BlackFeather: A framework for Background Noise Forensics

By:
Qi Li (University of Guelph), Giuliano Sovernigo (University of Guelph), and Xiaodong Lin (University of Guelph)

From the proceedings of
The Digital Forensic Research Conference
DFRWS USA 2022
July 11-14, 2022

DFRWS is dedicated to the sharing of knowledge and ideas about digital forensics research. Ever since it organized the first open workshop devoted to digital forensics in 2001, DFRWS continues to bring academics and practitioners together in an informal environment.
As a non-profit, volunteer organization, DFRWS sponsors technical working groups, annual conferences and challenges to help drive the direction of research and development.
https://dfrws.org

DFRWS 2022 VIRTUAL USA

BlackFeather: A Framework for Background Noise Forensics

Qi Li, Giuliano Sovernigo, Xiaodong Lin

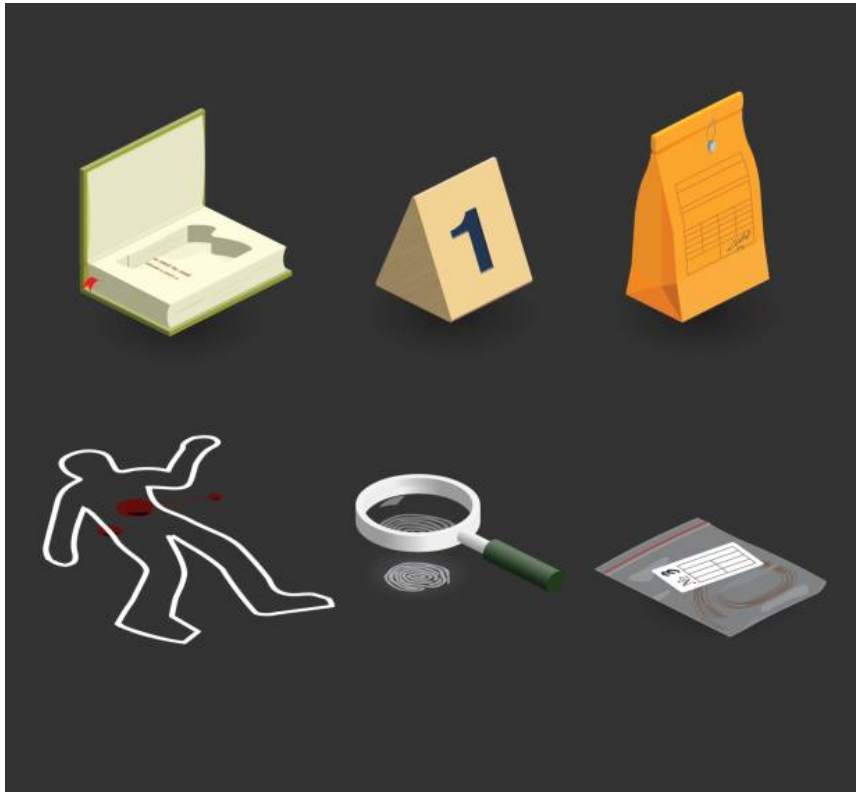
School of Computer Science

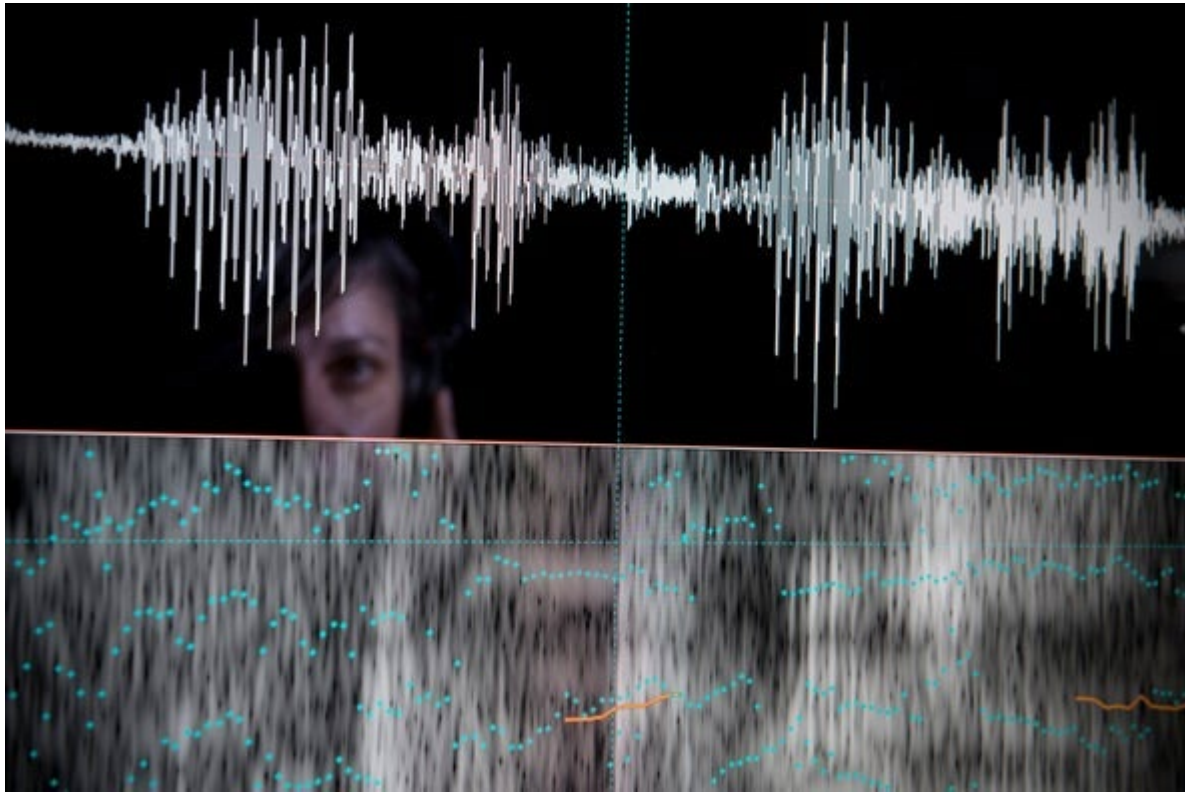
University of Guelph

Content

- Background
- Related work
- Framework
- Experiments results
- Conclusion

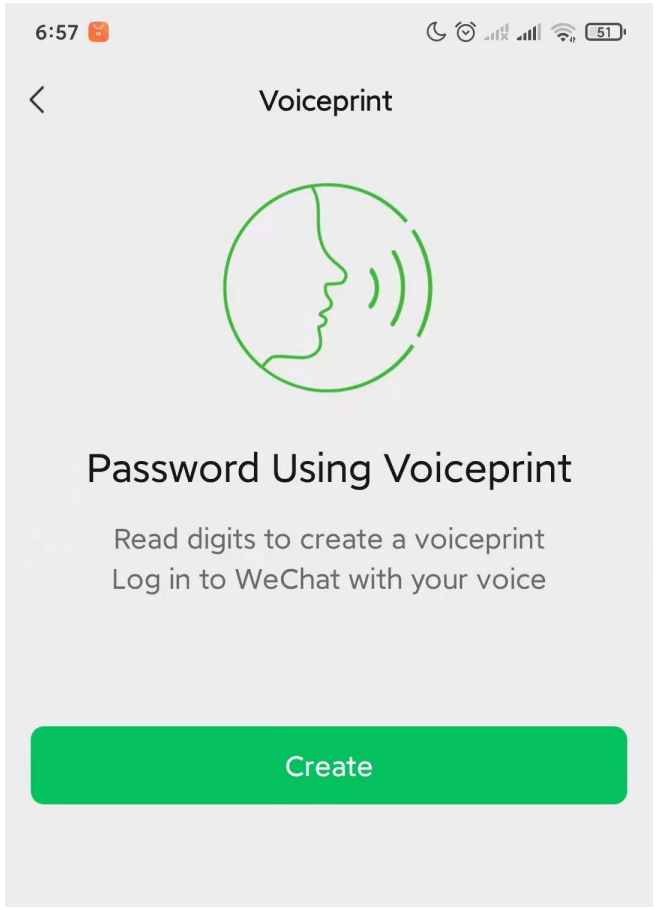
Background

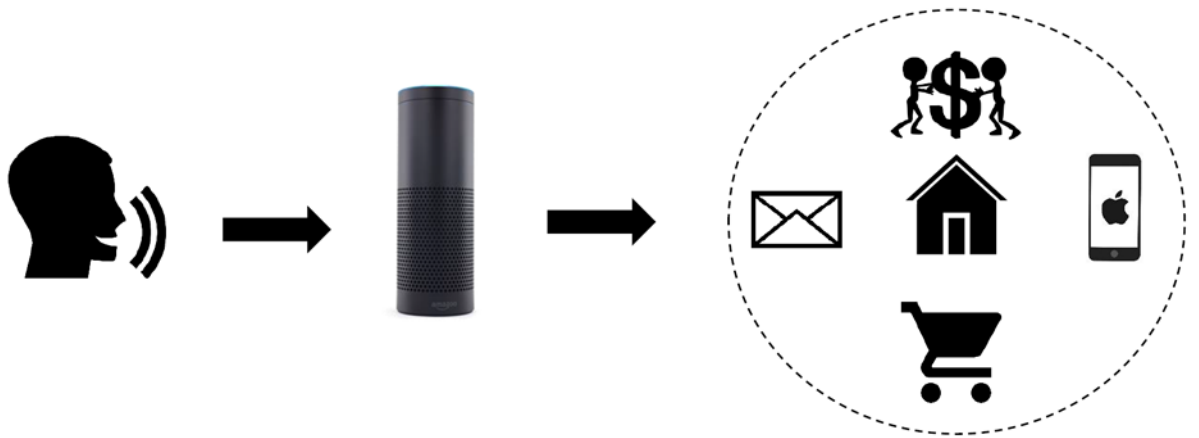




Background

Hey Siri





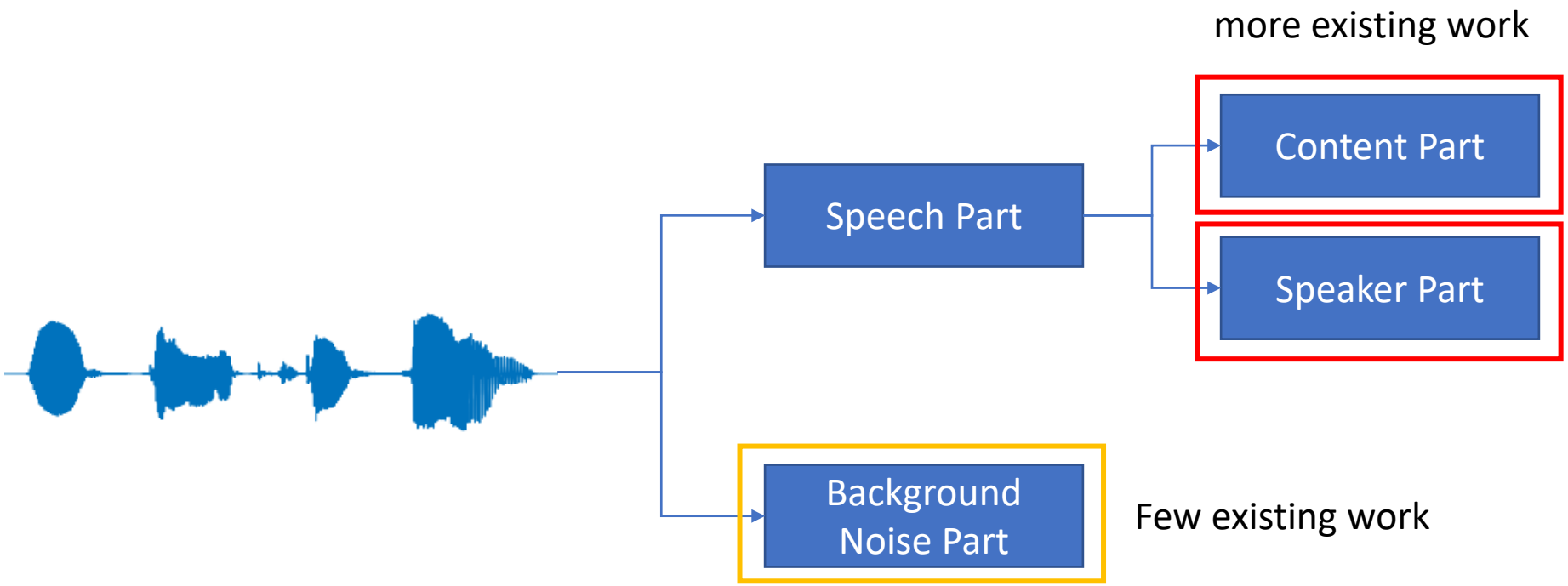
Smart home voice assistant use to grow 1000% by 2023

Thursday, 28 June, 2018

Share Tweet Share Email

quite easy to collect voice!

Background


more existing work
Content Part
Speech Part
Speaker Part
Background
Noise Part
Few existing work

Background

Bus engine

Dog, bird, cat bark

Background

- To replace human-being
- To get high accuracy
- To help forensic investigators
- To provide evidence and clue

A blind Sherlock Holmes: Fighting crime with acute listening

By Dan Bilefsky

Oct. 29, 2007

Our contribution

A. Examine the extent to which overlapping, raw background noise can be used to determine the probable location of the recording.

B. Present the first overall comprehensive scheme for background noise forensics.

C. The scheme is divided into two steps. In the first step, the background noise is extracted, separated and identified. In the second step, the identification information is modeled or used by experts to get the environment information.

1

Our contribution

A. This paper is the first forensic work for audio containing multiple, overlapping background noise sources as well as speakers' voices.
B. We introduce the noise extractor and the multi-speaker separation in speaker recognition to the background noise and achieve good results.

