Concept Embeddings

Austin Leung, Tyson Wang, Amir Zadeh

Carnegie Mellon University

Problem

When you think of the word an **owl hooting**, what comes to mind?

Is it the **words** describing how it is doing it?

Definition of *hoot* (Entry 1 of 3)

intransitive verb

1 : to shout or laugh usually derisively

2 : to make the natural throat noise of an owl or a similar cry

Is it the **image** of an owl with its mouth open?

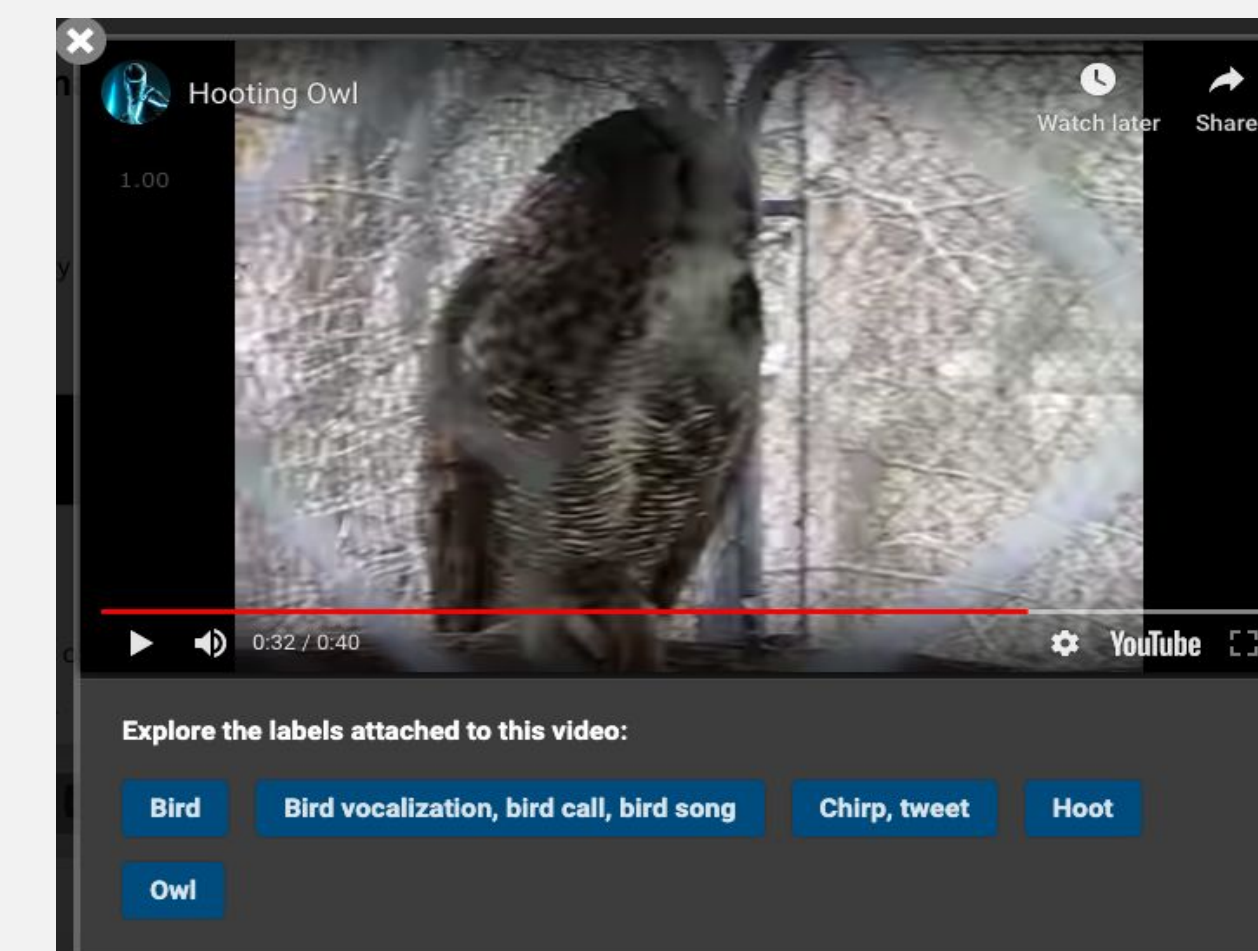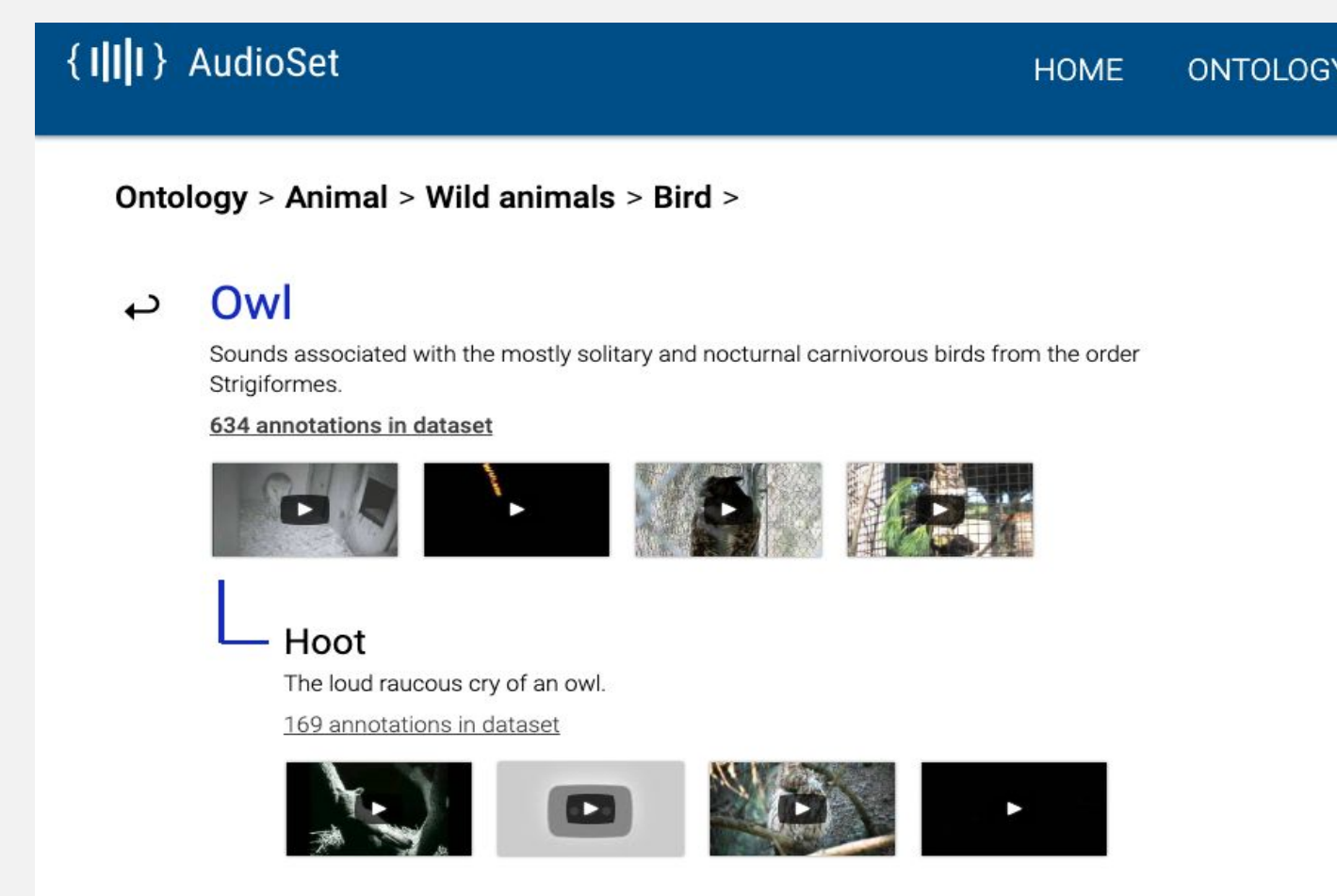Is it the **motion** of an owl opening its mouth?

