

Chirping Up the Right Tree:

Incorporating biological taxonomies into deep bioacoustic classifiers



Jason Cramer



Vincent Lostanlen



Andrew Farnsworth



Justin Salamon

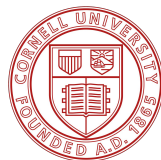


Juan Pablo Bello



NYU

TANDON SCHOOL
OF ENGINEERING



TheCornellLab
of Ornithology



Adobe Research

Motivation

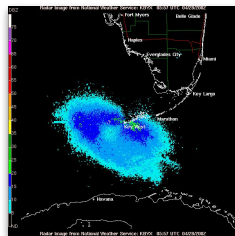
Migratory birds are in peril!

- Phenomena like deforestation and climate change threaten many avian habitats
- Tracking migratory changes can help conservation efforts

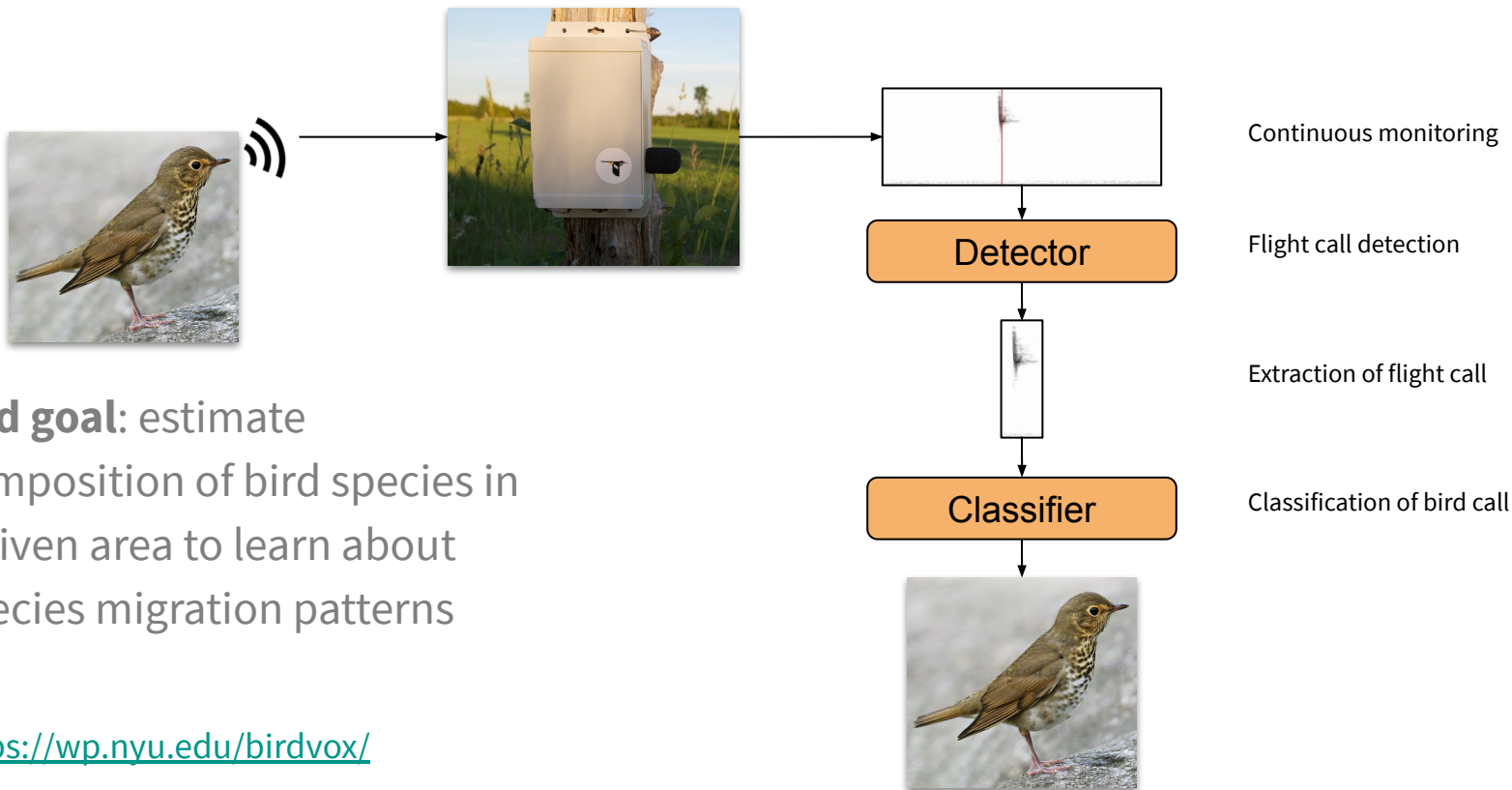


How do we track migratory birds?

- Birdwatchers: **wide geographical coverage**, but **limited to daytime**
- Weather radar data: **tracks avian biomass at high spatio-temporal resolution**, but **unknown species composition**
- Autonomous recording units: **24/7 high temporal resolution** with **potential for automatic detection/classification**



Practical Use Case: BirdVox Project



Machine listening relies on class hierarchies

Tend to be based on similarity judgements that may vary across **individuals**, **cultures**, and **use cases**

Human sounds

- Human voice
- Whistling
- Respiratory sounds
- Human locomotion
- Digestive
- Hands
- Heart sounds, heartbeat
- Otoacoustic emission
- Human group actions

Animal sounds

- Domestic animals, pets
- Livestock, farm animals, working animals
- Wild animals

Natural sounds

- Wind
- Thunderstorm
- Water
- Fire

Music

- Musical instrument
- Music genre
- Musical concepts
- Music role
- Music mood

Sounds of things

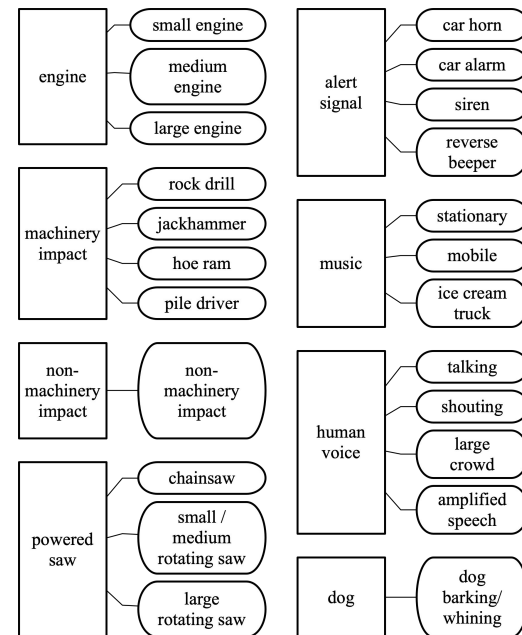
- Vehicle
- Engine
- Domestic sounds, home sounds
- Bell
- Alarm
- Mechanisms
- Tools
- Explosion
- Wood
- Glass
- Liquid
- Miscellaneous sources
- Specific impact sounds

