NEC

NEC Laboratories **America**

SEED: SOUND EVENT EARLY DETECTION VIA EVIDENTIAL UNCERTAINTY

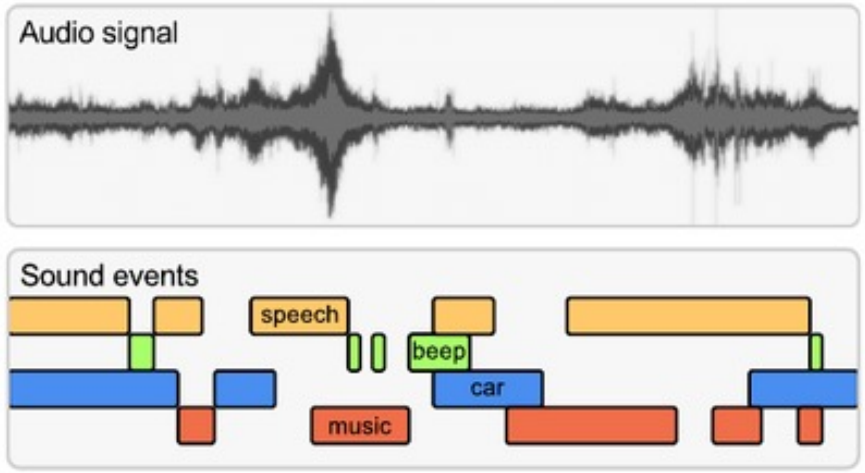
Xujiang Zhao[1], Xuchao Zhang[2], Wei Cheng[2], Wenchao Yu[2], Yuncong Chen[2] , Haifeng Chen[2] , Feng Chen[1]

[1]The University of Texas at Dallas, [2]NEC Laboratories America

IEEE ICASSP 2022

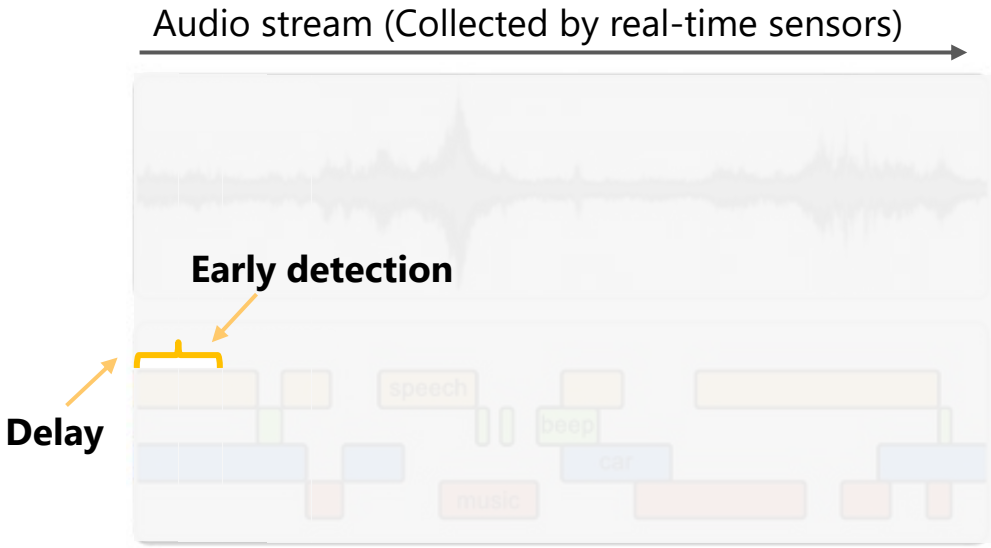
Motivation

- Sound is everywhere. Sound event would happen in many environments: domestic environment, plant environment, etc.
- **Real-time response after the event is happened.**



Sound Event Early Detection (SEED)

