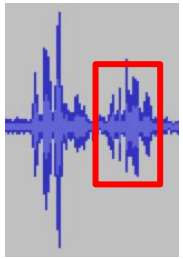
Audio Content Replication Detection

Yutong Wen

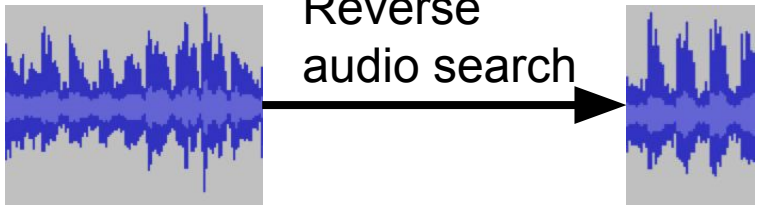
Task and Motivation

Given two audio clips, we want to know if one is a modified version of the other.

Violating content



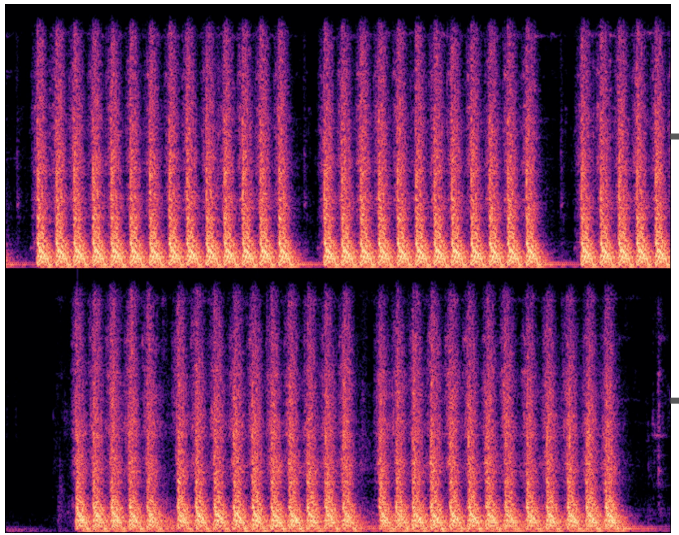
Copyright protected



Different Kinds of Copies

There are mainly two types of copies:

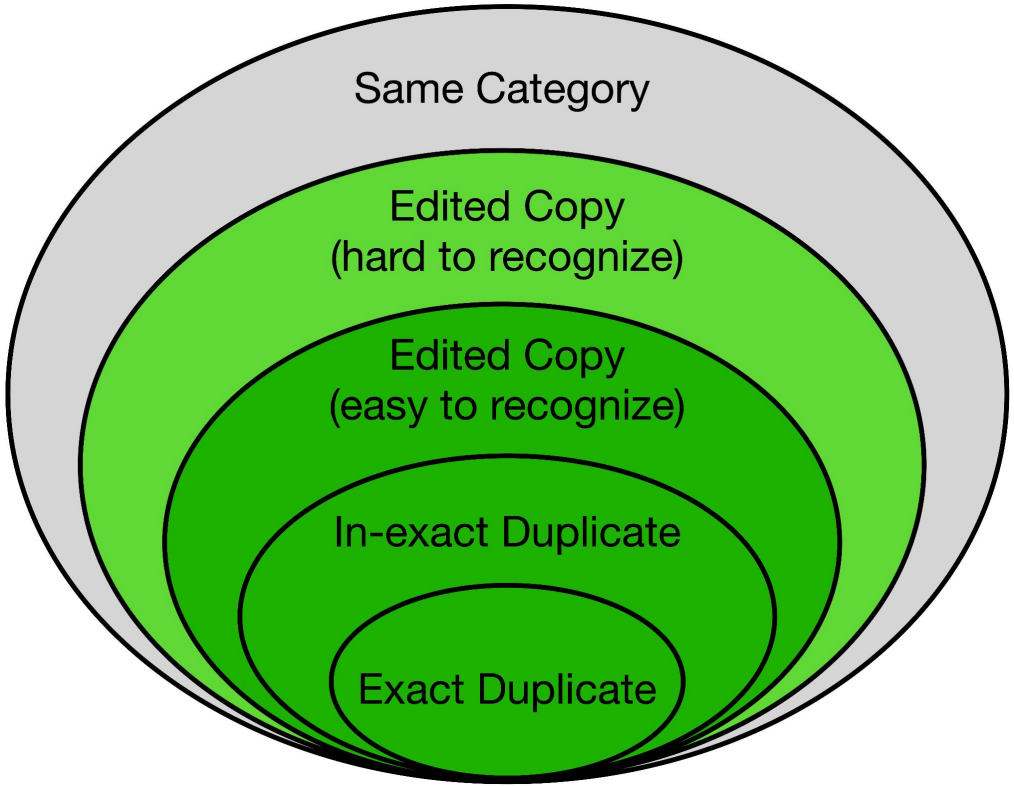
- Two audio clips share high perceptual similarity
- One audio clip is a modified version of the other



High perceptual similarity.
What if speed up 5x?