Source-ambiguous sounds

- Generic impact sounds
- Surface contact
- Deformable shell
- Onomatopoeia
- Silence
- Other sourceless

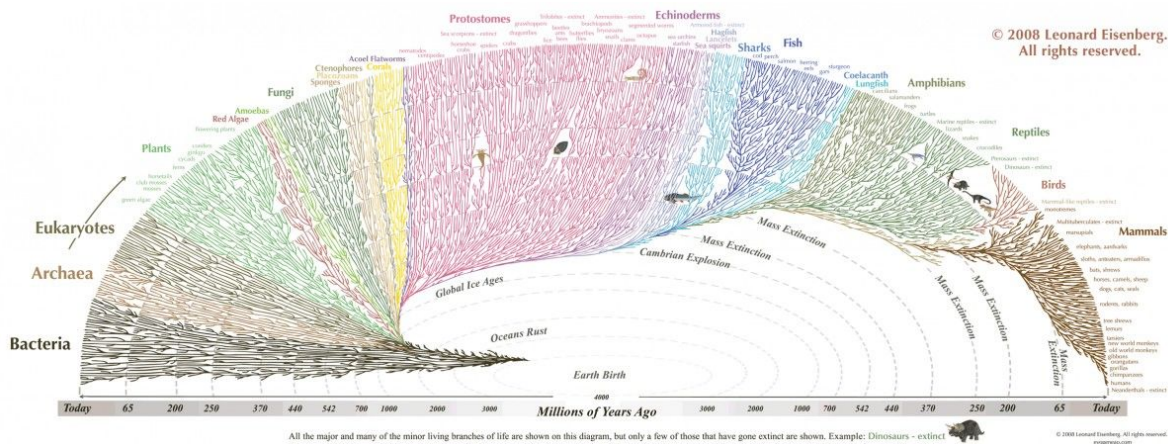
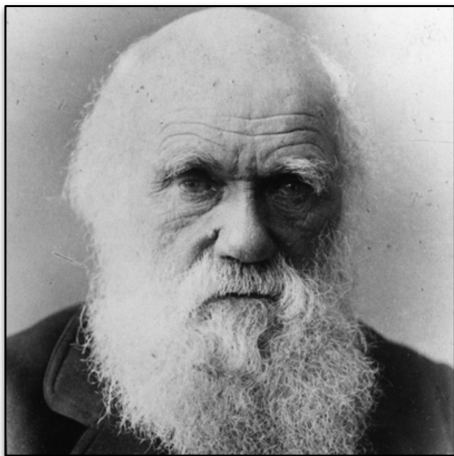
Channel, environment and background

- Acoustic environment
- Noise
- Sound reproduction



Evolutionary biology can provide structure!

- Darwinian classification yields a systematic phylogentic hierarchy
- We can try to leverage these biological taxonomies in machine listening!



Big idea

- We leverage biological hierarchies to impose inductive bias on bioacoustic classifiers to make more effective use of data
- Achieve this via **multi-task training** and **hierarchical model architectures**
- Benefits:
 - Encourage model to learn hierarchical relationships
 - In conservation science, most insightful taxonomic level may not be known a priori
 - Train one model instead of multiple for multi-level prediction