- Input: stream audio
- Detect each event and reduce the detection delay



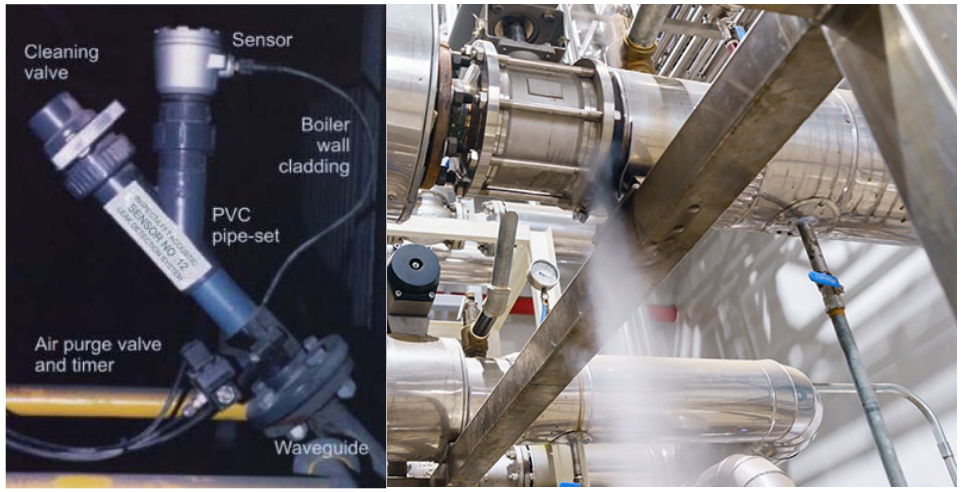
Sound Event Early Detection would focus on early detection

Motivation (Example)

- Audio Surveillance: Steam Leak Detection in Thermal Power Plant

  Purpose: Detecting the sound waves emanating from the steam leak.

- Benefits
  - Increase operating profits
  - Ensure personal safety
  - Avoidance of unscheduled outages



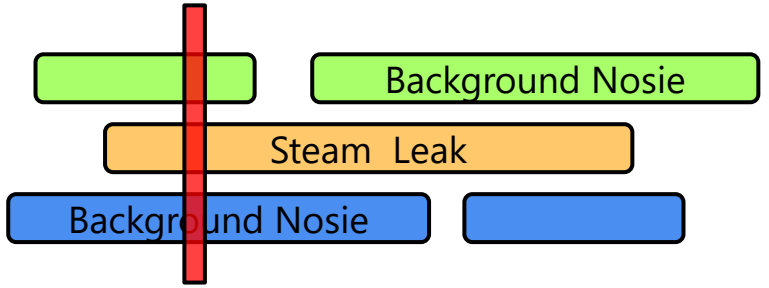
Sai et al. Application of Acoustic Techniques in Thermal Power Plants

Impact of Sound Event Early Detection

Challenge

I. Overlapping sound event

- It is difficult to detect target event due to overlapping of polyphony sound.



II. Difficult to detect sound event in early stage

- Early stage: 1-2 small detection windows (60ms).

III. Real-time inference

- Inference time < detection window.