2

Our contribution

A. The top-K selection for classification, the proposed combined datasets such as MixEsc50.

B. Detailed experiments show that our scheme is effective, fast and accurate.

3

Related work

- Digital audio forensics using background noise (Ikram and Malik [1], 2010)
  - Used spectral subtraction
  - Noise Estimation based on GA
  - No classification of background noise
  - Use five classifications to conduct the experiments
  - Used to help authenticate recording originality
- Background Sound Classification in Speech Audio Segments (Singh and Joshi [2], 2020)
  - Background sound classification
  - Using convolutional neural networks and a youtube dataset (but it just has 10 classifications).
  - Could not analyze multi-noise situation
- Background/Foreground Classification (Thorogood et al. [3], 2015)
  - background/foreground sound classification
  - Important note: how does clip duration (called “window” in the paper) affect accuracy
  - Model performed in 80-95% range for accuracy (notably similar to human performance)
  - Not classify background noise for specific background noise

[1] Sohaib Ikram, Hafiz Malik. Digital audio forensics using background noise. ICME 2010: 106-110
[2] Janvijay Singh, Raviraj Joshi. Background Sound Classification in Speech Audio Segments. SpeD 2019: 1-6
[3] Miles Thorogood, Jianyu Fan, Philippe Pasquier. BF-Classifier: Background/Foreground Classification and Segmentation of Soundscape Recordings. Audio Mostly Conference 2015CCF none: 32:1-32:6

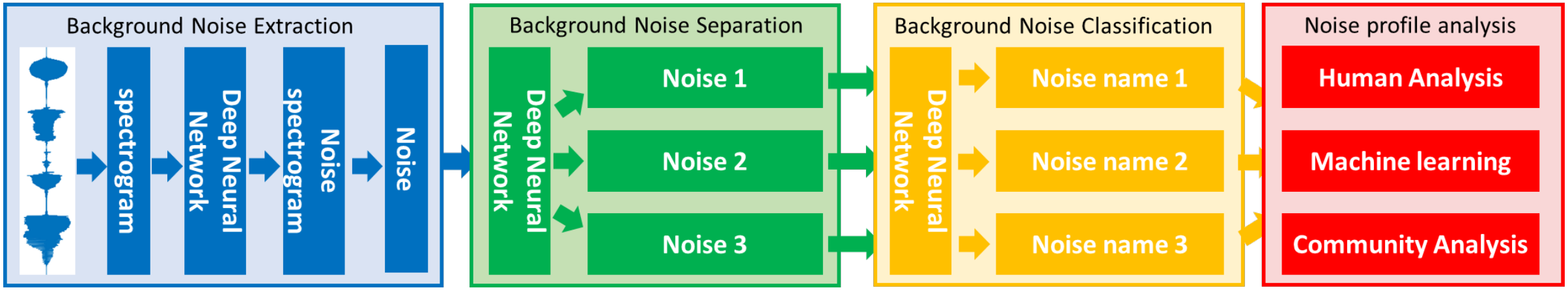
Related work

- Digital Audio Forensics: Microphone and Environment Classification Using Deep Learning (Qamhan et al. 2021)
  - 3 environments, 4 classes of recording devices
  - No human speech
- Detection and Classification of Acoustic Scenes and Events, DCASE task-acoustic-scene-classification(2016 - now)
  - 10 different acoustic scenarios
  - No interpretability
  - No human speech

[4] M. A. Qamhan, H. Altaheri, A. H. Meftah, G. Muhammad and Y. A. Alotaibi, "Digital Audio Forensics: Microphone and Environment Classification Using Deep Learning," in IEEE Access, vol. 9, pp. 62719-62733, 2021, doi: 10.1109/ACCESS.2021.3073786.
[5] http://dcase.community/challenge2020/task-acoustic-scene-classification

Proposed framework - Overview

- 2 steps: noise processing, noise profile analysis(location processing)
- 3 parts in step1