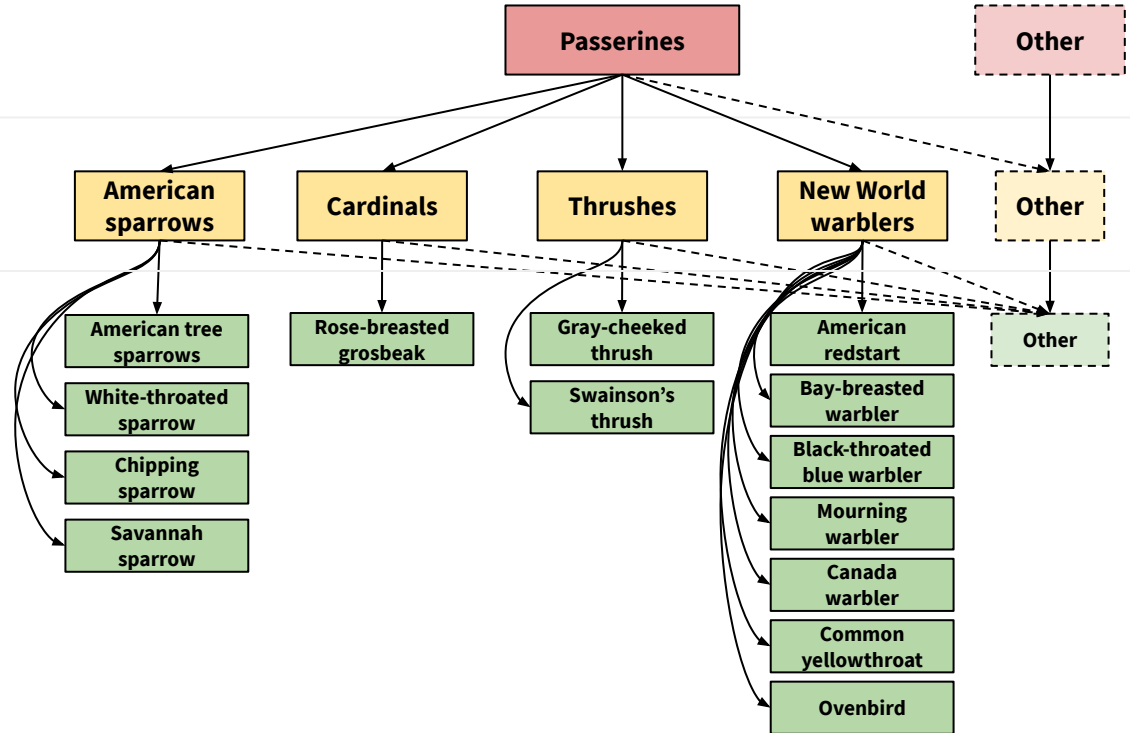
Methods

Our Taxonomy

Coarse Level: Order

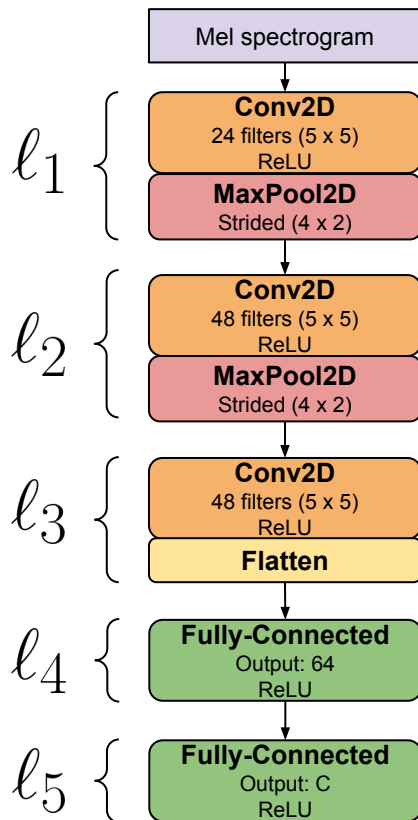
Medium Level: Family

Fine Level: Species



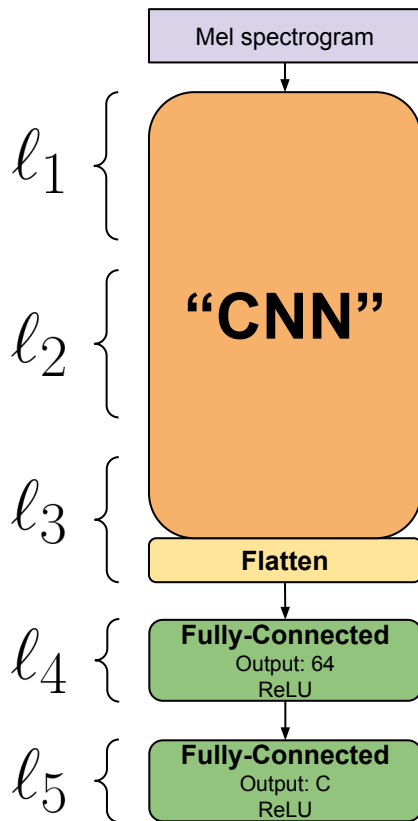
Base Model (Salamon et al. 2017)

Architecture used in state
of the art flight call
classification method

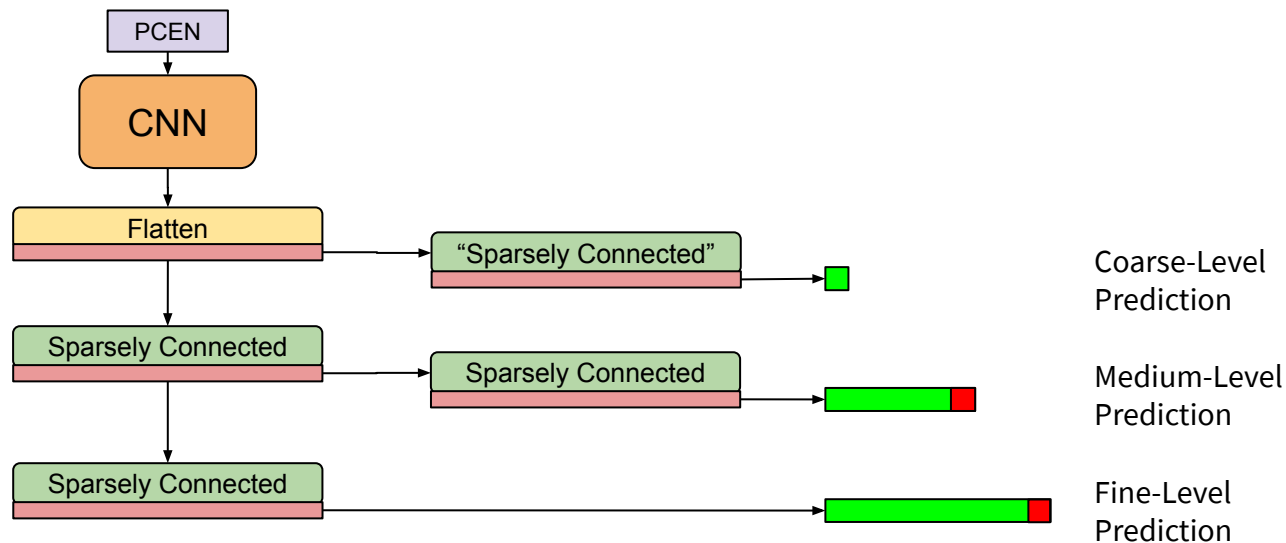


Base Model (Salamon et al. 2017)