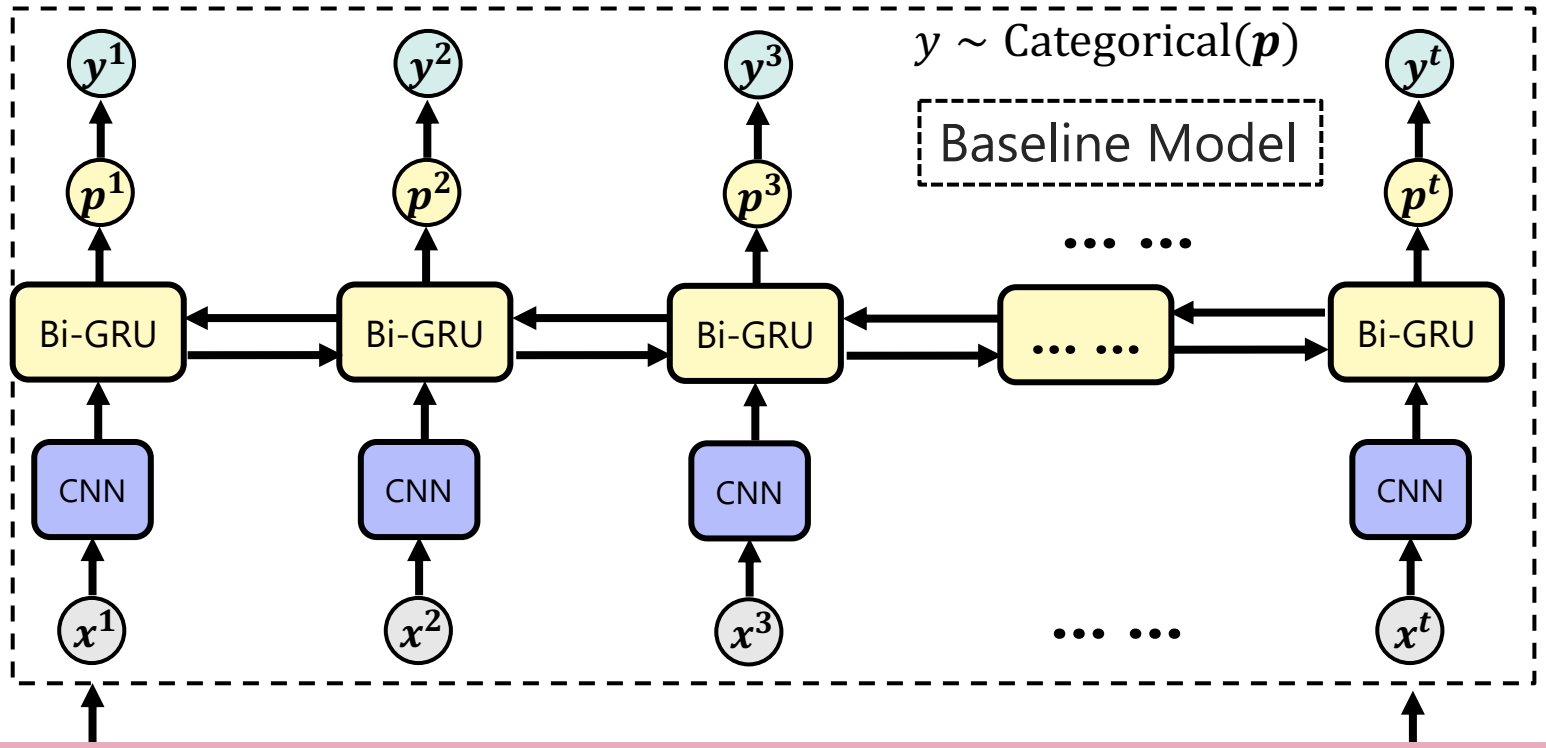
Challenge of Sound event early detection

Contribution

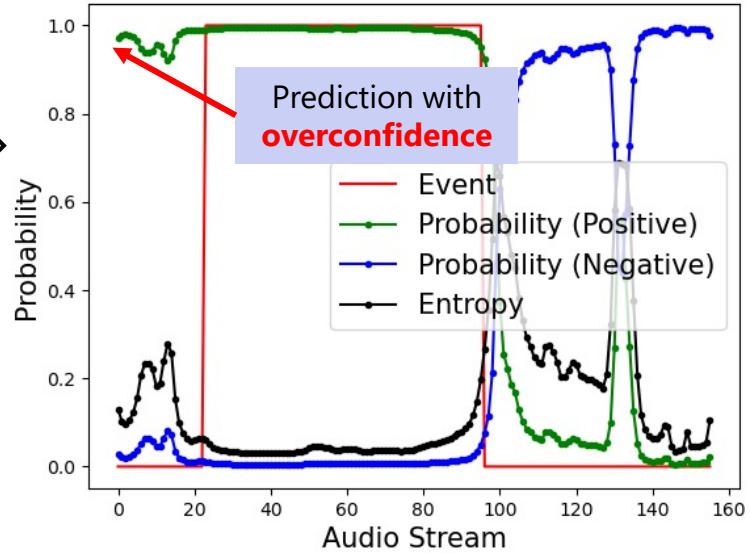
- We proposed a Multi-label Evidential neural network to solve SEED and provide a realizable prediction based on uncertainty scores.
- Our model can significantly reduce the detection delay and improve the prediction accuracy.
- The evidence information (include belief, disbelief and uncertainty) can help human being to make better decision.

