

Machine Listening in SONYC

Justin Salamon, Jason Cramer, Ho-Hsiang Wu,
Peter Li, Mark Cartwright, Brian McFee, Juan Pablo Bello

Justin Salamon^{1,2},

Senior Research Scientist

¹Music and Audio Research Laboratory

²Center for Urban Science and Progress

New York University, USA

 @justin_salamon

 justin.salamon@nyu.edu

 www.justinsalamon.com

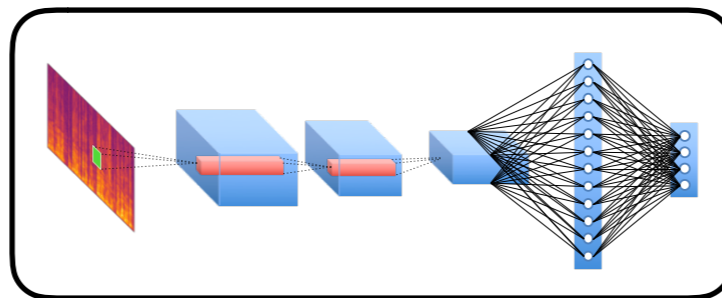


Machine Listening

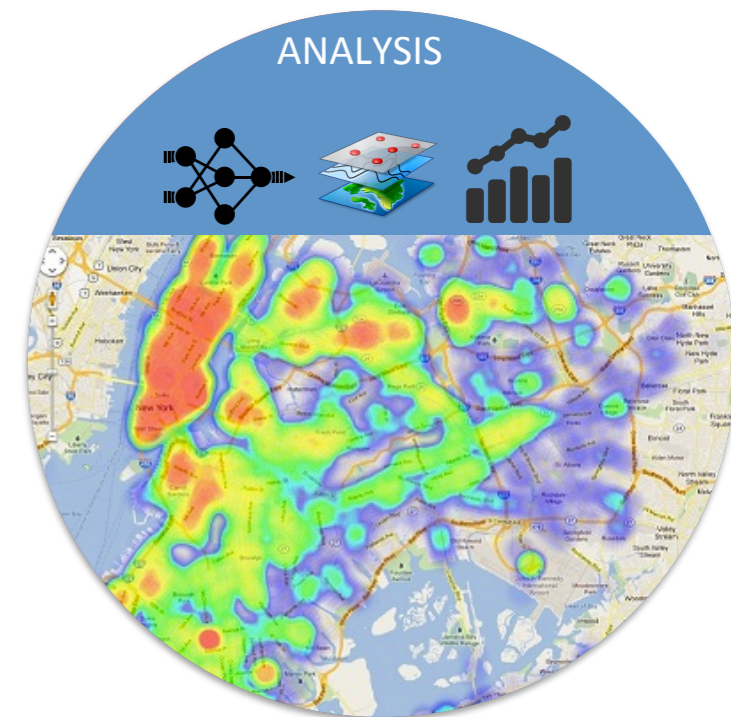
Acoustic Sensor Network



Machine Listening



Data Science



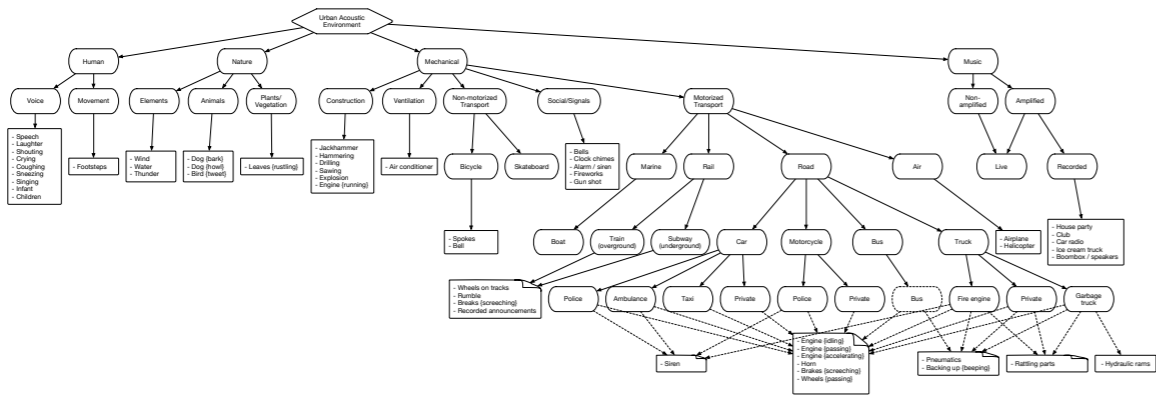
Motivation: noise monitoring

- SPL measurements alone cannot identify the **source** of the noise
- Noise source is relevant to identification of possible noise code violation

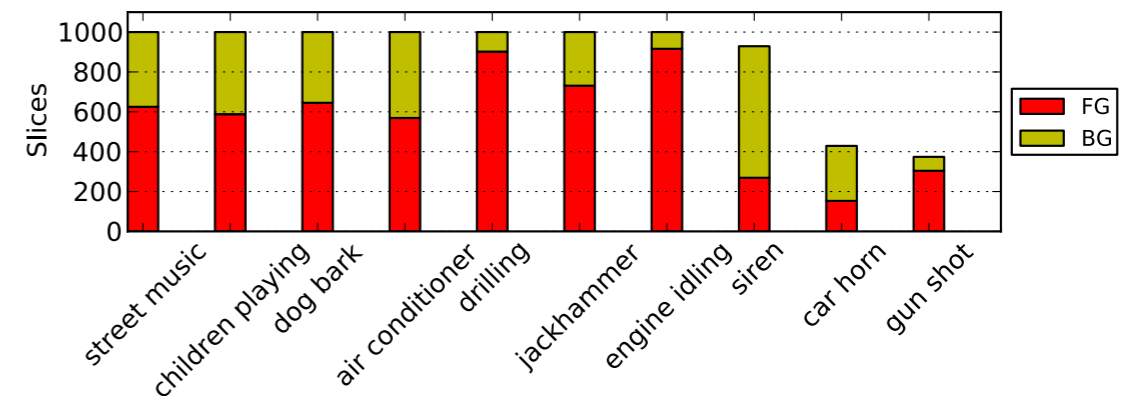


Where we left off...