Architecture used in state
of the art flight call
classification method



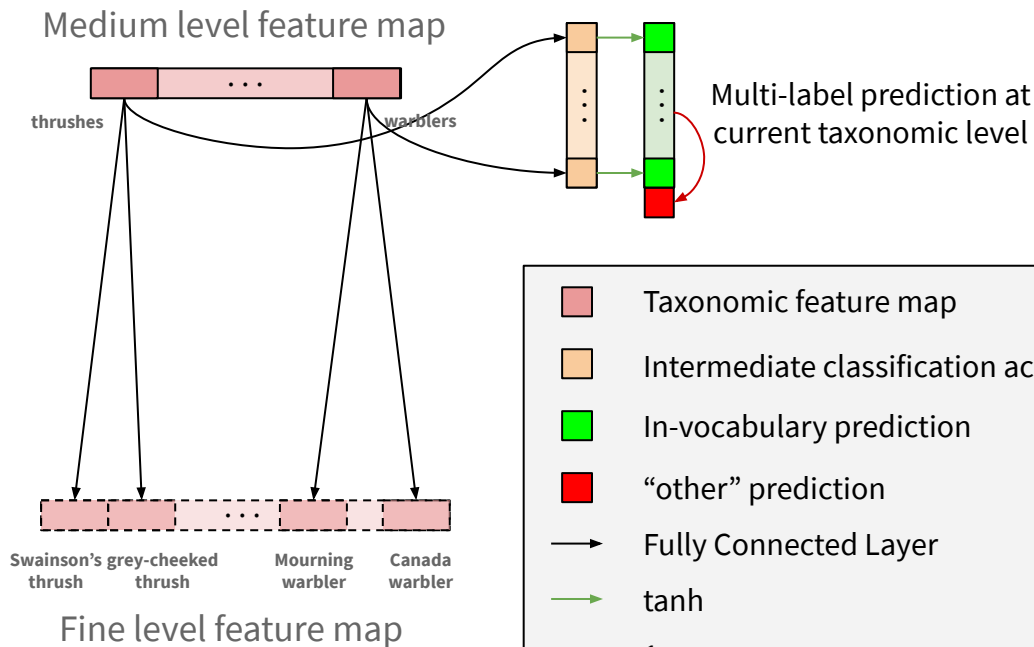
TaxoNet



TaxoNet Hierarchical Composition Block

Feature maps at layers are partitioned, with **each partition mapping to a member of the current taxonomic level**

Partitions for a taxa member in the current layer **only connect to partitions corresponding to descendants** in the next taxonomic level, **mirroring the structure of the taxonomy**



Experimental Design - New Datasets

- American Northwest Avian Flight Call Classification (**ANAFCC**)
 - <https://doi.org/10.5281/zenodo.3666782>
 - **Aggregation of datasets:** BirdVox-70k, CLO-43SD, CLO-SWTH, CLO-WTSP, Macaulay Library, Xeno-Canto, Old Bird
 - **Verified and re-annotated** by expert ornithologist (Andrew Farnsworth)
 - Used for **training** and **validation**
- BirdVox 14 Species Dataset (**BirdVox-14SD**)
 - <https://doi.org/10.5281/zenodo.3667094>
 - Sampled from 10 autonomous recording units over a **full migration season**, annotated for **14 species and “other”**
 - Used for **testing**
- Enforce consistent distribution of species across splits using knapsack solver

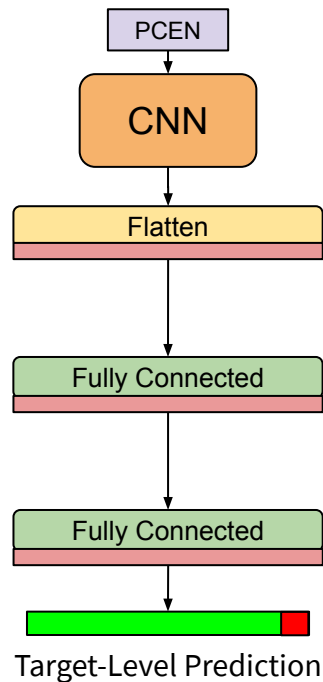
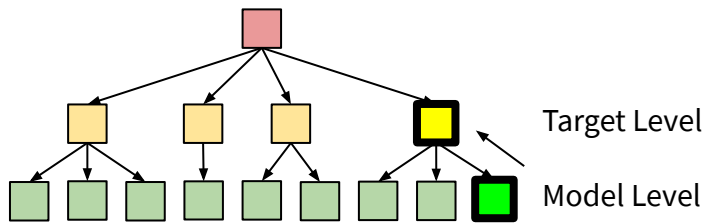
Experimental Design - Training

- Input: **log-scale mel-frequency spectrogram** with **per-channel energy normalization (PCEN)** applied
- **Data augmentation:** pitch shifting, time-stretching, additive background noise
- **Uniformly sampled** with respect to **fine-level classes** using pescador
- Equally weighted **multi-task training** used to train at multiple taxonomic levels
- Evaluate using **micro-averaged** (class independent) **accuracy**

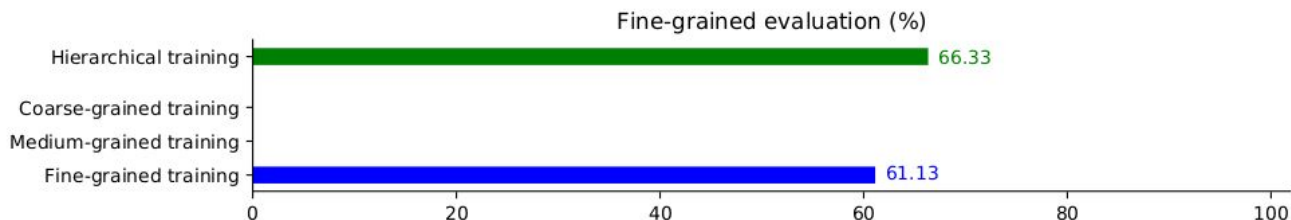
Evaluation

Baseline Models - Flat Single-Task

- Train a model specialized for each taxonomic level
- Two main strategies:
 - **“Specialist”** - at a given level, predict using a model trained at that level
 - **“Coarsening”** - at a given level, predict using a model trained at a finer level and project up to the corresponding level

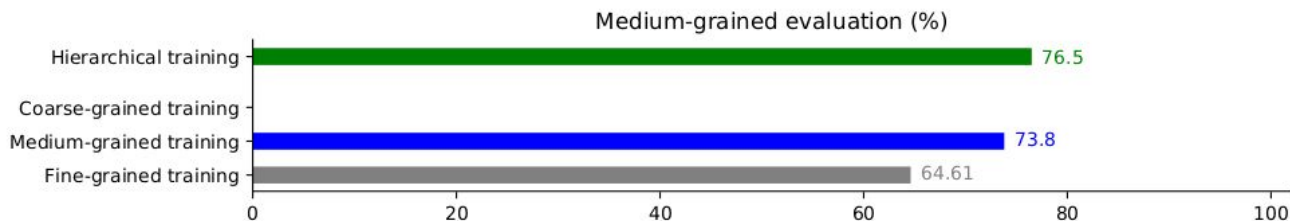



Results: Fine-Level Prediction

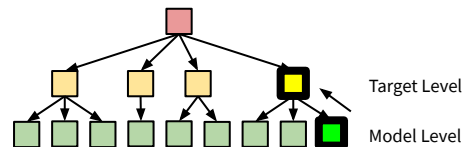


- TaxoNet **outperforms** the specialist baseline at the fine level!
- **Leveraging hierarchical classification improves performance** even if we only care about the finest level!

