occurrence time of **novel** sound events (e.g., new species) given a **few examples.**

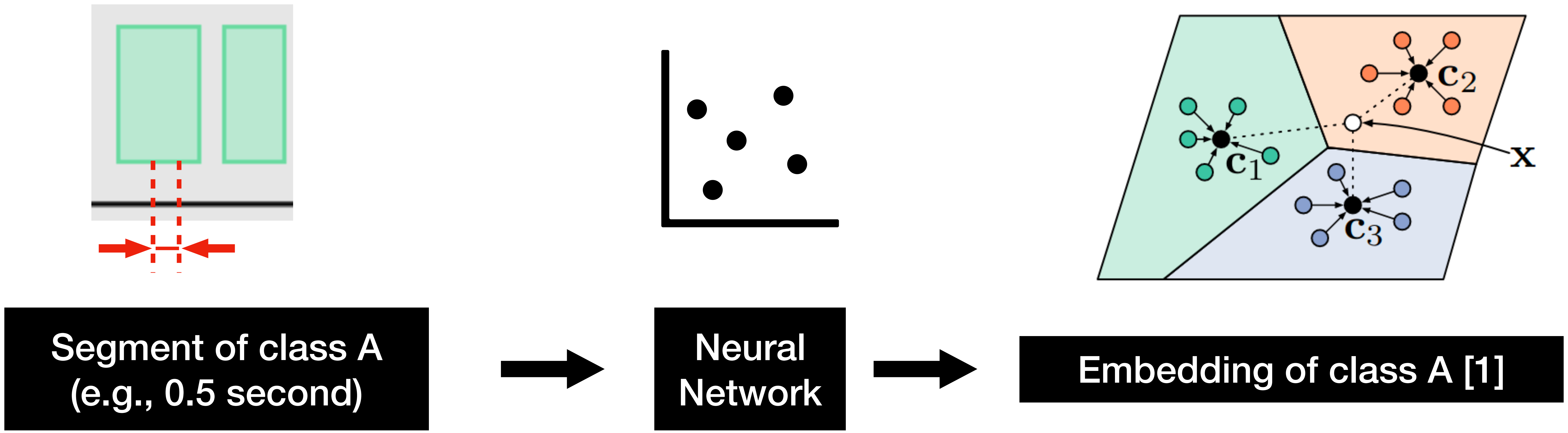


Picture taken from: https://github.com/c4dm/dcase-few-shot-bioacoustic

Introduction

Previous studies

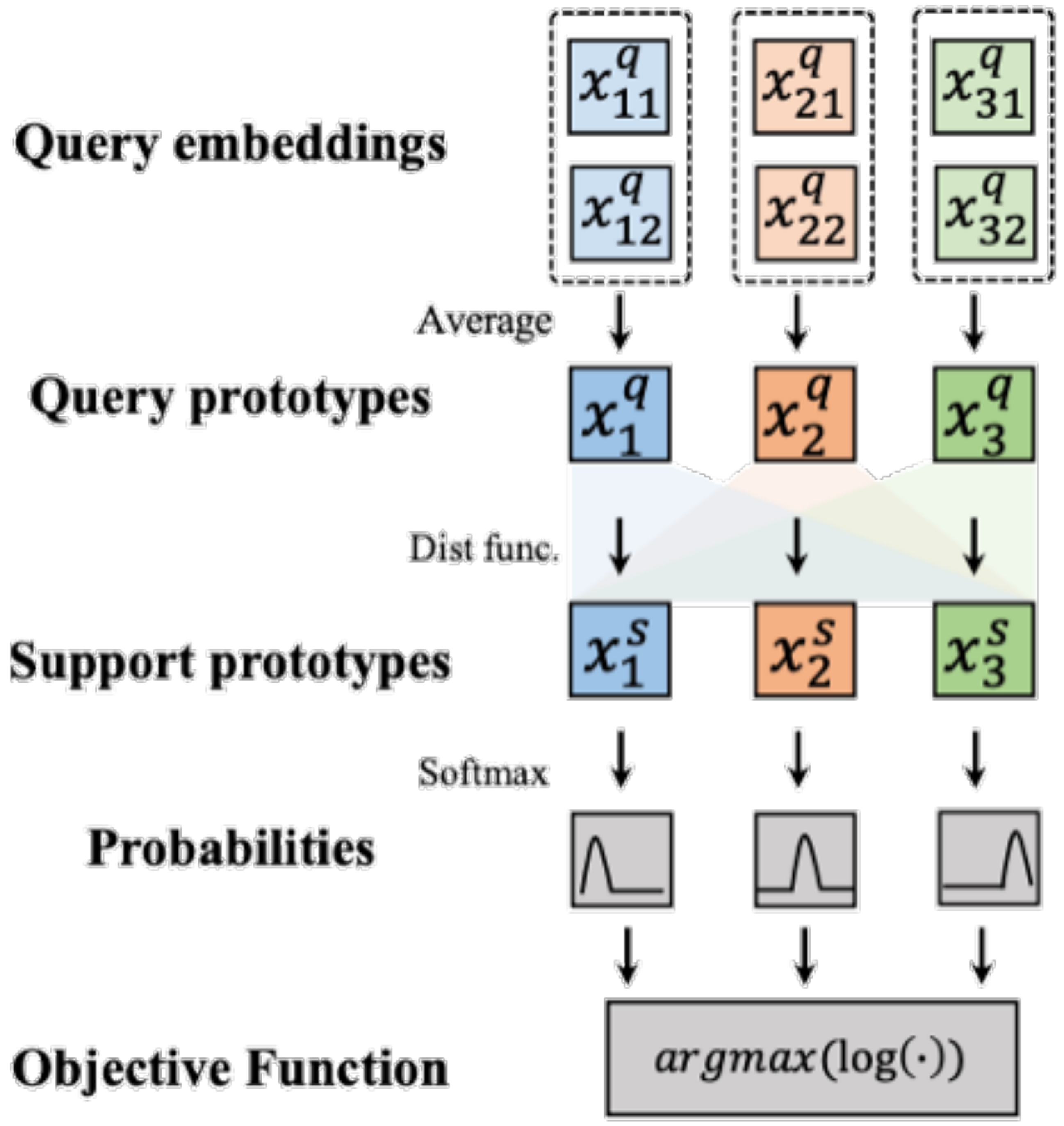
- Metric learning → Prototypical network [1] → learn a latent space.
- In the latent space, the embeddings of different audio segments are expected to be
  - closer (the same class) or further apart (different classes).



Segment of class A (e.g., 0.5 second) → Neural Network → Embedding of class A [1]

[1] Snell, Jake, Kevin Swersky, and Richard Zemel. "Prototypical networks for few-shot learning." *Advances in neural information processing systems* 30 (2017).