Urban Sound Taxonomy

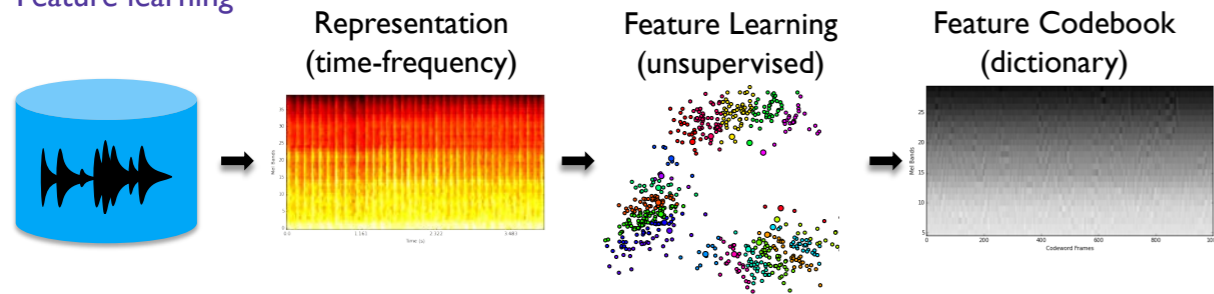


UrbanSound8K dataset

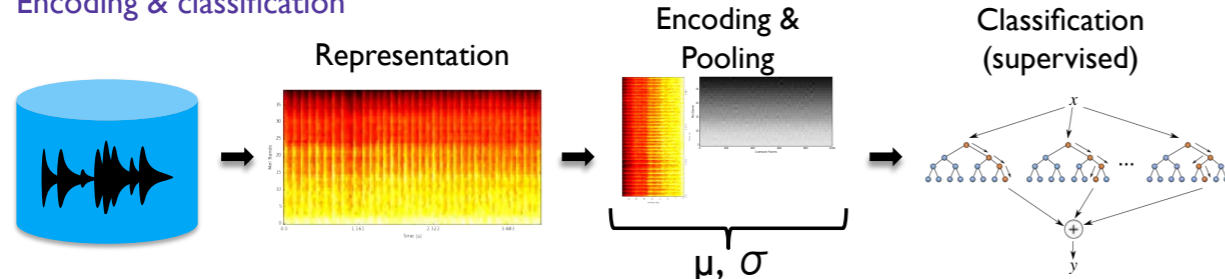


Unsupervised Feature Learning

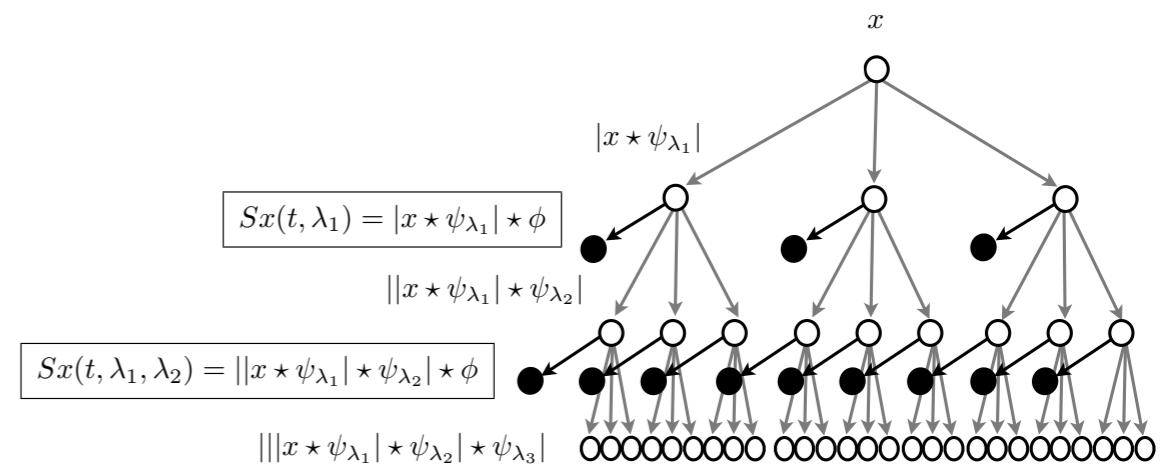
Feature learning



Encoding & classification

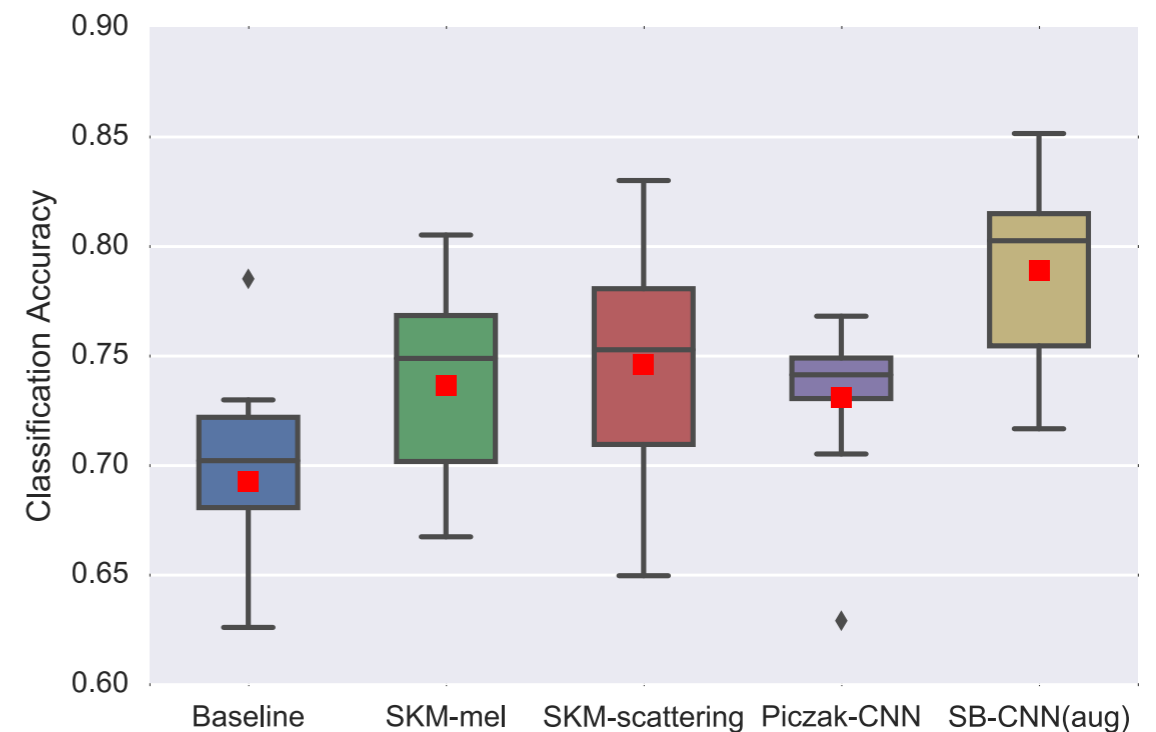


Deep Scattering Signal Representations

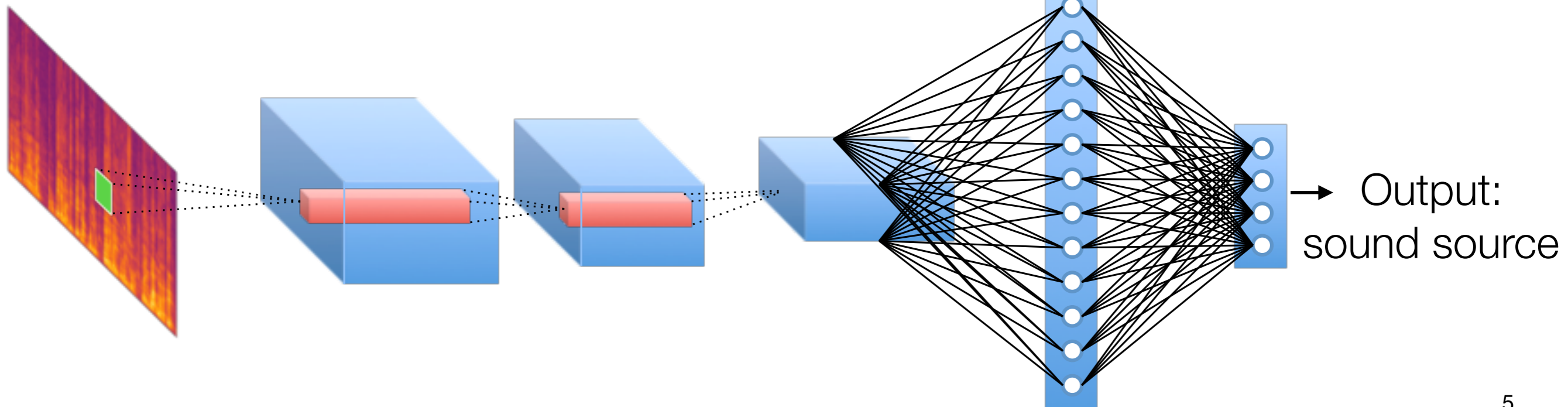


Deep Learning with Convolutional Networks and Data Augmentation

- J. Salamon and J. P. Bello, “Deep Convolutional Neural Networks and Data Augmentation For Environmental Sound Classification”, IEEE Signal Processing Letters, 24(3), pages 279 - 283, 2017.