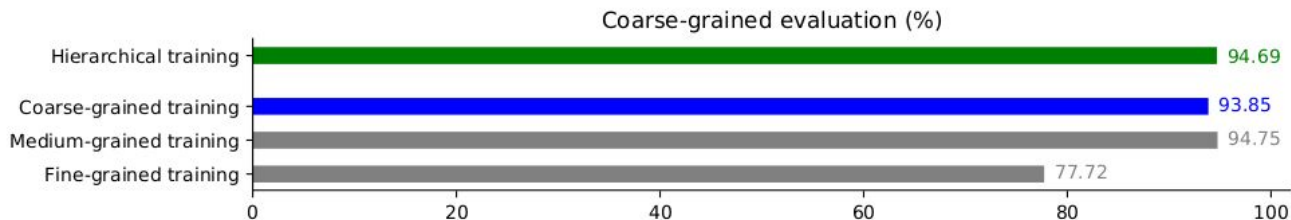
Results: Medium-Level Prediction



- **TaxoNet still outperforms specialist strategy** at the medium level
 - **TaxoNet significantly outperforms coarsening strategy**
 - **TaxoNet is capable of multi-tasking!**
- 
- ```
graph TD; A[Red] --> B[Yellow]; A --> C[Yellow]; B --> D[Green]; B --> E[Green]; B --> F[Green]; C --> G[Green]; C --> H[Green];
```



# Results: Coarse-Level Prediction

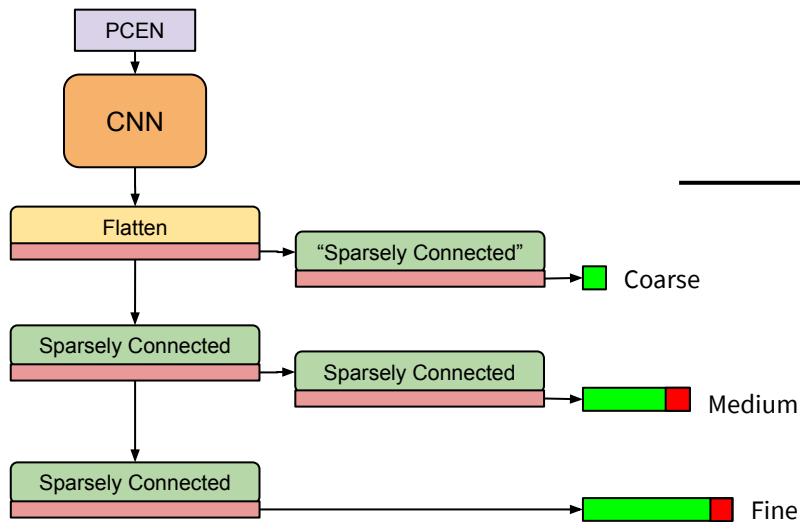


- Important to note that there is **severe class imbalance** at the coarse level
- **TaxoNet matches specialist strategy** performance **and coarsening** from the **medium level**
- **TaxoNet significantly outperforms coarsening** from the **fine level**