$$b = Extract(x),$$

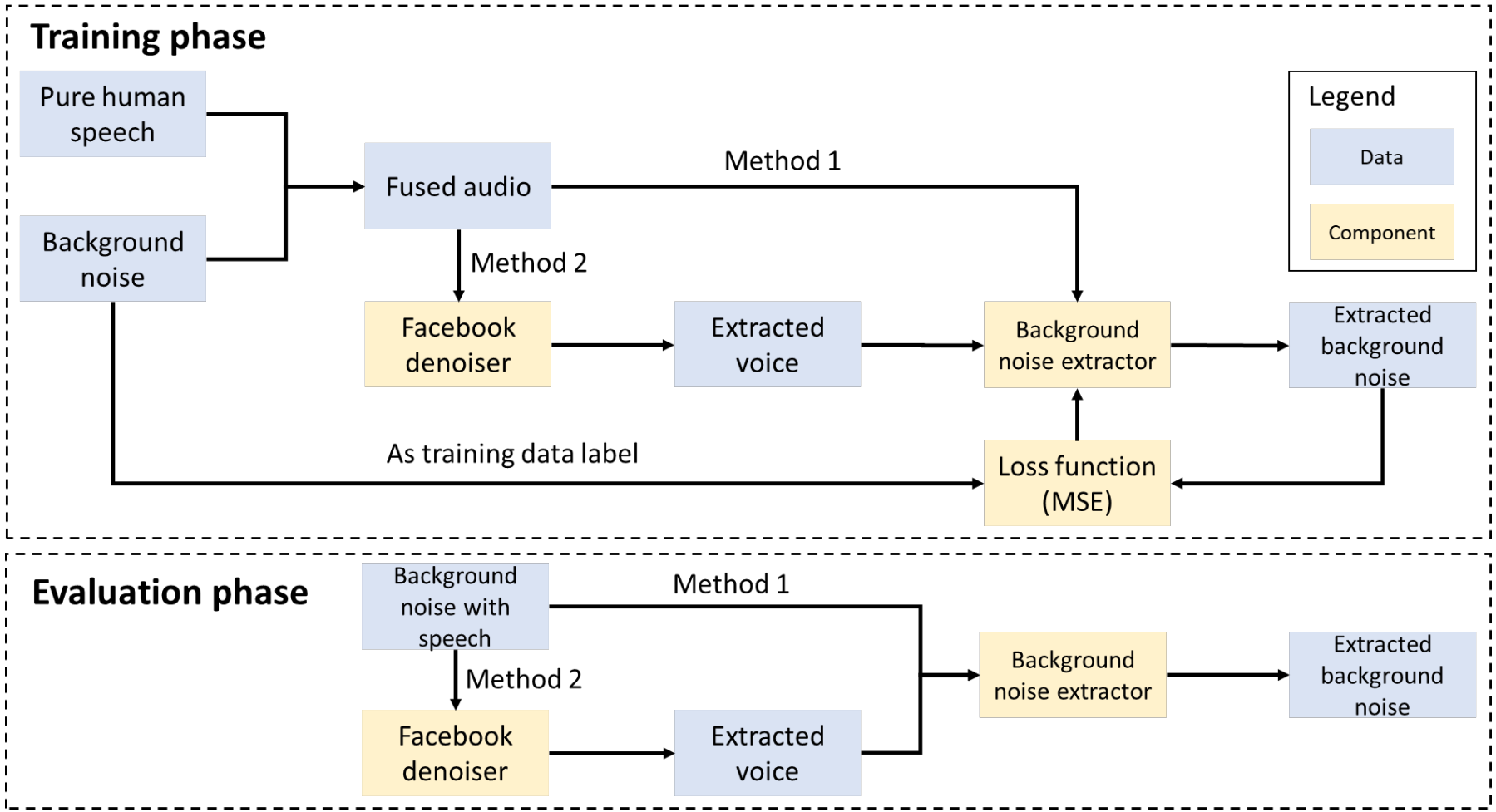
$$[b_1, \cdots, b_k] = Separate(b),$$

$$[n_1, \cdots, n_k] = Classify([b_1, \cdots, b_k]),$$

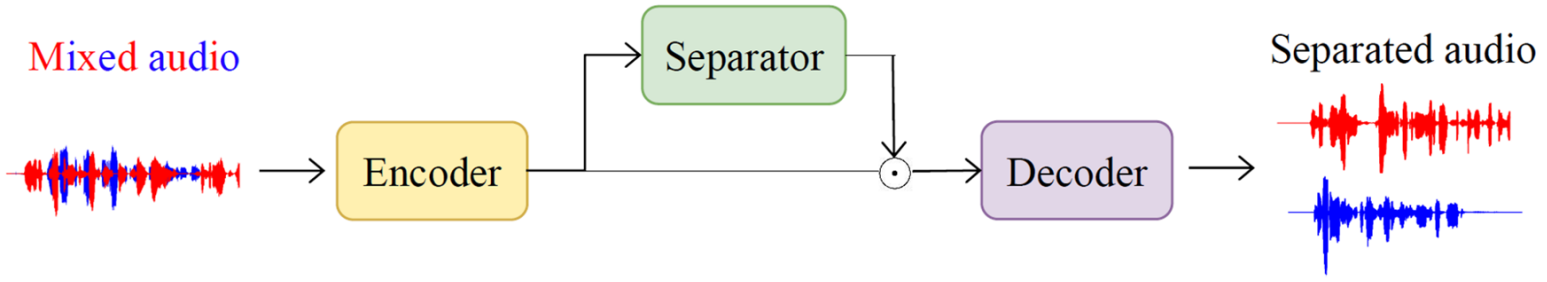
$$d = Analyze([n_1, \cdots, n_k]).$$

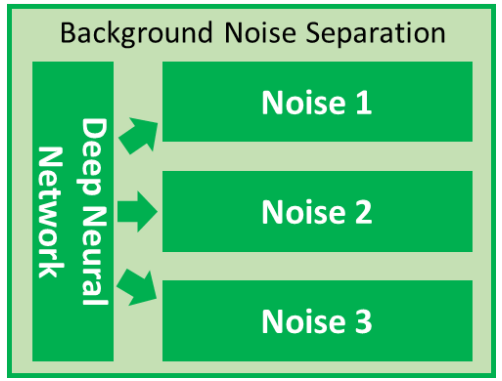
Framework – Background Noise Extraction

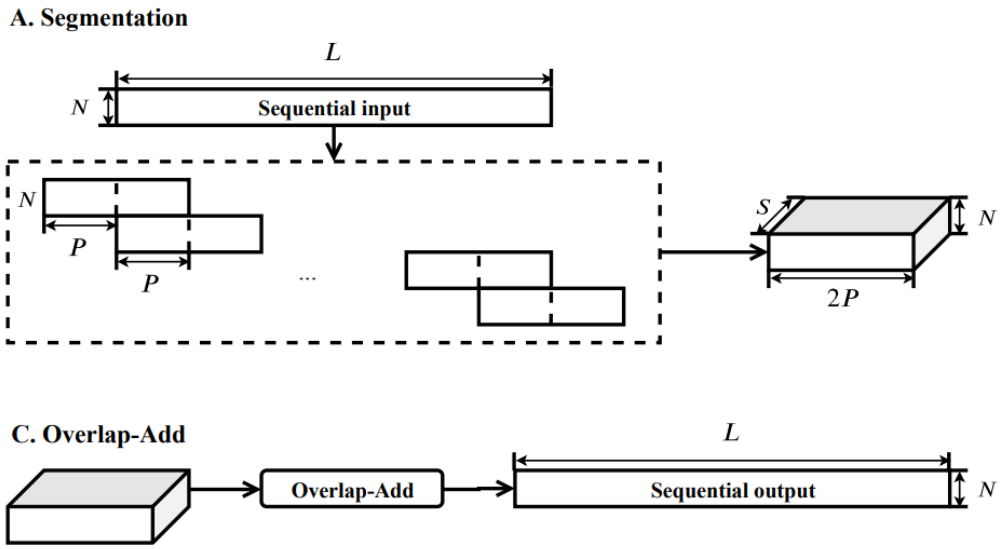
- Using Transformer directly