Input:
audio recording



UrbanEars v0.2

Air conditioner

Car horn

Children

Dog bark

Drilling

Engine idling

Gun shot

Jackhammer

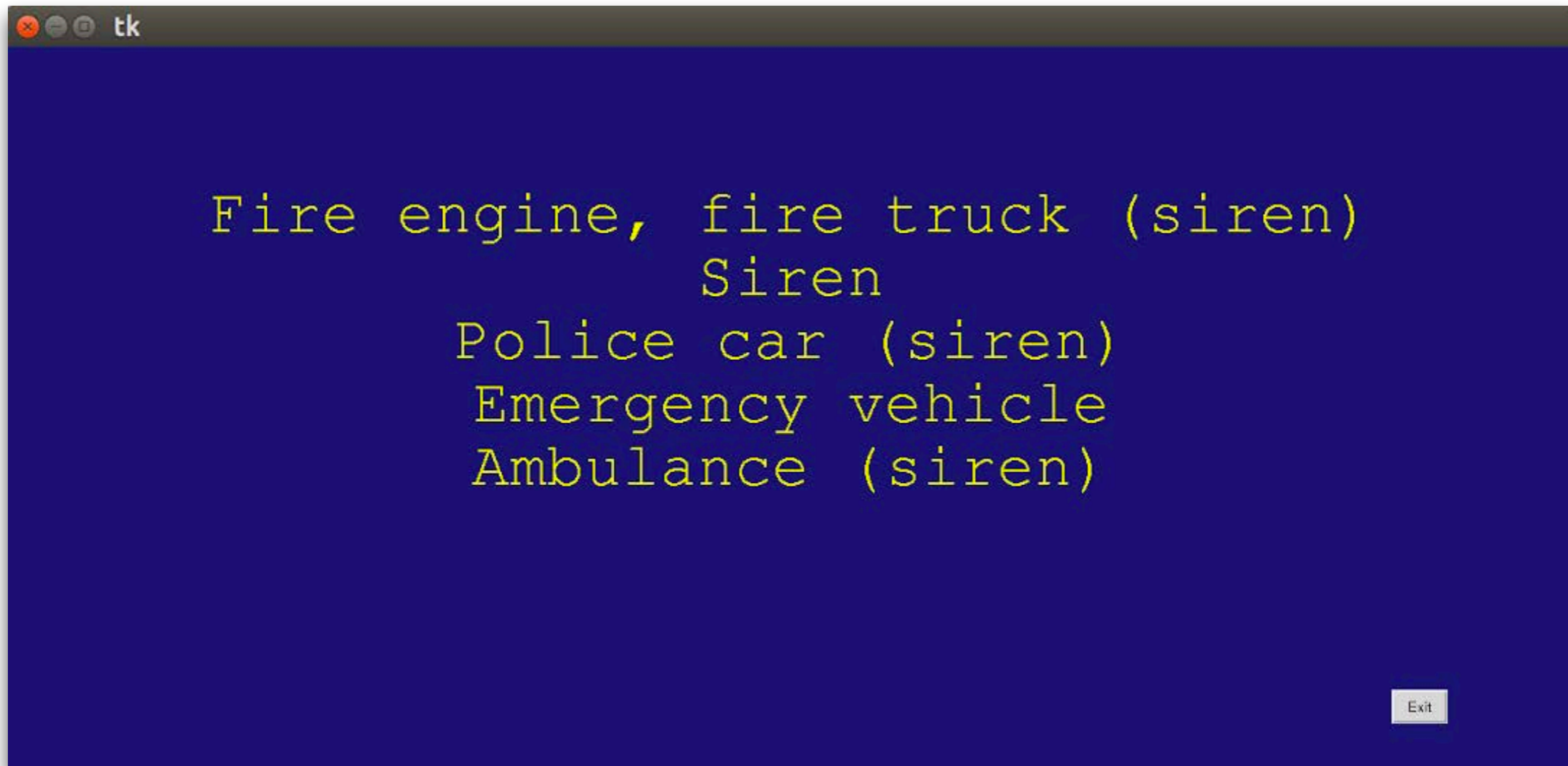
Siren

Street music



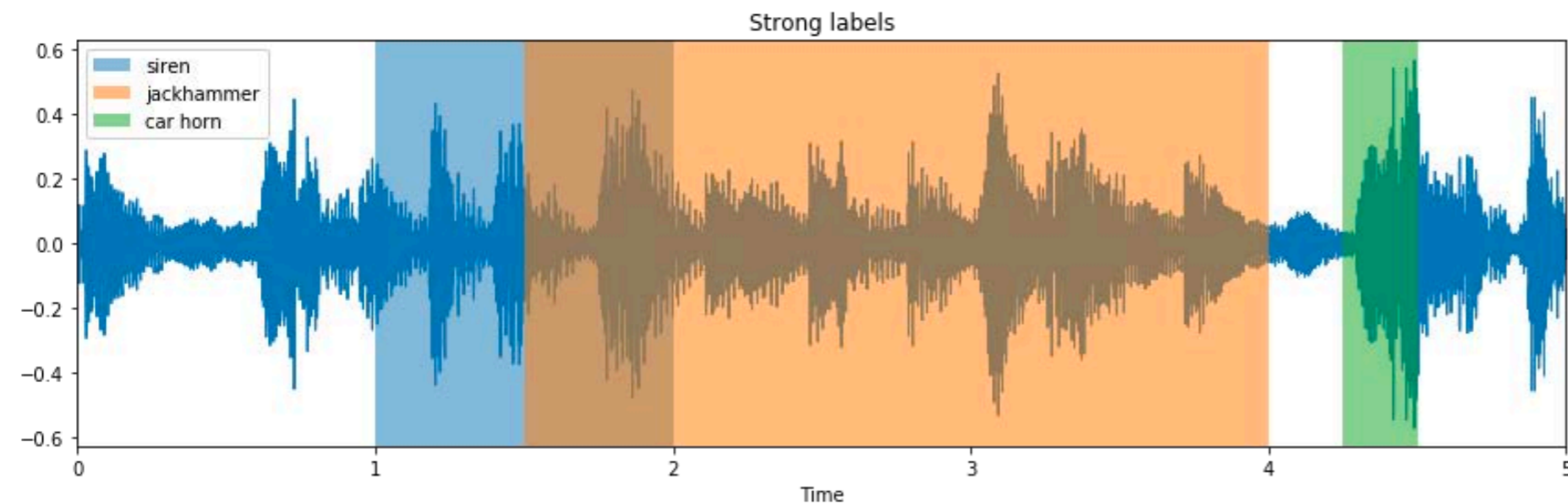
RMS

Status: gun_shot: 1.00 (rms:0.0000)



The Challenge

- Training supervised models requires **labeled** data
- Strong label = [start time, end time, source label]



[1.0, 2.0, "siren"]
[1.5, 4.0, "jackhammer"]
[4.2, 4.5, "car horn"]

- Strongly labelled data is:
 1. Scarce
 2. Costly to produce

Solutions

1. Crowdsourcing → Harness the crowd to label data!
2. Data synthesis → Create artificial data!
3. Multiple instance learning → Do more with less!
4. Deep audio embeddings → Exploit structure in unlabeled data!

Crowdsourcing: harness the crowd to label data

- Next talk... stay tuned!

Data synthesis: create artificial data



Data synthesis: *Scaper*

- Open source python library for soundscape synthesis (IEEE WASPAA 2017)
- Returns soundscape audio + annotation containing strong labels
- github.com/justinsalamon/scaper