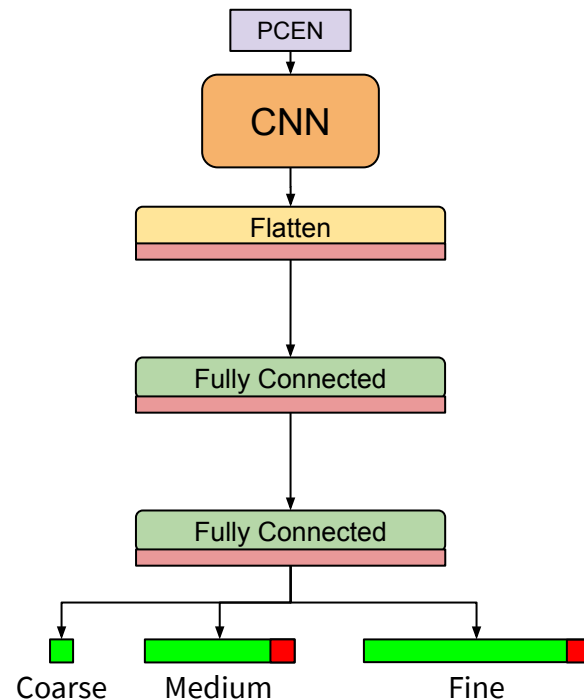
# Ablation Studies

# 1. Non-hierarchical multi-task model

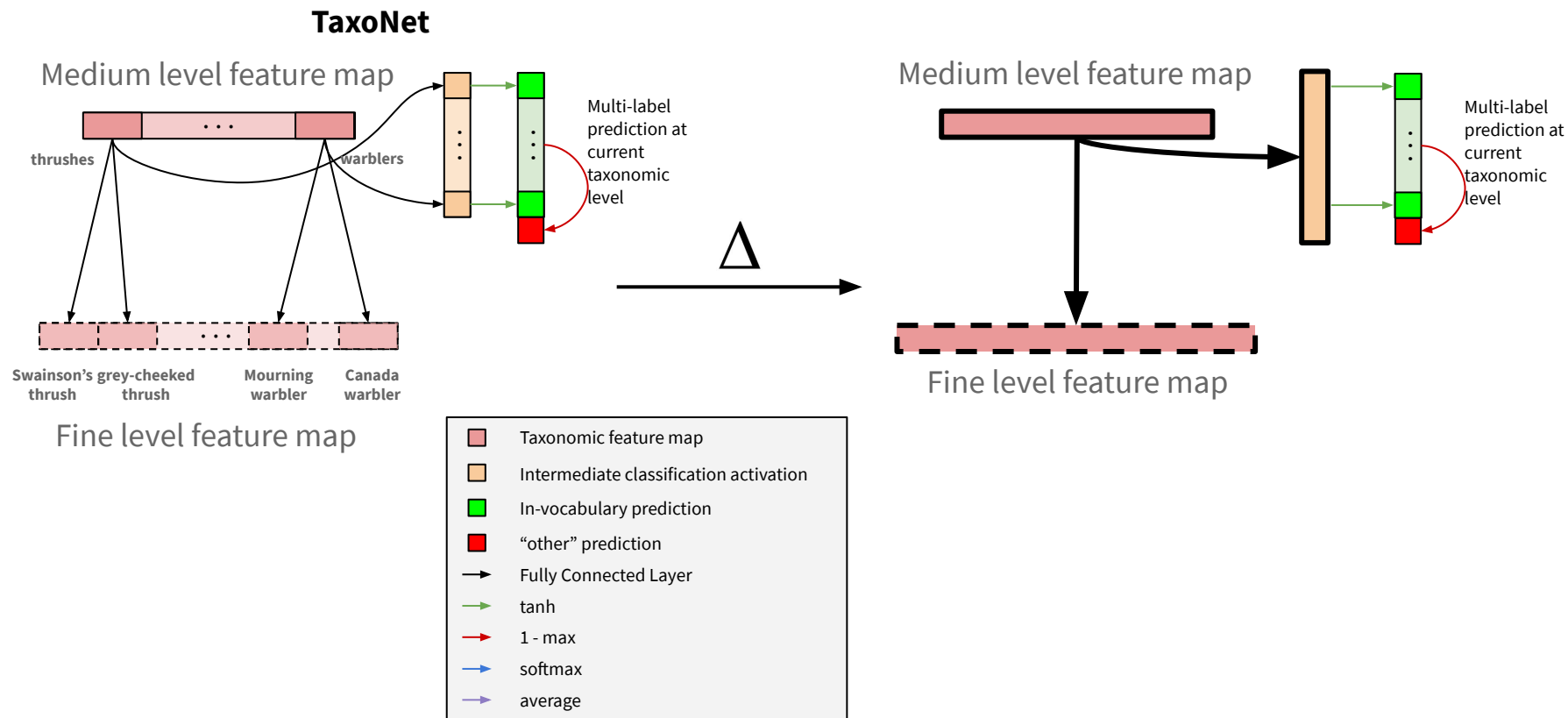
TaxoNet



$\Delta$



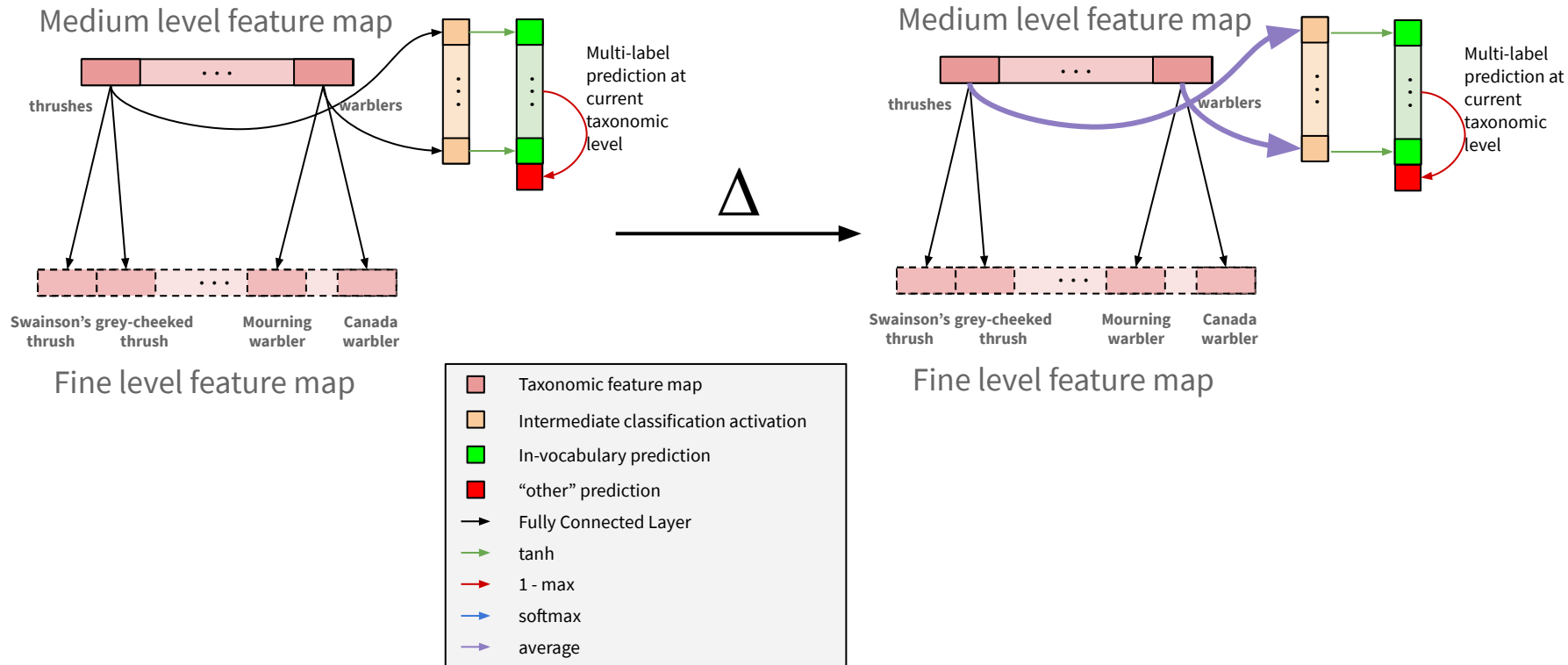
## 2. Hierarchical baseline model - block