Contribution

Limitation of existing method



The prediction is **not reliable** at the early stage



We consider Evidential Uncertainty to improve the performance of early sound event detection!

Background of Evidential Uncertainty

We obtain the empirical evidence of the test image from training set (most similar training images):

Test image: 

Empirical evidence:

Dog ($e_1 = 3$ images)

Cat ($e_2 = 3$ images)

Pig ($e_3 = 3$ images)

$e = [e_1, e_2, e_3]$

In deep learning, the researchers are interested in predicting class probabilities

(Class probabilities)

$$P_k = \frac{e_k}{\sum_k e_k}$$

$$p = [p_1, p_2, p_3] = \left[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right]$$

(Probabilistic uncertainty)

$$\text{Entropy(p)} = -\frac{1}{3}\sum_{k=1}^{3} \log\frac{1}{3}$$

Evidence: a measure of the amount of support for a certain class

Background of Evidential Uncertainty

We obtain the empirical evidence of the test image from training set (most similar training images):

Test image: 

Empirical evidence:   

Dog ($e_1 = 3$ images) Cat ($e_2 = 3$ images) Pig ($e_3 = 3$ images)

$e = [e_1, e_2, e_3]$

In the belief theory domain (e.g., subjective logic), the researchers are interested in predicting subjective opinions

(Subjective Opinion ω)

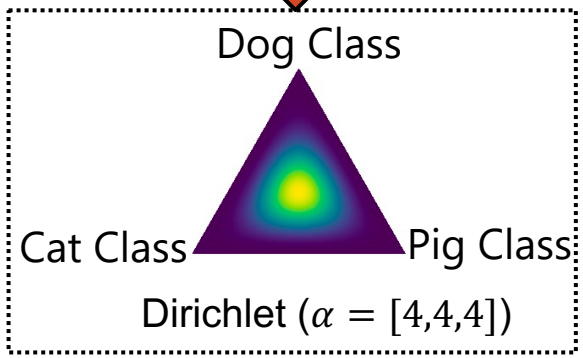
$$\omega = [b_1, b_2, b_3, u] = [0.25, 0.25, 0.25, 0.25]$$

$$K = 3$$

$$b_k = \frac{e_i}{\sum_{i=1}^{K} e_i + K},$$

$$u = \frac{K}{\sum_{i=1}^{K} e_i + K}$$ **Vacuity**

$\boldsymbol{\alpha} = \boldsymbol{e} + \mathbf{1} = [\mathbf{4}, \mathbf{4}, \mathbf{4}]$



Background of Evidential Uncertainty

We obtain the empirical evidence of the test image from training set (most similar training images):

Test image: 

Empirical evidence:   

Dog ($e_1 = 3$ images) Cat ($e_2 = 3$ images) Pig ($e_3 = 3$ images)

$e = [e_1, e_2, e_3]$

In deep learning, we are interested in predicting class probabilities

(Class Probabilities)

$$p = [p_1, p_2, p_3] = \left[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right]$$

$y \sim \text{Categorical}(\boldsymbol{p})$

$\boldsymbol{p} \sim \text{Dir}(\boldsymbol{\alpha})$

$\boldsymbol{\alpha}$

p can be visualized as a point estimator in the 2-D simplex

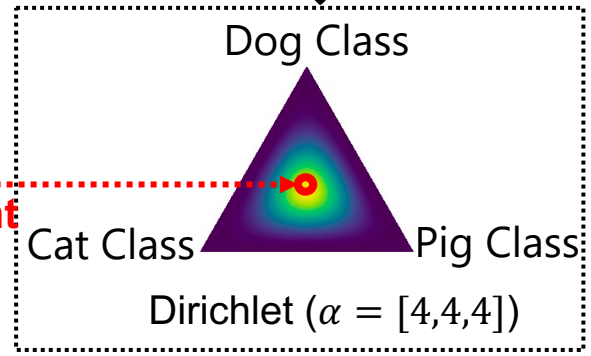
$e = [e_1, e_2, e_3]$

In the belief theory domain (e.g., subjective logic), we are interested in predicting subjective opinions