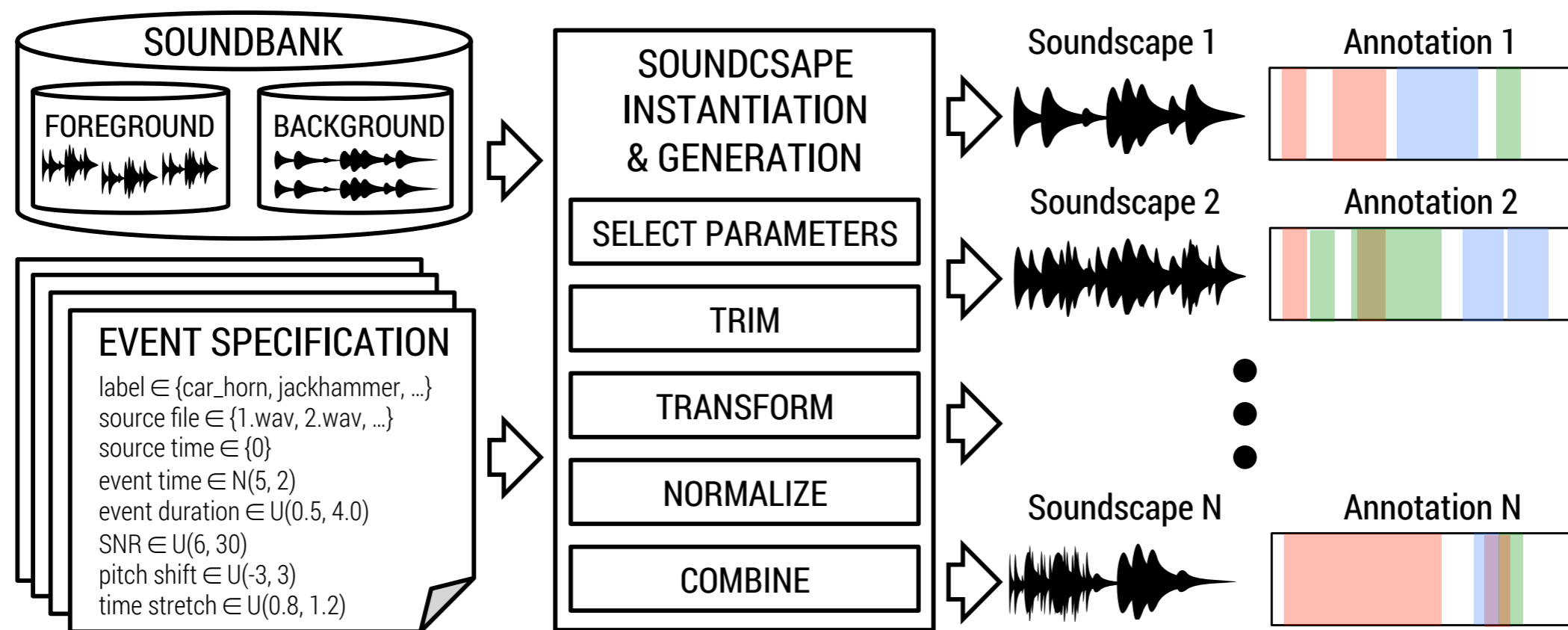
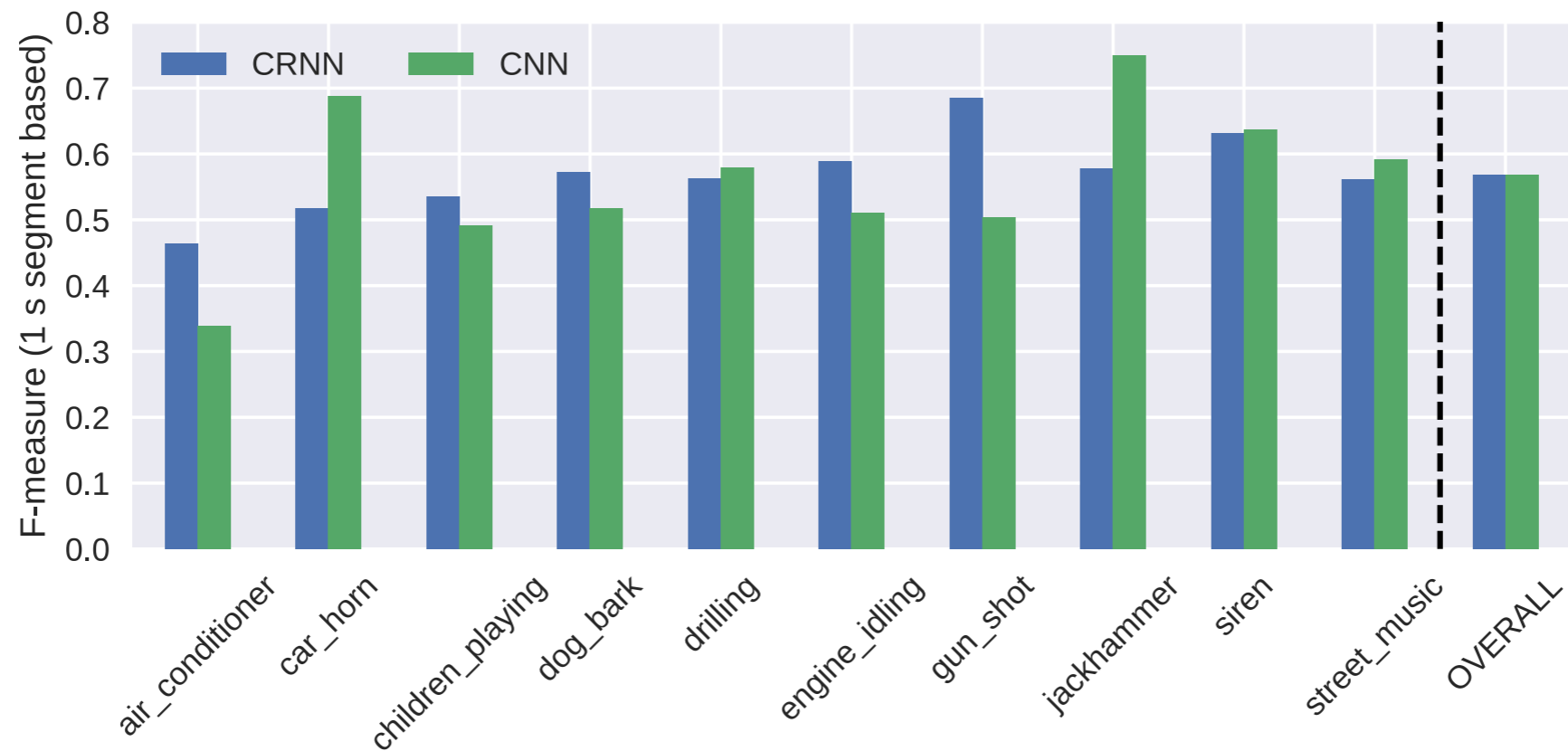


Figure 1: Block diagram of the Scaper synthesis pipeline.

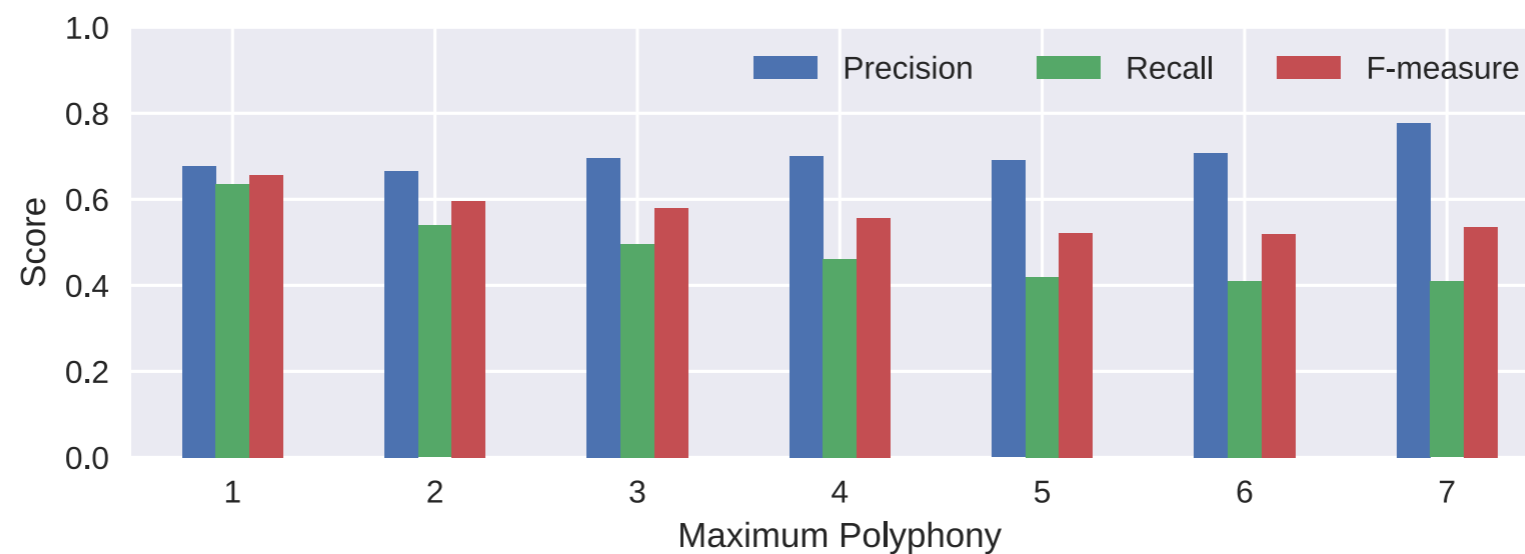
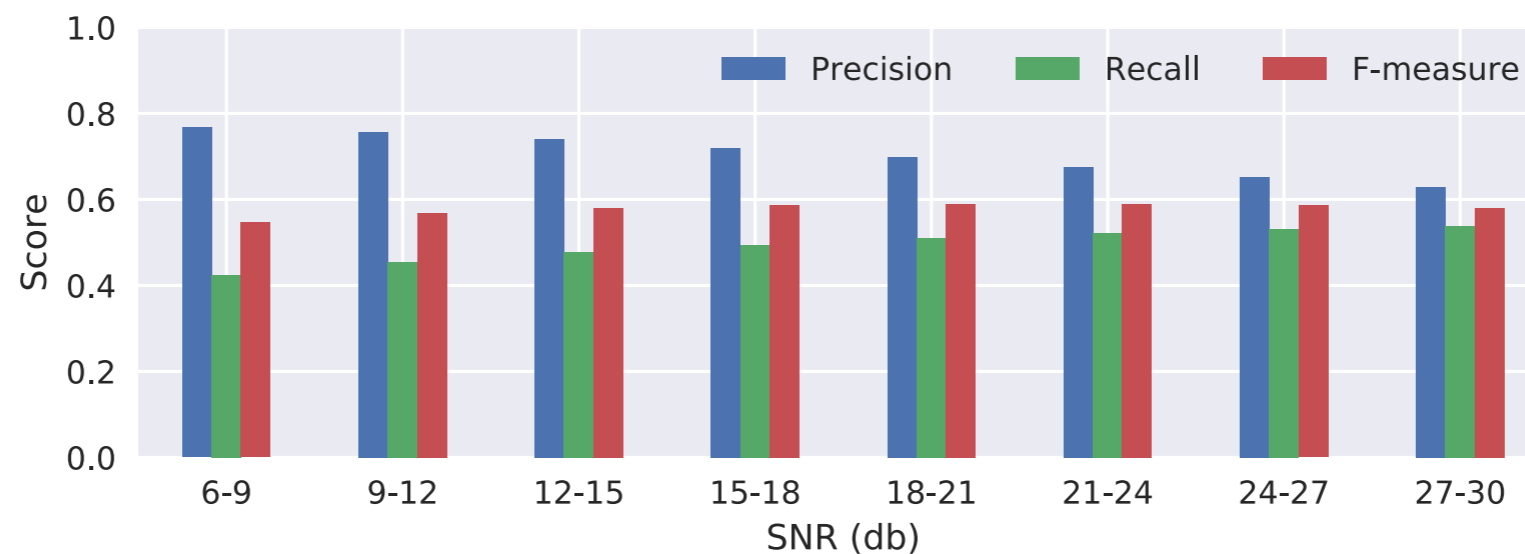
Data synthesis: *Scaper*

