Engineering and Physical Sciences Research Council

FEW-SHOT BIOACOUSTIC EVENT DETECTION: ENHANCED CLASSIFIERS FOR PROTOTYPICAL NETWORKS

Author(s):

Ren Li

Jinhua Liang

Huy Phan

Queen Mary University of London

centre for digital music

By Jinhua Liang

Nov. 3, 2022

Introduction

Why use few-shot learning approach?

- Manual data labeling is time-consuming
- Collecting certain animal sounds is not always feasible in practice
- Few-shot learning can recognize new classes only with a few unseen samples (Morfi et al. 2021)

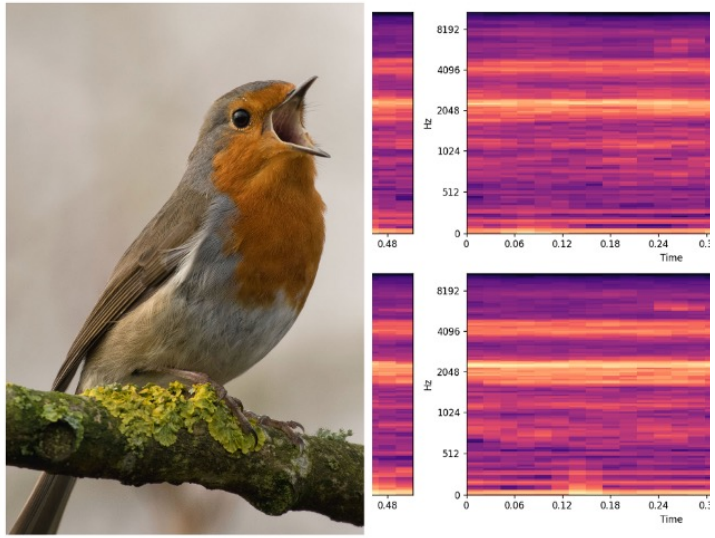


Fig. 1. Bird sounds detection from audio (So et al. 2020)

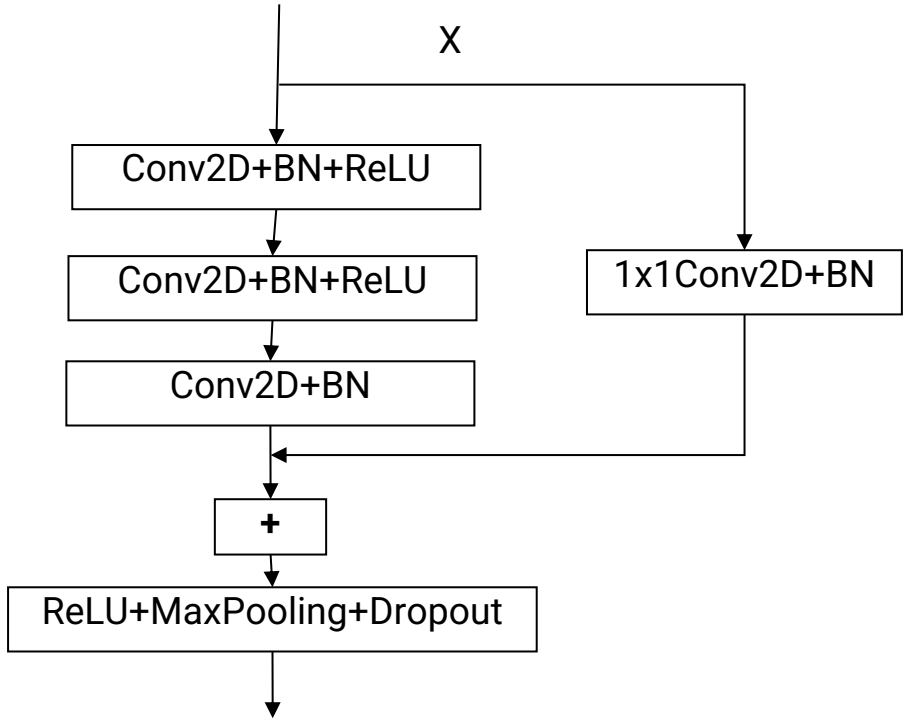
Our proposed methods: enhance the Prototypical network. ProtoNet by using ResNet and Autoencoder as the embedding networks.