Two ideas are used in different practical scenarios.

Audio Pair Similarity Level



We use commonly used audio transformations to create these edited copies and in-exact duplicates.

Audio Similarity Dataset

We construct a dataset to mimic real-world scenarios.

- Reference set
- Query set
  - True queries
  - Distractors
- Training set

We only use the Epidemic Sound dataset currently

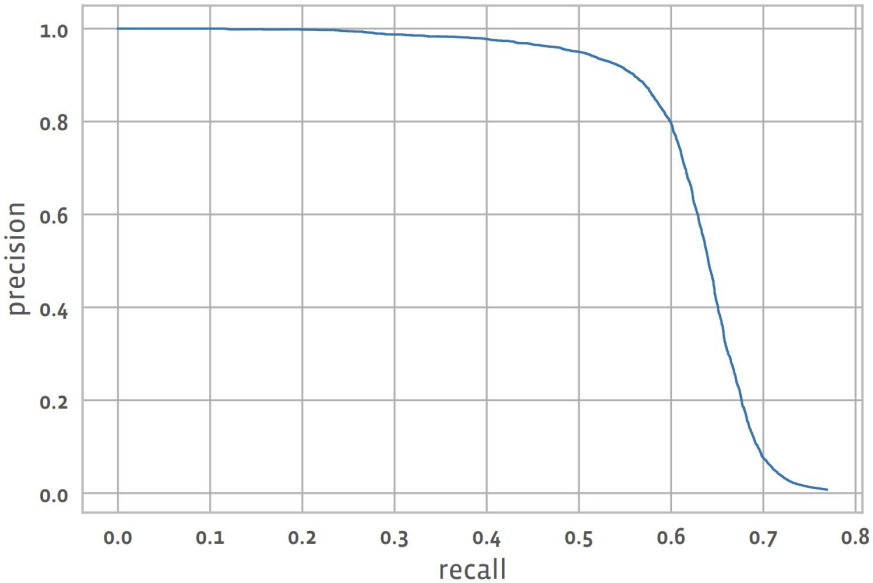
- Reference set (37072)
- Query set (1852)
  - True queries (370)
  - Distractors (1482)
- Training set (37072)

16kHz, mono, 4 seconds

Evaluation Metric

The algorithm is asked to output identified pairs of copies with a confidence level.

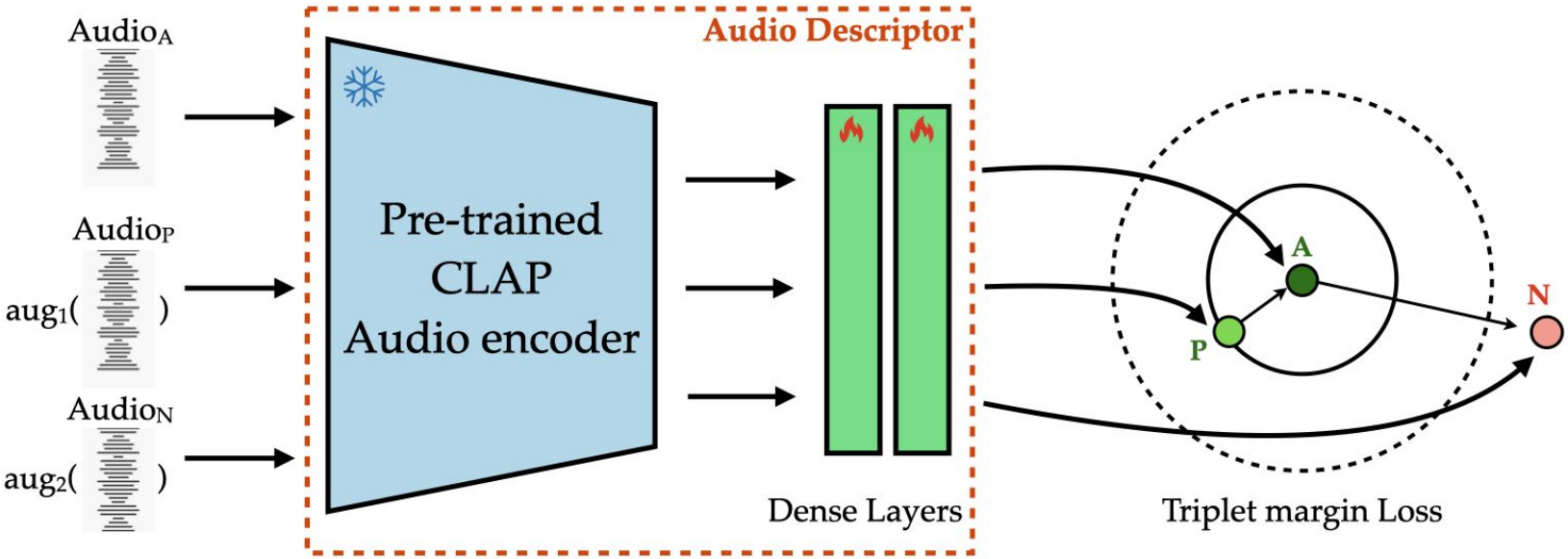
Micro average precision, the area under the precision-recall curve.



