



Engineering and
Physical Sciences
Research Council

Leveraging Label Hierarchies for Few-shot Everyday Sound Recognition

Author(s):

Jinhua Liang

Huy Phan

Emmanouil Benetos



centre for digital music

By Jinhua Liang

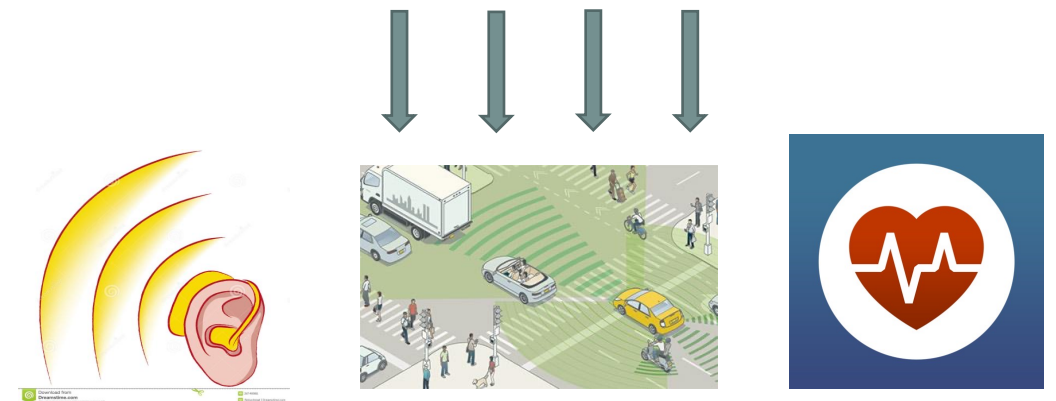
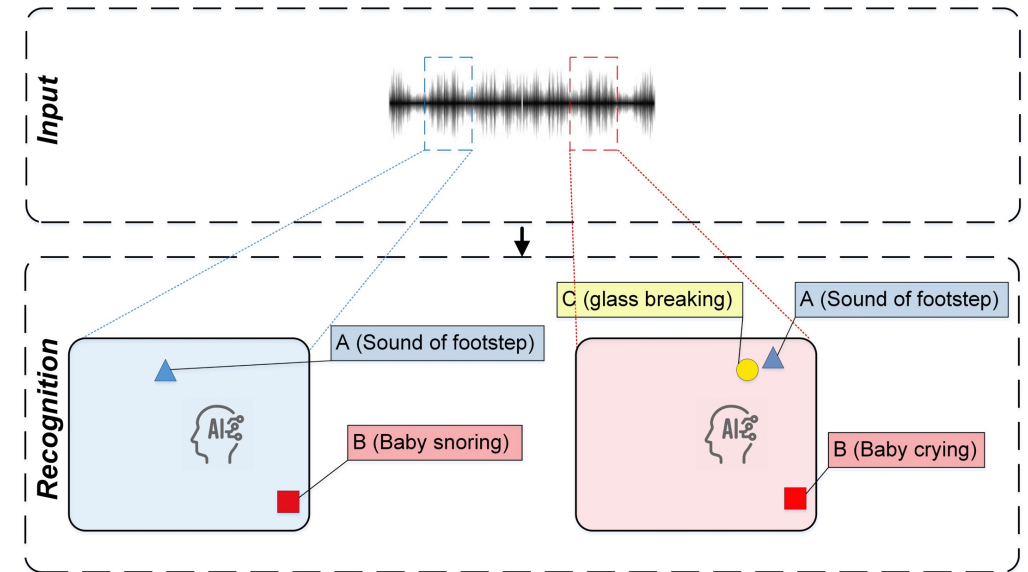
Nov. 3, 2022

Introduction

Everyday sound recognition suffers problems of **limited labelled data**...

- Everyday sounds covers a great amount of sound events
- Recordings of a specific sound event are scarce
- Sound annotation is costly

This work thus explore **few-shot learning in everyday sound recognition by leveraging audio taxonomy knowledge**



Some applications of everyday sound recognition

Methodology

Hierarchical prototypical networks (HPNs) are trained in a multi-task framework:

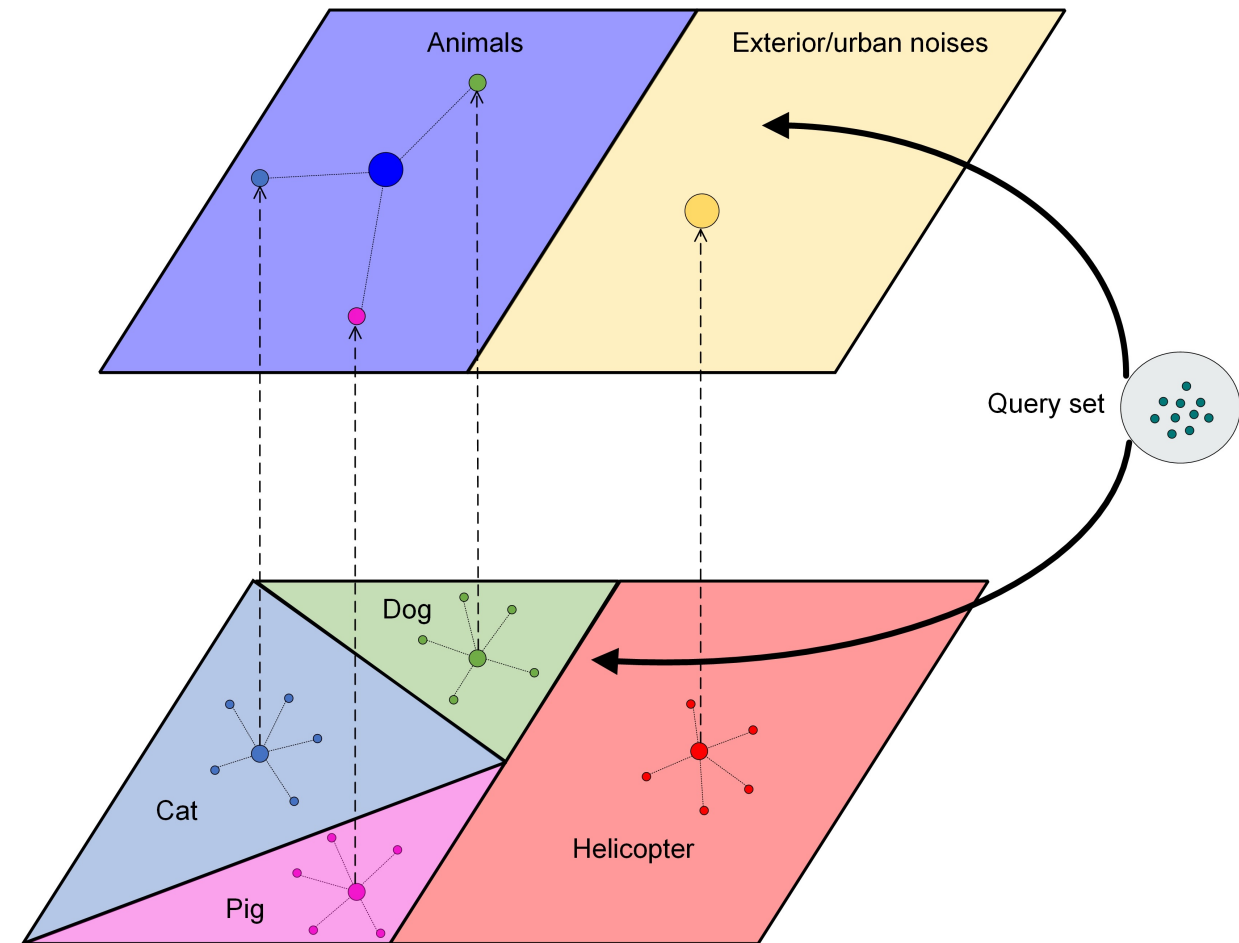
- Prototypes at the bottom are calculated by averaging support embedding.
- Prototypes of a higher level are generated by clustering their children's prototypes.

HPNs take the prototype of the child class when its ancestor contains one node only.

Structure loss:

$$L_{\text{hierarchical}} = \sum^H e^{\alpha h} L_{CE}^{(h)}$$

$$L_{CE} = - \sum_{k=0}^{K-1} p(x_i) \log q(x_i)$$



Training HPNs in a multi-task scenario

Experiments

	<i>Accuracy (%)</i>	<i>F₁ (%)</i>
Transfer Learning	72.90	72.87
Proto	77.70	77.52
Matching	71.81	71.75
HPN (ours)	78.65	78.51

HPNs yield better performance than other few-shot methods

- audio taxonomy knowledge can help an encoder to learn a better embedding space

	<i>Accuracy (%)</i>	<i>F₁ (%)</i>
Random	74.59	74.59
Parent	73.35	73.35
Uniform	77.70	77.52

Performance of models drops when the gap between the distribution of a base split and a novel split gets bigger.

Takeaways...

- ❖ ***Knowledge in audio taxonomy helps model to learn an unfamiliar class with a few labelled data***
- ❖ ***A hierarchical prototypical network (HPN) is adopted in a multi-task framework where each level of the tree structure is treated as an independent classification task in the training stage.***
- ❖ ***The approach to split the label set has great impact on the performance of few-shot model even in the same database.***

Takeaways...

- ❖ ***Knowledge in audio taxonomy helps model to learn an unfamiliar class with a few labelled data***
- ❖ ***A hierarchical prototypical network (HPN) is adopted in a multi-task framework where each level of the tree structure is treated as an independent classification task in the training stage.***
- ❖ ***The approach to split the label set has great impact on the performance of few-shot model even in the same database.***

E-mail: jinhua.liang@qmul.ac.uk

ORCID: <https://orcid.org/0000-0002-4570-0735>

Homepage: <https://jinhualiang.github.io/>, or:



- By Jinhua Liang
- Nov. 3, 2022