There are countless media by which concepts are expressed. We would like to connect these different modalities.

Dataset

Google Images is used as a large, descriptive dataset that we can use to learn visual connections between concepts (e.g. running involves humans, shoes, etc.)

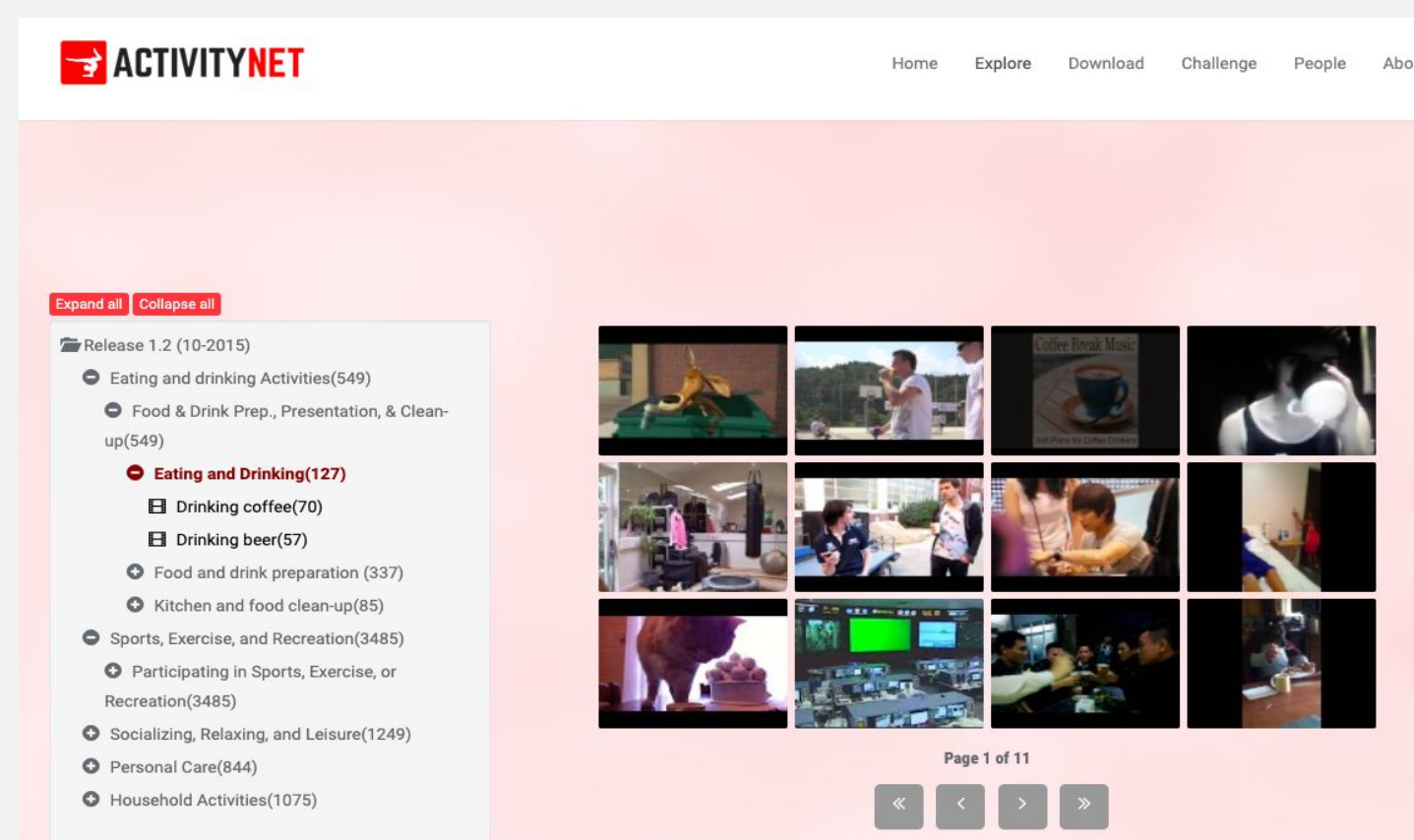|			
Owl - Wikipedia	Barn Owl	Audubon Field Guide	15 Mysterious Facts About Owls	Mental ...	Similar Species to Great Horned...
en.wikipedia.org	audubon.org	mentalfloss.com	allaboutbirds.org

Google's Audioset has over 2 million human-labeled 10 second sound clips from Youtube videos. It is being downloaded to understand the connection between sound, video, and concepts on a large scale.

Goals

As we continue in collecting different types of data, we are interested in downloading ActivityNet, a video dataset containing complex human activities of daily living.

We are also interested in MIT's Moments in Time dataset which will help us understand actions in dynamic 3 minute video clips.

