



AI:Mechanic

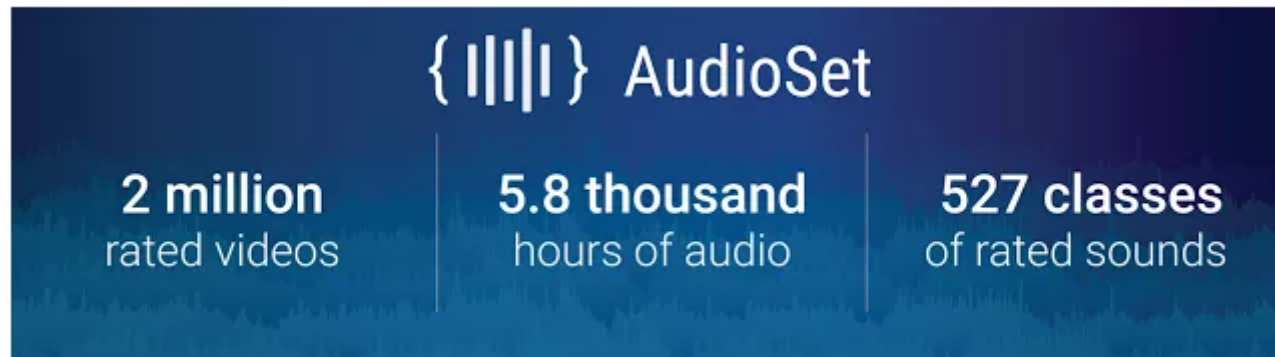
March 31, 2017

d411: AudioSet Audio Events Dataset

AudioSet Audio Events Dataset: <https://research.googleblog.com/2017/03/announcing-audioset-dataset-for-audio.html>

AudioSet a large-scale dataset of manually annotated audio events: <https://research.google.com/audioset/>

Systems able to recognize sounds familiar to human listeners have a wide range of applications, from adding sound effect information to automatic video captions, to potentially allowing you to search videos for specific audio events. Building Deep Learning systems to do this relies heavily on both a large quantity of computing (often from highly parallel GPUs), and also – and perhaps more importantly – on significant amounts of accurately-labeled training data. However, research in environmental sound recognition is limited by currently available public datasets.



A sound vocabulary and audio dataset:

AudioSet consists of an expanding ontology of 632 audio event classes and a collection of 2,084,320 human-labeled 10-second sound clips drawn from YouTube videos. The ontology is specified as a hierarchical graph of event categories, covering a wide range of human and animal sounds, musical instruments and genres, and common everyday environmental sounds.

Based on 2,084,320 YouTube videos containing 527 labels:

Human sounds

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

Source-ambiguous sounds

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

Animal

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

Sounds of things

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

Music

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

Natural sounds

- Wind
- Thunderstorm
- Water
- Fire

Channel, environment and background

- Acoustic environment
- Noise
- Sound reproduction

The top two levels of the AudioSet [ontology](#).

- Discussion: https://www.reddit.com/r/MachineLearning/comments/62g3xw/n_announcing_audio_set_a_dataset_for_audio_event/

- AudioSet download: <https://research.google.com/audioset/download.html>
 - Evaluation – eval_segments.csv – 20,383 segments from distinct videos, providing at least 59 examples for each of the 527 sound classes that are used. Because of label co-occurrence, many classes have more examples.
 - Balanced train – balanced_train_segments.csv – 22,176 segments from distinct videos chosen with the same criteria: providing at least 59 examples per class with the fewest number of total segments.
 - Unbalanced train – unbalanced_train_segments.csv – 2,042,985 segments from distinct videos, representing the remainder of the dataset
 - Features:
 - US: storage.googleapis.com/us_audioset/youtube_corpus/v1/features/features.tar.gz
 - EU: storage.googleapis.com/eu_audioset/youtube_corpus/v1/features/features.tar.gz
 - ASIA: storage.googleapis.com/asia_audioset/youtube_corpus/v1/features/features.tar.gz
- AudioSet
Ontology: <https://github.com/audioset/ontology> | <https://github.com/audioset/ontology/blob/master/ontology.json>
- Related paper: “**Audio Set: An ontology and human-labeled dataset for audio events**” – <https://research.google.com/pubs/pub45857.html>:

Abstract: Audio event recognition, the human-like ability to identify and relate sounds from audio, is a nascent problem in machine perception. Comparable problems such as object detection in images have reaped enormous benefits from comprehensive datasets — principally ImageNet. This paper describes the creation of Audio Set, a large-scale dataset of manually-annotated audio events that endeavors to bridge the gap in data availability between image and audio research. Using a carefully structured hierarchical ontology of 635 audio classes

guided by the literature and manual curation, we collect data from human labelers to probe the presence of specific audio classes in 10 second segments of YouTube videos. Segments are proposed for labeling using searches based on metadata, context (e.g., links), and content analysis. The result is a dataset of unprecedented breadth and size that will, we hope, substantially stimulate the development of high-performance audio event recognizers.

Share this:



Related:

d301: Machine Learning Datasets

December 12, 2016

In "News"

d583: NSynth Dataset - dataset of annotated musical notes

September 21, 2017

In "Sources"

d227: YouTube-8M: A Large and Diverse Labeled Video Dataset for Video Understanding Research

September 29, 2016

In "Sources"

📁 Sources

🔗 Dataset, Google

Previous

d410: DeepMind Neural Episodic Control explanation

Next

d412: Machine Learning and AI Jobs, March 2017

Proudly powered by WordPress | Theme: Afterlight by WordPress.com.