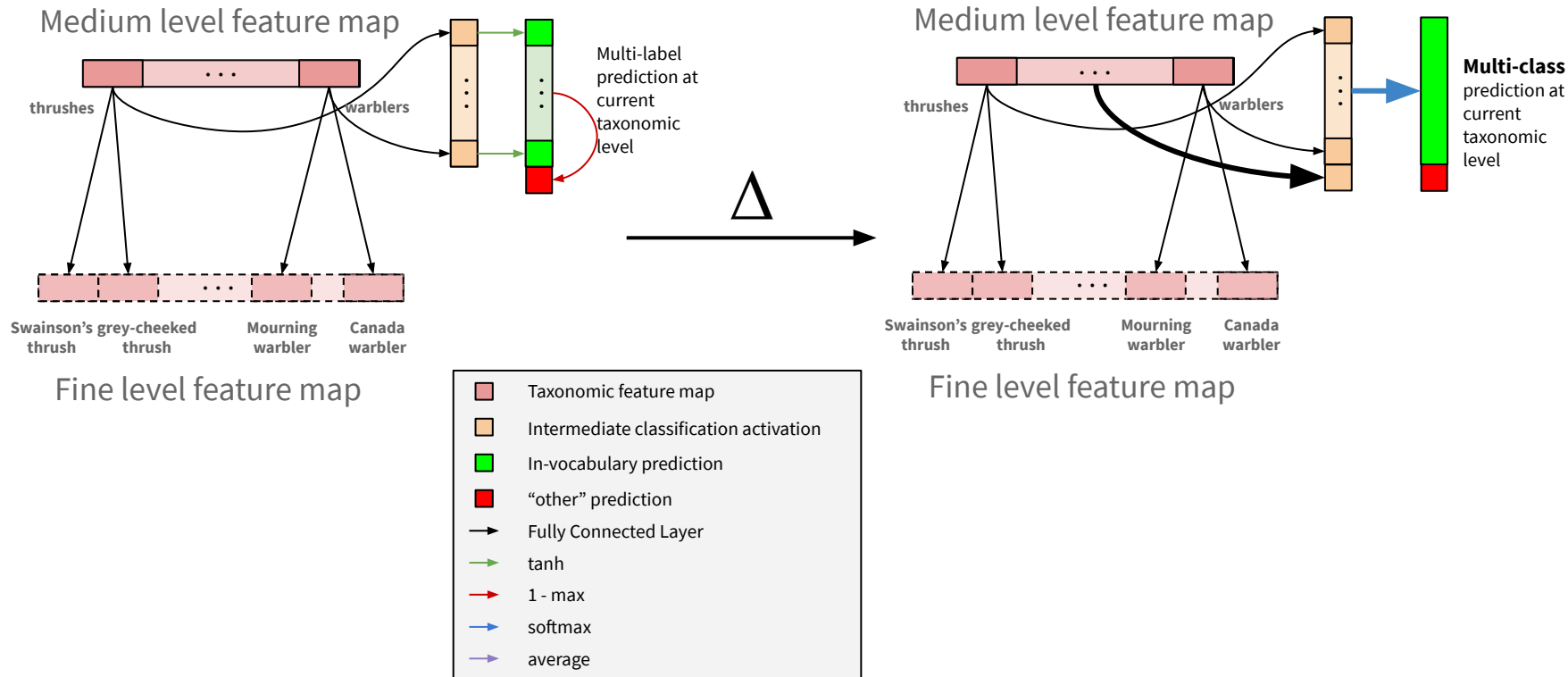
# 3. Hierarchical containment model - block

TaxoNet



# 4. Hierarchical composition multiclass - block

## TaxoNet



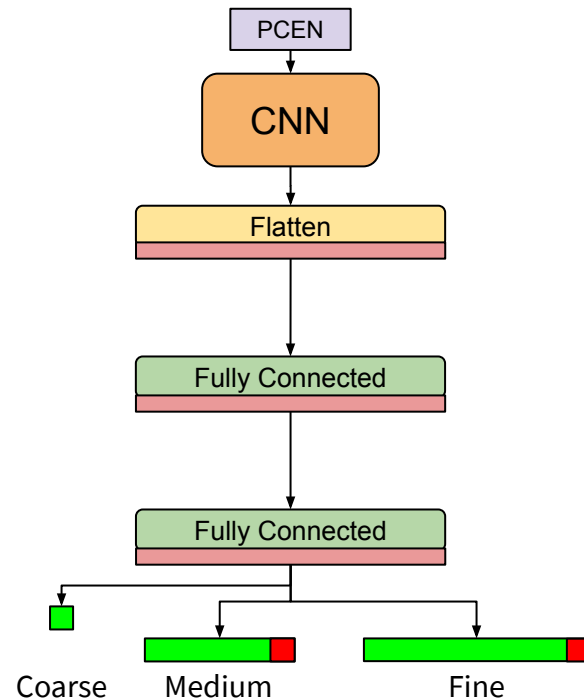
# Ablation Comparisons

- TaxoNet performs best at fine-level
- TaxoNet performs second best at medium level
- All ablations perform similarly to TaxoNet at the coarse level

| Model                                | Fine-Level Accuracy | Medium-Level Accuracy | Coarse-Level Accuracy |
|--------------------------------------|---------------------|-----------------------|-----------------------|
| <u>TaxoNet</u>                       | <b>66.33</b>        | 76.50                 | 94.69                 |
| Non-Hierarchical Multi-Task          | 61.82               | 75.10                 | 94.39                 |
| Hierarchical Baseline                | 58.74               | 75.83                 | 94.54                 |
| Hierarchical Containment             | 63.47               | <b>79.36</b>          | 94.75                 |
| Hierarchical Composition Multi-Class | 60.39               | 75.94                 | 94.67                 |

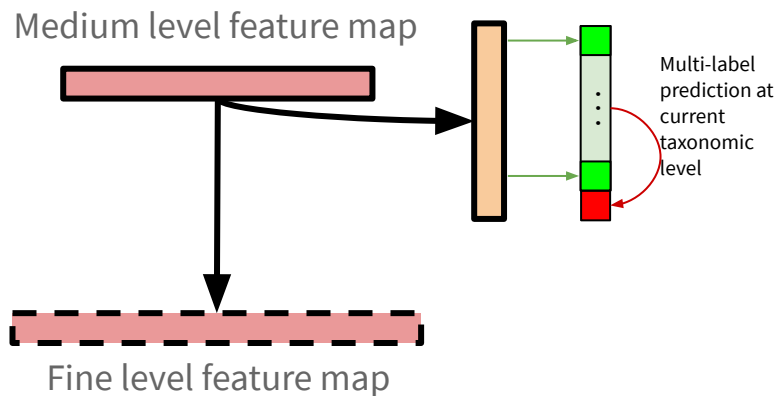
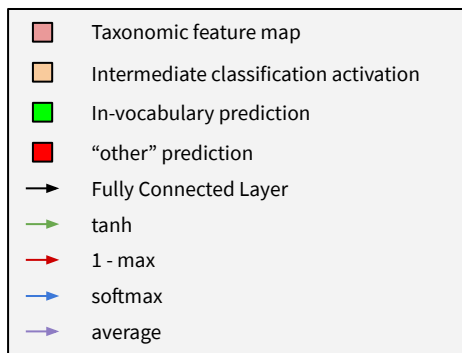
# 1. Non-hierarchical multi-task model

| Model          | Fine-Level Accuracy | Medium-Level Accuracy | Coarse-Level Accuracy |
|----------------|---------------------|-----------------------|-----------------------|
| <u>TaxoNet</u> | <b>66.33</b>        | 76.50                 | 94.69                 |
| 1.             | 61.82               | 75.10                 | 94.39                 |