- URBAN-SED: new dataset of 10,000 soundscapes created using Scaper



Data synthesis: *Scaper*

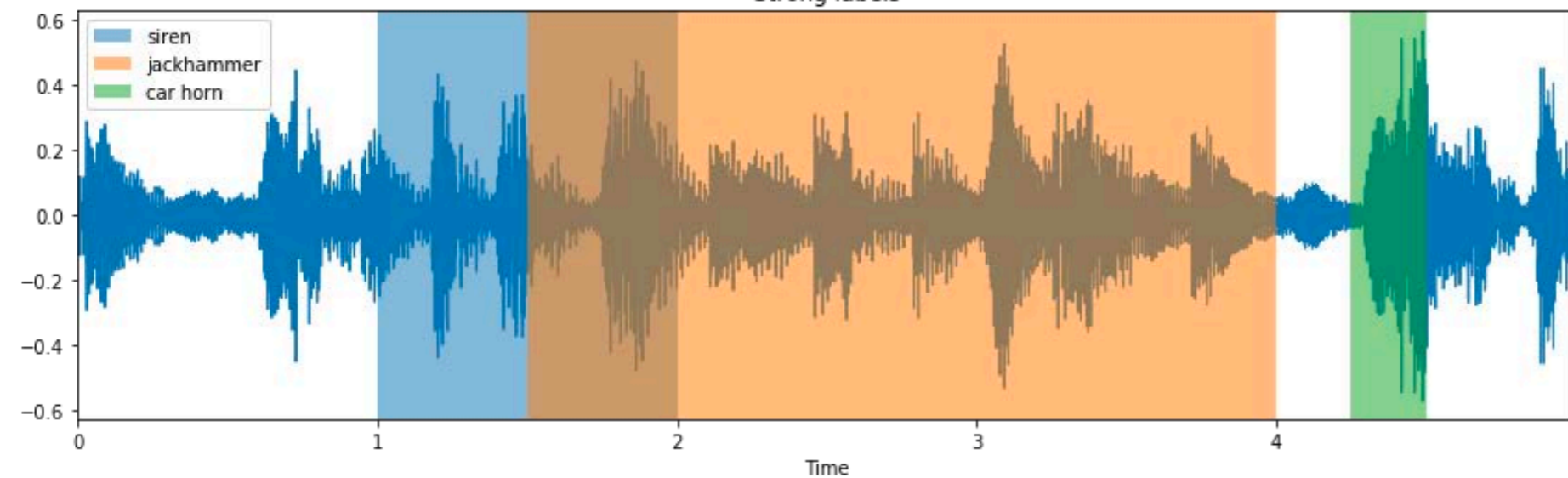
- URBAN-SED: new dataset of 10,000 soundscapes created using Scaper



Multiple Instance Learning: do more with less

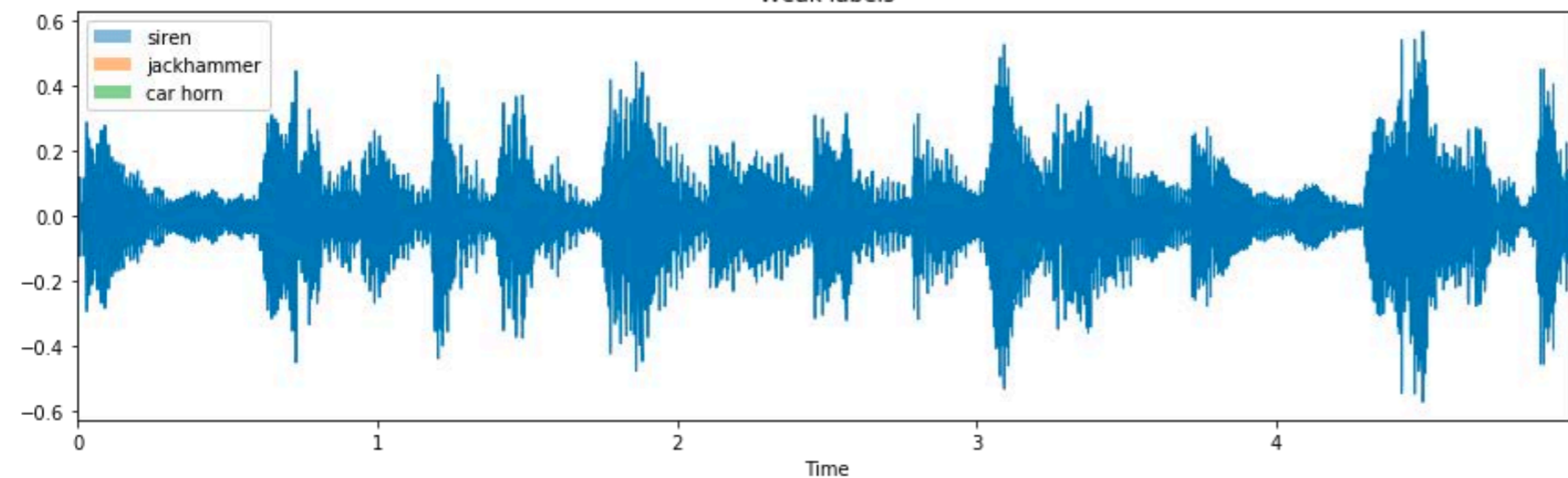
- Strong labels (start time, end time, label) are scarce and costly
- But... **weak** labels (no timing information) are easier to come by and generate