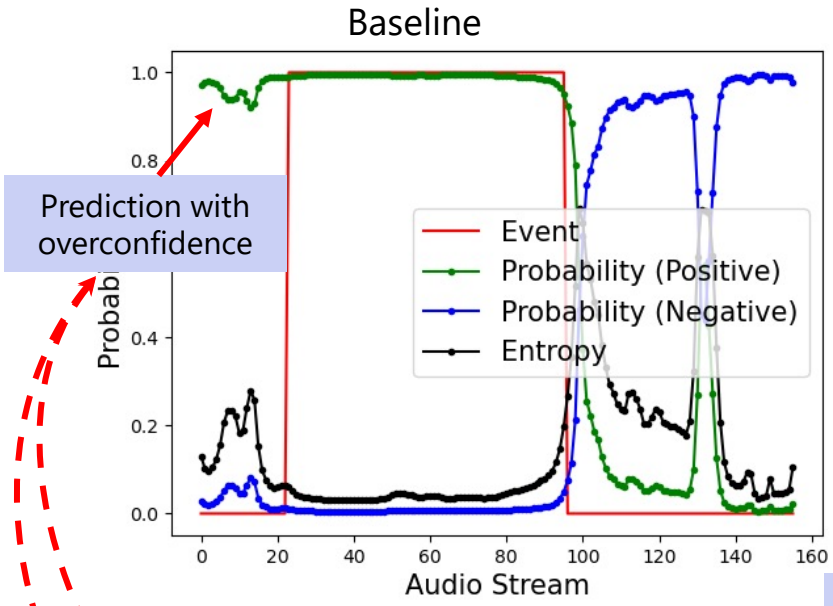
(Subjective Opinion)

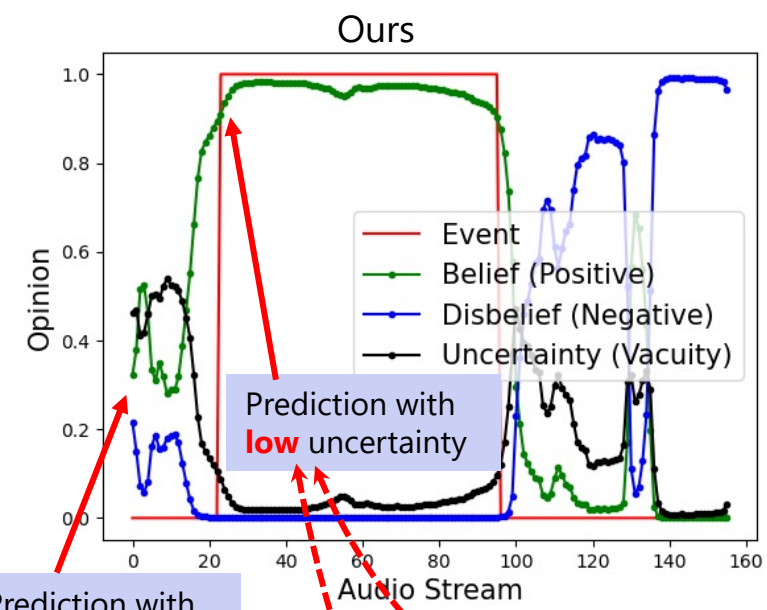
$$\omega = [b_1, b_2, b_3, u] = [0.25, 0.25, 0.25, 0.25]$$

