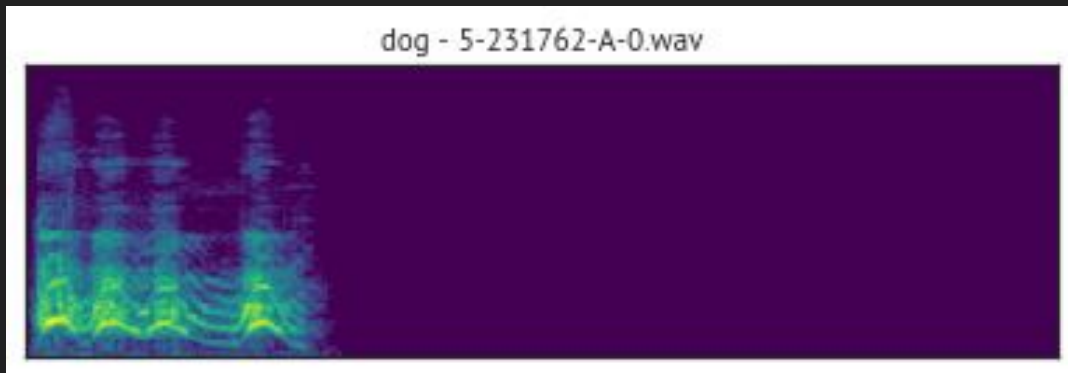


# ESC-50: Dataset

## Environmental Sound Classification



# Overview

- collection of 2000 audio recordings
- used for environmental sound classification
- overall size ~600 MB (.zip archive)
- all files are in WAV format (44.1 kHz, mono)
- sources of all sounds: [Freesound.org](https://www.freesound.org)
- licensed under CC BY

 1-100032-A-0.wav

 1-100038-A-14.wav

 1-100210-A-36.wav

 1-100210-B-36.wav

 1-101296-A-19.wav

 1-101296-B-19.wav

 1-101336-A-30.wav

# Classes

- 5 major classes:
  - Animals
  - Natural soundscapes & water sounds
  - Human, non-speech sounds
  - Interior/domestic sounds
  - Urban noises
- 50 semantic classes (10 for each major class)
- 40 samples for each semantic class



# Classes full overview

Animals	Natural soundscapes & water sounds	Human, non-speech sounds	Interior/domestic sounds	Exterior/urban noises
Dog	Rain	Crying baby	Door knock	Helicopter
Rooster	Sea waves	Sneezing	Mouse click	Chainsaw
Pig	Crackling fire	Clapping	Keyboard typing	Siren
Cow	Crickets	Breathing	Door, wood creaks	Car horn
Frog	Chirping birds	Coughing	Can opening	Engine
Cat	Water drops	Footsteps	Washing machine	Train
Hen	Wind	Laughing	Vacuum cleaner	Church bells
Insects (flying)	Pouring water	Brushing teeth	Clock alarm	Airplane
Sheep	Toilet flush	Snoring	Clock tick	Fireworks
Crow	Thunderstorm	Drinking, sipping	Glass breaking	Hand saw

# Content structure

- Each sample following convention:

{FOLD}-{CLIP\_ID}-{TAKE}-{TARGET}.wav

- {FOLD} – the index of the cross-validation fold
  - {CLIP\_ID} – ID
  - {TAKE} – symbolize different takes from the same original sample
  - {TARGET} – target class number [0, 49]
- 
- Additionally have .csv file `meta/esc50.csv` with all dataset metadata