$$\text{Precision} = \frac{TP}{TP + FP}$$

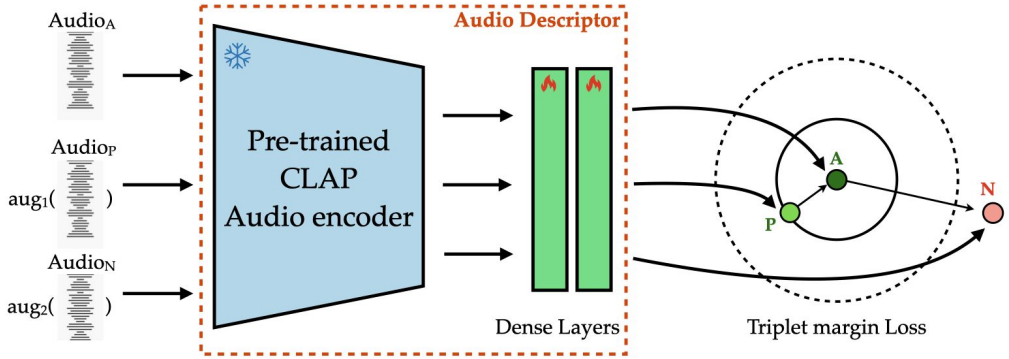
$$\text{Recall} = \frac{TP}{TP + FN}$$

Baseline Method



We freeze the pre-trained CLAP audio encoder and fine-tune the audio descriptor with the triplet margin loss.

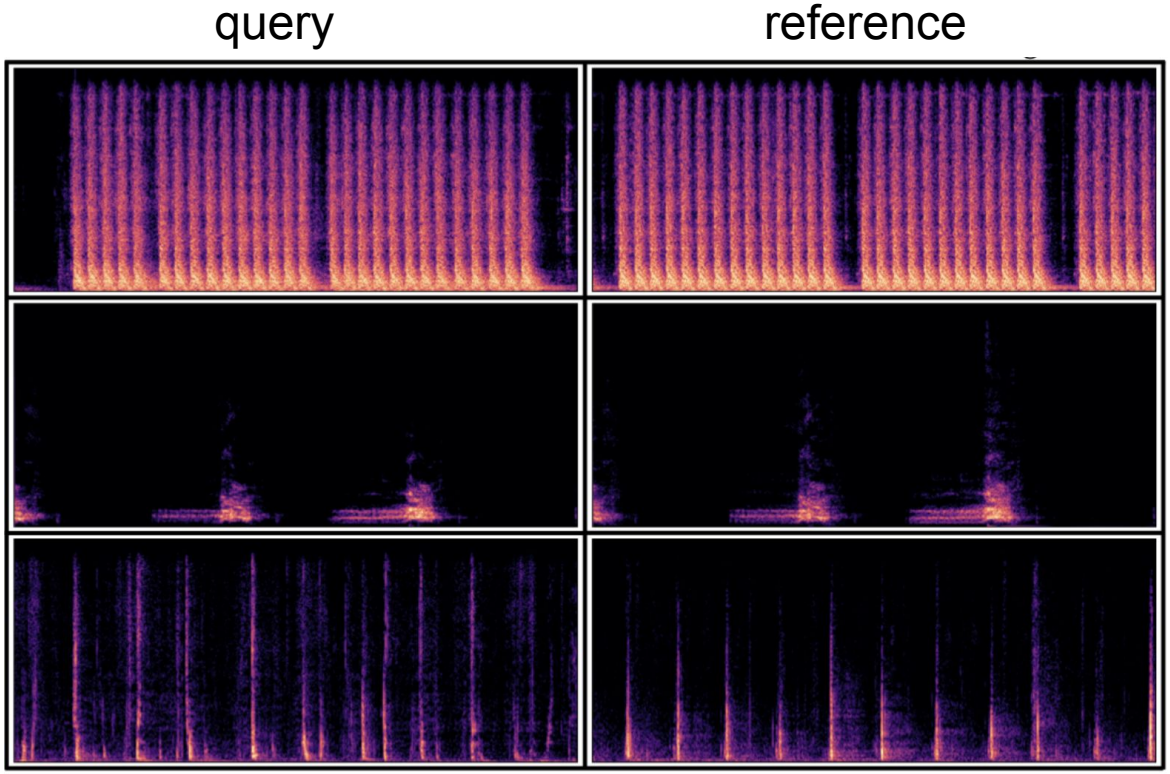
Baseline Method



- Anchor
- Positive: augmented anchor
- Negative: augmented sample other than anchor

The data augmentation includes random injection of Gaussian noise, amplitude scaling, and temporal shifting.

Examples



Conclusion and Future Directions

- Introduce the task of audio copy detection, and propose an audio similarity dataset to tackle and assess this task;
- introduce a baseline method to do this task;
- Future directions lie in selecting appropriate range of audio transformation parameters to tailor the dataset to real-world scenarios, and at the same time has reasonable difficulty.