$$\boldsymbol{\alpha} = \boldsymbol{e} + 1 = [4,4,4]$$



Advantage of Evidence-based model





A binary case with two beliefs

Prediction with High vacuity

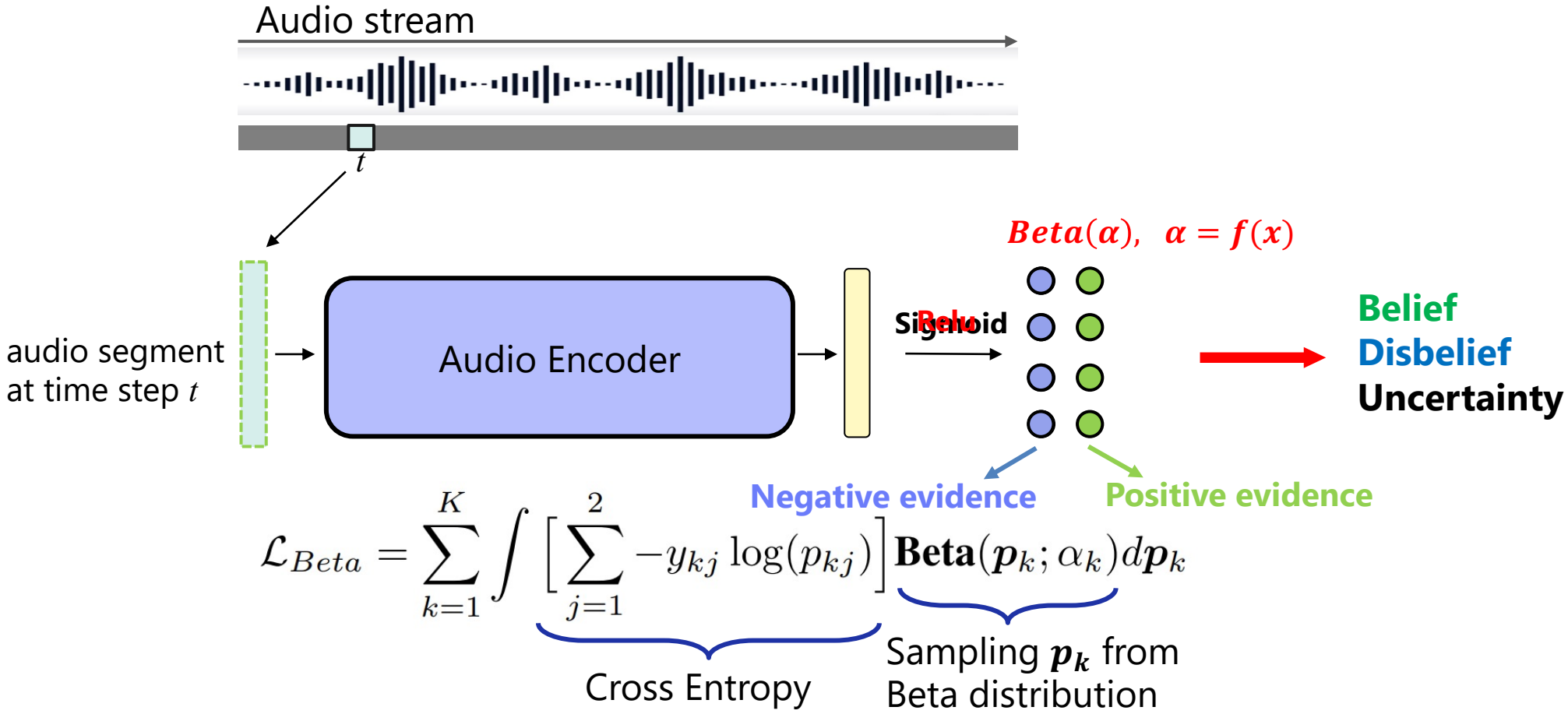
Evidence	$e = [1, 4]$
Probability	$p = [0.2, 0.8]$
Entropy	Low
Vacuity	High