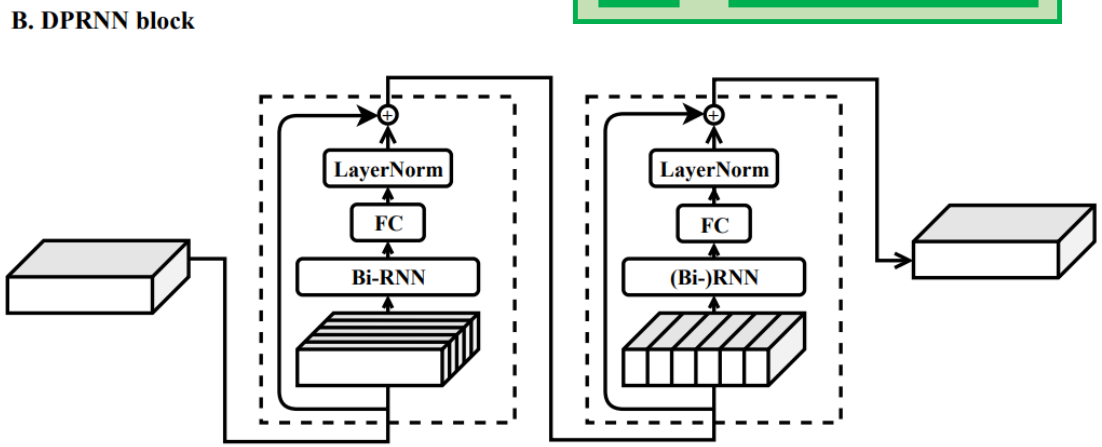


Framework – Background Noise Separation







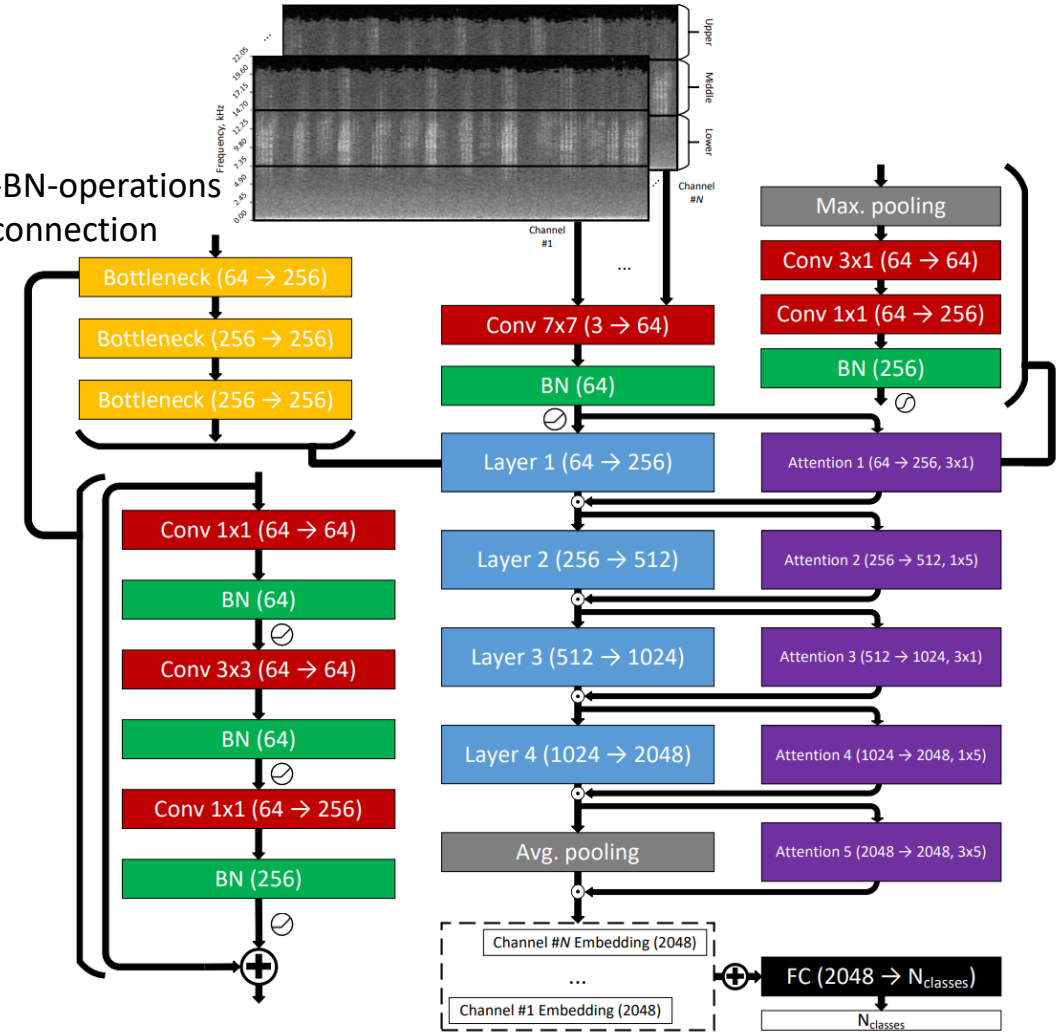


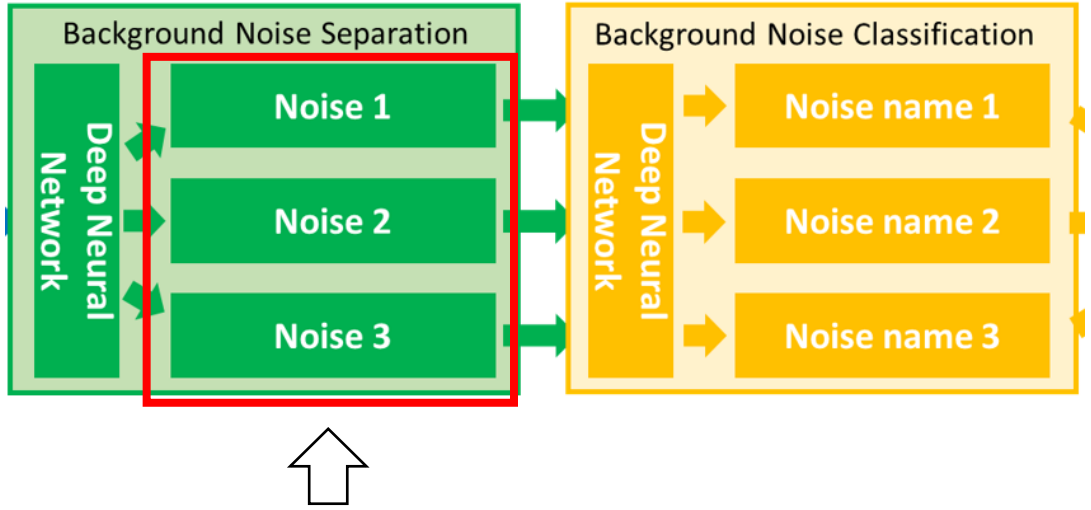
https://github.com/JusperLee/Dual-path-RNN-Pytorch

Luo Y, Chen Z, Yoshioka T. Dual-path RNN: efficient long sequence modeling for time-domain single-channel speech separation[J]. arXiv preprint arXiv:1910.06379, 2019.