We hope to combine all of these datasets into a model that will make sense of it all.
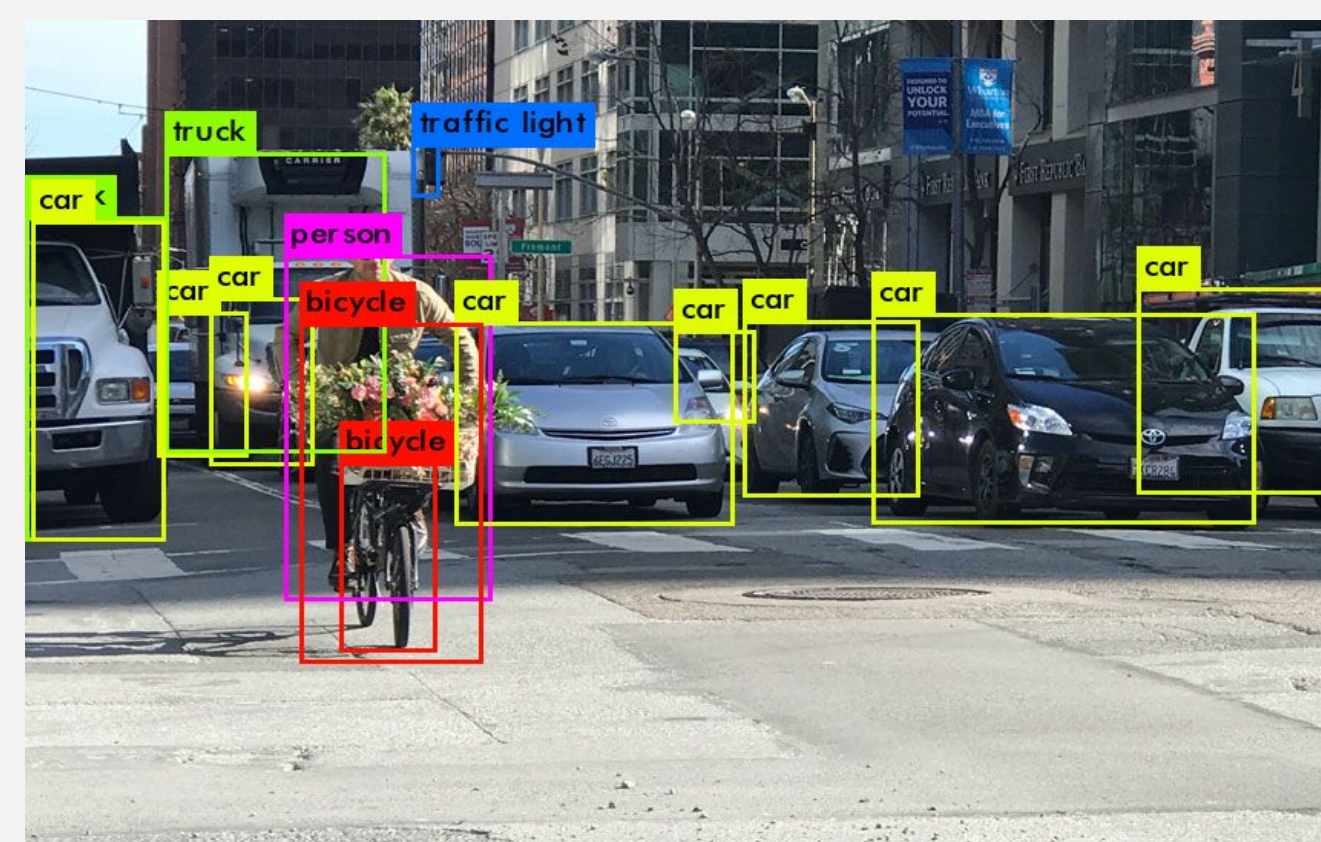
Milestones

We were able to complete a full download of Audioset for initial video analysis and further classification by the Audiovisual team. This data was downloaded into a hierarchy to mirror the Audioset ontology.

Index of /verifications/data

Name	Last modified	Size	Description
Parent Directory		-	
Animal/	2019-09-26 23:31	-	
Channel, environment and background/	2019-08-09 02:57	-	
Human sounds/	2019-08-09 02:57	-	
Music/	2019-09-26 23:31	-	
Natural sounds/	2019-08-09 02:57	-	
Sounds of things/	2019-08-09 02:57	-	
Source-ambiguous sounds/	2019-08-09 02:57	-	

We also finished running object detection algorithms on our dataset of pictures.