Evidence	$e = [4, 200]$
Probability	$p = [0.02, 0.98]$
Entropy	Low
Vacuity	Low

The uncertainty (opinion) information is important for early detection

Evidential Sound Event Early Detection

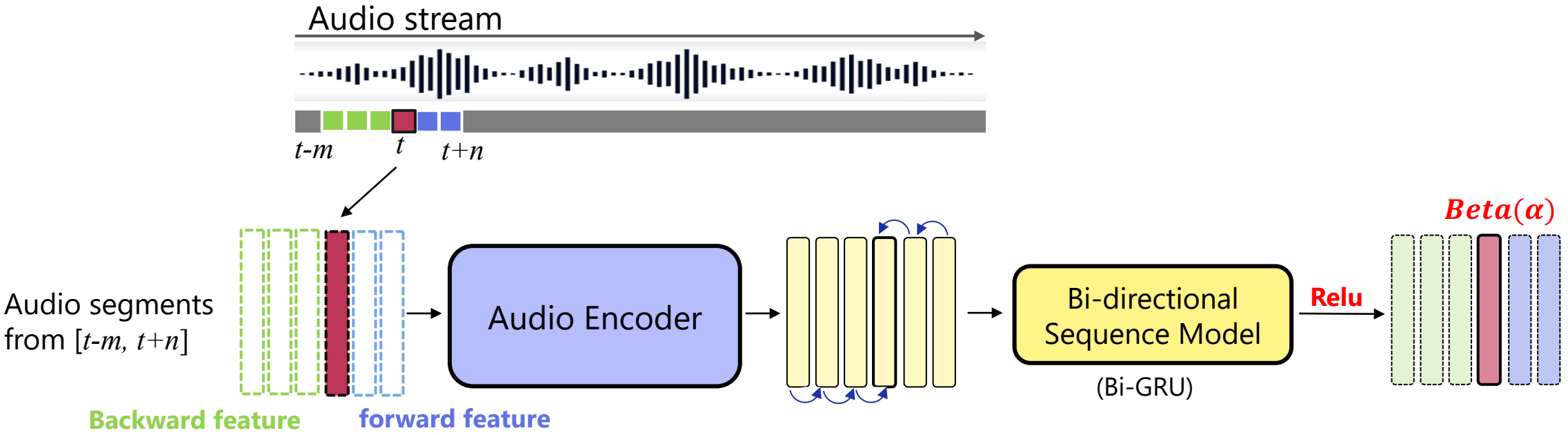
- Multi-label Evidential neural network (ML-ENN)



Provide uncertainty information for over lapping sound event (challenge I)

Evidential Sound Event Early Detection

- Bi-directional Multi-shift inference



- Using both ***forward*** and ***backward*** audio information to estimate evidence more accurately
- A balance between delay and accuracy when using ***forward*** information

Multi-shift training can overcome the challenge II of small detection window

Results

- DESED Dataset: composed of audio clips recorded in domestic environments (Focus on 10 classes such as Speech, Running water and Dishes)
- Metrics: Delay & Event F1

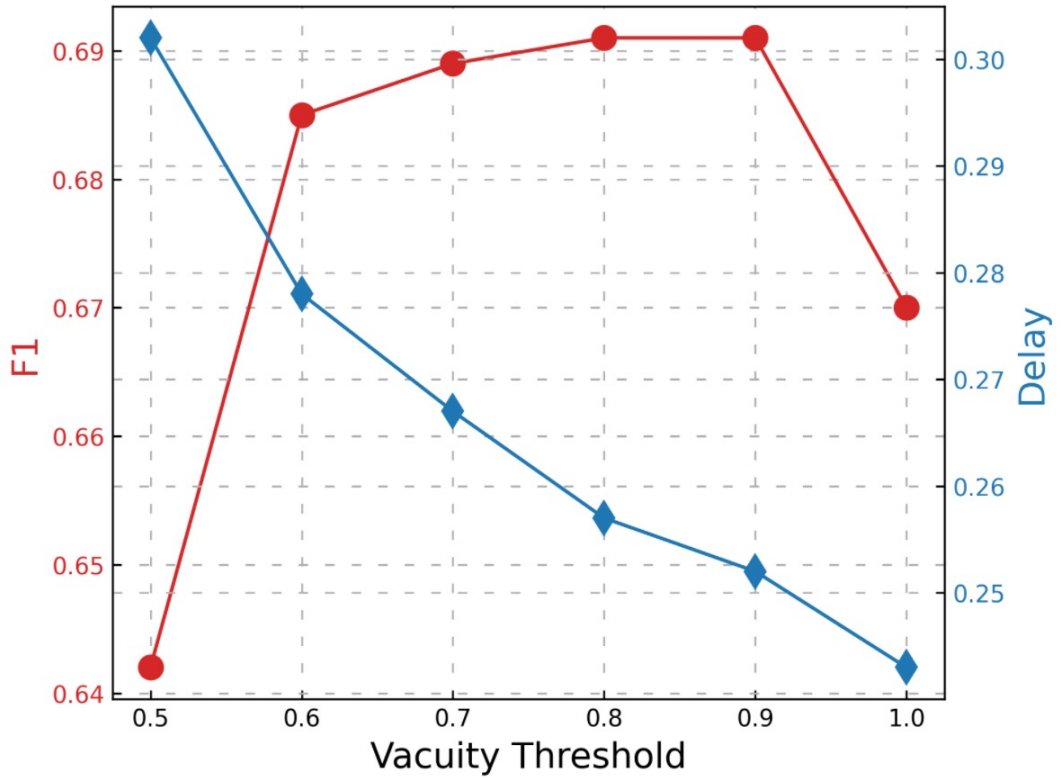
	Delay (seconds)	Event F1
Conformer	0.372	0.639
CRNN	0.284	0.687
Ours (delay)	**0.247**	0.670
Ours (balanced)	0.252	0.691
Ours (accuracy)	0.310	**0.725**