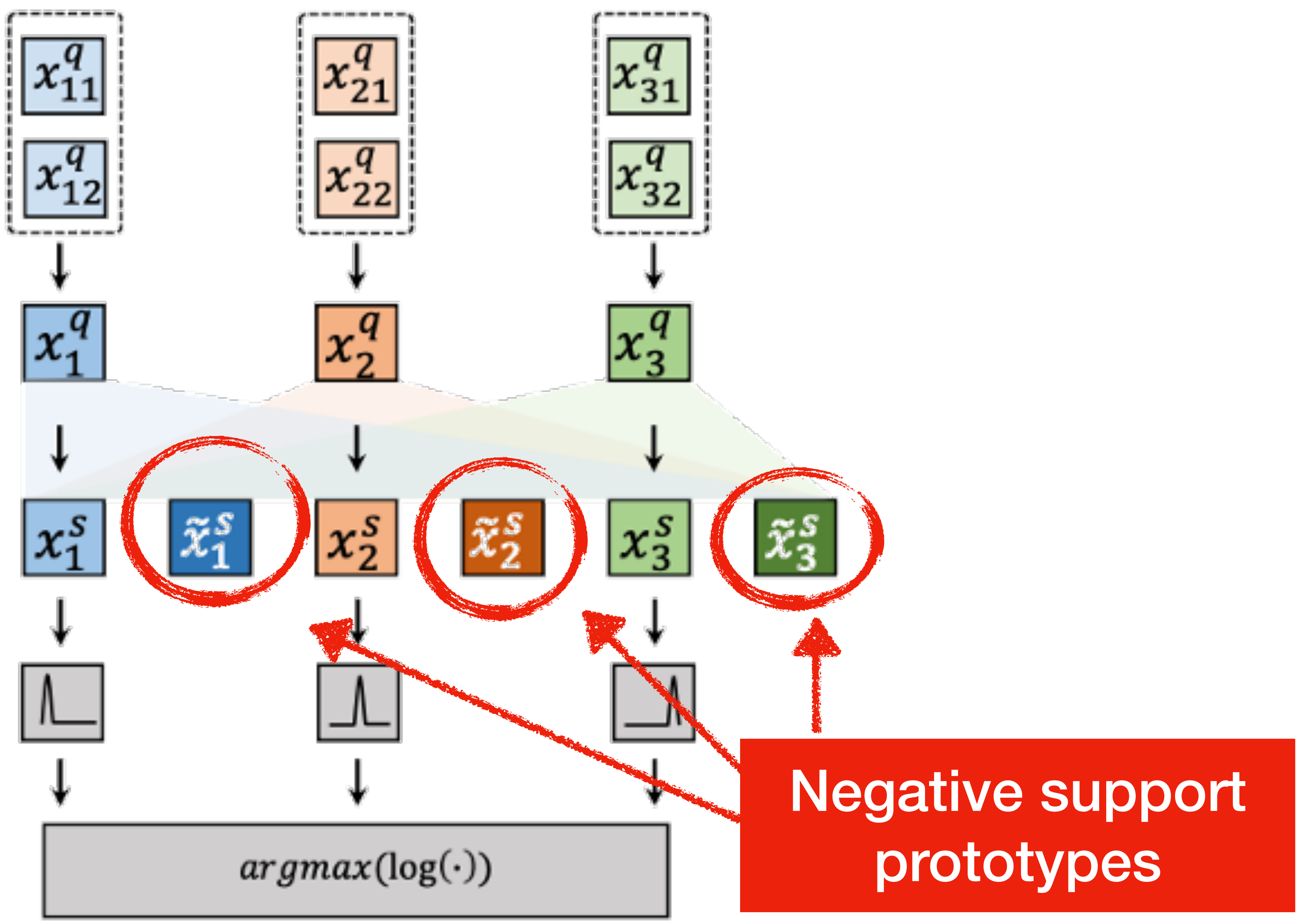
Proposed method



Previous studies

Training with
positive events
(8.7% of the training data)

Contrastive learning with negative segments



Negative support prototypes

Proposed learning with negative segments

Training with
positive and negative events
(100% of the training data)