Strong labels



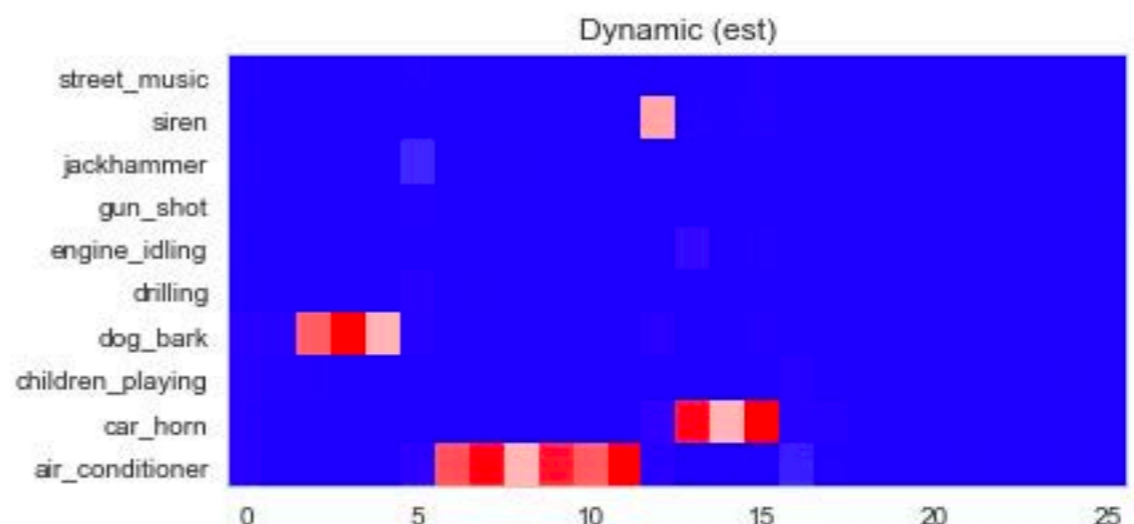
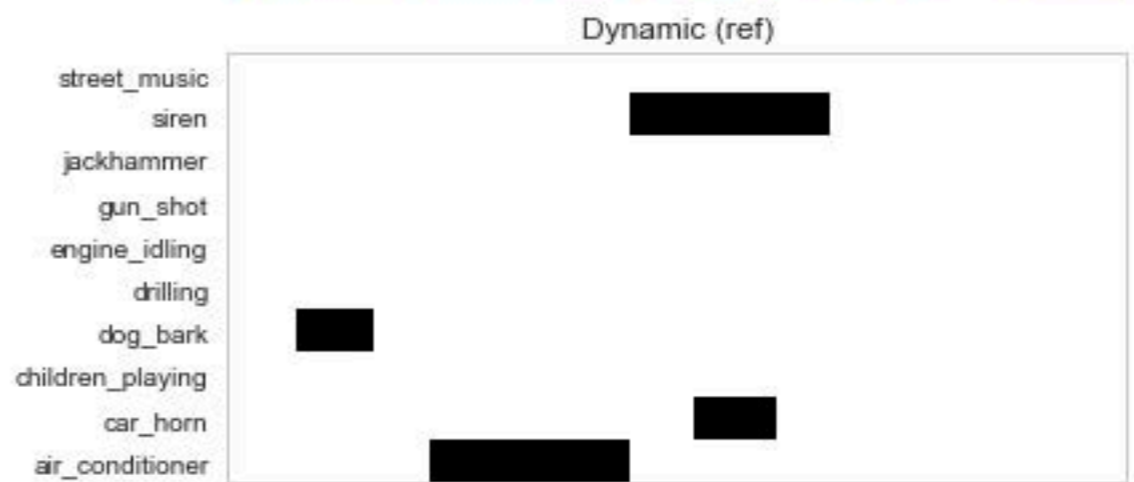
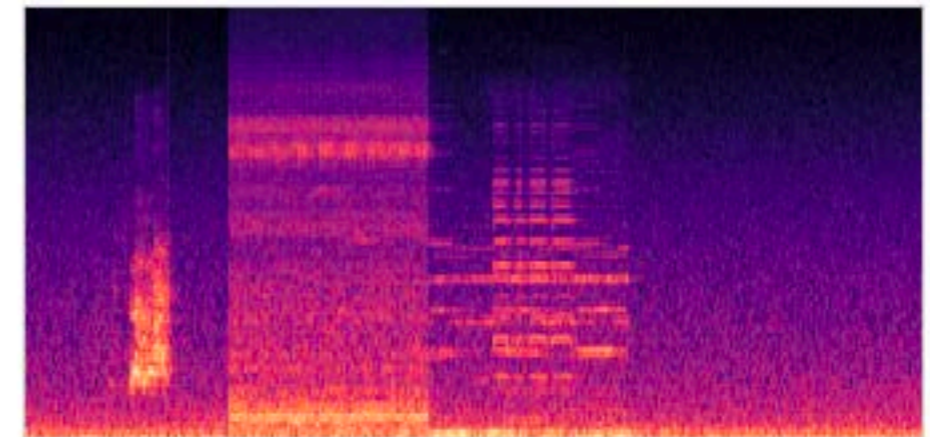
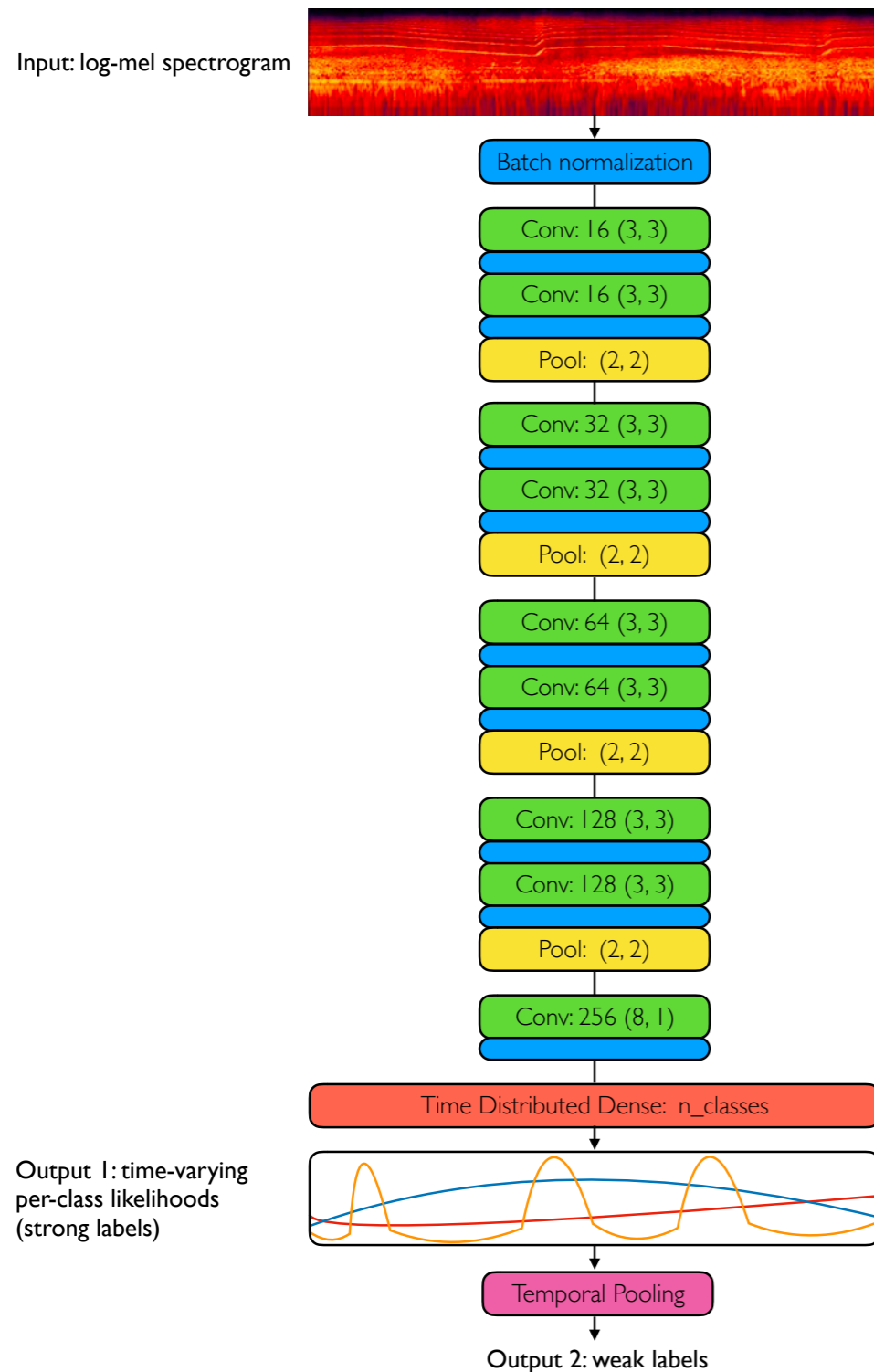
[1.0, 2.0, "siren"]
[1.5, 4.0, "jackhammer"]
[4.2, 4.5, "car horn"]