Framework - classification

Conv-BN-operations
Skip-connection

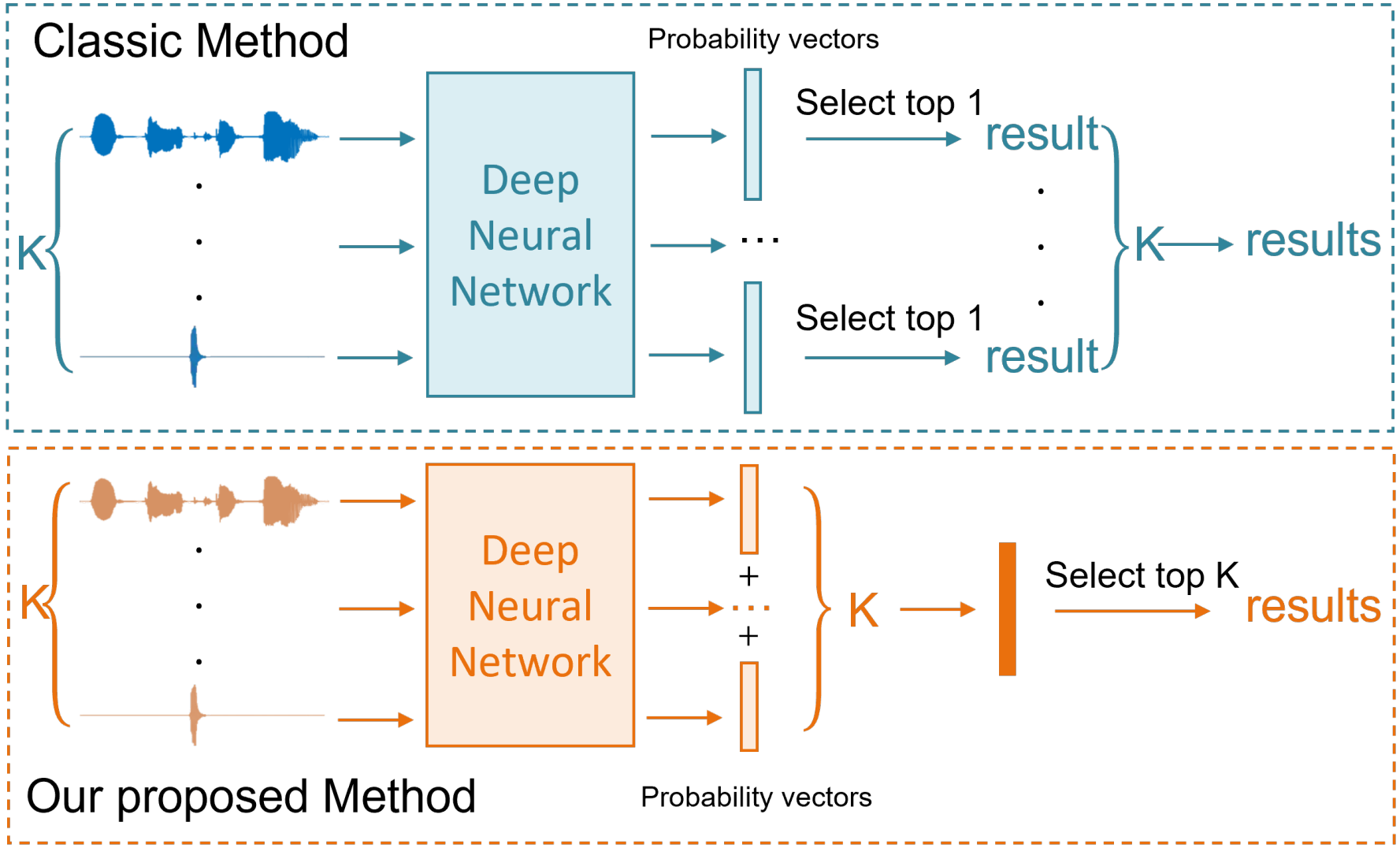




Any limitations?

[1] Andrey Guzhov, Federico Raue, Jörn Hees, Andreas Dengel, ESResNe(X)t-fbsp: Learning Robust Time-Frequency Transformation of Audio. IJCNN, 1-8

Framework – classification – new method



Use leaked information!

Framework - Classification – new method

Method 1

$$F = \sum_{i}^{K} f_i.$$

$$P = softmax(F).$$

Method 2

$$p_i = softmax(f_i), i \in \{1, 2, \cdots, K\}.$$

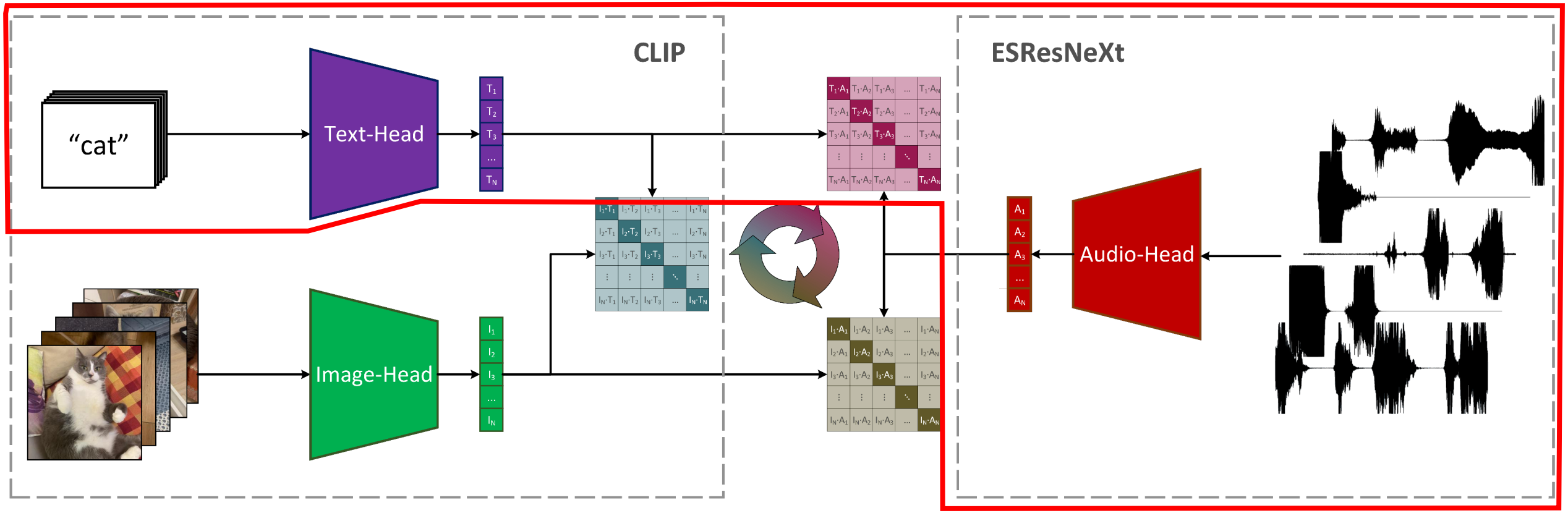
$$P = \sum_{i}^{K} p_i.$$

Assume x, y are two vectors, x_i, y_i are the i-th value of x, y. If $y = softmax(x)$, then $\forall i$, $y_i = \frac{e^{x_i}}{\sum_j e^{x_j}}$.

Framework – classification

- Customized classification – AudioCLIP[1]

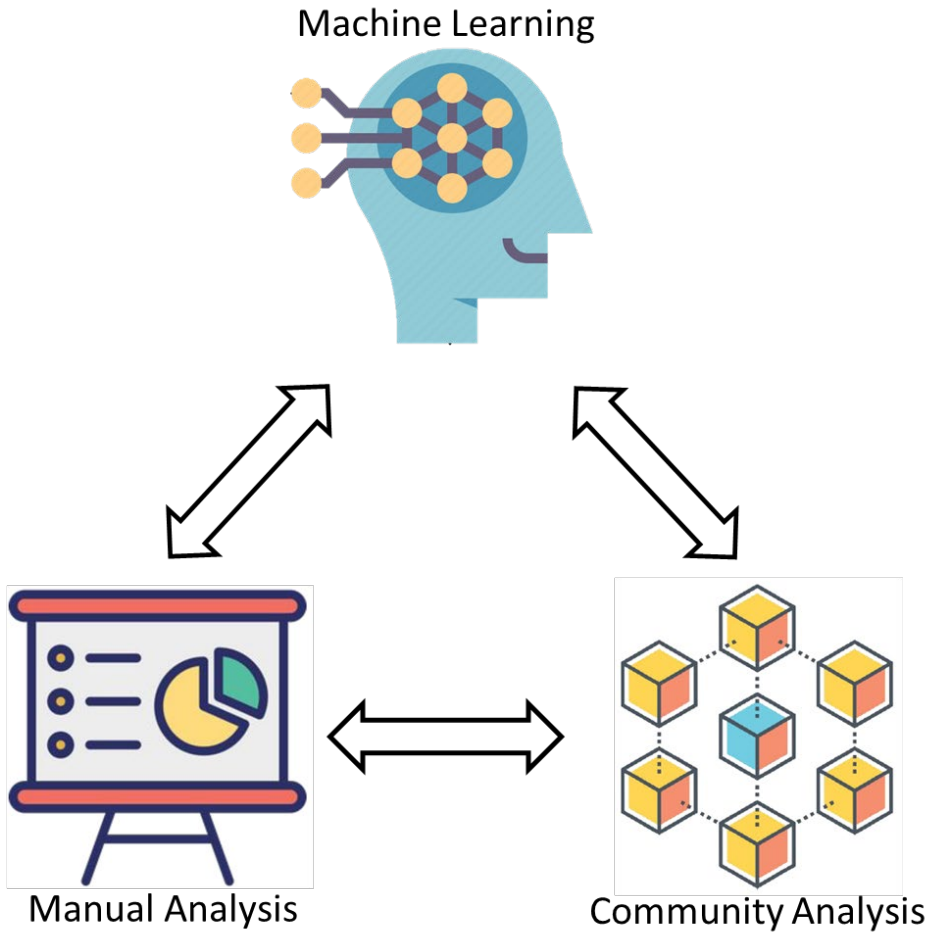
Achieve flexible classification numbers



[1] Andrey Guzhov, Federico Raue, Jörn Hees, Andreas Dengel: Audioclip: Extending Clip to Image, Text and Audio. ICASSP 2022