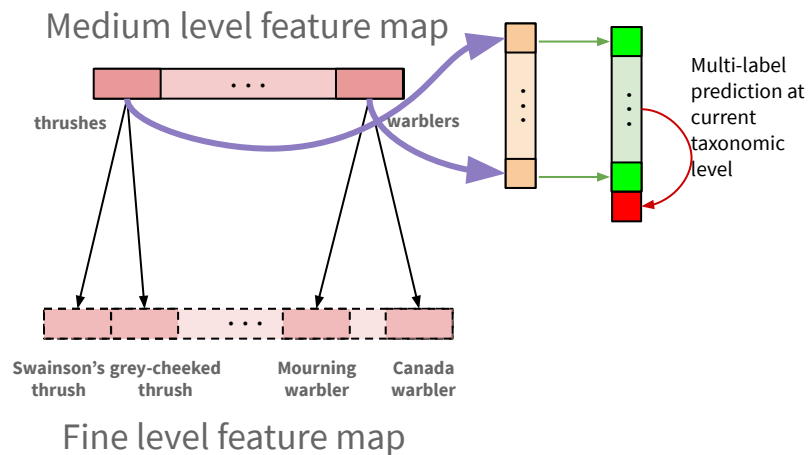
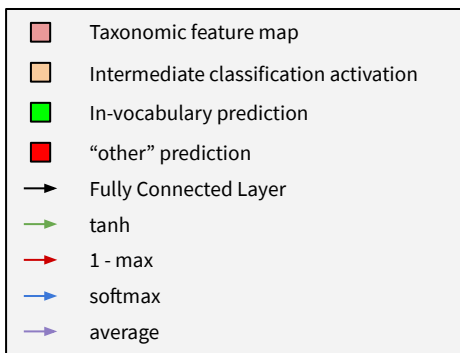
## 2. Hierarchical baseline model - block

| Model          | Fine-Level Accuracy | Medium-Level Accuracy | Coarse-Level Accuracy |
|----------------|---------------------|-----------------------|-----------------------|
| <u>TaxoNet</u> | <b>66.33</b>        | 76.50                 | 94.69                 |
| 2.             | 58.74               | 75.83                 | 94.54                 |



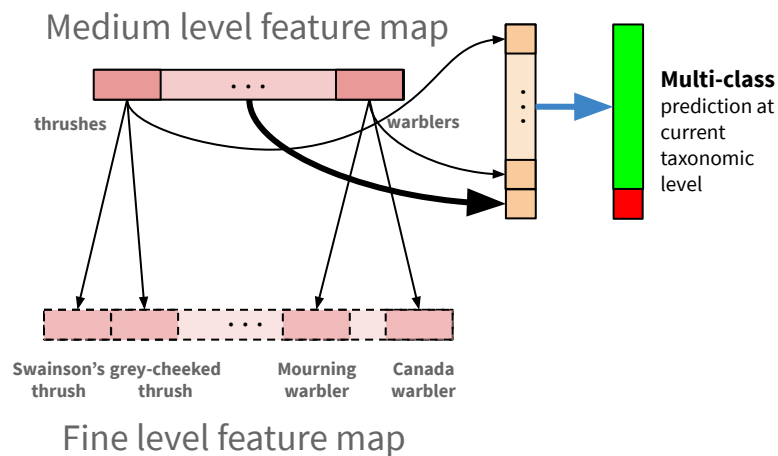
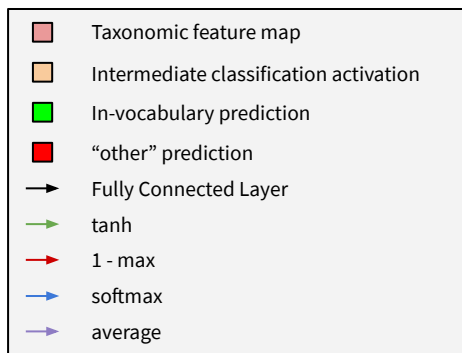
### 3. Hierarchical containment model - block

| Model          | Fine-Level Accuracy | Medium-Level Accuracy | Coarse-Level Accuracy |
|----------------|---------------------|-----------------------|-----------------------|
| <u>TaxoNet</u> | <b>66.33</b>        | 76.50                 | 94.69                 |
| 3.             | 63.47               | <b>79.36</b>          | 94.75                 |



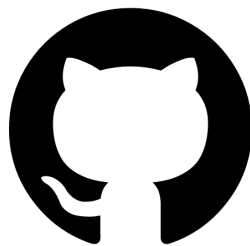
# 4. Hierarchical composition multiclass - block

| Model          | Fine-Level Accuracy | Medium-Level Accuracy | Coarse-Level Accuracy |
|----------------|---------------------|-----------------------|-----------------------|
| <u>TaxoNet</u> | <b>66.33</b>        | 76.50                 | 94.69                 |
| 4.             | 60.39               | 75.94                 | 94.67                 |



# In summary:

- **TaxoNet matches or outperforms specialist and coarsening strategies**
- Deep hierarchical classification **exploits taxonomic relationships** to better leverage the training data
- **New datasets:** ANAFCC and BirdVox-14SD
- **Open-source flight call classification:** BirdVoxClassify
  - Integrated with BirdVoxDetect: `pip install birdvoxdetect`
- **// TODO:**
  - Enforcing hierarchical consistency
  - Exploring the effect of the depth and width of taxonomies



<https://github.com/BirdVox/birdvoxclassify>  
<https://github.com/BirdVox/cramer2020icassp>

# Thank you!