Weak labels

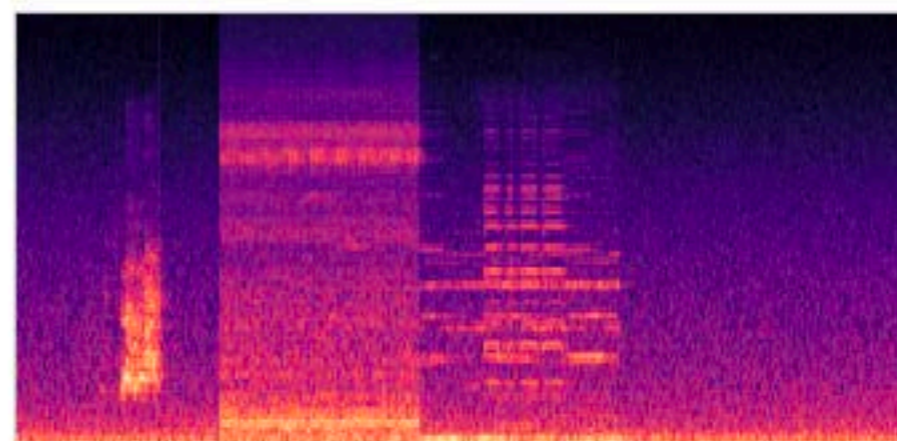


["siren"]
["jackhammer"]
["car horn"]

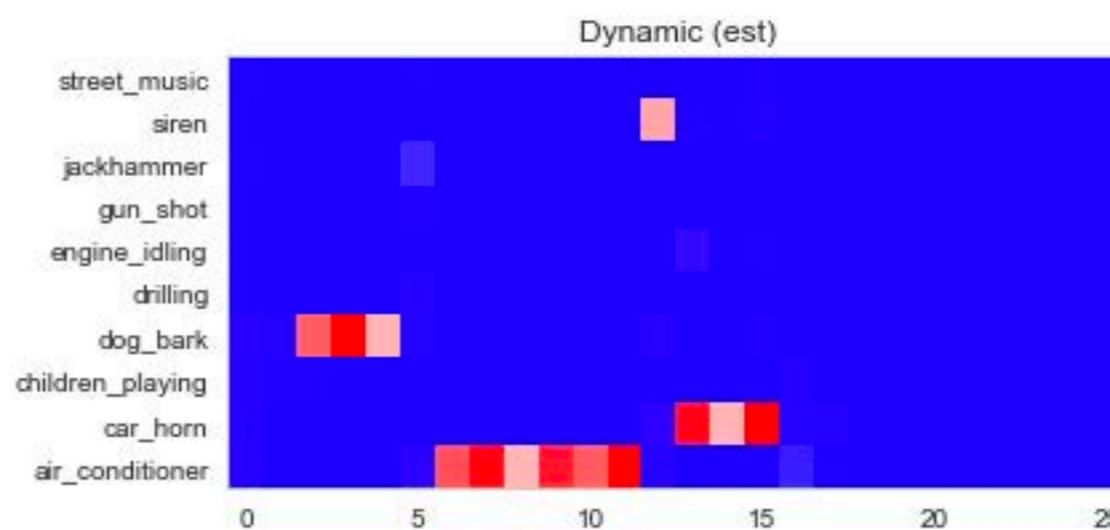
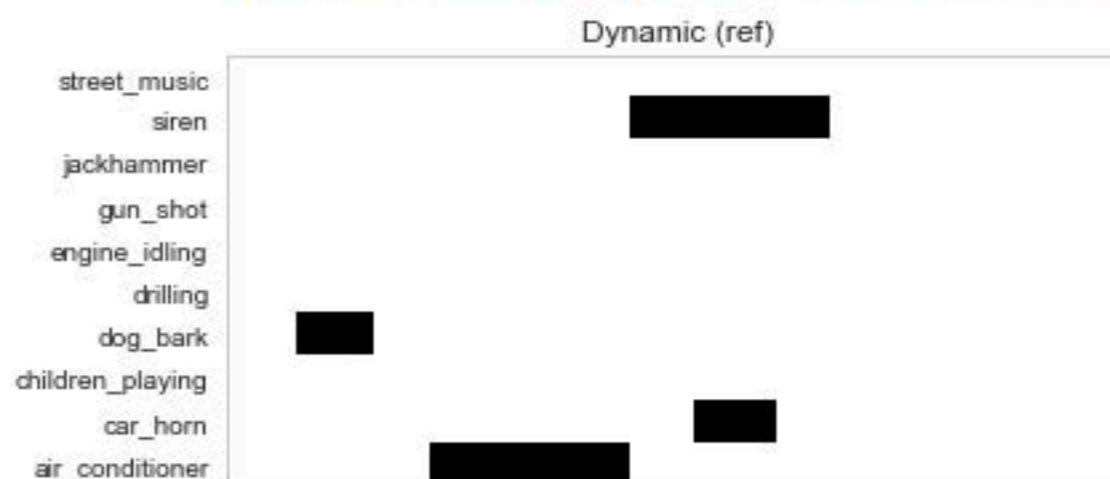
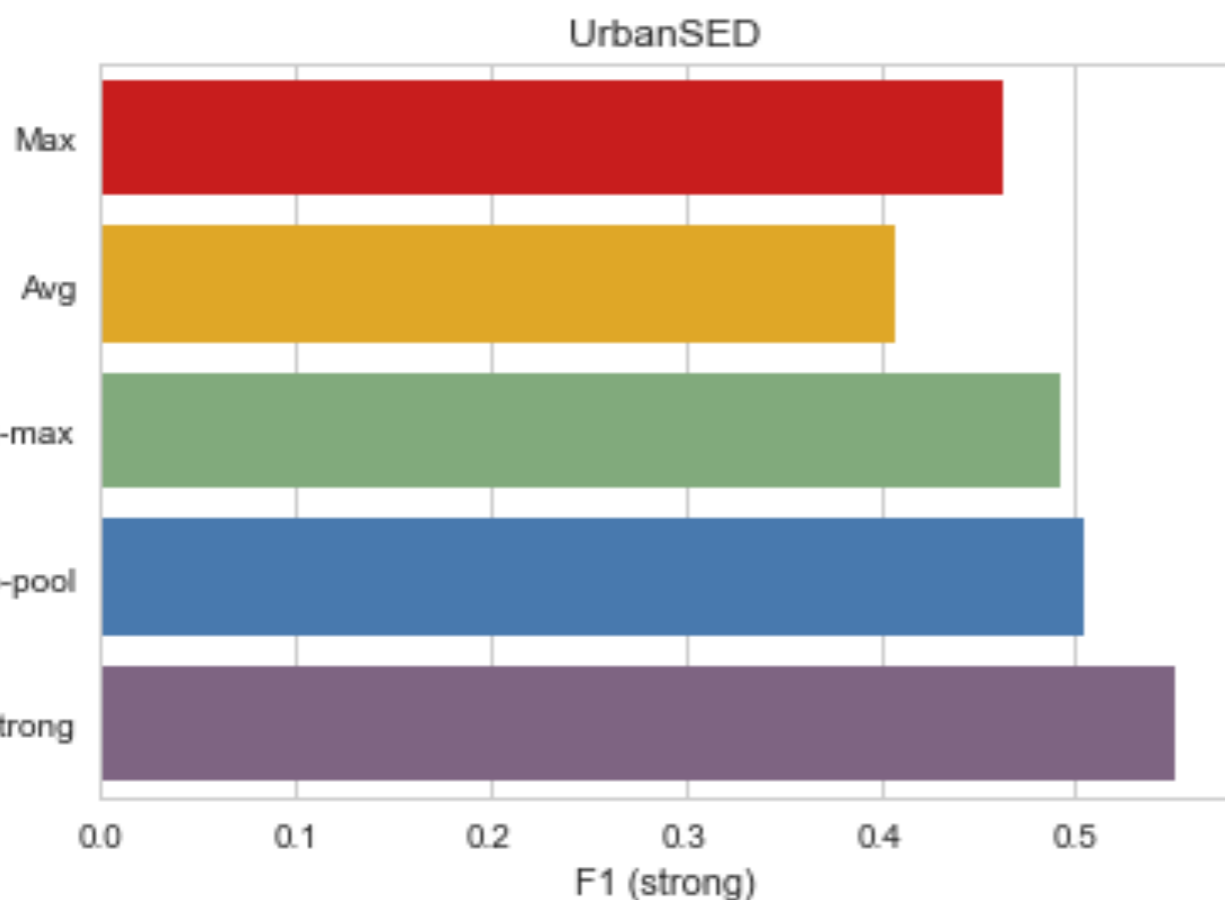
Multiple Instance Learning: do more with less



