

# DCASE 2019 CHALLENGE TASK 5: CNN+VGGISH

## Technical Report

*Daniel Tompkins*

Microsoft  
Dynamics 365 AI Research  
Redmond, WA 98052, USA  
daniel.tompkins@microsoft.com

*Eric Nichols*

Microsoft  
Dynamics 365 AI Research  
Redmond, WA 98052, USA  
eric.nichols@microsoft.com

### ABSTRACT

We trained a model for multi-label audio classification on Task 5 of the DCASE 2019 Challenge [1]. The model is composed of a preprocessing layer that converts audio to a log-mel spectrogram, a VGG-inspired Convolutional Neural Network (CNN) that generates an embedding for the spectrogram, the pre-trained VGGish network [2] that generates a separate audio embedding, and finally a series of fully-connected layers that converts these two embeddings (concatenated) into a multi-label classification. This model directly outputs both “fine” and “coarse” labels; it treats the task as a 37-way multi-label classification problem. One version of this network did better at the coarse labels (submission 1); another did better with fine labels on Micro AUPRC (submission 2).

A separate family of CNNs models, one per coarse label, was also trained to take into account the hierarchical nature of the labels (submission 3), but the single model solution performed slightly better.

**Index Terms**—audio, classification, CNN

### 1. INTRODUCTION

TODO

### 2. RELATED WORK

TODO

### 3. MODEL

TODO

### 4. RESULTS

#### 4.1. Validation Set Results

#### 4.1.1. Submission 1

Fine level evaluation:

```
=====
* Micro AUPRC:      0.6464364540680165
* Micro F1-score (@0.5): 0.48275862068965525
* Macro AUPRC:      0.42517821727445076
* Coarse Tag AUPRC:
- 1: 0.6452856315691152
- 2: 0.15554209969203026
- 3: 0.402088600041743
- 4: 0.3985514806702355
- 5: 0.6655855442707173
- 6: 0.251310185085831
- 7: 0.8777335154047885
- 8: 0.005328681461145428
```

Coarse level evaluation:

```
=====
* Micro AUPRC:      0.7866166576551467
* Micro F1-score (@0.5): 0.6088560885608856
* Macro AUPRC:      0.5789550808992179
* Coarse Tag AUPRC:
- 1: 0.8353122731547575
- 2: 0.4327822490122867
- 3: 0.40356804631393906
- 4: 0.7018285524935755
- 5: 0.8556952720570562
- 6: 0.45479025811812124
- 7: 0.9422679583899046
- 8: 0.005396037654102171
```

#### 4.1.2. Submission 2

Fine level evaluation:

```
=====
* Micro AUPRC:      0.6560867707337048
* Micro F1-score (@0.5): 0.3982202447163515
* Macro AUPRC:      0.40059566061982355
* Coarse Tag AUPRC:
- 1: 0.6405304300056462
```

- 2: 0.21380721739093014  
 - 3: 0.3336595526102611  
 - 4: 0.2761679795550256  
 - 5: 0.70229675965  
 - 6: 0.1563059224638172  
 - 7: 0.8819974232829082  
 - 8: 0.0

- 3: 0.40356804631393906  
 - 4: 0.7018285524935755  
 - 5: 0.8556952720570562  
 - 6: 0.45479025811812124  
 - 7: 0.942311730093642  
 - 8: 0.005396037654102171

Coarse level evaluation:

=====

\* Micro AUPRC: 0.7687132764444584  
 \* Micro F1-score (@0.5): 0.5328109696376102  
 \* Macro AUPRC: 0.5552582865865167  
 \* Coarse Tag AUPRC:  
 - 1: 0.8178125638838614  
 - 2: 0.42810526358580453  
 - 3: 0.3332328931822347  
 - 4: 0.6770075945049289  
 - 5: 0.8674275089257218  
 - 6: 0.37801009220794807  
 - 7: 0.9404703764016337  
 - 8: 0.0

#### 4.1.3. Submission 3

Fine level evaluation:

=====

\* Micro AUPRC: 0.6427645749560833  
 \* Micro F1-score (@0.5): 0.48983739837398366  
 \* Macro AUPRC: 0.41428465325770225  
 \* Coarse Tag AUPRC:  
 - 1: 0.626624522289203  
 - 2: 0.18453628320402102  
 - 3: 0.4020905123120714  
 - 4: 0.3795611579853643  
 - 5: 0.6698543764513578  
 - 6: 0.2293541518972007  
 - 7: 0.8169275404612543  
 - 8: 0.005328681461145428

Coarse level evaluation:

=====

\* Micro AUPRC: 0.7866247951279302  
 \* Micro F1-score (@0.5): 0.6088560885608856  
 \* Macro AUPRC: 0.5789589780231039  
 \* Coarse Tag AUPRC:  
 - 1: 0.8353362187853394  
 - 2: 0.4327457086690561

## 5. CONCLUSIONS

TODO

## 6. REFERENCES

- [1] <http://dcase.community/workshop2019/>.
- [2] Hershey et. al. "CNN Architectures for Large-Scale Audio Classification". *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.