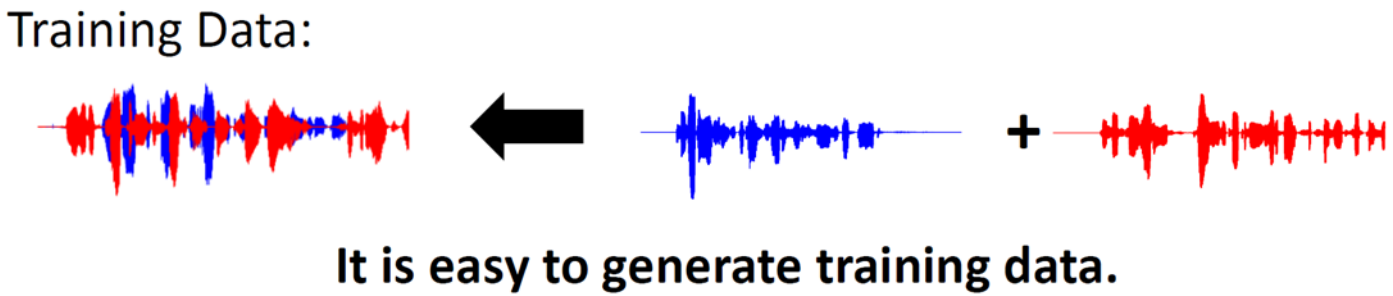
Framework – noise profile and location analysis

- Naïve Bayesian model
- Dataset: 1721 pairs
  - 1376 pairs training
  - 345 pairs test
  - 95.8% accuracy
  - 37 unique classification
  - 112 kinds of noises



Experiments Results

- Windows PC with NVIDIA p5000 GPU
- Dataset
  - Audioset
  - Librispeech
  - ESC-50
    - MixESC50
    - The k voices are extracted from 50 categories and intercepted for 10s each to get $50k$ voices. Finally, I cross-mix these voices to get $\sum_{i=1}^{49} ik^2 = 1225k^2$ voices.
    - $k = 5$, 80% for training, 20% for testing
- Learning rate:1e-4



Experiments Results

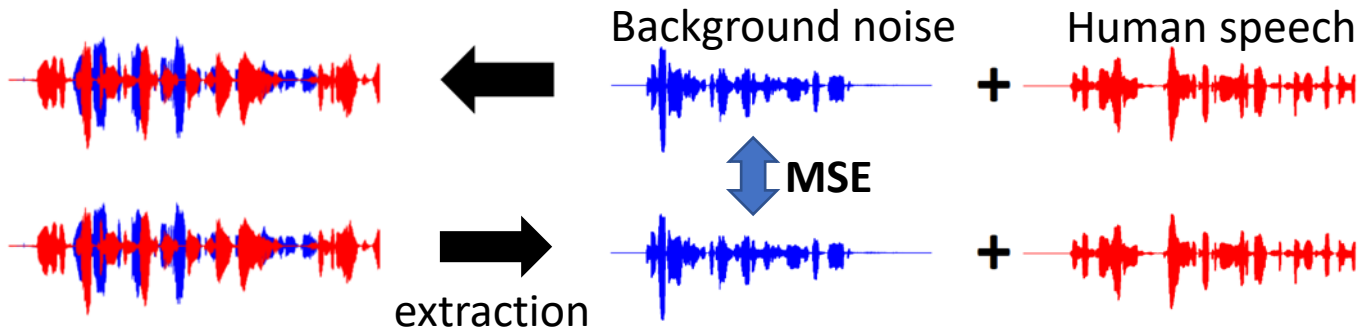


Table 1: Background noise extraction results

MSE	CombinedAudioSet	CombinedESC50
Without Denoiser	2.28	2.19
With Denoiser	1.80	1.68
Original Distance	3062.74	2944.50