Proposed method

Other Highlights

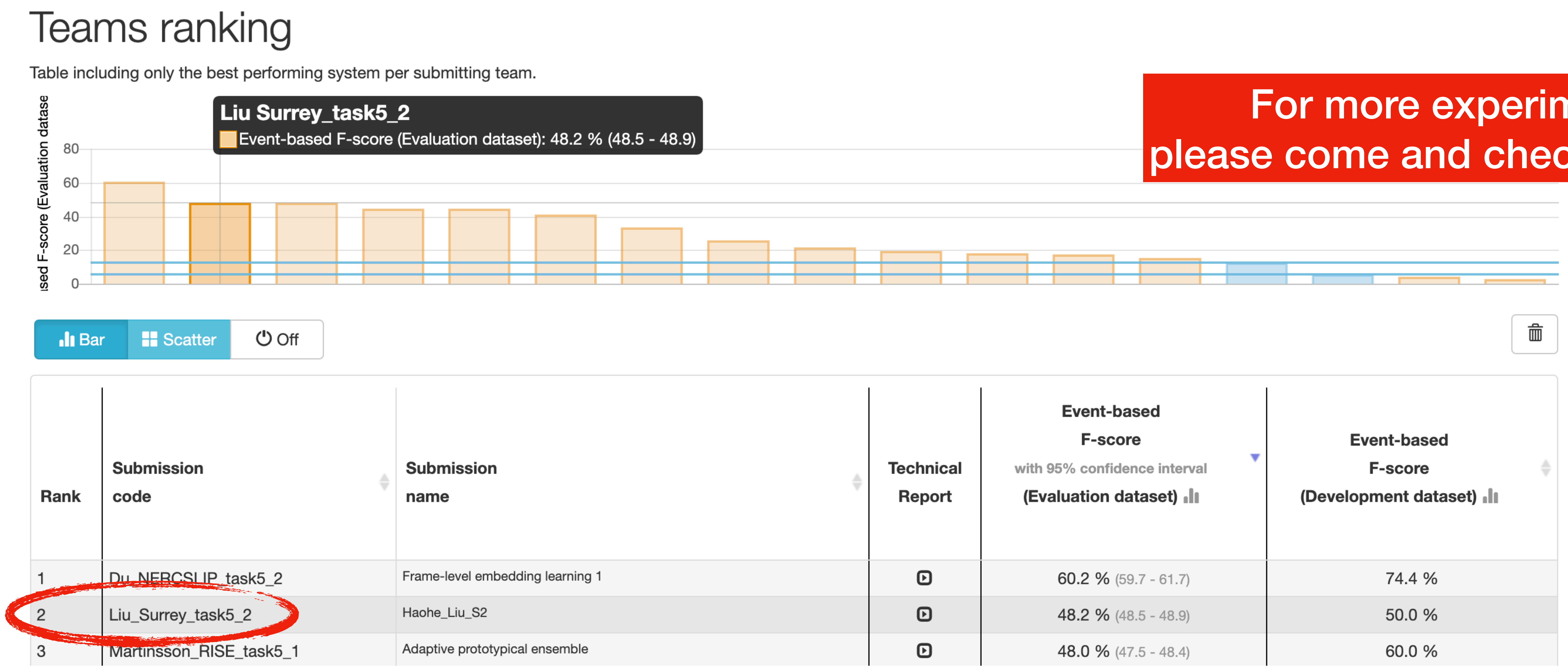
- **How to better adapt to the evaluation data?** → Transductive learning
- **Can more training data help?** → AudioSet strongly-labeled animal sound
- **Which feature is the best?** → Perform feature engineering.
- **Is F-score the ideal metric for evaluation?** → Evaluate with Polyphonic Sound Detection Score (PSDS).
- Other tricks applied: (i) Negative sample searching algorithm; (ii) Adaptive segment length; (iii) Augment training data; (iv) Post-processing.

Result

DCASE2022-T5 Evaluation Set

- Based on the method proposed in this study, our system ranks 2nd in the DCASE 2022 Challenge Task 5: Few-shot Bio-acoustic Detection with an F-score of 48.2.

For more experimental results please come and checkout our poster :-)



Teams ranking

Table including only the best performing system per submitting team.

Liu Surrey_task5_2
Event-based F-score (Evaluation dataset): 48.2 % (48.5 - 48.9)
ased F-score (Evaluation datase
80
60
40
20
0
Bar
Scatter
Off

Rank	Submission code	Submission name	Technical Report	Event-based F-score with 95% confidence interval (Evaluation dataset)	Event-based F-score (Development dataset)
1	Du_NERCSLIP_task5_2	Frame-level embedding learning 1		60.2 % (59.7 - 61.7)	74.4 %
2	Liu_Surrey_task5_2	Haohe_Liu_S2		48.2 % (48.5 - 48.9)	50.0 %
3	Martinsson_RISE_task5_1	Adaptive prototypical ensemble		48.0 % (47.5 - 48.4)	60.0 %

https://dcase.community/challenge2022/task-few-shot-bioacoustic-event-detection-results

Thanks for your listening!

In conclusion, the following points are helpful for few-shot bioacoustic detection: (1) Contrastive learning with negative segments; (2) Transductive learning; (3) Data augmentations; (4) Feature engineering; and (5) External data from AudioSet.

- Paper: https://arxiv.org/abs/2207.07773
- Open-sourced code: https://github.com/haoheliu/DCASE_2022_Task_5

AI for sound

CVSSP Centre for Vision, Speech and Signal Processing

UNIVERSITY OF SURREY

ByteDance