Polyphonic sound event detection for highly dense birdsong scenes

Alberto Garcia Arroba Parrilla, Dan Stowell
Tilburg University

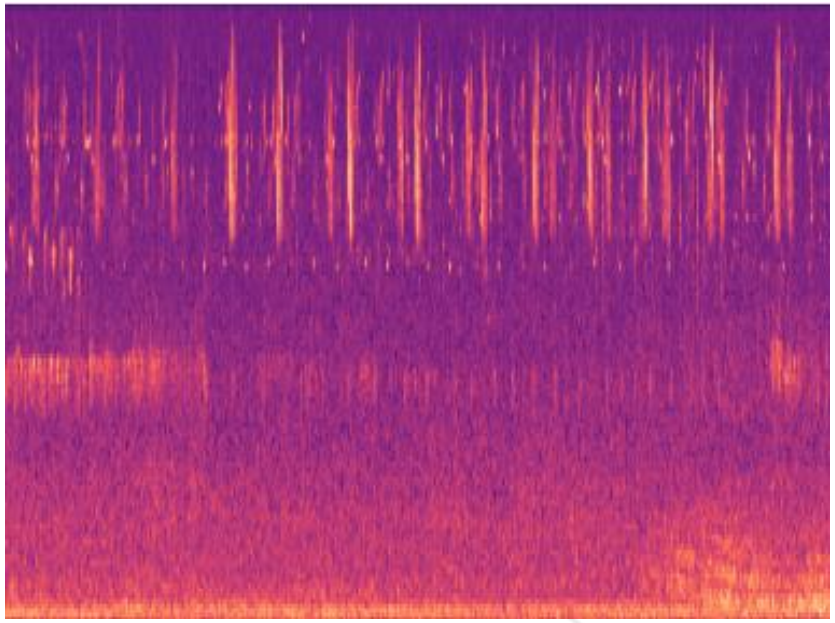
DCASE Workshop, 2022

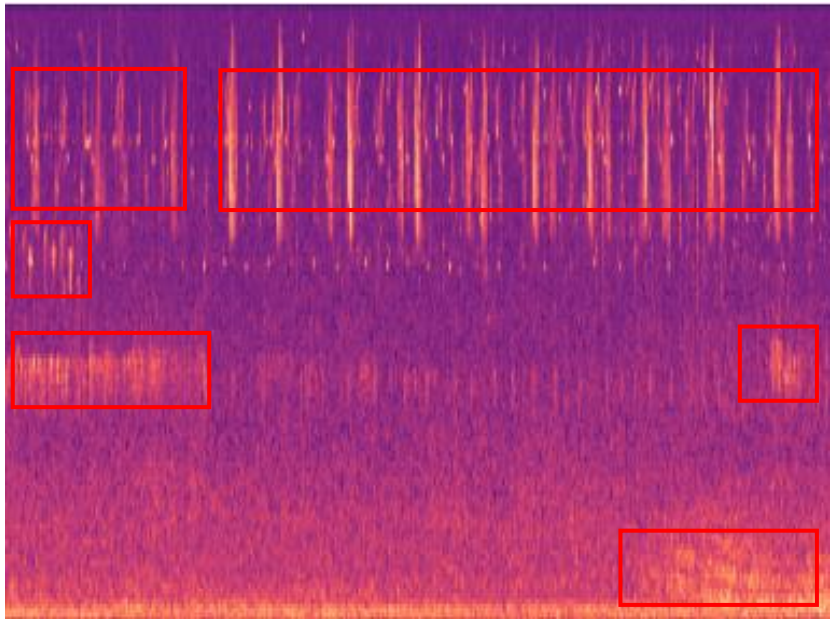
Project Aim

- Study *Sound Event Detection (*SED) on scenes with **high polyphony**
- High polyphony is common, but not extendedly studied
  - Mostly in other sound domains
  - Doesn’t normally reach more than **6 overlapping sounds**
- Focus on the **maximum polyphony reached**
  - Maximum overlapping sounds

Project Aim

- Birdsong scenarios are very **dense** (i.e. high polyphony)
  - **Target scenario**: *Dawn chorus (30+ species)*





*Example with maximum **polyphony** of 3*

To understand…

- How accurate can a model detect events in dense scenes?
- Does a model train with denser samples learn faster than a model trained with simpler samples?
- How does a model perform on samples with *unseen* polyphony?

Want to know more?

You are welcome to ask anything at the poster!