- **Inference time (Challenge III):** 5ms << 60ms (detection window)

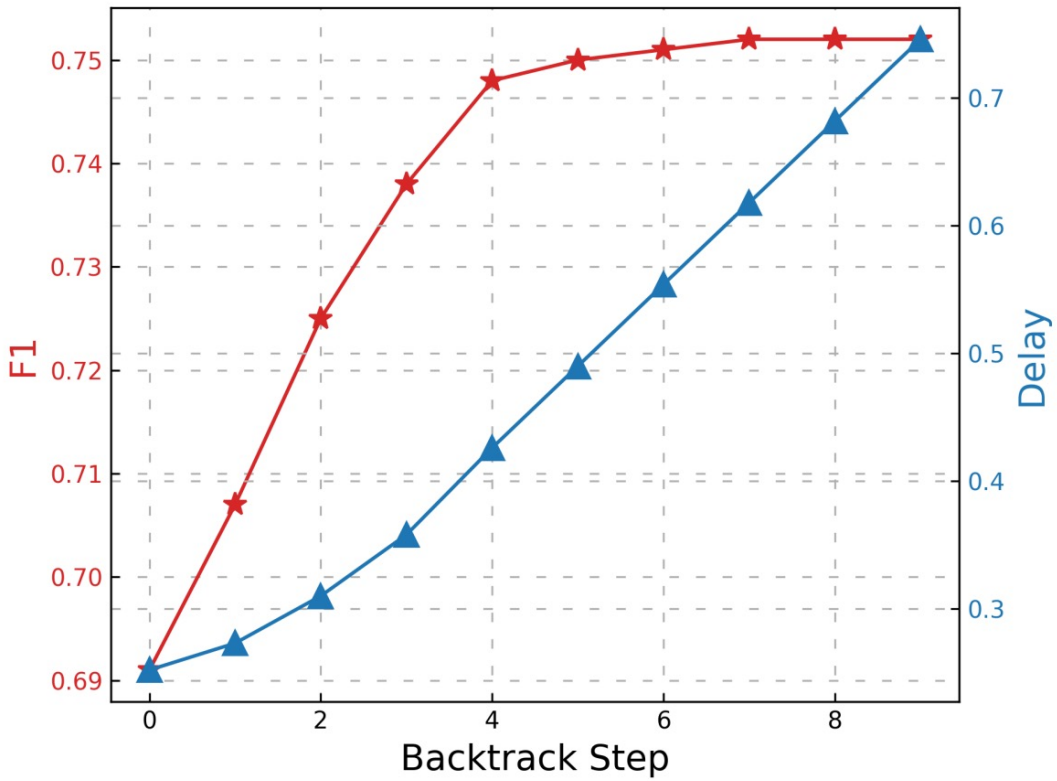
Our Method can early detect sound events and satisfy the real-time requirement.

Results

- Effect of vacuity threshold



- Multi-shift inference



A balance between delay and accuracy when using forward information

Future Work

- Consider sequential uncertainty at the inference stage.
- Consider multi-label dependency for sound event overlapping.
- Apply our model in more real-world datasets.

Sequential Uncertainty and large dataset


Q&A