Multiple Instance Learning: do more with less



Strong Label Prediction Performance



Deep Audio Embeddings: exploit structure in unlabeled data

- Some types of data, even without any labels, have **structure**

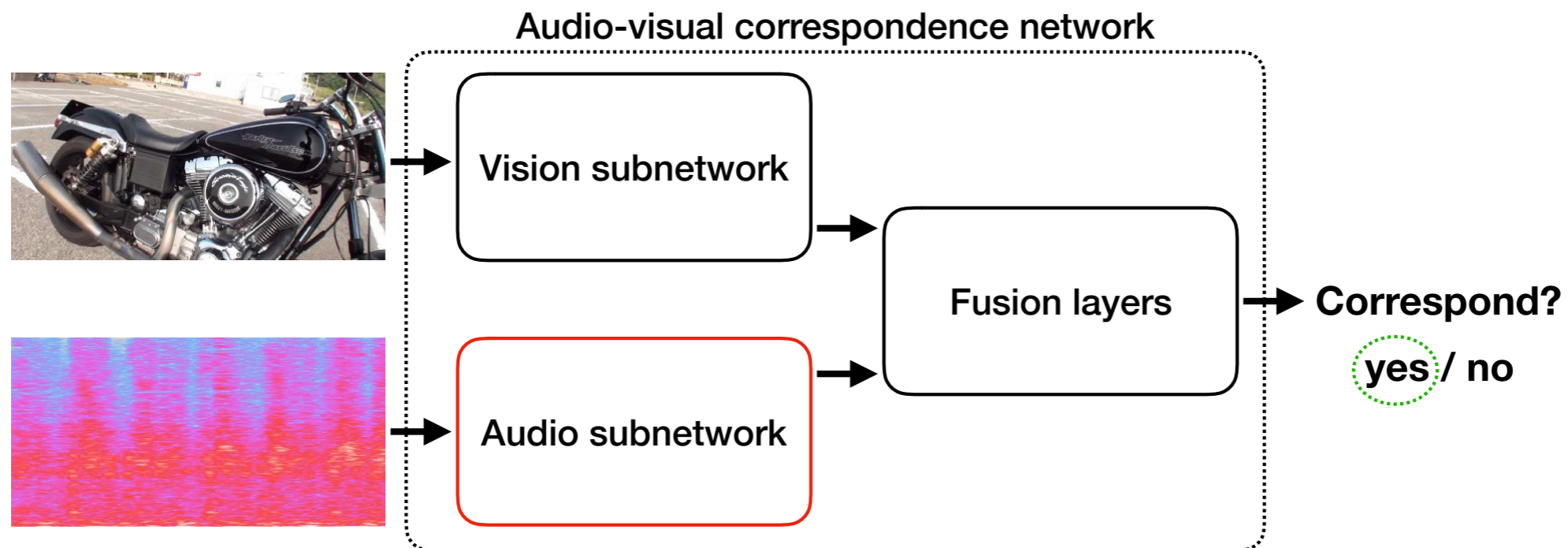
“The quick brown fox jumps over the lazy dog”



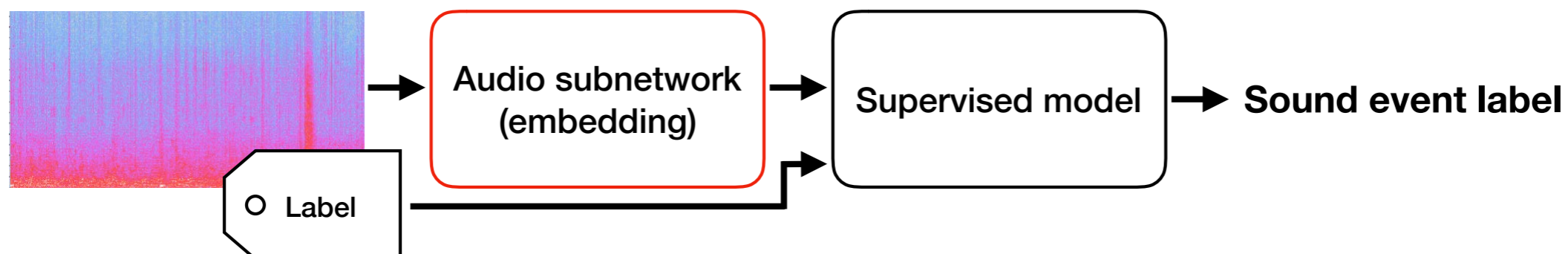
Deep Audio Embeddings: exploit structure in unlabeled data

- Arandjelovic & Zisserman, “Look, listen and learn”, ICCV 2017

Self-supervised training (learning an embedding):



Supervised training using the learned audio embedding



Publications

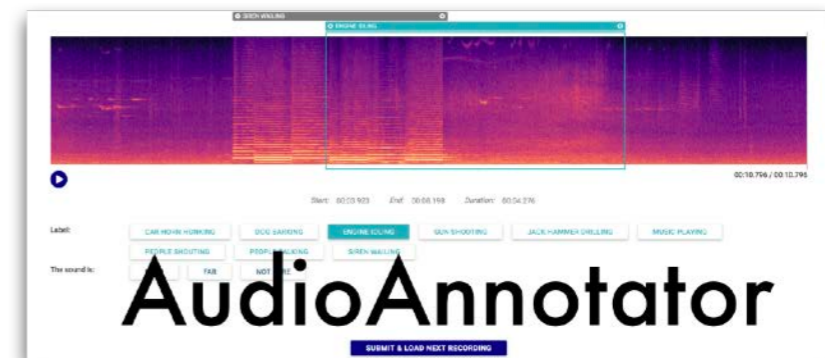
- J. P. Bello, C. Mydlarz, and J. Salamon. **Sound analysis in smart cities**. In T. Virtanen, M. D. Plumbley, and D. P. W. Ellis, editors, Computational Analysis of Sound Scenes and Events, pages 373–397. Springer International Publishing, 2018.
- M. Cartwright, J. Salamon, A. Seals, O. Nov, and J. P. Bello, **Investigating the effect of sound-event loudness on crowdsourced audio annotations**. In IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Calgary, Canada, Apr. 2018.
- M. Cartwright, A. Seals, J. Salamon, A. Williams, S. Mikloska, D. MacConnell, E. Law, J. Bello, and O. Nov. **Seeing sound: Investigating the effects of visualizations and complexity on crowdsourced audio annotations**. Proceedings of the ACM on Human-Computer Interaction, 1(2), 2017.
- J. Salamon, D. MacConnell, M. Cartwright, P. Li, and J. P. Bello. **Scaper: A library for soundscape synthesis and augmentation**. In IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, Oct. 2017.
- J. Salamon and J. P. Bello. **Deep convolutional neural networks and data augmentation for environmental sound classification**. IEEE Signal Processing Letters, 24(3):279–283, Mar. 2017.
- J. Salamon and J. P. Bello. **Feature learning with deep scattering for urban sound analysis**. In 2015 European Signal Processing Conference, Nice, France, Aug. 2015.
- J. Salamon and J. P. Bello. **Unsupervised feature learning for urban sound classification**. In IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), pages 171–175, Brisbane, Australia, Apr. 2015.
- J. Salamon, C. Jacoby, and J. P. Bello. **A dataset and taxonomy for urban sound research**. In 22nd ACM International Conference on Multimedia (ACM-MM'14), pages 1041–1044, Orlando, FL, USA, Nov. 2014.

Open & Reproducible Science

- Datasets



- Open-source code



SKM

envsoundcnn

MILSED

AudioSetDL

l3embedding

What next...?

- Deep embeddings based on Audio-Visual Correspondence
- Deep embeddings based on **spatio-temporal structure in SONYC audio data**
- New multi-label sound recognition model based on deep embeddings and new **taxonomy** targeting sources of interest to the city agencies
- Model compression and **deployment in SONYC sensors**

- Posters & demos!
 - Deep embeddings (Jason Cramer & Ho-Hsiang Lee)
 - Multiple Instance Learning (Brian McFee)
 - Real-time sound classification (Peter Li & Justin Salamon)
 - Scaper (Justin Salamon)

Thanks!