Experiments Results

$$SISDR = 10\log_{10}\frac{\|X_T\|^2}{\|X_E\|^2}$$

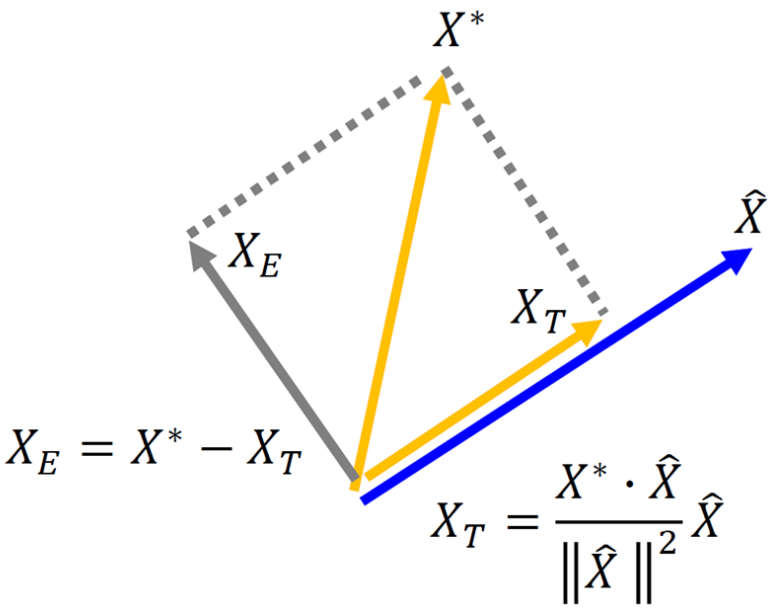
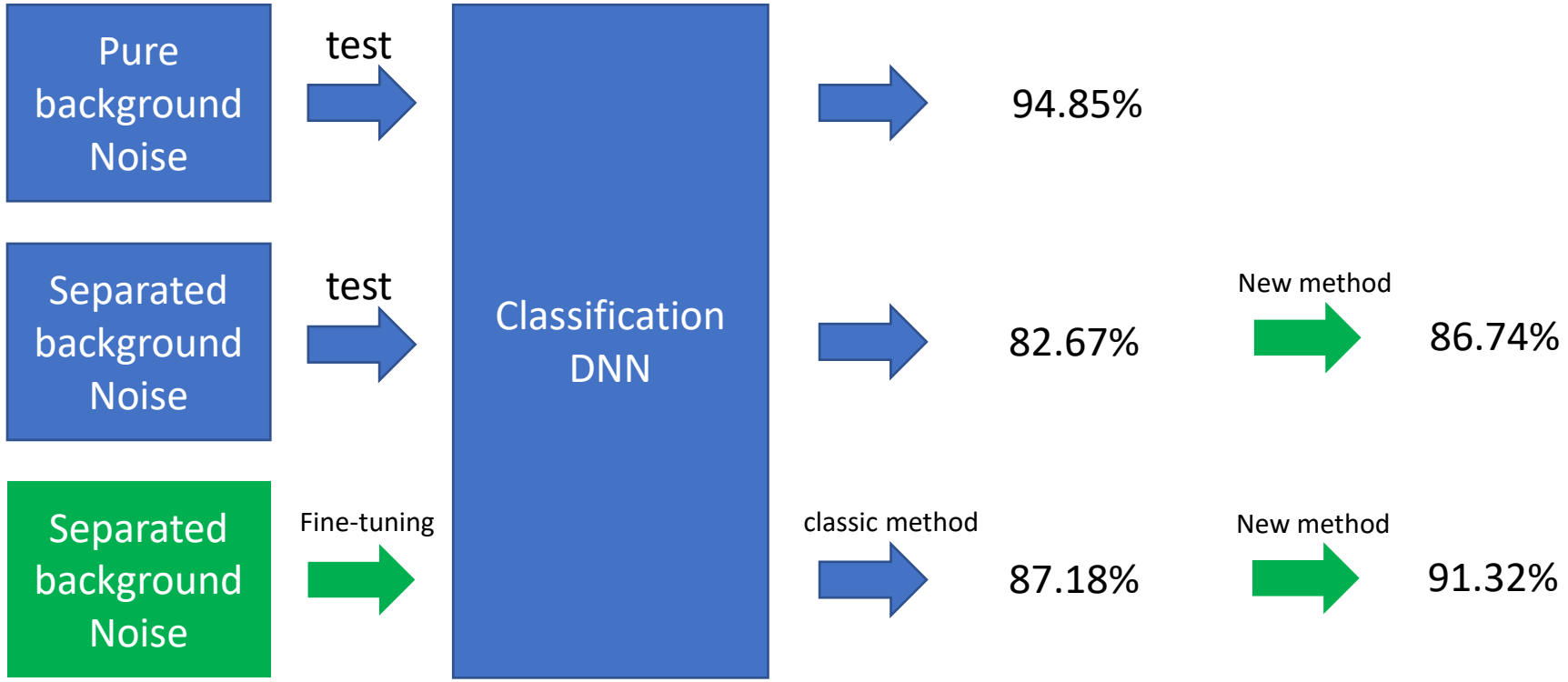


Table 2: Background noise separation results

	SI-SDR
Human speech [8]	19.10
Background noise with training	7.64
Background noise without training	-25.4

Experiments Results



Experiments Results

Table 3: Classification results of Method 1 and Method 2 in MixESC50

	Top 1	Top 3	Top 5
Old method	82.67%	93.45%	97.53%
Method 1	84.91%	94.39%	97.84%
Method 2	86.74%	95.16%	98.22%

Table 4: Full pipeline classification results of Method 1 and Method 2

	Old method	Method 1	Method 2
Pure single noise	94.85%	–	–
Separated noise	82.67%	84.91%	86.74%
Fine-tuned model	87.18%	90.15%	91.32%

Table 5: Classification results of BlackFeather and Original AudioClip[14]

	BlackFeather	AudioClip[14]
Accuracy	91.32%	46.3%

Select top2 as results

Experiments Results

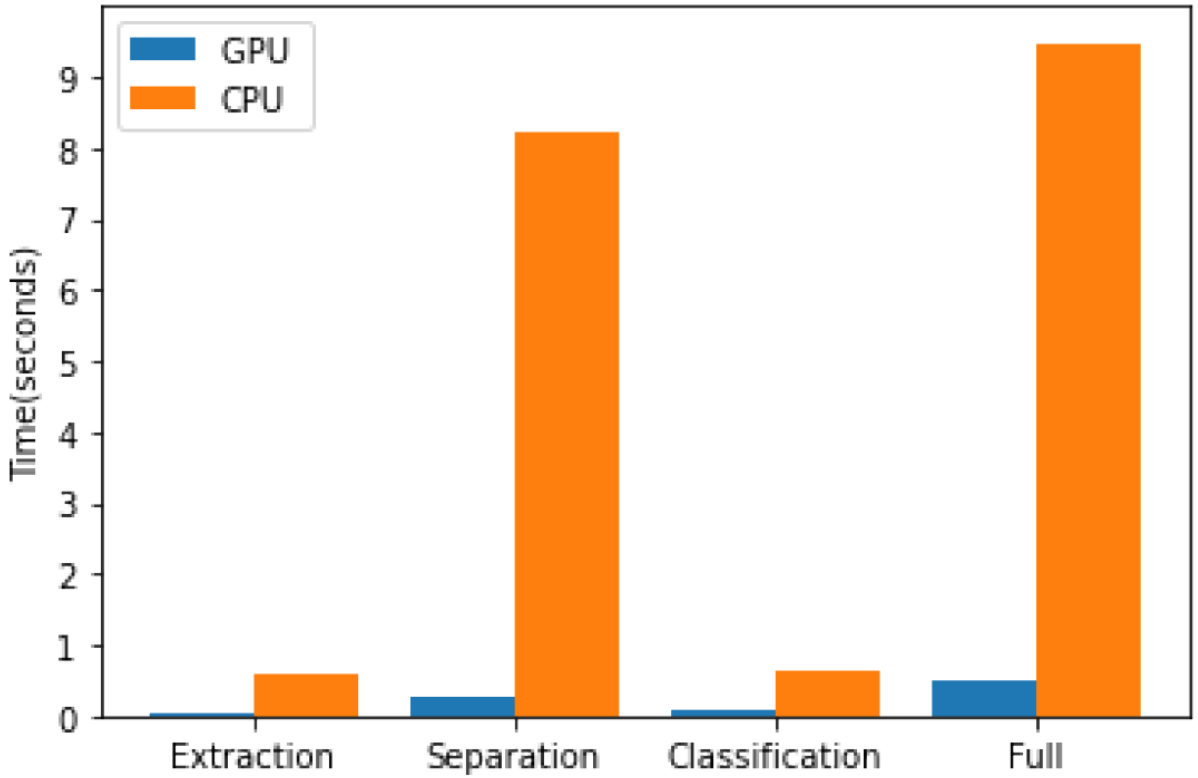


Figure 5: Time overhead of BlackFeather.

Application

- Environmental Inference
  - Extract sensitive information
  - suggest environments where the sound could not have been recorded
- Temporal Inference
  - narrow recording time down to minutes
  - disproving or confirming alibis

Conclusion

- We propose **a systematic framework** to address the challenge of **automated background noise forensics**. And in the paper, we use **location information** as an example to achieve good results.
- We have carefully designed, constructed and experimented for each module of the proposed framework. This research can be **a good replacement** for existing manual or **assisted** manual forensic work.
- Our work can also **be extended** and **fine-tuned** to accommodate forensic tasks **beyond environment**.

Q & A