Methodology

1. ResNet-based Prototypical Network

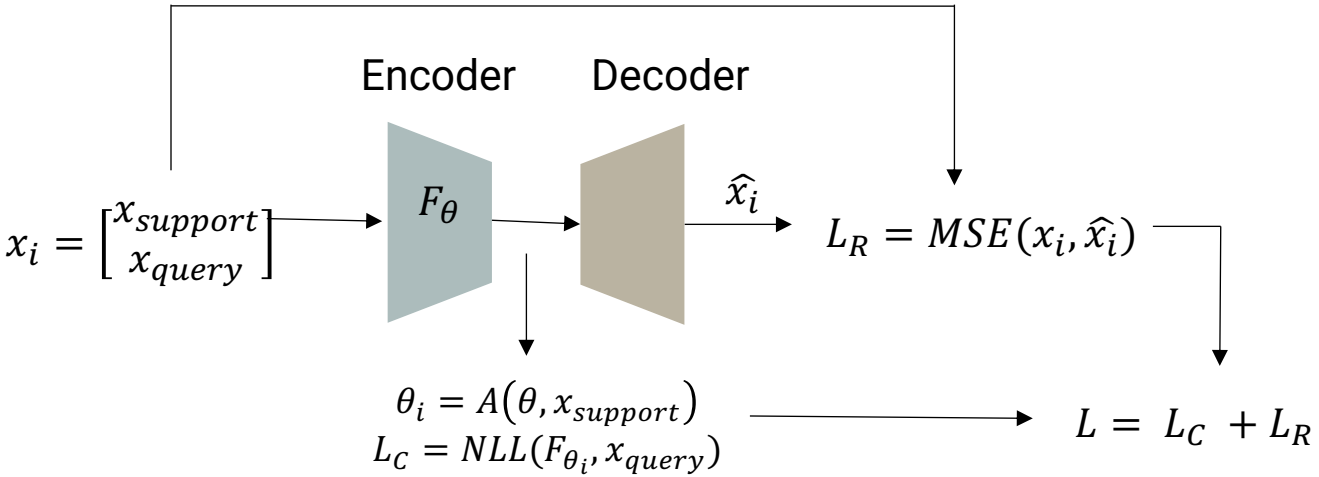
Residual blocks use skip connections to carry more information to a deeper layer.

Layers	*Channels*	*Kernel Size*
Conv2D+ BatchNorm +ReLU	16	3 × 3
Residual Block	64	3 × 3
Residual Block	128	3 × 3
Residual Block	64	3 × 3
Adaptive AvgPooling +SoftMax	-	3 × 3



Methodology

2. Combine Autoencoder and ProtoNet to learn the low-dimensional features



Conv Block		*Conv Transpose Block*	
Layers	*Kernel Size*	*Layers*	*Kernel Size*
Conv2D	3 × 3	Conv2D Transpose	2 × 2, stride 2
BatchNorm	-	BatchNorm	-
ReLU	-	Conv2D	3 × 3
Max pool	2 × 2	ReLU	-

Experiments

Model components			*Validation set scores (%)*			*Subset F1-score (%)*		
Exp No.	*Classifer*	*Feature*	*F1-score*	*Precision*	*Recall*	*HB*	*ME*	*PB*
1	CNN (Baseline)	PCEN	29.59	36.34	24.96	/	/	/
2	ResNet	PCEN	45.64	48.34	43.22	50.00	57.14	**26.18**
3	Autoencoder	PCEN	37.94	38.95	36.97	44.53	52.05	25.68
4	CNN	PCEN+Augment	37.16	42.09	33.26	38.86	**72.01**	15.33
5	ResNet	PCEN+Augment	**47.88**	**52.11**	44.30	**53.45**	50.98	17.65
6	Autoencoder	PCEN+Augment	47.61	50.18	**45.34**	52.68	53.10	22.44

- Exps. 1 and 2: the ResNet model outperformed the baseline with an improvement of more than 15%
- Exps. 1, 2 and 3: the autoencoder produced about 8% improvement, which was slightly lower than the ResNet
- Exps. 4, 5 and 6: data augmentation helps models perform better than those without data augmentation

Takeaways…

- Using **ResNet** and **Autoencoder** (or construction loss) contributes to learn a more advanced embedding space and improve the performance of Prototypical Network.
- Our training framework has not changed, and **it still relies too much on one-sided input, making it prone to overfitting.**

Takeaways…

- Using **ResNet** and **Autoencoder** (or construction loss) contributes to learn a more advanced embedding space and improve the performance of Prototypical Network.
- Our training framework has not changed, and **it still relies too much on one-sided input, making it prone to overfitting.**

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